The admixture histories of Cabo Verde

Authors:
Romain Laurent[1], Zachary A. Szpiech[2,3], Sergio S. da Costa[1], Valentin Thouzeau[4,5], Cesar A. Fortes-Lima[6], Françoise Dessarps-Freichey[1], Laure Lémée[7], José Utgé[1], Noah A. Rosenberg[8], Marlyse Baptista[9,10], and Paul Verdu[1,*].

Authors' affiliations :
1. UMR7206 Eco-anthropologie, CNRS-MNHN-Université Paris Cité, Paris, France
2. Department of Biology, Pennsylvania State University, University Park, PA 16802, USA
3. Institute for Computational and Data Sciences, Pennsylvania State University, University Park, PA 16802, USA
4. UMR 7534 Centre de Recherche en Mathématiques de la Décision, CNRS-Université Paris-Dauphine-PSL University, Paris, France
5. Département d'Etudes Cognitives, Laboratoire de Sciences Cognitives et Psycholinguistique, ENS-PSL University-EHESS-CNRS, Paris, France
6. Department of Organismal Biology, Sub-department of Human Evolution, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden
7. Plate-forme Technologique Biomics–Centre de Ressources et Recherches Technologiques (C2RT), Institut Pasteur, Paris, France.
8. Department of Biology, Stanford University, Stanford, CA, USA
9. Department of Linguistics, University of Michigan, Ann Arbor, MI, USA
10. Department of Afroamerican and African Studies, University of Michigan, Ann Arbor, MI, USA

***Corresponding author:**
Paul Verdu, UMR7206 Eco-anthropologie, CNRS-MNHN-Université Paris Cité, Musée de l'Homme, 17, place du Trocadéro, 75016 Paris, France. Email : paul.verdu@mnhn.fr. Tel : +33 144057317

Authors' contributions:
PV, MB, NAR designed sampling strategy
PV, MB, SSC, VT, CFL conducted sampling field-work
FD, PV, LL, JU conducted molecular-genetics experiments
RL, PV conducted genetic data quality control and merging
SSC conducted linguistic transcriptions
MB, SSC conducted linguistic data categorization
RL, PV conducted linguistic data parsing
RL, PV, ZAS, MB, SSC, VT conducted genetic and linguistic statistical descriptions
RL, PV conducted genetic inferences
PV, RL, MB, ZAS wrote the paper with the help of all authors
PV, MB, and NAR supervised the research

Ethics statement
Research sampling protocols followed the Declaration of Helsinki guidelines and the French laws of scientific research deontology (Loi n° 2016-483 du 20 avril 2016). Research and ethics authorizations were provided by the Ministério da Saúde de Cabo Verde (228/DGS/11), Stanford University IRB (Protocol ID n°23194–IRB n°349), University of Michigan IRB (n°HUM00079335), and the French ethics committees and CNIL (Declaration n°1972648). All volunteer participants provided written and video-recorded informed consent.

Data availability

The novel genetic data presented here can be accessed via the European Genome-phenome Archive (EGA) database accession number #XX (pending) upon request to PV. Linguistic data presented here can be obtained directly upon request to the corresponding author (PV). Both datasets can be shared provided that future envisioned studies comply with the informed consents provided by the participants, and in agreement with institutional ethics committee's recommendations applying to this data.

Software availability

MetHis is an open source C software available with user manual on GitHub at https://github.com/romain-laurent/MetHis (Fortes-Lima et al. 2021).

Keywords

Admixture; population genetics; historical inference; machine-learning; approximate Bayesian computations; genetic diversity; linguistic diversity; anthropology; slave-trade

ABSTRACT

From the 15th to the 19th century, the Trans-Atlantic Slave-Trade influenced the genetic and cultural diversity of numerous populations. We explore genomic and linguistic data from the nine islands of Cabo Verde, the earliest European colony of the era in Africa, a major Slave-Trade platform between the 16th and 19th centuries, and a previously uninhabited location ideal for investigating early admixture events between Europeans and Africans. We find that genetic admixture in Cabo Verde occurred primarily between Iberian and Senegambian populations, although forced and voluntary migrations to the archipelago involved numerous other populations. Inter-individual genetic and linguistic variation recapitulate geographic distribution of individuals' birth-places across Cabo Verdean islands, suggesting that Kriolu language variants have developed together with genetic divergences. Furthermore, we find that admixture occurred both early in each island, long before the 18th-century massive TAST deportations triggered by the expansion of plantation economy in Africa and the Americas, and after this era mostly during the abolition of the TAST and of slavery in European colonial empires. Our results illustrate how shifting socio-cultural relationships between enslaved and non-enslaved communities shaped enslaved-African descendants' genomic diversity and structure on both sides of the Atlantic.

INTRODUCTION

Between the 15th and 19th centuries, European colonization and the Trans-Atlantic Slave-Trade (TAST) put into contact groups of individuals previously isolated genetically and culturally. These forced or voluntary migrations profoundly influenced the descent of numerous European, African, and American populations, creating new cultures, languages, and genetic patterns (1,2).

Population geneticists have extensively described genetic admixture patterns in enslaved-African descendants in the Americas, and mapped their genomes for regions of recently shared ancestry with continental Africa and Europe (3,4). This allowed for reconstructing their detailed possible origins, as this knowledge is often intractable with genealogical records alone (1). Furthermore, genetic admixture-mapping methods have been used to identify genetic variation underlying phenotypic variation (5,6), and to identify post-admixture natural selection signatures (7), thus revealing how admixture shaped human populations' recent evolution. Maximum-likelihood approaches based on linkage-disequilibrium (LD) patterns of admixed individuals(8,9) have repeatedly highlighted the diversity of admixture processes experienced by populations historically related to the TAST. In particular, they identified different European, African, and American populations respectively at the source of genetic admixture patterns, sometimes consistent with the preferred commercial routes of each European empire (8,10). Furthermore, they identified variable timing of admixture events during and after the TAST, sometimes consistent with major socio-historical events such as the expansion of the plantation economic system or the abolition of slavery (11,12). From a cultural perspective, linguists have shown that novel contact-languages, such as creole languages (13,14), emerged from recurring interactions between socio-economically dominant Europeans with Africans and Americans. Furthermore, they identified the languages of origin of numerous linguistic traits in several creole languages (15–17), and emphasized the complex histories of contacts that shaped language diversity on both sides of the Atlantic.

Numerous questions remain unsolved and novel interrogations have emerged concerning the history of admixture during and after the TAST. First, while the genetic history of enslaved-African descendants in the Americas has been extensively studied, the influence of the TAST on genetic admixture in Africa remains under-investigated. Nevertheless, studying these questions in Africa would provide invaluable information about the influence of the onset and early stages of the TAST and the subsequent expansion of European empires on genetic admixture patterns on both sides of the ocean. Second, while admixture-LD inference methods have repeatedly brought novel insights into the admixture processes experienced by enslaved-African descendant populations, they could only explore historical models with one or two pulses of admixture, a methodological limitation extensively discussed by the authors (8,9). Nevertheless, complex admixture histories may be expected as a result of the recurring flows of enslaved-Africans forcibly displaced between and within continents, changes of social relationships among enslaved and non-enslaved communities, and variable assimilation of new migrants in pre-existing communities, during and after the TAST (18,19). Finally, while the comparison of genetic and linguistic diversities has been the focus of numerous evolutionary anthropology studies at large geographical scales, it has rarely been endeavored for creole-speaking populations in the historical context of the TAST (20–22).

Here, we propose to reconstruct the detailed genetic and linguistic admixture histories of Cabo Verde, as this archipelago represents an ideal case to address these three understudied aspects of TAST history. First, Cabo Verde is the first European peopling colony in Sub-Saharan Africa, located 500 kms West of Senegal in Western Africa (**Figure 1**), and settled in the 1460s by Portuguese immigrants and enslaved-Africans forcibly removed from the continent. After Columbus, and in particular after the 17th century expansion of plantation economy in the Americas, Cabo Verde served as a major slave-trade platform between continents (23). Second, Cabo Verde forms an archipelago of nine inhabited islands that were settled over the course of three centuries due to the changing political, commercial, and forced or voluntary migratory contexts (24–26). Therefore,

studying the admixture history of Cabo Verde will provide unique insights into the onset of the TAST before Columbus, and into the history of slavery thereafter. This setting further promises to illustrate, at a micro-geographical scale island-per-island, the fundamental socio-historical and serial founding migrations mechanisms having influenced genomic patterns in admixed populations throughout the TAST. Finally, Cabo Verdean Kriolu is the first creole language of the TAST, born from contacts between the Portuguese language and a variety of African languages (15,17,27,28). The archipelago thus represents a unique opportunity to investigate, jointly, genetic and linguistic admixture histories and their interactions since the mid-15th century.

Previous genetic studies exploring sex-specific or genome-wide markers from several islands of the archipelago (29–32), attested to the dual, sex-biased, origins of the Cabo Verdean gene-pool, resulting mainly from admixture between African females and European males. Furthermore, these studies described variable admixture patterns across islands. Interestingly, adaptive-introgression signals for malaria resistance in Santiago island were recently identified as a result of migrations and genetic admixture during the TAST (33). Finally, joint analyses of genetic and linguistic diversities from Santiago showed that genetic and linguistic admixture histories possibly occurred in parallel (21).

Based on these previous studies, we propose to first determine which continental African and European populations in fact contributed to the genetic landscape of each island today. Indeed, which enslaved-African populations only briefly transited through the archipelago, and which remained for longer periods is largely debated by historians (23–25); and, while Portuguese influence is clear, further details about which European migrations genetically influenced Cabo Verde remain to be assessed (34). These aspects are often crucial for understanding the genetic history of enslaved-African descendant populations on either side of the Atlantic (1,12,35,36). Second, we propose to further evaluate the possible parallels between genetic and linguistic admixture histories in each island to better understand how contacts shaped cultural variation during the TAST (21). Finally, we reconstruct the detailed history of admixture dynamics in each island since the 15th century. Altogether, this will highlight the socio-historical mechanisms that shaped the genetic and linguistic diversity of this population, the first to be born from the TAST.

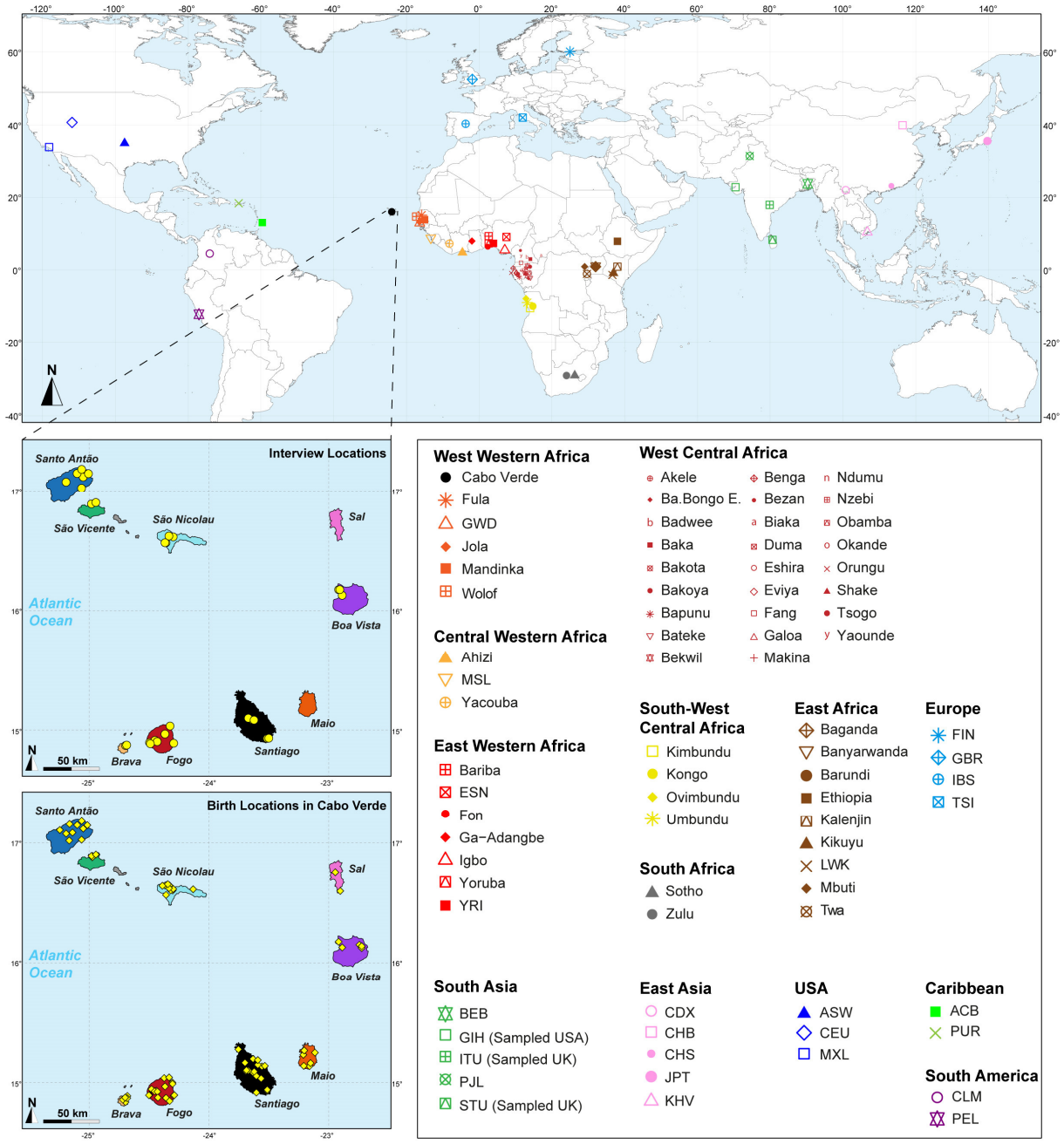


Figure F1:

Sampling and birth place locations of 225 unrelated Cabo Verdean-born individuals, merged with 77 worldwide populations. Linguistic and familial anthropology interviews, and genetic sampling for Cabo Verde participants were conducted during six separate interdisciplinary fieldworks between 2010 and 2018. Further details about populations are provided in **Table ST1**.

RESULTS

We investigate genetic and linguistic variation in 261 Kriolu speakers from the nine Cabo Verdean islands (Brava, Fogo, Santiago, Maio, Sal, Boa Vista, São Nicolau, São Vicente, Santo Antão, **Figure F1**). With novel genome-wide genotyping autosomal data (**Figure SF1**), we first describe genetic differentiation patterns in Cabo Verde and other enslaved-African descendants in the Americas from previous datasets, in particular with respect to continental Africa and Europe. Second, we deploy local-ancestry inferences and determine the best proxy source-populations for admixture patterns in each Cabo Verde island. Then, we describe Runs of Homozygosity and genetic isolation-by-distance patterns at reduced geographical scale within Cabo Verde. We also investigate Kriolu utterance diversity with respect to geography and socio-cultural co-variables and, then, investigate jointly genetic and linguistic admixture patterns throughout the archipelago. Finally, we infer the detailed genetic admixture history of each island using the machine-learning *MetHis*-Approximate Bayesian Computation (ABC) approach (37).

1. Cabo Verde and other admixed populations related to the TAST in the worldwide genetic context

We explored genetic diversity patterns captured along the first three axes of the Multi-Dimensional Scaling (MDS) projection of individual pairwise Allele Sharing Dissimilarities (ASD (38)), computed from different individual subsets (**Table ST1**). Detailed ASD-MDS decompositions are provided in **Supplementary Text** and **Figures SF2-6**. Note that we considered seven geographical regions in Africa shown in **Figure F1**.

Figure F2 show that the second ASD-MDS axis distinguishes West Western African Senegambian populations from East Western Africans, while Central Western Africans are at intermediate distances between these two clusters. Moreover, the third MDS axis separates Northern and Southern European populations; the British-GBR and USA-CEU individuals clustering at intermediate distances between the Finnish-FIN, and a cluster represented by Iberian-IBS and Tuscan-TSI Western Mediterranean individuals. Consistently with previous results (3,21), on the first MDS axis, Cabo Verdean individuals cluster almost exclusively along a trajectory going from the Southern European cluster to Senegambia (**Figure F2A-C**), with little traces of affinity with other African populations. Instead, the USA African-American ASW (**Figure F2D-F**) and Barbadian-ACB (**Figure F2G-I**) cluster along a trajectory going from the GBR and CEU cluster to Central and East Western Africa; and Puerto Ricans-PUR cluster along a trajectory going from the Southern European cluster to the Central Western African cluster (**Figure F2J-L**).

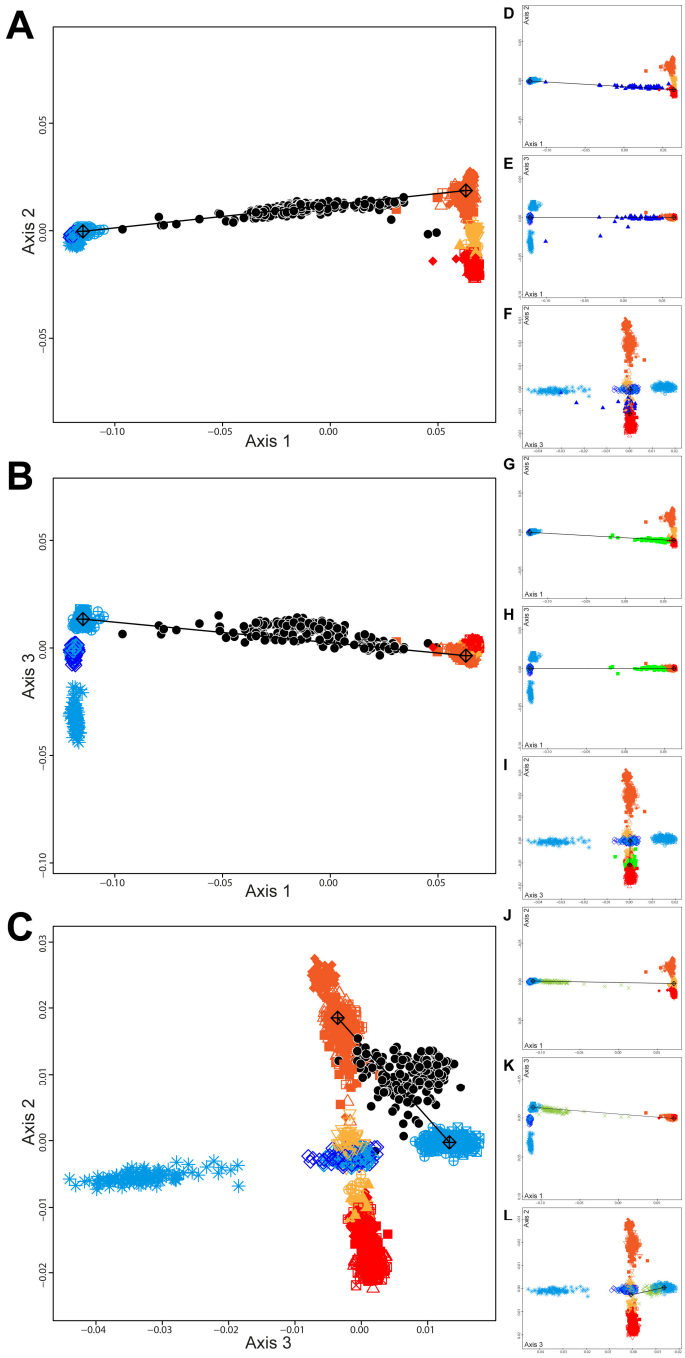


Figure F2:

Multidimensional Scaling projections of pairwise Allele Sharing Dissimilarities in Cabo Verdeans and continental African and European populations.

A-C) 3-Dimensional MDS projection of ASD computed among 233 unrelated Cabo Verdeans and other continental African and European populations using 445,705 autosomal SNPs. Cabo Verdean patterns in panels A-C are compared to **D-F** the USA African-Americans ASW, **G-I** the Barbadians-ACB, or **J-L** the Puerto Ricans-PUR in the same African and European contexts. We computed the Spearman correlation between 3D MDS projections and the original ASD matrix to evaluate the precision of the dimensionality reduction. We find significant ($p<2.2\text{x}10^{-16}$) Spearman $\rho=0.9635$ for the Cabo Verde analysis (A-C); 0.9383 for the ASW (D-F); 0.9306 for the ACB (G-I); and 0.9437 for the PUR (J--L). Each individual is represented by a single point. See **Table ST1** for the populations used in these analyses, and **Figure F1** for sample locations and symbols.

205 **2. Genetic structure in Cabo Verde and other admixed populations related to the TAST**
206 Based on these results, we further investigated patterns of individual genetic structure among Cabo Verde-
207 born individuals, ASW, and ACB populations with respect to European and Western, Central, South-Western,
208 and Southern African populations (**Table ST1**), using ADMIXTURE (39). Extended descriptions of the results
209 are presented in **Supplementary Text**.
210 At *K*=2, the orange genetic-cluster in **Figure F3A** is maximized in African individuals while the blue
211 alternative cluster is maximized in Europeans. Cabo Verdean, ASW and ACB individuals exhibit intermediate
212 genotype-membership proportions between the two clusters, consistently with patterns expected for European-
213 African admixed individuals. ACB individuals show, on average, the highest membership to the orange "Af-
214 rican" cluster (88.23%, SD=7.33%), followed by the ASW (78.00%, SD=10.88%), and Cabo Verdeans
215 (59.01%, SD=11.97%). Furthermore, membership proportions to this cluster are highly variable across Cabo
216 Verdean islands, with highest average memberships for Santiago-born individuals (71.45%, SD=10.39%) and
217 Maio (70.39%, SD=5.26%), and lowest for Fogo (48.10%, SD=6.89%) and Brava (50.61%, SD=5.80%). Fur-
218 thermore, we found that inter-individual memberships variation within Cabo Verde islands captured as
219 F_{ST}/F_{ST}^{max} values (40), are significantly different across pairs of islands for 32 out of 36 comparisons (Wilcoxon
220 rank-sum test $p<3.96\text{x}10^{-8}$), with variability across islands ranging from a lowest value of 0.010 in individuals
221 from Santo Antão to a highest value of 0.0519 in Santiago (**Figure SF7**).
222 At *K*=5, the new red cluster is maximized in the YRI, Igbo and ESN populations, distinct from Western
223 and Southern African orange and gray clusters, respectively. Note that the former orange cluster is almost
224 completely replaced with red membership in the ACB and ASW populations. Conversely, Cabo Verde-born
225 individuals exhibit small to virtually no membership to this East Western African red cluster, but their African
226 component, instead, resembled genetic patterns found among West Western Africans.
227 Ultimately, at *K*=10, the light-green cluster is maximized in Cabo Verdean individuals born on Maio, Boa
228 Vista, Sal, and São Nicolau, distinct form the dark green cluster maximized in individuals born on Santo Antão
229 and São Vicente, and hence producing three distinct ADMIXTURE patterns among Cabo Verdean birth-is-
230 lands. Furthermore, an alternative mode at *K*=10 shows (**Figure SF8**), that Cabo Verde-born individuals re-
231 semble more IBS and TSI patterns for their European-like membership than ASW and ACB individuals who,
232 instead, resemble more CEU and GBR patterns, consistently with ASD-MDS results and previous studies
233 (10,21).
234
235

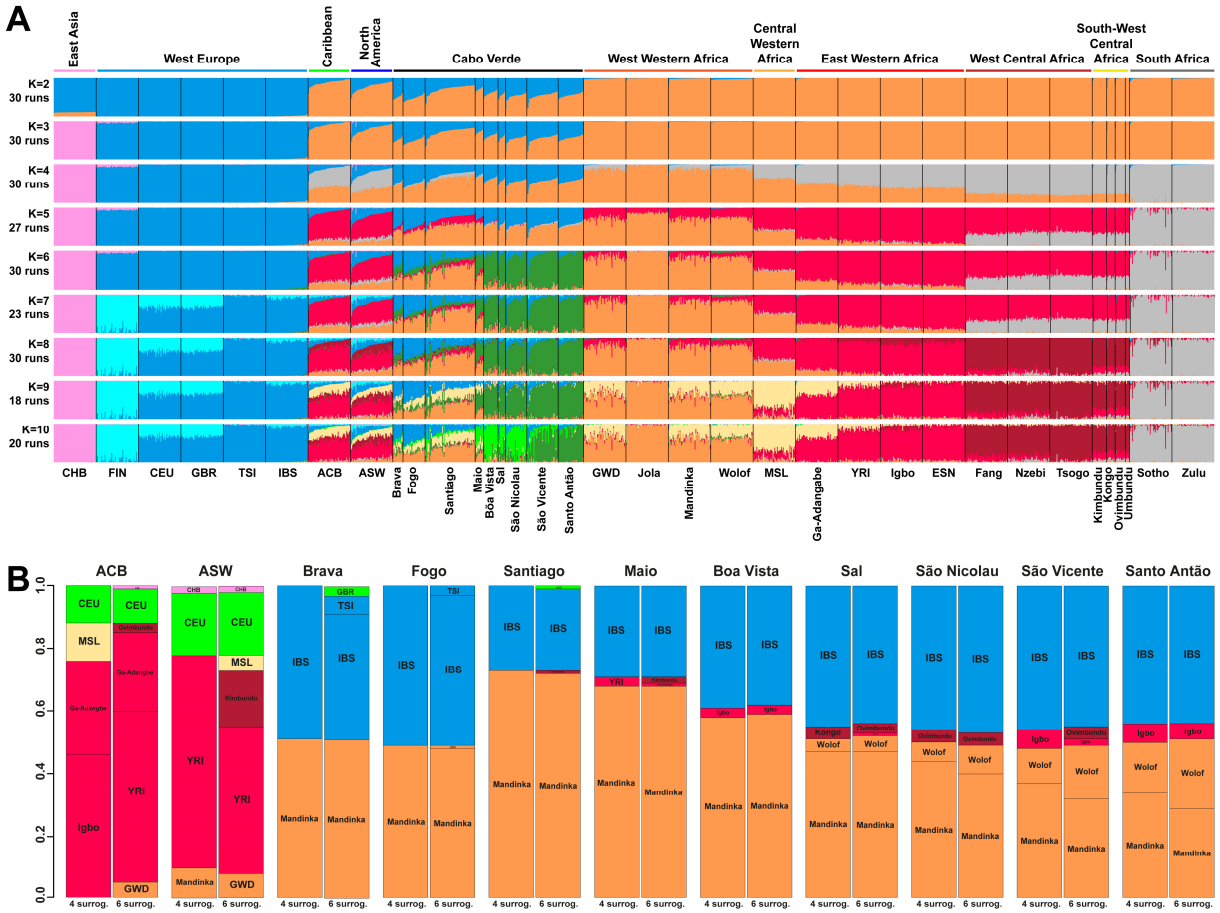


Figure F3:

Individual genetic structure and haplotypic local ancestry inference among Cabo Verdean, Barbadian-ACB and African-American ASW populations.

A) Unsupervised ADMIXTURE analyses using randomly resampled individual sets for populations originally containing more than 50 individuals (**Table ST1**). 225 unrelated Cabo Verdean-born individuals are comprised in the analysis and grouped by birth-island. Number of runs highly resembling one-another using CLUMPP are indicated below each *K*-value. All other modes are presented in **Figure SF8**. **B)** SOURCEFIND results for each eleven target admixed populations (ASW, ACB, each nine Cabo Verde birth-island), considering respectively 4 or 6 possible source surrogate populations (abbreviated "surrog.") among the 24 possible European, African, and Asian populations considered in the ADMIXTURE analyses. The cumulated average African admixture levels in each admixed population was highly consistent between SOURCEFIND estimates and ADMIXTURE results at K=2 (Spearman ρ=0.99772, $p<8\times10^{-12}$, and 0.98861, $p<2\times10^{-8}$, for 6 or 4 surrogates respectively). Furthermore, individual admixture levels estimated using an ASD-MDS based approach were highly consistent with estimates based on ADMIXTURE at K=2 (ρ=0.99725; $p<2.2\times10^{-16}$ for Cabo Verde; ρ=0.99568; $p<2.2\times10^{-16}$ for ASW; ρ =0.99366; $p<2.2\times10^{-16}$ for ACB, **Supplementary Text**).

253 **3. Local-ancestry in Cabo Verde and other admixed populations related to the TAST**
254 ASD-MDS and ADMIXTURE descriptive analyses do not formally test admixture and putative source popu-
255 lations of origins, they rather disentangle genetic-structure resemblances among individuals (39,41–44). There-
256 fore, we conducted admixture-LD haplotypic local-ancestry inferences with the SHAPEIT2-CHROMO-
257 PAINTER-SOURCEFIND pipeline (45–47), to identify the possible European and African populations at the
258 sources of admixture patterns.
259 **Figure F3B** shows striking differences concerning both the European and the African source populations
260 involved in the admixture history of ACB, ASW, and individuals born on different Cabo Verdean islands. We
261 find that individuals from all Cabo Verdean islands share almost all their European haplotypic ancestry with
262 the Iberian-IBS population rather than other European populations. Santiago-born individuals present the
263 smallest (27%) average haplotypic ancestry shared with IBS, and Fogo-born the highest (51%). Conversely,
264 the ASW and ACB both share haplotypic ancestries only with the USA-CEU of North-Western European
265 origin (20% and 12% respectively).
266 Furthermore, we find that all Cabo Verdeans almost exclusively share African haplotypic ancestries with
267 two Senegambian populations (Mandinka and Wolof) and very reduced to no shared ancestries with other
268 regions of Africa. More specifically, we find that the Mandinka from Senegal are virtually the sole African
269 population to share haplotypic ancestry with Cabo Verdeans born on Brava, Fogo, Santiago, Maio, and Boa
270 Vista, and the majority of the African shared ancestry for Sal, São Nicolau, São Vicente, and Santo Antão. In
271 individuals from these four latter islands, we find shared haplotypic ancestry with the Wolof population rang-
272 ing from 5%-4% for individuals born on Sal (considering six or four possible sources, respectively), up to
273 22%-16% for Santo Antão. Finally, we find limited (1%-6%) shared haplotypic ancestry with East Western
274 (Igbo, YRI, or Ga-Adangbe) or South-West Central (Kimbundu, Kongo, or Ovimbundu) African populations
275 in all Cabo Verdean islands, except Fogo and Brava, and the specific populations identified and their relative
276 proportion of shared haplotypic ancestries varies across analyses. Conversely, we find that the ASW and ACB
277 populations share African haplotypic ancestries in majority with East Western African populations (YRI, Ga-
278 Adangbe, and Igbo), and substantial shared ancestries with Senegambian populations (5-10%), the MSL from
279 Sierra Leone in Central Western Africa (5-12%), and South-West Central African populations (3-18%), albeit
280 variable depending on the number of putative sources considered.
281
282 **4. Genetic and linguistic diversity and admixture within Cabo Verde**
283 *Runs of Homozygosity (ROH) patterns within Cabo Verde*
284 Runs of homozygosity (ROH) are Identical-By-Descent haplotypes resulting from recent parental relatedness
285 and present in individuals as long stretches of homozygous genotypes. Their length and abundance can reflect
286 demographic events, such as population bottlenecks and founder events, natural selection, and cultural
287 preferences for endogamy (48–50).
288 We find higher levels of long ROH (≥1cM) in Cabo Verdeans compared to most other analyzed
289 populations, including ASW and ACB (**Figures F4A, SF9**). We find the highest levels of long-ROH in
290 individuals born on Maio, Boa Vista, Sal, São Nicolau, São Vincente, and Santo Antão, with a mean individual
291 length of long-ROHs around 3cM (**Figure F4B**), and the lowest levels of long-ROH in Santiago and Brava-
292 born individuals. Among long ROH (**Figure SF10C**), we find little to no correlation with total non-ROH levels
293 for African local-ancestry segments (Pearson ρ=-0.06689, p=0.3179), European (ρ=0.1551, p=0.01989), or
294 Asian (ρ=0.06239, p=0.3516).
295 In admixed populations, we expected that some of the long ROH spanned local ancestry breakpoint
296 switches (see **Materials and Methods**), indicating that the most recent common ancestor existed after the
297 initial admixture event having generated local-ancestry patterns. Furthermore, we expected that these
298 "spanning" ROH should be among the longest ROH observed if admixture occurred only in the past few
299 generations. We find that (**Figure F4C**), almost uniformly across Cabo Verde, the longest ROH identified

300 indeed spanned at least one ancestry breakpoint, excluding the very few Asian ancestry regions identified.
301 Furthermore, correcting ancestry-specific long-ROH sizes (**Figure SF11**) for individual's total ancestry
302 fraction of that ancestry, we find that individuals born in Brava and Fogo have, on average, an apparent
303 overrepresentation of European ancestry in long-ROH (**Figure F4D**), and that individuals from Sal and São
304 Vincente have, conversely, an apparent overrepresentation of African ancestry. Finally, we find that
305 individuals from Santiago, Maio, Boa Vista, São Nicolau, and Santo Antão have relatively similar long-ROH
306 levels in either African or European segments. This latter pattern may be consistent with these populations
307 being founded by admixed individuals, while the former patterns could indicate, in addition to such admixture
308 founding effect, more recent or recurring contributions from either source, respectively (Mooney et al. 2018).
309

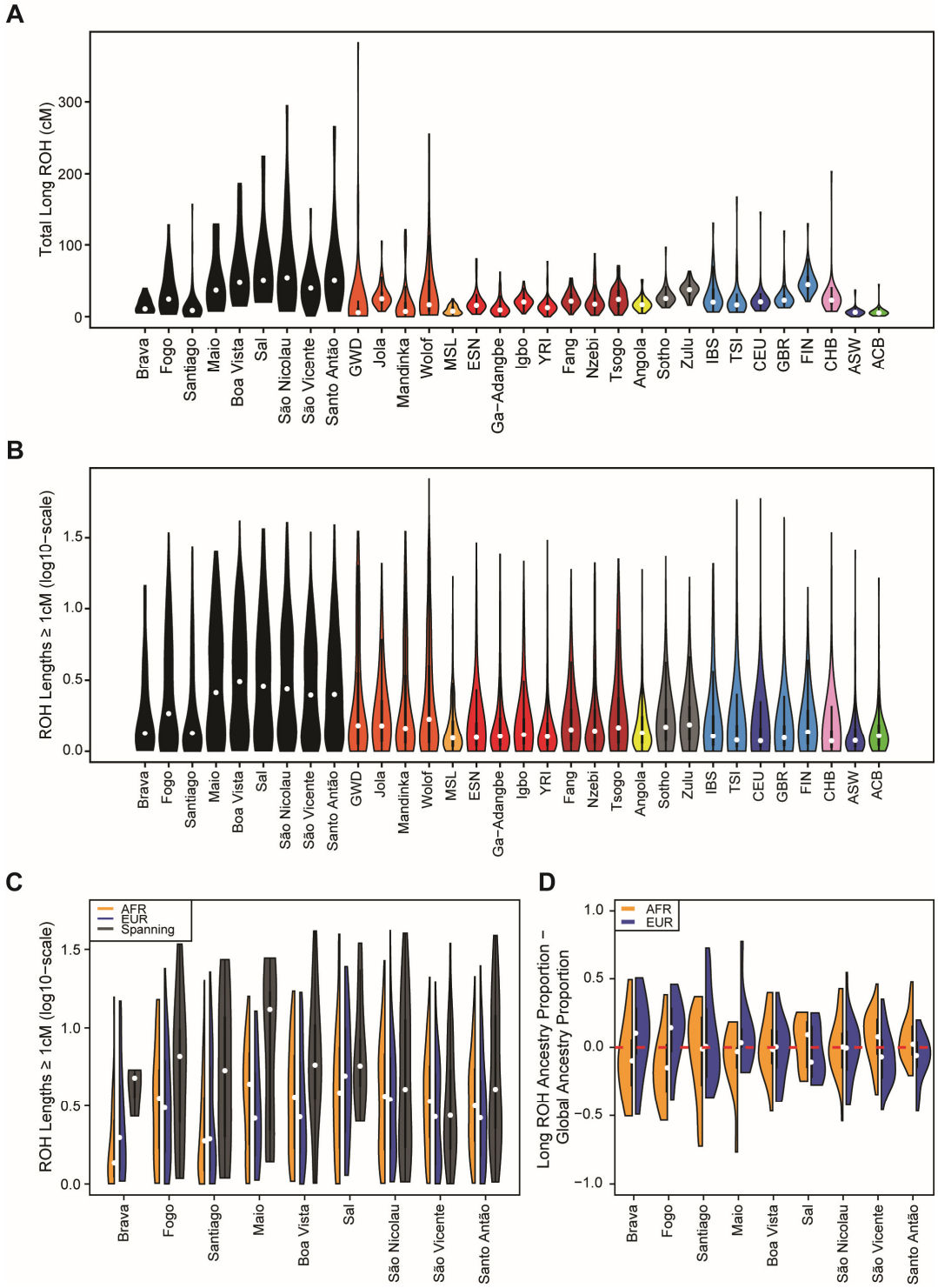


Figure F4:

Distributions of long ROH in Cabo Verde.

A) The distribution of total long-ROH per person for each Cabo Verdean birth-island and other populations. **B)** The distribution of individual long-ROH lengths (log10-scale). **C)** The distribution of ancestry-specific and ancestry-spanning individual long-ROH lengths for each Cabo Verdean birth-island. **D)** The distribution of differences between individuals' long-ROH ancestry proportion and their global ancestry proportion, for African and European ancestries separately and for each Cabo Verdean birth-island.

ASD and utterance diversity within Cabo Verde

Together with the above results, previous studies suggested substantial genetic and linguistic isolation across certain Cabo Verdean groups of islands (15,17,21,27,28,32). We aim at further characterizing possible patterns of isolation-by-distance among and within islands, both genetically and linguistically. We explored both pairwise ASD and inter-individual variation in Kriolu utterance frequencies (**Material and Methods**), in the same set of 225 Cabo Verde-born individuals, using MDS and Mantel testing (**Figure F5**, **Table T1**, and **Table ST14**).

The first ASD-MDS axis mainly differentiates individuals born on Brava and Fogo compared to Santiago (**Figure F5A-B**). The second axis differentiates mainly individuals from Santiago, Fogo, and Brava from all other islands, while the third axis differentiates individuals from Boa Vista, São Nicolau, Sal, and Maio from all other birth-islands. Furthermore, we find a significant positive correlation between ASD and actual individual birth-locations across Cabo Verde (**Table T1**; Spearman ρ=0.2916, two-sided Mantel $p<2x10^{-4}$). This correlation increases within-islands excluding inter-island distances (Spearman ρ=0.3460, two-sided Mantel $p<2x10^{-4}$), thus illustrating the strong signal of genetic isolation-by-distance (51) within Cabo Verde at very reduced geographical scales.

Furthermore, the first utterance-MDS axis of pairwise inter-individual Euclidean distances between utterance frequencies mainly differentiates Santiago and Santo Antão/São Vicente-born individuals' speech-varieties; all other Cabo Verdeans clustering intermediately (**Figure F5C-D**). The third axis further separates speech-varieties recorded in individuals from Fogo, Maio, and Brava. Analogously to genetic differentiation patterns, we find a positive correlation between differences in utterance frequencies and actual birth-places' distances (Spearman ρ=0.2794, two-sided Mantel $p<2x10^{-4}$), as well as paternal and maternal birth-places respectively (**Table T1**). However, contrarily to ASD, we find that utterance-frequencies differences stem from inter-birth-islands' distances, rather than shorter distances within islands only. Consistently with previous results from Santiago only (21), these results altogether show that speech-varieties are significantly transmitted from one generation to the next throughout Cabo Verde, anchored in individuals' birth-places, albeit not accounting for the majority of observed linguistic variation in our dataset. Indeed, we find that age-differences also substantially correlate with utterance-frequency differences even correcting for individual birth-places (Spearman ρ=0.2294, two-sided partial-Mantel $p<2x10^{-4}$). Finally, while we might intuitively expect that academic education influences idiolects, we find instead that differences in education-duration do not correlate with Kriolu utterance-frequencies differences, whether correcting for residence or birth-places distances, or not (**Table T1**). This shows the modest influence of academic education on Kriolu variation. Altogether, our results highlight strong genetic and linguistic isolation-by-distance patterns at reduced geographic distances within Cabo Verde.

Genetic and linguistic admixture within Cabo Verde

Interestingly, we find that absolute differences in inter-individual genetic admixture levels from Africa, estimated with ADMIXTURE or ASD-MDS, significantly correlate with actual birth-places distances across islands (Spearman ρ=0.1865, two-sided Mantel $p<2x10^{-4}$ and ρ=0.1813, $p<2x10^{-4}$, respectively), but not within-islands only (ρ=0.0342 p=0.3094 and ρ=0.0282 p=0.3385, respectively), indicating that two individuals born on far-away islands are more likely to differ in African-admixture levels.

We explored inter-individual variation in Kriolu utterance frequencies specifically for uttered items of clearly African and dual European-African origins (utterance categories A and B; **Material and Methods**) providing an estimate of individual African linguistic-admixture scores (21). We find that African linguistic-admixture score differences significantly correlate with actual birth-places' distances throughout Cabo Verde (Spearman ρ=0.1297, two-sided Mantel $p<2x10^{-4}$), and even marginally significantly correlate with birth-places' distances at short distances within birth-islands (Spearman ρ=0.1209, two-sided Mantel p=0.0419).

366 Finally, we find a significant positive correlation (Spearman ρ=0.2070, p=0.0018) between genetic and
367 linguistic admixture in Cabo Verde (**Figure 5E**), indicating that individuals who frequently use African-related
368 utterances in their manner of speaking Kriolu are more likely to exhibit higher levels of African genetic-ad-
369 mixture. This correlation remains, respectively, marginally significant and significant when considering utter-
370 ances of strictly African-origin (Category A) or utterances with a dual European-African etymology (Category
371 B) separately (Spearman ρ=0.1631, p=0.0143, and ρ=0.1829, p=0.0059, respectively).
372 Altogether, our genetic and linguistic diversity results generalize our previous results obtained in Santiago
373 only (21), and further suggest that genetic and linguistic admixture histories occurred in parallel in Cabo Verde.
374 Together with the above genetic and linguistic isolation-by-distance signals and ROH results, this strongly
375 suggests that different founding events and admixture histories are at the root of genetic and linguistic diversity
376 on each Cabo Verde islands.
377
378

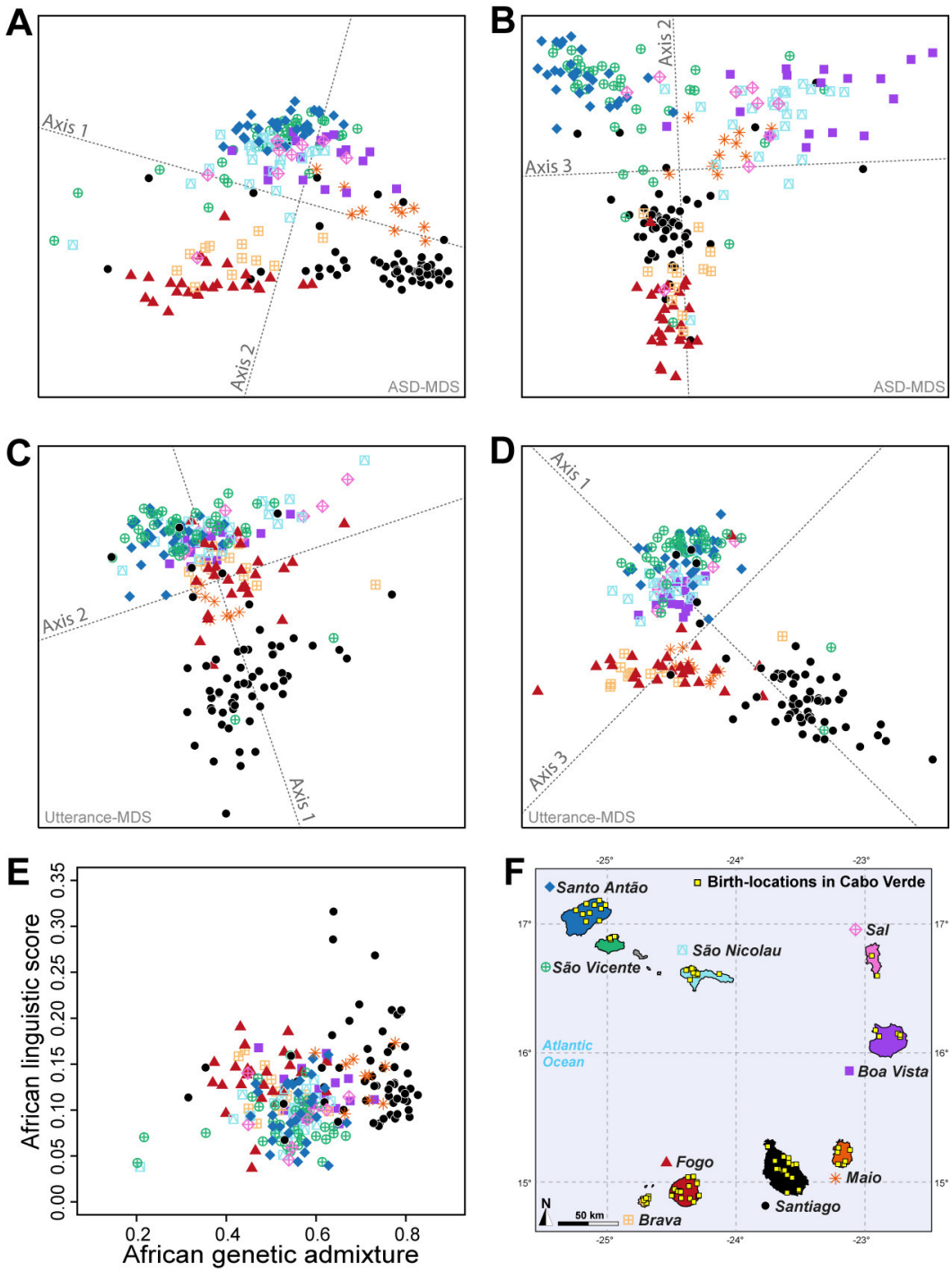


Figure F5:

Utterance and genetic diversity and admixture within Cabo Verde.

A-B) 3D MDS projection of Allele Sharing Dissimilarities computed among 225 unrelated Cabo-Verde-born individuals using 1,899,878 autosomal SNPs. Three-dimensional Euclidean distances between pairs of individuals in this MDS significantly correlated with ASD (Spearman ρ=0.6863; p<2.2x10^{-16}). **C-D)** 3D MDS projection of individual pairwise Euclidean distances between uttered linguistic items frequencies based on the 4831 unique uttered items obtained from semi-spontaneous discourses. Three-dimensional Euclidean distances between pairs of individuals in this MDS significantly correlated with the utterance-frequencies distances (Spearman ρ=0.8647; p<2.2x10^{-16}). **E)** Spearman correlation between individual African utterance scores and individual genetic African admixture rates obtained with ADMIXTURE at K=2. **F)** Birth-locations of 225 individuals in Cabo Verde. Symbols for individuals' birth-island in panels **A-D** are shown in panel **F**. Panel **A-D** were procrusted according to individual actual birth-places' geographical locations in panel **F** (52).

Mantel variable	Partial-Mantel control	n	Geographic scale	Utterance-frequency Euclidean distances - 4831 uttered items		Genetic ASD - 1,899,978 SNPs	
				Spearman rho	10000 Mantel two-sided permutation p	Spearman rho	10000 Mantel two-sided permutation p
abs(Age difference)	--	225	within and between islands	**0.2215**	**< 2.10^{-4}**	**0.1303**	**< 2.10^{-4}**
abs(Age difference)	log(Birth-loc. dist.)	225	within and between islands	**0.2294**	**< 2.10^{-4}**	**0.1348**	**< 2.10^{-4}**
log(Birth-loc. dist.)	--	225	within and between islands	**0.2794**	**< 2.10^{-4}**	**0.2916**	**< 2.10^{-4}**
log(Birth-loc. dist.)	abs(Age difference)	225	within and between islands	**0.2855**	**< 2.10^{-4}**	**0.2935**	**< 2.10^{-4}**
abs(Education duration difference)	--	186	within and between islands	*0.0962*	*0.0024*	*0.0168*	*0.2730*
abs(Education duration difference)	log(Birth-loc. dist.)	186	within and between islands	*0.0834*	*0.0071*	*-0.0023*	*0.4900*
abs(Education duration difference)	--	185	within and between islands	*0.1001*	*0.0014*	*0.0159*	*0.2825*
abs(Education duration difference)	log(Residence dist.)	185	within and between islands	*0.0824*	*0.0068*	*-0.0041*	*0.4651*
log(Residence dist.)	--	224	within and between islands	**0.2145**	**< 2.10^{-4}**	**0.1658**	**< 2.10^{-4}**
log(Residence dist.)	log(Birth-loc. dist.)	224	within and between islands	*0.0306*	*0.0682*	**-0.0488**	**0.0005**
log(Birth-loc. dist.)	--	224	within and between islands	**0.2800**	**< 2.10^{-4}**	**0.2889**	**< 2.10^{-4}**
log(Birth-loc. dist.)	log(Residence dist.)	224	within and between islands	**0.1863**	**< 2.10^{-4}**	**0.2445**	**< 2.10^{-4}**
log(Father Birth-loc. dist.)	--	222	within and between islands	**0.1704**	**< 2.10^{-4}**	**0.2424**	**< 2.10^{-4}**
log(Father Birth-loc. dist.)	log(Birth-loc. dist.)	222	within and between islands	*0.0066*	*0.3915*	*0.0846*	*0.0014*
log(Mother Birth-loc. dist.)	--	224	within and between islands	**0.2634**	**< 2.10^{-4}**	**0.2619**	**< 2.10^{-4}**
log(Mother Birth-loc. dist.)	log(Birth-loc. dist.)	224	within and between islands	*0.0853*	*0.0071*	*0.0748*	*0.0057*
abs(Age difference)	--	225	within islands only	**0.2727**	**< 2.10^{-4}**	**0.2124**	**0.0006**
abs(Age difference)	log(Birth-loc. dist.)	225	within islands only	**0.2546**	**< 2.10^{-4}**	*0.1648*	*0.0041*
log(Birth-loc. dist.)	--	225	within islands only	*0.1412*	*0.0401*	**0.3460**	**< 2.10^{-4}**
log(Birth-loc. dist.)	abs(Age difference)	225	within islands only	*0.0990*	*0.1030*	**0.3212**	**< 2.10^{-4}**
abs(Education duration difference)	--	186	within islands only	*0.1287*	*0.0440*	*-0.0370*	*0.3077*
abs(Education duration difference)	log(Birth-loc. dist.)	186	within islands only	*0.1239*	*0.0496*	*-0.0537*	*0.2330*
abs(Education duration difference)	--	185	within islands only	*0.1421*	*0.0292*	*-0.0382*	*0.3037*
abs(Education duration difference)	log(Residence dist.)	185	within islands only	*0.1202*	*0.0546*	*-0.0491*	*0.2566*
log(Residence dist.)	--	224	within islands only	*0.0982*	*0.0911*	*-0.0667*	*0.1907*
log(Residence dist.)	log(Birth-loc. dist.)	224	within islands only	*0.1063*	*0.0704*	*-0.0549*	*0.2319*
log(Birth-loc. dist.)	--	224	within islands only	*0.1537*	*0.0282*	**0.3465**	**< 2.10^{-4}**
log(Birth-loc. dist.)	log(Residence dist.)	224	within islands only	*0.1589*	*0.0230*	**0.3446**	**< 2.10^{-4}**
log(Father Birth-loc. dist.)	--	222	within islands only	*0.0160*	*0.4123*	**0.2660**	**0.0006**
log(Father Birth-loc. dist.)	log(Birth-loc. dist.)	222	within islands only	*-0.0111*	*0.4546*	*0.2187*	*0.0045*
log(Mother Birth-loc. dist.)	--	224	within islands only	*0.1283*	*0.0423*	*0.2240*	*0.0034*
log(Mother Birth-loc. dist.)	log(Birth-loc. dist.)	224	within islands only	*0.1000*	*0.0925*	*0.1563*	*0.0303*

Table T1:
Mantel and partial-Mantel correlations between utterance frequency differences and covariables, and between genetic ASD and the same covariables, in 225 genetically unrelated Cabo Verde-born Kriolu-speaking individuals.
Spearman correlation ρ are indicated in bold when significant after Bonferroni correction for multiple-testing (α<1%), and in italics when not significant after such correction. Spearman correlations and Mantel-tests among covariables are provided in **Table ST14**.

5. Genetic admixture histories in Cabo Verde inferred with *MetHis*-ABC

MetHis–ABC prior checking

We considered four competing genetic-admixture scenarios (**Figure F6**; **Table T2**), tested separately for individuals born on each Cabo Verdean island with *MetHis*–ABC machine-learning scenario-choice and posterior parameter inferences (37,53,54). ABC inferences are based on 42 summary statistics (**Table ST3**) calculated from 60,000 independent autosomal SNPs in individuals form each Cabo Verdean island, and in the African Mandinka and European Iberian-IBS proxy source populations.

We find that the summary-statistics calculated from the observed datasets fall well within the space of summary-statistics obtained from 10,000 simulated-datasets under each of the four competing scenarios (**Figures SF13-22**), considering non-significant (α>5%) goodness-of-fit, visual inspection of summary-statistics PCA-projections, and each summary-statistics' distribution, for each Cabo Verdean birth-island separately. Prior-checks thus demonstrate that *MetHis* simulations are appropriate for further ABC scenario-choice and posterior parameter inferences.

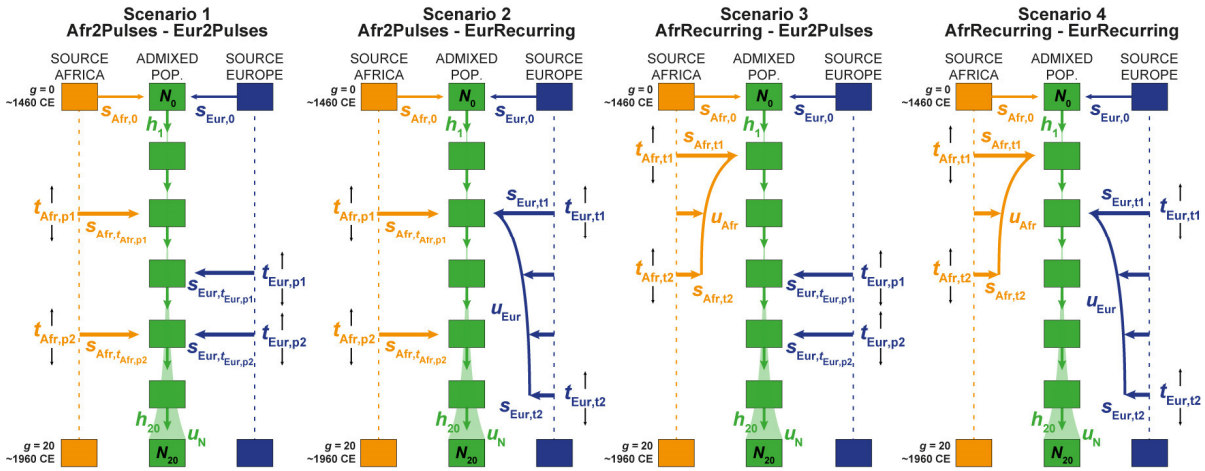


Figure F6:

Four competing scenarios for the genetic admixture histories of each Cabo Verde island.

For all scenarios, the duration of the admixture process is set to 20 generations after the initial founding admixture event occurring at generation 0, which corresponds roughly to the initial peopling of Cabo Verde in the 1460's, considering 25 years per generation and sampled individuals born on average between the 1960s and 1980's.

Scenario 1 Afr2Pulses-Eur2Pulses: after the initial founding pulse of admixture, the admixed population receives two separate introgression pulses from the African and European sources, respectively. ***Scenario 2 Afr2Pulses-EurRecurring***: after the initial founding pulse of admixture, the admixed population receives two separate introgression pulses from the African source, and a period of monotonically constant or decreasing recurring introgression from the European source. ***Scenario 3 AfrRecurring-Eur2Pulses***: after the initial founding pulse of admixture, the admixed population receives a period of monotonically constant or decreasing recurring introgression from the African source, and two separate introgression pulses from the European source. ***Scenario 4 AfrRecurring-EurRecurring***: after the initial founding pulse of admixture, the admixed population receives a period of monotonically constant or decreasing recurring introgression from the African source, and, separately, a period of monotonically constant or decreasing recurring introgression from the European source. For all scenarios, we consider the same demographic models with a possible increase in reproductive population size N_g between the founding event and present. See **Table T2** for parameter prior distributions, and **Materials and Methods** for detailed descriptions of model-parameters.

Description	Scenario	Model Parameter	Prior	Conditions
African admixture-pulse times	1, 2	$t_{Afr,p1}$	Uniform [1..20] in generations, a range corresponding to between ~1485 and ~1960 in years CE	$t_{Afr,p1} \neq t_{Afr,p2}$
		$t_{Afr,p2}$		
European admixture-pulse times	1, 3	$t_{Eur,p1}$	Uniform [1..20] in generations, a range corresponding to between ~1485 and ~1960 in years CE	$t_{Eur,p1} \neq t_{Eur,p2}$
		$t_{Eur,p2}$		
African admixture period start and end times	2, 4	$t_{Afr,t1}$	Uniform [1..20] in generations, a range corresponding to between ~1485 and ~1960 in years CE	$t_{Afr,t1} \neq t_{Afr,t2}$
		$t_{Afr,t2}$		
European admixture period start and end times	3, 4	$t_{Eur,t1}$	Uniform [1..20] in generations, a range corresponding to between ~1485 and ~1960 in years CE	$t_{Eur,t1} \neq t_{Eur,t2}$
		$t_{Eur,t2}$		
African admixture-pulse intensities	1, 2	$s_{Afr,tAfr,p1}$	Uniform [0,1]	$s_{Afr,g} + s_{Eur,g} = 1 - h_g$, with h_g in [0,1]
		$s_{Afr,tAfr,p2}$		
European admixture-pulse intensities	1, 3	$s_{Eur,tEur,p1}$	Uniform [0,1]	
		$s_{Eur,tEur,p2}$		
African admixture period intensity parameters	2, 4	$s_{Afr,tAfr,t1}$	Uniform [0,1]	$s_{Afr,tAfr,t1} \geq s_{Afr,tAfr,t2}$
		$s_{Afr,tAfr,t2}$	Uniform [0,1]	$s_{Afr,g} + s_{Eur,g} = 1 - h_g$, with h_g in [0,1]
		u_{Afr}	Uniform [0 , 0.5]	
European admixture period intensity parameters	3, 4	$s_{Eur,tEur,t1}$	Uniform [0,1]	$s_{Eur,tEur,t1} \geq s_{Eur,tEur,t2}$
		$s_{Eur,tEur,t2}$	Uniform [0,1]	$s_{Afr,g} + s_{Eur,g} = 1 - h_g$, with h_g in [0,1]
		u_{Eur}	Uniform [0 , 0.5]	
Founding reproductive population size	1, 2, 3, 4	N_0	Uniform [10 , 1000]	
Current reproductive population size	1, 2, 3, 4	N_{20}	Uniform [100 , 100000]	$N_0 < N_{20}$
Steepness of the reproductive population size increase	1, 2, 3, 4	u_N	Uniform [0 , 0.5]	

Table T2:
Prior distributions for the parameters of four competing scenarios for the admixture history of Cabo Verde islands.
Parameters are presented in **Figure F6** and described in **Materials and Methods**.

MetHis–Random Forest-ABC scenario-choices

Overall (**Figure F7B**), *MetHis*-RF-ABC scenario-choices indicate that multiple pulses of admixture from the European and African source populations (after the founding admixture pulse, two independent admixture pulses from both Africa and Europe: "Afr2Pulses-Eur2Pulses" scenarios, **Figure F4**), best explain the genetic history of individuals born on six out of nine Cabo Verdean islands. Furthermore, we find that even more complex scenarios involving a period of recurring admixture from either source best explain the history of the remaining three islands. Conversely, scenarios with periods of recurring admixture processes from both Africa and Europe ("AfrRecurring-EurRecurring") are the least favored throughout Cabo Verde.

RF-ABC cross-validation prior-errors for each of the 40,000 simulations used, in-turn, as pseudo-observed data indicate a reasonably good, albeit not perfect, discriminatory power of the RF (**Figure SF23A**). We find a majority of correct scenario assignation for most scenarios and most islands and asymmetrical scenarios are the least confused with one-another (AfrRecurring-Eur2Pulses vs Afr2Pulses-EurRecurring, or Afr2Pulses-Eur2Pulses vs AfrRecurring-EurRecurring). These results show that scenario-nestedness induce, as expected, increased assignation-errors in large parts of the parameter-space where scenarios are, by definition, biologically equivalent (37,55). Finally (**Figure SF23B**), the mean, variance, skewness, kurtosis, minimum, and maximum of individual admixture proportions' distributions are the most informative summary-statistics for RF-ABC scenario-choice in every island, consistently with previous studies (37,56).

MetHis–Neural Network-ABC posterior parameter estimations

For individuals born on each Cabo Verdean separately, we performed NN-ABC joint posterior parameter estimation based on 100,000 *MetHis* simulations under the winning scenario (37). For each island separately, detailed posterior parameters' distributions, Credibility Intervals (CI), and cross-validation errors are provided in **Figure SF24-26** and **Tables ST5-13**. We synthesized our results considering median point-estimates and 50%-CI for each scenario parameter in the admixture history of each island in **Figure F7C-D**. We detailed our results and discussion for admixture history inferences for each island separately in **Supplementary Text,** in the light of historical data about the peopling of Cabo Verde (**Table ST2**).

Figure F7C shows that the reproductive population size of Cabo Verde islands remained very low until a strong increase in the last three generations for all islands but Santiago and Brava. In Santiago the population expansion was more linear since the founding of Cabo Verde in the 1460's, while the reproductive size of Brava remained low until today.

In brief summary, **Figure F7D** shows that European and African admixture events throughout the archipelago occurred first during the early peopling history of each island, before the mid-17th century massive expansion of the TAST due to the expansion of plantation economy (1). We find that other admixture events from Europe or Africa, or both, likely occurred much later during, or immediately after, the 19th century abolition of the TAST and of slavery in European colonial empires. Altogether our *MetHis*-ABC results support limited historical admixture having occurred in Cabo Verde during the most intense periods of the TAST between the mid-17th and early 19th centuries.

Furthermore, note that we find admixture events often earlier than, or concomitant with, the first perennial peopling of an island. For the islands of Santiago, Fogo, Santo Antão and, to a lesser extent, São Nicolau, initial historical admixture events occurred synchronically to the first perennial settlement of the island. For the islands of Brava, Maio, Boa Vista, and São Vicente, early admixture events occurred long before their first perennial peopling, thus showing that their founding was already largely composed of admixed individuals, likely originated from other Cabo Verde islands. Finally, our *MetHis*-NN-ABC posterior parameter inferences cannot infer all scenario-parameters accurately, as some parameters hardly depart from their respective priors (**Figure SF24-26**), and the admixture history of the island of Sal remains overall poorly inferred.

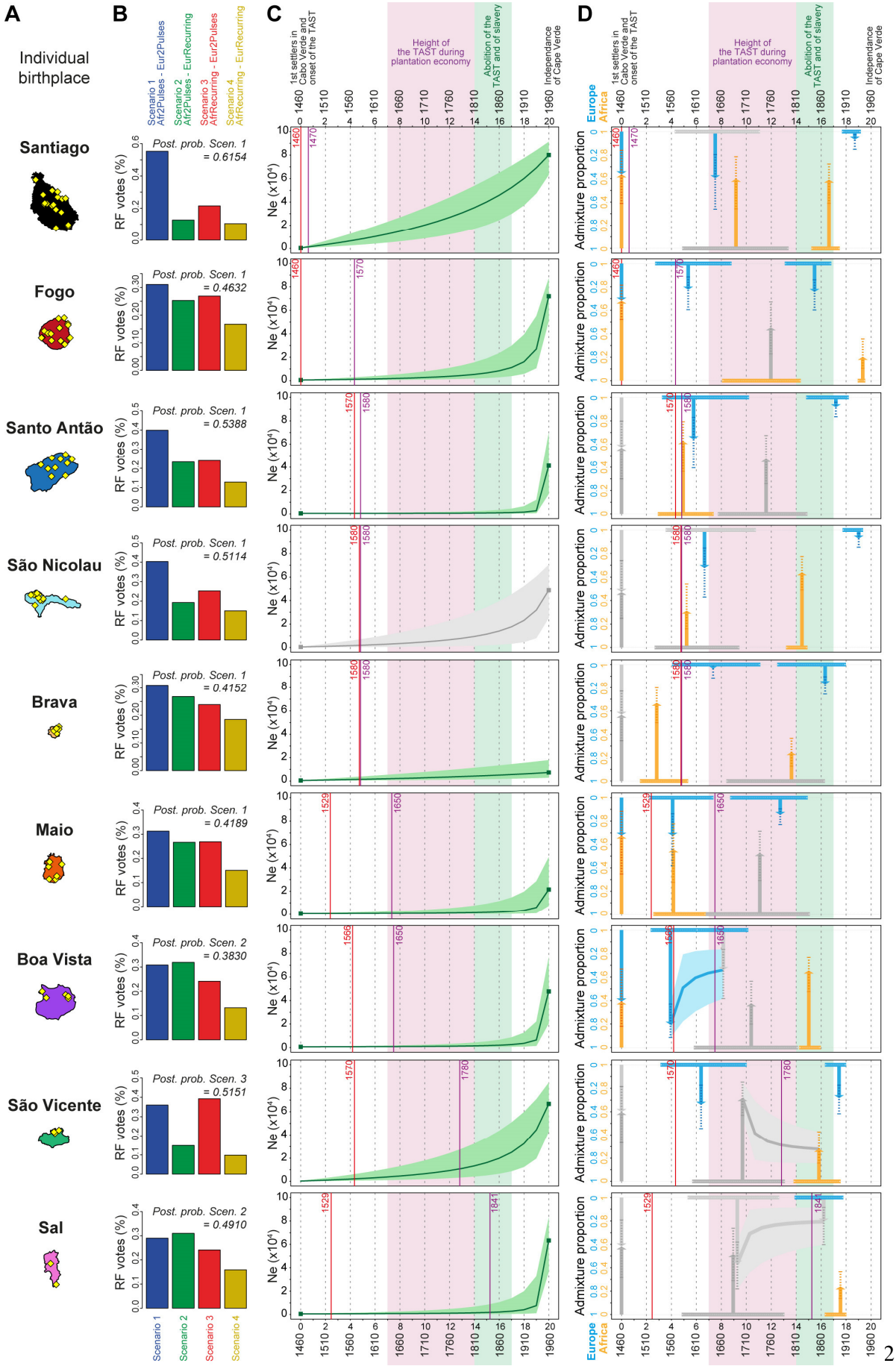

A
Individual birthplace
B
Scenario 1 Afr2Pulses - Eur2Pulses
Scenario 2 Afr2Pulses - EurRecurring
Scenario 3 AfrRecurring - Eur2Pulses
Scenario 4 AfrRecurring - EurRecurring
C
1st settlers in Cabo Verde and onset of the TAST
Height of the TAST during plantation economy
Abolition of the TAST and of slavery
Independance of Cape Verde
D
Europe
Africa
Santiago
Post. prob. Scen. 1 = 0.6154
Fogo
Post. prob. Scen. 1 = 0.4632
Santo Antão
Post. prob. Scen. 1 = 0.5388
São Nicolau
Post. prob. Scen. 1 = 0.5114
Brava
Post. prob. Scen. 1 = 0.4152
Maio
Post. prob. Scen. 1 = 0.4189
Boa Vista
Post. prob. Scen. 2 = 0.3830
São Vicente
Post. prob. Scen. 3 = 0.5151
Sal
Post. prob. Scen. 2 = 0.4910
RF votes (%)
Ne (x10⁴)
Admixture proportion

Figure F7:

Genetic admixture histories of Cabo Verde islands.

Elements of the peopling-history of Cabo Verde islands are synthesized in **Table ST2**, stemming from historical work cited therein. Islands are ordered from top to bottom in the chronological order for the first historical census perennially above 100 individuals within an island, indicated with the purple vertical lines. First historical record of the settlement of an island as reported in political and administrative historical documents, tax-records, and/or parishes establishments, are indicated with the red vertical lines.

A) Within-island birth-places of 225 Cabo-Verde-born individuals. **B)** *MetHis*-Random Forest-ABC scenario-choice vote results for each island separately. Posterior probabilities of correctly identifying a scenario if correct are indicated for the winning scenario. **C)** *MetHis*-Neural Network-ABC posterior parameter distributions with 50% Credibility Intervals for the reproductive population size history of each birth-island separately. **D)** Synthesis of *MetHis*-NN-ABC posterior parameter median point-estimates and associated 50%CI, for the admixture history of each island under the winning scenario identified with RF-ABC. European admixture history appears in blue, African admixture history in orange. Horizontal bars indicate 50% CI for the corresponding admixture time parameters, vertical arrows correspond to median admixture intensity estimates.

For **C)** and **D)**, posterior parameter distributions showing limited departure from their respective priors and large Credibility Intervals are grayed, as they were largely unidentifiable in our ABC procedures. Detailed parameter posterior distributions, 95%CI, and cross-validation errors are provided in **Figure SF24-26** and **Tables ST5-13**. Detailed results description for each island are provided in **Supplementary Text**.

All ABC inference rely on forward-in-time *MetHis* simulations over 21 generations of history. Individuals in our sample set were born on average between the 1960's and the 1980's. We translated *MetHis* generations in Gregorian calendar dates by considering an average generation time of 25 generations. We considered the onset of the massive increase of TAST African-enslaved individuals' deportations concomitant to the expansion of plantation economy in the 1630's (1,2). We considered the abolition of the TAST starting around the 1810's, as the British empire abolition of the TAST in 1807 marked the onset of TAST abolition in all other European empires over the following 30 years (1). We considered the abolition of slavery in the 1885's, as slavery was abolished in Portugal and it's colonial empire over the course of more than 20 years between 1856 and 1878, and further took several years to be fully enacted in all regions of the empire (23). The independence of Cabo Verde occurred in 1975.

514 **DISCUSSION**
515
516 **Which African and European populations contributed genetically to Cabo Verde?**
517 *The genetic heritage of continental Africa in Cabo Verde*
518 Numerous enslaved-African populations from Western, Central, and South-Western Central Africa have been
519 forcibly deported during the TAST to both Cabo Verde and the Americas, as shown by historical demographic
520 records (1,23). There is still extensive debate about whether enslaved-Africans remained or more briefly trans-
521 ited in Cabo Verde during the most intense period of the TAST, in the 18th and 19th centuries, when the archi-
522 pelago served as a slave-trade platform between continents (23,57,58); a question also of major interest
523 throughout the Americas (18,35).
524 In this context, our genetic results favor scenarios where mostly certain Senegambian populations (Man-
525 dinka and Wolof in our study) contribute to the genetic makeup of Cabo Verde (**Figures F2** and **F3**). Other
526 Western, Central, and South-Western African populations historically forcibly deported during the TAST seem
527 to have had only very limited contributions to the genomic diversity of Cabo Verde, and only in certain islands.
528 This could be due to Cabo Verde being only a temporarily waypoint for these latter enslaved-African popula-
529 tions between Africa, the Americas, and Europe, but would also be consistent with additional socio-historical
530 processes (see below). Interestingly, and further echoing these genetic results, the Cabo Verdean Kriolu lan-
531 guage carries specific signatures mainly from the Mande language-family, and Wolof and Temne languages
532 from Western Africa, and largely more limited signatures of Kikongo and Kimbundu Bantu languages from
533 Central and South-Western Africa (15,28).
534 These results contrast with the admixture patterns identified in other enslaved-African descendant popu-
535 lations in the Americas in our dataset (African-American and Barbadian, **Figures F2 and F3**), and in previous
536 studies (3,4,10,36,59). Indeed, the origins of African ancestries in numerous populations throughout the Car-
537 ibbean and the Americas traced to varied continental African regions, from Western to South-Western Africa,
538 thus qualitatively consistent with the known diversity of slave-trade routes used between continents and within
539 the Americas after the Middle Passage.
540
541 *The genetic heritage of continental Europe in Cabo Verde*
542 After the initial settlement of Cabo Verde by Portuguese migrants, temporary changes of European dominion
543 in certain islands, newly developing commercial opportunities, and intense piracy during the 16th and 17th
544 centuries have triggered different European populations to migrate to the archipelago (24,25,34). Nevertheless,
545 these latter historical events do not seem to have left a major signature in the genetic landscape of Cabo Verde
546 (**Figures F2 and F3**). Instead, we find that Cabo Verdean-born individuals in our dataset share virtually all
547 their European genetic ancestry with Iberian populations, with extremely limited evidence of contributions
548 from other European regions, consistent with previous studies (3,21). Interestingly, the reduced diversity of
549 European populations' contributions to the genomic landscape of Cabo Verde is also identified in other en-
550 slaved-African descendant populations in our study, as well as in previous studies in Caribbean and American
551 populations (11,12,36). Our results thus show that European admixture in enslaved-African descendant popu-
552 lations on both sides of the Atlantic mainly stem from the gene-pool of the European empires historically and
553 socio-economically dominant locally, rather than subsequent European migrations (2).
554
555 **Genetic and linguistic isolation-by-distance and difficult demographic expansion in Cabo Verde**
556 A scenario of island peopling via series of founding events followed by slow-growing population sizes and
557 isolation would consistently explain the increasing differentiation of island-specific patterns with increasing
558 values of *K* found with ADMIXTURE, ASD-MDS and isolation-by-distance results (**Figures F2-F5**, **Table**
559 **T1**), as well as *MetHis*-ABC demographic inferences (**Figure F7C**).

Indeed, *MetHis*-ABC results (**Figure F7C**) show the long period of small reproductive sizes until very recent increases in most Cabo Verdean islands; with the notable exception of Santiago, the political and commercial center of Cabo Verde throughout the colonial history of the archipelago. This result was expected as the dry Sahelian climate of Cabo Verde with scarcely accessible water resources, recurring famines and epidemic outbreaks, and the Portuguese crown maintaining a strong control over population movements within Cabo Verde, rendered difficult the perennial peopling of most islands (see **Table ST2** and references therein, **Supplementary Text**).

Furthermore, such demographic scenario is also consistent with long-ROH patterns reflecting isolation on each Cabo Verdean islands, whereas elevated shorter ancestry-specific ROH patterns likely stemmed from admixture (**Figure F4**), similarly to our results in ASW and ACB populations and as previously identified (50). Interestingly, we also found substantial signals of isolation-by-distance among Kriolu idiolects (i.e. individual manners of uttering Cabo Verdean Kriolu), also anchored in individual birth-places (**Figure F5**, **Table T1**), thus showing striking parallels between the history of biological and cultural isolation in Cabo Verde at a micro-geographical scale.

Importantly, we considered only random matting processes in our inferences and interpretations, albeit our ROH results together with recent work (60), may suggest that non-random matting processes significantly influenced Cabo Verdean genetic patterns. Therefore, future studies will need to evaluate how possible deviations from random-mating in Cabo Verde, such as assortative mating (61), may have influenced our results and interpretations. Note that this is a conceptually particularly challenging task in a small census-size population with strong marital stratification where mate-choices have been, by definition, limited during most of the TAST (62). Nevertheless, such complex processes may also underlie the correlations found here between linguistic and genetic admixture patterns, and inter-island differences in both genetic and linguistic admixture levels (**Figure F5**); and would explain that genetic and linguistic histories of admixture apparently occurred in parallel in Cabo Verde.

Histories of genetic admixture in Cabo Verde

Early admixture in Cabo Verde, limited admixture during the plantation economy era

While we expected recurring African admixture processes due to the known history of regular forced migrations from Africa during most of Cabo Verde history (23), our *MetHis*–ABC scenario-choices indicate that, qualitatively, these demographic migrations did not necessarily translate into clearly recurring gene-flow processes to shape genetic patterns in most of Cabo Verde islands (**Figure F7B**). Indeed, African admixture processes in all islands, except São Vicente, seem to have occurred during much more punctual periods of Cabo Verdean history. Our *MetHis*–ABC posterior parameter estimations further highlighted often largely differing admixture histories across Cabo Verde islands (**Figure F7** and **Supplementary Text**).

We find that admixture from continental Europe and Africa occurred first early during the TAST history, concomitantly with the successive settlement of each Cabo Verdean island between the 15th and the early 17th centuries (**Figure F7D**). Furthermore, we find that the most intense period of enslaved-African deportations during the TAST via Cabo Verde, between the mid-17th and early-19th centuries during the expansion of the plantation economic system in the Americas and Africa (1,23–25), seem to have left a limited signature in the admixture patterns of most Cabo Verdean islands today. Interestingly, previous studies also highlighted that African admixture in enslaved-African descendants in North America and the Caribbean may have been limited during the height of the plantation economy era (3,12). Therefore, our results illustrate the apparent discrepancy between intense demographic forced migrations during the TAST and genetic admixture signatures on both sides of the Atlantic.

Historians reported, in Cabo Verde and other European colonies in the Americas, that relationships between enslaved and non-enslaved communities largely changed with the expansion of plantation economy at the end of the 17th century. These changes are often referred to as the shift from Societies-with-Slaves to Slave-

Societies (1,19,63). Slave-Societies legally enforced the socio-marital and economic segregation between communities, and coercively controlled relationships between novel enslaved-African migrants and pre-existing enslaved-African descendant communities more systematically and violently than Societies-with-Slaves (19)[-p.15-46,95-108,](23)[-p.281-319]. The high prevalence of segregation during the height of plantation economy could have limited genetic admixture between enslaved-African descendants and non-enslaved communities of European origin, as well as admixture between new migrants, forced or voluntary, and pre-existing Cabo Verdeans, notwithstanding the known history of dramatic sexual abuses during and before this era. This could consistently explain our observations of a relative lack of diverse African or European origins in Cabo Verdean genomes despite the known geographical diversity of populations deported and emigrated to the archipelago throughout the TAST. Furthermore, with legally enforced segregation, we might expect more marital pairings than before to occur among individuals with common origins; i.e. between two individuals with the same continental African or European origin. Such ancestry-specific marriages triggered by socio-cultural segregation would be consistent with our ROH patterns (**Figure F4**).

In this context, the diversified African ancestries here found in the Americas (**Figure F2** and **F3**), consistently with previous studies also showing relatively limited admixture events during the height of plantation economy in the Americas (3,4,12), would suggest that the gene-pool of enslaved-Africans communities admixing with Europeans in the Americas during the 16th and early 17th centuries already often involved, at a local scale, multiple African source populations, thus reflecting the multiple slave-trade routes between continents. Conversely in Cabo Verde, the early onset of the TAST during the 15th century likely privileged commercial routes with nearby Senegambia (23)[-p.31-54,281-319], thus favoring almost exclusively admixture events with individuals from this region. Altogether, our results in Cabo Verde contrasting with other enslaved-African descendant populations in the Americas, highlight the importance of early admixture processes and socio-cultural constraints changes on intermarriages throughout the TAST, which likely durably influenced genomic diversities in descendant populations locally, on both sides of the Atlantic.

Admixture in Cabo Verde after the abolition of the TAST and of Slavery

Finally, we find that recent European and African admixture in Cabo Verde occurred mainly during the complex historical transition after the abolition of the TAST in European colonial empires in the 1800's and the subsequent abolition of slavery between the 1830's and the 1870's (**Figure F7D**). Historians showed that these major historical shifts strongly disrupted pre-existing segregation systems between enslaved and non-enslaved communities (1)[-p.271-290,](23)[-p.335-362,](64). In addition, illegal slave-trade has flourished during this era, bringing numerous enslaved-Africans to Cabo Verde outside of the official routes (23)[-p.363-384]. Altogether, our results indicate increased signals of European and African admixture events in almost every island of Cabo Verde during this period, and were thus consistent with a change of the social constraints regarding admixture having prevailed over the preceding 200 years of the TAST. These results were largely consistent with previous studies elsewhere in the Caribbean (11,65), and Central and South America (4,36); showing, at a local scale, the major influence of this recent and global socio-historical change in inter-community relationships in European colonial empires on either sides of the ocean.

Perspectives

Altogether, our results highlight both the unity and diversity of the genetic peopling and admixture histories of Cabo Verde islands, the first colony peopled by Europeans in Sub-Saharan Africa, resulting from the socio-cultural and historical complexity of the Trans-Atlantic Slave Trade and European colonial expansions since the 15th century. Our results obtained at a micro-geographical scale reveal the fundamental importance of the early TAST history, before the expansion of plantation economy, in durably shaping the genomic and cultural diversities of enslaved-African descendant populations in both Africa and the Americas.

Importantly, we considered only the genome-wide autosomal admixture history of Cabo Verde in this study, and therefore did not explore possible sex-biased admixture processes. However, previous studies demonstrated the strong sex-biased admixture processes involved in Cabo Verde using sex-specific genetic markers (30,31), similarly as in most enslaved-African descendant populations in the Americas (2). Furthermore, previous theoretical work considering sex-specific mechanistic admixture models showed that sex-biased admixture processes may possibly bias historical inferences based only on autosomal data (66). It will thus be imperative, in future studies, to explore how this sex-biased admixture history may have influenced the ABC inferences here conducted; for instance, via sex-specific developments of *MetHis*-ABC.

Our novel results together with their methodological limitations massively beg for future work further complexifying the genetic admixture models here considered, as well as incorporating other summary-statistics such as admixture-LD and sex-specific statistics. This will allow to further dissect the admixture processes having given birth to enslaved-African descendant populations, on both sides of the Atlantic.

667 **MATERIALS AND METHODS**
668
669 **1. Cabo Verde genetic and linguistic datasets**
670 We conducted joint sampling of anthropological, genetic, and linguistic data in Cabo Verde with the only
671 inclusion criteria that volunteer-participants would be healthy adults with Cabo Verdean citizenship and self-
672 reporting speaking Kriolu (21). Between 2010 and 2018, six interdisciplinary sampling-trips were conducted
673 to interview 261 participants from more than thirty interview-locations throughout the archipelago (**Figure**
674 **F1**).
675
676 **1.a. Familial anthropology and geography data**
677 The 261 Cabo Verdean individuals were each interviewed to record a classical familial anthropology question-
678 naire informing self-reported life-history mobility. In particular, we recorded primary residence location, birth
679 location, parental and grand-parental birth and residence locations, and history of islands visited in Cabo Verde
680 and foreign experiences (21). Furthermore, we also recorded age, sex, marital status, and cumulated years of
681 schooling and higher education (for 211 individuals only), and languages known and their contexts of use.
682 GPS coordinates for each reported location was acquired on site during field-work, supplemented by pa-
683 per maps and Google Earth™ for foreign locations and islands that we did not visit (Sal and Maio). While
684 participants' birth-locations were often precise, increasing levels of uncertainty arose for the reported parental
685 and grand-parental locations. We arbitrarily assigned the GPS coordinates of the main population center of an
686 island when only the island of birth or residence could be assessed with some certainty by participants. All
687 other uncertain locations where recorded as missing data. Figure maps were designed with the software *Quan-*
688 *tumGIS* v3.10 "București" and using Natural Earth free vector and raster map data (https://www.natura-
689 learthdata.com).
690
691 **1.b. Genome-wide genotyping data**
692 The 261 participants each provided 2mL saliva samples collected with DNAGenotek™ OG-250 kits, and com-
693 plete DNA was extracted following manufacturer's recommendation. DNA samples were genotyped using an
694 Illumina HumanOmni2.5Million-BeadChip genotyping array following manufacturer's instructions. We fol-
695 lowed the quality-control procedures for genotypic data curation using Illumina's GenomeStudio® Genotyping
696 Module v2.0 described in (21). Genotype-calling and population-level quality-controls, and merging proce-
697 dures are detailed in **Figure SF1**.
698 In summary, we extracted a preliminary dataset of 259 Cabo Verdean Kriolu-speaking individuals, in-
699 cluding relatives, genotyped at 2,118,835 polymorphic di-allelic autosomal SNPs genome-wide. We then
700 merged this dataset with 2,504 worldwide samples from the 1000 Genomes Project Phase 3 (67); with 1,307
701 continental African samples from the African Genome Variation Project (68) (EGA accession number
702 EGAD00001000959); and with 1,235 African samples (7) (EGA accession number EGAS00001002078). We
703 retained only autosomal SNPs common to all data sets, and excluded one individual for each pair of individuals
704 related at the 2nd degree (at least one grand-parent in common) as inferred with KING (69), following previous
705 procedure (21).
706 After merging all datasets (**Figure SF1**), we considered a final working-dataset of 5,157 worldwide un-
707 related samples, including 233 unrelated Cabo Verdean Kriolu-speaking individuals of which 225 were Cabo-
708 Verde-born, genotyped at 455,705 autosomal bi-allelic SNPs (**Figure F1**; **Table ST1**).
709
710 **1.c. Individual utterances of Kriolu**
711 We collected linguistic data for each of the 261 Cabo Verdean individuals using anthropological linguistics
712 questionnaires, and semi-directed interviews. Each participant was shown a brief (~6 min) speech-less movie,
713 *"The Pear Story"* (70), after which they were asked to narrate the story as they wanted in "the Kriolu they

speak every day at home". Discourses were fully recorded without interruption, whether individuals' discourses were related to the movie or not. Then each discourse was separately fully transcribed using the orthographic convention of Cabo Verdean Kriolu official alphabet "Alfabeto Unificado para a Escrita da Língua Cabo-verdiana (ALUPEC)" (Decreto-Lei nº 67/98, 31 de Dezembro 1998, I Série nº 48, Sup. B. O. da República de Cabo Verde).

Building on Verdu *et al.* (2017) approach, we were interested in inter-individual variation of "ways of speaking Kriolu" rather than in a prescriptivist approach to the Kriolu language. Thus, we considered each utterance as defined in (71)[-p.107]: "a particular instance of actually-occurring language as it is pronounced, grammatically structured, and semantically interpreted in its context". Transcripts were parsed together and revealed 4831 (L=4831) unique uttered items among the 225 discourses from the genetically-unrelated Cabo Verde-born individuals, for 92,432 uttered items in total. To obtain these counts, we considered phonetic, morphological, and syntactic variation of the same lexical root items that were uttered/pronounced differently, and we excluded from the utterance-counts onomatopoeia, interjections, and names. Note that we were here interested in the diversity of realizations in the Kriolu lexicon, including within the same individuals. In other words, we are interested in both between-speaker and within-speaker variation. Also note that a very limited number of English words were pronounced by particular individuals (10 utterances each occurring only once), and were kept in utterance-counts.

1.d. Individual Kriolu utterance frequencies

The list of unique uttered items was then used to compute individual's specific vectors of uttered items' relative frequencies as, for each genetically unrelated Cabo Verde-born individual i (in [1, 225]) and each unique uttered item l (in [1, L=4831]), $f_{i,l} = n_{i,l}/\sum_{j=1}^{L} n_{i,j}$, where $n_{i,l}$ is the absolute number of times individual i uttered the unique item l over her/his entire discourse, f_i being the vector $(f_{i,1}, f_{i,2}, \dots, f_{i,L})$. We compared vectors of individuals' utterance-frequencies by computing the inter-individual pairwise Euclidean distance matrix as, for all pairs of individuals i and j, $d(f_i, f_j) = \sqrt{\sum_{l=1}^{L}(f_{i,l} - f_{j,l})^2}$ (21).

1.e. African origin of Kriolu utterances

We categorized each of the 4831 unique uttered items separately into five linguistic categories (21). Category A included only unique utterances directly tracing to a known African language and comprised 88 unique items occurring 3803 times out of the 92,432 utterances. Category B included only items with a dual African and European etymology, attesting to the intense linguistic contacts at the origins of Cabo Verdean Kriolu, and comprised 254 unique items occurring 6960 times. Category C included 4432 items (occurring 73,799 times) with strictly Portuguese origin and not carrying identifiable traces of significant African linguistic origin or influence. Category D included 26 items occurring 6762 times with possible traces of African languages' phonetic or morphologic influences, albeit without certainty. Finally, Category U included the 10 English utterances occurring 10 times and the 21 unique utterances occurring 1089 times which could not be traced to African or European languages.

Following (21), we defined an "African-utterances score" based, conservatively, on the utterance frequencies obtained for items in Categories A and/or B as, for individual i and the set of utterances in each category denoted Cat (in [A; B; A&B]), $Z_{i,Cat} = \sum_{l=1}^{L_{Cat}} f_{i,l}$, with L_{Cat} the number of uttered items in the corresponding category, and $f_{i,l}$ defined as previously.

2. Population genetics descriptions

2.a. Allele Sharing Dissimilarities, Multidimensional Scaling, and ASD-MDS admixture estimates

759 We calculated pairwise Allele Sharing Dissimilarities (38) using the *asd* software (v1.1.0a;
760 https://github.com/szpiech/asd), among the 5,157 individuals in our worldwide dataset (**Figure F1**; **Table**
761 **ST1**), using 455,705 autosomal SNPs, considering, for a given pair of individuals, only the SNPs with no
762 missing data. We then projected this dissimilarity matrix in three dimensions with metric Multidimensional
763 Scaling using the *cmdscale* function in R (72). We conducted successive MDS analyses on different individual
764 subsets of the original ASD matrix, by excluding, in turn, groups of individuals and recomputing each MDS
765 separately. Lists of populations included in each analysis can be found in **Table ST1**. 3D MDS animated plots
766 in *gif* format (**Figures SF2-6**) were obtained with the *plot3d* and *movie3d* functions from the R packages *rgl*
767 and *magick*.
768 Recently admixed individuals are intuitively expected to be at intermediate distances between the clusters
769 formed by their putative proxy source populations on ASD-MDS two-dimensional plots. A putative estimate
770 of individual admixture rates can then be obtained by projecting admixed individuals orthogonally on the line
771 joining the respective centroids of each proxy-source populations and, then, calculating the distance between
772 the projected points and either centroid, scaled by the distance between the two centroids (**Figure SF12**;
773 (21,37)). We estimated such putative individual admixture rates in Cabo Verdean, ASW, and ACB individuals,
774 considering sets of individuals for the putative proxy-source populations identified visually and resulting from
775 our ASD-MDS decomposition (**Supplementary Text**).
776
777 **2.b. ADMIXTURE-CLUMPP-DISTRUCT analyses**
778 Based on ASD-MDS explorations, we focused on the genetic structure of individuals born in Cabo Verde
779 compared to that of other admixed populations related to TAST migrations. Therefore, we conducted ADMIX-
780 TURE analyses (39) using 1,100 individuals from 22 populations: four from Europe, 14 from Africa, the USA-
781 CEU, the African-American ASW, the Barbadian-ACB populations, and the North Chinese-CHB population
782 as an outgroup (**Table ST1**). To limit clustering biases due to uneven sampling, we randomly resampled 50
783 individuals, once, for each one of these 22 populations. Furthermore, we also included all 44 Angolan individ-
784 uals from four populations in the analyses, as no other samples from the region were available in our dataset.
785 In addition to the 1,100 individuals hence obtained, we included the 225 Cabo Verde-born individuals.
786 Following constructor recommendations, we pruned the initial 455,705 SNPs set for low LD using *plink*
787 (73) function *--indep-pairwise* for a 50 SNPs-window moving in increments of 10 SNPs and a r^2 threshold of
788 0.1. We thus conducted all subsequent ADMIXTURE analyses considering 1,369 individuals genotyped at
789 102,543 independent autosomal SNPs.
790 We performed 30 independent runs of ADMIXTURE for values of *K* ranging from 2 to 10. For each value
791 of *K* separately, we identified groups of runs providing highly similar results (ADMIXTURE "modes"), with
792 Symmetric Similarity Coefficient strictly above 99.9% for all pairs of runs within a mode, using CLUMPP
793 (74). We plotted each modal result comprising two ADMIXTURE runs or more, for each value of *K* separately,
794 using DISTRUCT (75). Finally, we evaluated within-population variance of individual membership propor-
795 tions as F_{ST}/F_{ST}^{max} values using FSTruct (40) with 1000 Bootstrap replicates per population, for the ADMIX-
796 TURE mode result at $K=2$ (**Figure SF7**).
797
798 **2.c. Local-ancestry inferences**
799 To identify all populations sharing a likely common ancestry with the Cabo Verdean, ASW or ACB individuals
800 in our dataset using local-ancestry haplotypic inferences, we considered the same sample-set as for the AD-
801 MIXTURE analysis (**Table ST1**), including all 455,705 SNPs from the merged dataset.
802
803 *Phasing with ShapeIT2*

804 We first phased individual genotypes using SHAPEIT2 (46) for the 22 autosomal chromosomes separately
805 using the joint Hap Map Phase 3 Build GRCh38 genetic recombination map (76). We considered default pa-
806 rameters using phasing windows of 2Mb and 100 states per SNP. We considered by default 7 burn-in iterations,
807 8 pruning iterations, 20 main iterations, and missing SNPs were imputed as monomorphic. Finally, we consid-
808 ered a "Ne" parameter of 30,000, and all individuals were considered unrelated.
809
810 *Chromosome painting with ChromoPainter2*
811 We determined the possible origins of each Cabo Verdean, ASW, and ACB individual phased haplotypes
812 among European, African, USA-CEU, and Chinese-CHB populations using CHROMOPAINTER v2 (45) with
813 the same recombination map used for phasing. Following authors' recommendations, we conducted a first set
814 of 10 replicated analyses on a random subset of 10% of the individuals for chromosomes 2, 5, 12, and 19,
815 which provided a posteriori Ne=233.933 and θ=0.0004755376, on average by chromosome weighted by chro-
816 mosome sizes, to be used in the subsequent analysis. Then, we conducted a full CHROMOPAINTER analysis
817 using these parameters to paint all individuals in our dataset, in turn set as Donor and Recipient. Finally, we
818 combined painted chromosomes for each individual in the Cabo Verdean, ASW, and ACB population, sepa-
819 rately.
820
821 *Estimating possible source populations for the Cabo Verde gene-pool using SOURCEFIND*
822 We used CHROMOPAINTER results aggregated for each Cabo Verdean, ACB, and ASW individual sepa-
823 rately and conducted two SOURCEFIND (47) analyses using all other populations in the dataset as a possible
824 source, separately for four or six possible source populations ("surrogates"). We considered 400,000 MCMC
825 iteration steps (including 100,000 burn-in) and kept only one MCMC step every 10,000 steps for final likeli-
826 hood estimation. Each individual genome was divided in 100 slots with possibly different ancestry, with an
827 expected number of surrogates equal to 0.5 times the number of surrogates, for each SOURCEFIND analysis.
828 We aggregated results obtained for all individuals in the ACB, ASW, and each Cabo Verdean birth-island,
829 separately. We present the highest likelihood results across 20 separate iterations of the SOURCEFIND anal-
830 ysis in **Figure F3B.** The second-best results were highly consistent and thus not shown.
831
832 **3. Genetic and linguistic diversities, and admixture within Cabo Verde**
833
834 **3.a. Runs of Homozygosity (ROH)**
835 Considering the same sample and SNP set as in the above local-ancestry analyses (**Table ST1**), we called ROH
836 with GARLIC (77). For each population separately, we ran GARLIC using the weighted logarithm of the odds
837 (wLOD) score-computation scheme, with a genotyping-error rate of $1e^{-3}$ (a likely overestimate), and using the
838 same recombination map as for phasing, window sizes ranging from 30 to 90 SNPs in increment of 10 SNPs,
839 100 resampling to estimate allele frequencies, and all other GARLIC parameters set to default values. We only
840 considered results obtained with a window size of 40 SNPs, which was the largest windows size associated
841 with a bimodal wLOD score distribution and a wLOD score cutoff between the two modes, for all populations.
842 For each populations and Cabo Verdean islands separately, we considered three classes of ROH that
843 correspond to the approximate time to the most recent common ancestor of the IBD haplotypes, which can be
844 estimated from the equation g=100/2L, where L is the ROH length in cM and g is the number of generations
845 to the most recent common ancestor of the haplotypes (48). Short ROH are less than 0.25cM reflecting
846 homozygosity of haplotypes from more than 200 generations ago; medium ROH are between 0.25cM and 1cM
847 reflecting demographic events having occurred between approximately 200 and 50 generations ago; and long
848 ROH are longer than 1cM reflecting haplotypes with a recent common ancestor less than 50 generations ago.
849 In **Figure F4A-B** and **Figure SF9**, we plot the distribution of the summed length of ROH per individual per
850 size-classes.

851
852 **3.b. Intersecting ROH and local ancestry painting**
853 Using the same phasing results as described above, we conducted 10 EM iterations of the RFMIX2 (78) algo-
854 rithm to assign, for each Cabo Verdean individual and each SNP on each chromosome, separately, it's putative
855 source population of origin among the 24 African, European, Chinese-CHB, and USA-CEU populations. We
856 collapsed the local ancestry assignments for each SNP in each Cabo Verdean individuals hence obtained into
857 three continental regions, representing broadly African, broadly European, and broadly Asian ancestries re-
858 spectively. We considered an approach similar to previous work (50), and intersected local ancestry calls with
859 ROH calls (**Figure 4C-D**).
860 RFMIX2 only makes local ancestry calls at loci that are present in the dataset. Therefore, a gap of unclas-
861 sified ancestry exists where inferred ancestry switches between two successive genotyped loci as called by
862 RFMIX2. These gaps necessarily each contain an odd number of ancestry switch points (≥1) absent from our
863 marker set. Therefore, when computing the total ancestry content of an ROH that overlaps one of these ancestry
864 breakpoints, we assign half of the length of this gap to each ancestry classification, effectively extending each
865 local ancestry segment to meet at the midpoint.
866 We then plotted the length distribution of long ROH for each island and we breakout the distributions by
867 ancestral haplotype background: those with only African ancestry, those with only European ancestry, and
868 those that span at least one ancestry breakpoint (**Figure F4C**). We excluded long ROH in Asian ancestry
869 segments from this and the following analyses, as we found overall very limited such ancestry in the samples
870 (**Figure SF10**).
871 Finally, we explored how total African/European ancestry in long ROH varies between islands. For each
872 individual we summed the total amount of each ancestry in long-ROH and plot the distributions across islands
873 (**Figure SF11**). High levels of a given ancestry in long ROH could stem from an overall high level of that
874 ancestry in that individual. Therefore, for each individual, we computed their global ancestry proportions by
875 summing up the total length of the genome inferred as a given ancestry and dividing by the length of the
876 genome. We then plotted (**Figure F4D**), the difference of an individual's long-ROH ancestry proportion and
877 their global ancestry proportion, for African and European ancestries separately. Values above zero indicated
878 that a given ancestry is overrepresented in long-ROH relative to genome-wide proportions of that ancestry.
879
880 **3.c. Genetic and linguistic diversity within Cabo Verde**
881 We explored genetic pairwise levels of differentiation calculated with ASD as previously, considering the
882 1,899,878 non-monomorphic SNPs within Cabo Verde obtained after QC Stage 3 (**Figure SF1**). We projected
883 the matrix for the 225 unrelated Cabo Verde-born individuals using metric MDS as above. Note that pruning
884 this data sets for very low LD using *plink* function *--indep-pairwise* for a 50 SNPs window moving in incre-
885 ments of 10 SNPs and a r^2 threshold of 0.025 resulted in 85,425 SNPs for which the ASD matrix was highly
886 correlated with the one using all SNPs (Spearman ρ=0.8745, $p<2.2\text{x}10^{-16}$). In parallel, we described the diver-
887 sity of Kriolu idiolects ("individuals' manners of speaking Kriolu") among the 225 genetically unrelated Cabo
888 Verde-born individuals by projecting, with metric MDS, the matrix of pairwise Euclidean distances between
889 vectors of individuals' utterance-frequencies (**Material and Methods 1.d**), considering the 4831 unique ut-
890 tered items.
891 Then, we conducted series of Mantel and partial Mantel correlation tests (Legendre and Legendre 1998),
892 using the *partial.mantel.test* function in the R package *ncf*, with Spearman correlation and 10,000 Mantel per-
893 mutations, to explore possible isolation-by-distance (51) patterns as well as correlations with other co-variables
894 of interest. We conducted correlation tests between either the pairwise Euclidean distances between vectors of
895 individuals' utterance-frequencies or genetic ASD separately, and individual pairwise matrices of *i)* absolute
896 age differences, *ii)* absolute differences in academic education duration, *iii)* geographic distances between
897 residence-locations (logarithmic scale), *iv)* between birth-locations, *v)* between mothers' birth-locations, *vi)*

between fathers' birth-locations (**Table T1**, **Table ST14**). All pairwise geographic distances were calculated with the Haversine great circle distance formulation (79), taking 6,371 km for the radius of the Earth, using the *rdist.earth* function in the R package *fields*. Before computing logarithmic distances, we added 1km to all pairwise distances.

3.d. Genetic and linguistic admixture within Cabo Verde

We further explored isolation-by-distance patterns within Cabo Verde specifically for African genetic and linguistic individual admixture levels. We considered African genetic admixture levels estimated from AD-MIXTURE at K=2 or ASD-MDS approaches (**Material and Methods 2.a** and **Supplementary Text**), and individual African linguistic admixture as "African-utterances scores" $Z_{i,Cat}$ as defined in **Material and Methods 1.e.** for utterance lists contained in Categories A and/or B (21). For genetic or linguistic admixture levels separately, we computed the pairwise matrix of individual absolute admixture levels differences, and conducted Mantel testing with the different geographical distance matrices as above. Finally, we compared African genetic and linguistic admixture scores using Spearman correlations, throughout Cabo Verde and within all birth-islands, separately.

4. Inferring genetic admixture histories in Cabo Verde with *MetHis*-ABC

We aimed at reconstructing the detailed genetic admixture history of each Cabo Verde island separately. To do so, we first design *MetHis* v1.0 (37) forward-in-time simulations of four competing complex admixture models. We then couple *MetHis* simulation and summary-statistics calculation tools with ABC Random-Forest scenario-choice implemented in the R package *abcrf* (54), followed by Neural-Network posterior parameter estimation with the R package *abc* (53), under the winning model for each island separately.

4.a. Four competing models of complex historical admixture for Cabo Verde Islands

Following our descriptive analyses results, we considered models with only one "European" and one "African" source population and each Cabo Verde island, separately, as the "Admixed" recipient population. This corresponds to the "two-source" version of the (56) general model also explored with *MetHis* in (37).

Foundation of the admixed population and duration of the admixture history

In all four scenarios (**Figure F6**; **Table T2**), we considered, forward-in-time, that the admixed population (Population H) is founded at generation 0, 21 generations before present, which roughly corresponds to the 15th century for an average generation time of 25 years and sampled individuals born before the 1990s and on average between the 1960's and the 1980's in our dataset. This is reasonable as historical records showed that Cabo Verde was likely un-inhabited before its initial colonial settlement established in the 1460s on Santiago (Albuquerque and Madeira Santos 1991). At generation 0, the admixed population is founded with a proportion $s_{\mathrm{Eur},0}$ of admixed individuals' parents originating from the European source and a proportion $s_{\mathrm{Afr},0}$ from the African source, with $s_{\mathrm{Eur},0} + s_{\mathrm{Afr},0}=1$. Note that simulations considering parameter-values of $s_{\mathrm{Afr},0}$, or $s_{\mathrm{Eur},0}$, close to 0 correspond to scenarios where the "admixed" population is initially founded by only one of the sources, and that genetic admixture *per se* may only occur at the following admixture event.

After the founding of Population H, we considered two different admixture scenarios for either source populations' contribution to the gene-pool of population H.

Two admixture pulses after foundation

For a given source population hereafter designated "Source" ("European" or "African" in our case), after founding at generation 0, we considered scenarios with two additional pulses of admixture ("Source"-2Pulses models). The two pulses occur, respectively, at generation $t_{\mathrm{Source,p1}}$ and $t_{\mathrm{Source,p2}}$ in [1,20], with $t_{\mathrm{Source,p1}} \neq t_{\mathrm{Source,p2}}$; and with intensity $s_{\mathrm{Source,tSource,p1}}$ and $s_{\mathrm{Source,tSource,p2}}$ in [0,1], respectively.

Note that simulations considering $t_{Source,p1}=t_{Source,p2}+1$, and simulations with either $s_{Source,tSource,p1}$ or $s_{Source,tSource,p2}$ close to 0, may strongly resemble those expected under scenarios with only one pulse of admixture after the founding of Population H.

A period of recurring admixture after foundation

For a given Source population, after founding at generation 0, we considered scenarios with a possible period of recurring admixture, where, during this period, admixture intensity followed a monotonically decreasing trend ("Source"-Recurring models). The period of admixture occurs between times $t_{Source,t1}$ and $t_{Source,t2}$ in [1,20] with $t_{Source,t2} \geq t_{Source,t1}+1$. The beginning of the admixture period at $t_{Source,t1}$ is associated with admixture intensity $s_{Source,tSource,t1}$ in [0,1]. The end of the admixture period at $t_{Source,t2}$ is associated with intensity $s_{Source,tSource,t2}$ in [0,1] such that $s_{Source,tSource,t1} \geq s_{Source,tSource,t2}$. In between, the admixture intensity values at each generation of the admixture period are the discrete numerical solutions of a decreasing rectangular hyperbola function of parameter u_{Source} in [0,1/2] (37).

Intuitively, a u parameter value close to 0 corresponds to a sharp pulse of admixture occurring at the beginning of the admixture period of intensity $s_{Source,tSource,t1}$, followed at the next generation by constant recurring admixture of intensity $s_{Source,tSource,t2}$ until the end of the admixture period. Alternatively, a u parameter value close to ½ corresponds to a linearly decreasing admixture at each generation of the admixture period from $s_{Source,tSource,t1}$ to $s_{Source,tSource,t2}$.

Note that in the limit when $s_{Source,tSource,t1} \sim s_{Source,tSource,t2}$, Recurring scenarios correspond to constant recurring admixture of that intensity. Furthermore, simulations with u and $s_{Source,tSource,t2}$ parameter values both close to 0 correspond to scenarios with a single pulse of admixture from a given source after the founding pulse, occurring at time $t_{Source,t1}$ and with intensity $s_{Source,tSource,t1}$.

We combined the 2Pulses and Recurring scenarios from either the African and European source populations in order to produce four competing scenarios of admixture for the genetic history of Cabo Verde (**Figure F6**), with the only constraint that at each generation g between 1 and 20, $s_{Afr,g}+s_{Eur,g}=1-h_g$, where h_g is the contribution of the admixed population H to itself at the following generation in [0,1].

Reproductive population size in the admixed population

We considered, for the admixed population H, a reproductive population size of N_0 diploid individuals at generation 0, with N_0 in [10,1000], and N_{20} in [100,100000] at generation 20, such that $N_0 < N_{20}$. In between these two values, we considered for N_g values the discrete numerical solution of an increasing rectangular hyperbola function of parameter u_N in [0,1/2] (37).

4.b. *MetHis*-ABC Random Forest scenario-choice and Neural Network posterior parameter estimations

To reconstruct highly complex admixture histories in Cabo Verde using genetic data, we conducted machine-learning Approximate Bayesian Computation inferences. They rely on simulations conducted with model-parameter values drawn randomly within prior distributions set explicitly by the user, and based on sets of summary-statistics calculated on each simulation and formally compared to the same statistics computed on the observed data (80,81). To do so, we conducted simulations using *MetHis* (37) forward-in-time simulator, followed by Random-Forest ABC scenario-choice (54), and Neural Network ABC posterior parameter inferences (53), for each Cabo Verde island separately. We followed the *MetHis*-ABC approach proposed in (37) for summary-statistics calculation, prior-checking, out-of-bag cross-validation, machine-learning ABC predictions and inferences parameterization, and posterior parameter cross-validation error calculations.

For concision here, all the details of these procedures can be found in corresponding **Supplementary Text**, **Tables ST3-13**, and **Figures SF13-26**. Briefly, we performed 10,000 simulations of 60,000 independent

SNPs under each of the four competing scenarios described above, drawing parameter values in prior distributions detailed in **Table T2**. We then computed 42 summary-statistics on each simulated dataset separately, described in **Supplementary Text** and **Table ST3**, by drawing sample-sets matching the observed sample-sizes of each Cabo Verde birth-island separately. Note that we considered summary-statistics specifically aiming at describing the distribution of individual admixture fractions in each island, known theoretically and empirically to be highly informative about complex admixture history parameters (37,56). Based on these vectors of summary-statistics, we conducted RF-ABC scenario-choice to identify, for each birth-island separately, the most likely admixture history scenario. Under the winning scenario in each birth-island separately, we performed an additional 90,000 simulations for further Neural Network ABC posterior parameter estimations and cross-validation parameter error calculations, thus each based on 100,000 total simulations.

ACKNOWLEDGEMENTS

The authors would like to warmly thank all Cabo Verdean participants to this study as well as the UniCV for facilitating the project. This project was funded in part by the French "Agence Nationale pour la Recherche" grant ANR METHIS 15-CE32-0009-1, and by a grant from the France-Stanford Center for Interdisciplinary Studies. MB was supported in part by the Linguistics Department at the University of Michigan (MI, USA). ZAS was supported in part by startup funds from the Department of Biology at the Pennsylvania State University (PA, USA). CAFL was supported in part by the Marcus Borgströms Foundation for Genetic Research and the Bertil Lundman Foundation for Anthropological Studies (Sweden). We thank the platform "Paléogénomique et génétique moléculaire" (P2GM) of the French Muséum National d'Histoire Naturelle at the Musée de l'Homme for support handling biological samples and generating genetic data. We thank the BIOMICS platform from the Pasteur institute for performing genotyping analyses. Finally, the authors warmly thank Frank Alvarez-Pereyre, Erkan O. Buzbas, Marta Ciccarella, Pierre Darlu, Ethan M. Jewett, Evelyne Heyer, Marie-France Mifune, Etienne Patin, Jorge M. Rocha, Lara Rubio Arauna, and Bruno Toupance, for useful comments and discussions about this work.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interests for this work.

1023 **REFERENCES**

1024

1025 1. Eltis D, Richardson D. Atlas of the transatlantic slave trade. 2015. 336 p.

1026 2. Fortes-Lima C, Verdu P. Anthropological genetics perspectives on the transatlantic slave
1027 trade. Human Molecular Genetics [Internet]. 2021 Apr 26 [cited 2021 Jun 4];30(R1):R79–87.
1028 Available from: https://academic.oup.com/hmg/article/30/R1/R79/6040749

1029 3. Micheletti SJ, Bryc K, Ancona Esselmann SG, Freyman WA, Moreno ME, Poznik GD, et al.
1030 Genetic Consequences of the Transatlantic Slave Trade in the Americas. The American
1031 Journal of Human Genetics [Internet]. 2020 Aug [cited 2021 Jun 4];107(2):265–77. Available
1032 from: https://linkinghub.elsevier.com/retrieve/pii/S0002929720302007

1033 4. Ongaro L, Scliar MO, Flores R, Raveane A, Marnetto D, Sarno S, et al. The Genomic Impact
1034 of European Colonization of the Americas. Current Biology [Internet]. 2019 Dec [cited 2022
1035 Apr 8];29(23):3974-3986.e4. Available from:
1036 https://linkinghub.elsevier.com/retrieve/pii/S0960982219313065

1037 5. Winkler CA, Nelson GW, Smith MW. Admixture Mapping Comes of Age. Annu Rev Genom
1038 Hum Genet [Internet]. 2010 Sep [cited 2021 Jun 4];11(1):65–89. Available from:
1039 http://www.annualreviews.org/doi/10.1146/annurev-genom-082509-141523

1040 6. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses
1041 of diverse populations improves discovery for complex traits. Nature [Internet]. 2019 Jun
1042 [cited 2021 Jun 4];570(7762):514–8. Available from: http://www.nature.com/articles/s41586-
1043 019-1310-4

1044 7. Patin E, Lopez M, Grollemund R, Verdu P, Harmant C, Quach H, et al. Dispersals and genetic
1045 adaptation of Bantu-speaking populations in Africa and North America. Science [Internet].
1046 2017 May 5 [cited 2021 Jun 4];356(6337):543–6. Available from:
1047 https://www.sciencemag.org/lookup/doi/10.1126/science.aal1988

1048 8. Gravel S. Population Genetics Models of Local Ancestry. Genetics [Internet]. 2012 Jun 1
1049 [cited 2021 Jun 4];191(2):607–19. Available from:
1050 https://academic.oup.com/genetics/article/191/2/607/5935164

1051 9. Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, et al. A Genetic Atlas of
1052 Human Admixture History. Science [Internet]. 2014 Feb 14 [cited 2021 Jun
1053 4];343(6172):747–51. Available from:
1054 https://www.sciencemag.org/lookup/doi/10.1126/science.1243518

1055 10. Mathias RA, Taub MA, Gignoux CR, Fu W, Musharoff S, O'Connor TD, et al. A continuum
1056 of admixture in the Western Hemisphere revealed by the African Diaspora genome. Nat
1057 Commun [Internet]. 2016 Nov [cited 2021 Jun 4];7(1):12522. Available from:
1058 http://www.nature.com/articles/ncomms12522

1059 11. Moreno-Estrada A, Gravel S, Zakharia F, McCauley JL, Byrnes JK, Gignoux CR, et al.
1060 Reconstructing the Population Genetic History of the Caribbean. Tarazona-Santos E, editor.
1061 PLoS Genet [Internet]. 2013 Nov 14 [cited 2021 Jun 4];9(11):e1003925. Available from:
1062 https://dx.plos.org/10.1371/journal.pgen.1003925

1063 12. Baharian S, Barakatt M, Gignoux CR, Shringarpure S, Errington J, Blot WJ, et al. The Great
1064 Migration and African-American Genomic Diversity. Gibson G, editor. PLoS Genet [Internet].

2016 May 27 [cited 2021 Jun 4];12(5):e1006059. Available from: https://dx.plos.org/10.1371/journal.pgen.1006059

13. Holm JA. An introduction to pidgin and creoles. Cambridge [England] ; New York: Cambridge University Press; 2000. 282 p. (Cambridge textbooks in linguistics).

14. Escure G, Schwegler A. Creoles, contact, and language change: linguistics and social implications. Amsterdam ; Philadelphia: John Benjamins Pub; 2004. 354 p. (Creole language library).

15. Quint N. Le cap-verdien: origines et devenir d'une langue métisse. Paris, France: L'Harmattan; 2000. 353 p.

16. Essegbey J, Migge B, Winford D. Cross-linguistic influence in language creation: Assessing the role of the Gbe languages in the formation of the Creoles of Suriname. Lingua [Internet]. 2013 May [cited 2021 Jun 4];129:1–8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S002438411300048X

17. Baptista M. Continuum and variation in Creoles: Out of many voices, one language. JPCL [Internet]. 2015 Oct 20 [cited 2021 Jun 4];30(2):225–64. Available from: http://www.jbe-platform.com/content/journals/10.1075/jpcl.30.2.02bap

18. Eltis D, editor. Coerced and free migration: global perspectives. Stanford, Calif: Stanford University Press; 2002. 447 p. (The making of modern freedom).

19. Berlin I. Many Thousands Gone: the First Two Centuries of Slavery in North America. [Internet]. Cambridge: Harvard University Press; 2009 [cited 2021 Jun 7]. 512 p. Available from: https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=3300728

20. Ansari-Pour N, Moñino Y, Duque C, Gallego N, Bedoya G, Thomas MG, et al. Palenque de San Basilio in Colombia: genetic data support an oral history of a paternal ancestry in Congo. Proc R Soc B [Internet]. 2016 Mar 30 [cited 2021 Jun 4];283(1827):20152980. Available from: https://royalsocietypublishing.org/doi/10.1098/rspb.2015.2980

21. Verdu P, Jewett EM, Pemberton TJ, Rosenberg NA, Baptista M. Parallel Trajectories of Genetic and Linguistic Admixture in a Genetically Admixed Creole Population. Current Biology [Internet]. 2017 Aug [cited 2022 Feb 2];27(16):2529-2535.e3. Available from: https://linkinghub.elsevier.com/retrieve/pii/S096098221730859X

22. Hagemeijer T, Rocha J. Creole languages and genes: the case of São Tomé and Príncipe. Faits Lang [Internet]. 2019 Aug 23 [cited 2021 Jun 4];49(1):167–82. Available from: https://brill.com/view/journals/fdl/49/1/article-p167_11.xml

23. Carreira A. Cabo Verde: formação e extinção de uma sociedade escravocrata (1460-1878). 3rd ed. Praia, Cabo Verde: IPC; 2000. 522 p. (Estudos e ensaios).

24. Albuquerque L de, Santos MEM, editors. História geral de Cabo Verde T1. Vol. 1. Lisboa : Praia [Cape Verde]: Centro de Estudos de História e Cartografia Antiga, Instituto de Investigação Científica Tropical ; Instituto Nacional da Cultura de Cabo Verde; 1991. 523 p.

25. Albuquerque L de, Santos MEM, editors. História geral de Cabo Verde T2. 2. ed. Vol. 2. Lisboa: Centro de Estudos de História e Cartogr. Antiga; 1995. 596 p.

26. Albuquerque L de, Santos MEM, editors. História concisa de Cabo Verde (Resumo da História geral de Cabo Verde T1-3). Lisboa: Inst. de Investigação Científica Tropical; 2007. 427 p.

27. Baptista M. The syntax of Cape Verdean Creole: the Sotavento varieties. Amsterdam ; Philadelphia: John Benjamins Pub; 2002. 289 p. (Linguistik aktuell = Linguistics today).

28. Lang J. Les langues des autres dans la créolisation: théorie et exemplification par le créole d'empreinte wolof à l'île Santiago du Cap Vert. Tübingen: Narr; 2009. 275 p.

29. Brehm A, Pereira L, Bandelt HJ, Prata MJ, Amorim A. Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. Ann Hum Genet [Internet]. 2002 Jan [cited 2021 Jun 4];66(1):49–60. Available from: http://doi.wiley.com/10.1017/S0003480001001002

30. Gonçalves R, Rosa A, Freitas A, Fernandes A, Kivisild T, Villems R, et al. Y-chromosome lineages in Cabo Verde Islands witness the diverse geographic origin of its first male settlers. Human Genetics [Internet]. 2003 Nov 1 [cited 2021 Jun 4];113(6):467–72. Available from: http://link.springer.com/10.1007/s00439-003-1007-4

31. Beleza S, Campos J, Lopes J, Araújo II, Hoppfer Almada A, e Silva AC, et al. The Admixture Structure and Genetic Variation of the Archipelago of Cape Verde and Its Implications for Admixture Mapping Studies. Relethford JH, editor. PLoS ONE [Internet]. 2012 Nov 30 [cited 2021 Jun 4];7(11):e51103. Available from: https://dx.plos.org/10.1371/journal.pone.0051103

32. Beleza S, Johnson NA, Candille SI, Absher DM, Coram MA, Lopes J, et al. Genetic Architecture of Skin and Eye Color in an African-European Admixed Population. Spritz RA, editor. PLoS Genet [Internet]. 2013 Mar 21 [cited 2021 Jun 4];9(3):e1003372. Available from: https://dx.plos.org/10.1371/journal.pgen.1003372

33. Hamid I, Korunes KL, Beleza S, Goldberg A. Rapid adaptation to malaria facilitated by admixture in the human population of Cabo Verde. eLife [Internet]. 2021 Jan 4 [cited 2021 Jun 4];10:e63177. Available from: https://elifesciences.org/articles/63177

34. Soares MJ. The British Presence on the Cape Verdean Archipelago (Sixteenth to Eighteenth centuries). African Economic History. 2011;39:129–46.

35. Berlin I. The making of African America: the four great migrations. New York: Penguin Books; 2010. 304 p.

36. Gouveia MH, Borda V, Leal TP, Moreira RG, Bergen AW, Kehdy FSG, et al. Origins, Admixture Dynamics, and Homogenization of the African Gene Pool in the Americas. Nielsen R, editor. Molecular Biology and Evolution. 2020 Jun 1;37(6):1647–56.

37. Fortes-Lima CA, Laurent R, Thouzeau V, Toupance B, Verdu P. Complex genetic admixture histories reconstructed with Approximate Bayesian Computation. Mol Ecol Resour [Internet]. 2021 May [cited 2021 Jun 4];21(4):1098–117. Available from: https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13325

38. Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL. High resolution of human evolutionary trees with polymorphic microsatellites. Nature [Internet]. 1994 Mar [cited 2022 Feb 2];368(6470):455–7. Available from: http://www.nature.com/articles/368455a0

39. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Research [Internet]. 2009 Sep 1 [cited 2021 Jun 4];19(9):1655–64. Available from: http://genome.cshlp.org/cgi/doi/10.1101/gr.094052.109

40. Morrison ML, Alcala N, Rosenberg NA. FSTruct: an FST-based tool for measuring ancestry variation in inference of population structure [Internet]. Evolutionary Biology; 2021 Sep [cited 2022 Feb 2]. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.09.24.461741

41. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000 Jun;155(2):945–59.

42. Rosenberg NA. Genetic Structure of Human Populations. Science [Internet]. 2002 Dec 20 [cited 2021 Jun 4];298(5602):2381–5. Available from: https://www.sciencemag.org/lookup/doi/10.1126/science.1078311

43. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 2003 Aug;164(4):1567–87.

44. Lawson DJ, van Dorp L, Falush D. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nat Commun [Internet]. 2018 Dec [cited 2021 Jun 4];9(1):3258. Available from: http://www.nature.com/articles/s41467-018-05257-7

45. Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of Population Structure using Dense Haplotype Data. Copenhaver GP, editor. PLoS Genet [Internet]. 2012 Jan 26 [cited 2021 Jun 4];8(1):e1002453. Available from: https://dx.plos.org/10.1371/journal.pgen.1002453

46. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods [Internet]. 2012 Feb [cited 2021 Jun 4];9(2):179–81. Available from: http://www.nature.com/articles/nmeth.1785

47. Chacón-Duque JC, Adhikari K, Fuentes-Guajardo M, Mendoza-Revilla J, Acuña-Alonzo V, Barquera R, et al. Latin Americans show wide-spread Converso ancestry and imprint of local Native ancestry on physical appearance. Nat Commun [Internet]. 2018 Dec [cited 2021 Jun 4];9(1):5388. Available from: http://www.nature.com/articles/s41467-018-07748-z

48. Thompson EA. Identity by Descent: Variation in Meiosis, Across Genomes, and in Populations. Genetics [Internet]. 2013 Jun 1 [cited 2021 Jun 4];194(2):301–26. Available from: https://academic.oup.com/genetics/article/194/2/301/6065409

49. Mooney JA, Huber CD, Service S, Sul JH, Marsden CD, Zhang Z, et al. Understanding the Hidden Complexity of Latin American Population Isolates. The American Journal of Human Genetics [Internet]. 2018 Nov [cited 2022 Feb 2];103(5):707–26. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0002929718303513

50. Szpiech ZA, Mak ACY, White MJ, Hu D, Eng C, Burchard EG, et al. Ancestry-Dependent Enrichment of Deleterious Homozygotes in Runs of Homozygosity. The American Journal of Human Genetics [Internet]. 2019 Oct [cited 2021 Jun 4];105(4):747–62. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0002929719303374

51. Rousset F. Genetic Differentiation and Estimation of Gene Flow from *F* -Statistics Under Isolation by Distance. Genetics [Internet]. 1997 Apr 1 [cited 2022 Feb 2];145(4):1219–28. Available from: https://academic.oup.com/genetics/article/145/4/1219/6018124

52. Wang C, Szpiech ZA, Degnan JH, Jakobsson M, Pemberton TJ, Hardy JA, et al. Comparing Spatial Maps of Human Population-Genetic Variation Using Procrustes Analysis. Statistical Applications in Genetics and Molecular Biology [Internet]. 2010 Jan 27 [cited 2022 Apr 27];9(1). Available from: https://www.degruyter.com/document/doi/10.2202/1544-6115.1493/html

53. Csilléry K, François O, Blum MGB. abc: an R package for approximate Bayesian computation (ABC): *R package: abc*. Methods in Ecology and Evolution [Internet]. 2012 Jun [cited 2021 Jun 4];3(3):475–9. Available from: http://doi.wiley.com/10.1111/j.2041-210X.2011.00179.x

54. Pudlo P, Marin JM, Estoup A, Cornuet JM, Gautier M, Robert CP. Reliable ABC model choice via random forests. Bioinformatics [Internet]. 2016 Mar 15 [cited 2021 Jun 4];32(6):859–66. Available from: https://academic.oup.com/bioinformatics/article/32/6/859/1744513

55. Robert CP, Mengersen K, Chen C. Model choice versus model criticism. Proceedings of the National Academy of Sciences [Internet]. 2010 Jan 19 [cited 2021 Jun 4];107(3):E5–E5. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0911260107

56. Verdu P, Rosenberg NA. A General Mechanistic Model for Admixture Histories of Hybrid Populations. Genetics [Internet]. 2011 Dec 1 [cited 2021 Jun 4];189(4):1413–26. Available from: https://academic.oup.com/genetics/article/189/4/1413/6089061

57. Patterson KD. Epidemics, Famines, and Population in the Cape Verde Islands, 1580-1900. The International Journal of African Historical Studies [Internet]. 1988 [cited 2021 Jun 4];21(2):291. Available from: https://www.jstor.org/stable/219938?origin=crossref

58. Brooks GE. Cabo Verde: Gulag of the South Atlantic: Racism, Fishing Prohibitions, and Famines. Hist Afr. 2006;33:101–35.

59. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. The American Journal of Human Genetics [Internet]. 2017 Apr [cited 2022 Apr 8];100(4):635–49. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0002929717301076

60. Korunes KL, Soares-Souza GB, Bobrek K, Tang H, Araújo II, Goldberg A, et al. Sex-biased admixture and assortative mating shape genetic variation and influence demographic inference in admixed Cabo Verdeans [Internet]. Evolutionary Biology; 2020 Dec [cited 2022 Feb 2]. Available from: http://biorxiv.org/lookup/doi/10.1101/2020.12.14.422766

61. 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 [Internet]. 2017 Jan 1 [cited 2021 Jun 4];205(1):375–83. Available from: https://academic.oup.com/genetics/article/205/1/375/6095598

62. Versluys TMM, Mas-Sandoval A, Flintham EO, Savolainen V. Why do we pick similar mates, or do we? Biol Lett [Internet]. 2021 Nov [cited 2022 Feb 2];17(11):20210463. Available from: https://royalsocietypublishing.org/doi/10.1098/rsbl.2021.0463

63. Chaudenson R, Mufwene SS. Creolization of language and culture. London ; New York: Routledge; 2001. 340 p.

64. Cooper F, Holt TC, Scott RJ. Beyond slavery: explorations of race, labor, and citizenship in

postemancipation societies. Chapel Hill: University of North Carolina Press; 2000. 198 p.

65. Fortes-Lima C, Bybjerg-Grauholm J, Marin-Padrón LC, Gomez-Cabezas EJ, Bækvad-Hansen M, Hansen CS, et al. Exploring Cuba's population structure and demographic history using genome-wide data. Sci Rep [Internet]. 2018 Dec [cited 2021 Jun 4];8(1):11422. Available from: http://www.nature.com/articles/s41598-018-29851-3

66. Goldberg A, Verdu P, Rosenberg NA. Autosomal Admixture Levels Are Informative About Sex Bias in Admixed Populations. Genetics [Internet]. 2014 Nov 1 [cited 2021 Jun 4];198(3):1209–29. Available from: https://academic.oup.com/genetics/article/198/3/1209/6065659

67. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015 Oct 1;526(7571):68–74.

68. Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K, et al. The African Genome Variation Project shapes medical genetics in Africa. Nature [Internet]. 2015 Jan 15 [cited 2021 Jun 4];517(7534):327–32. Available from: http://www.nature.com/articles/nature13997

69. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics [Internet]. 2010 Nov 15 [cited 2021 Jun 4];26(22):2867–73. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btq559

70. Chafe WL, editor. The Pear stories: cognitive, cultural, and linguistic aspects of narrative production. Norwood, N.J: Ablex Pub. Corp; 1980. 327 p. (Advances in discourse processes).

71. Croft W. Linguistic Selection: An Utterance-based Evolutionary Theory of Language Change. Nordic Journal of Linguistic [Internet]. 1996 Dec [cited 2021 Jun 4];19(2):99–139. Available from: https://www.cambridge.org/core/product/identifier/S0332586500003358/type/journal_article

72. R Development Core Team. R: A language and environment for statistical computing. [Internet]. R Foundation for Statistical Computing; 2020. Available from: http://www.R-project.org

73. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics [Internet]. 2007 Sep [cited 2021 Jun 4];81(3):559–75. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0002929707613524

74. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics [Internet]. 2007 Jul 15 [cited 2021 Jun 4];23(14):1801–6. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btm233

75. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes [Internet]. 2003 Dec 10 [cited 2021 Jun 4];4(1):137–8. Available from: http://doi.wiley.com/10.1046/j.1471-8286.2003.00566.x

76. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. Nature. 2010 Sep;467(7311):52–8.

77. Szpiech ZA, Blant A, Pemberton TJ. GARLIC: Genomic Autozygosity Regions Likelihood-based Inference and Classification. Berger B, editor. Bioinformatics [Internet]. 2017 Jul 1 [cited 2021 Jun 4];33(13):2059–62. Available from: https://academic.oup.com/bioinformatics/article/33/13/2059/3000374

78. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. The American Journal of Human Genetics [Internet]. 2013 Aug [cited 2022 Feb 2];93(2):278–88. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0002929713002899

79. Snyder JP. Map projections--a working manual [Internet]. 1987 [cited 2022 Feb 2]. Available from: https://purl.fdlp.gov/GPO/gpo93084

80. Tavaré S, Balding DJ, Griffiths RC, Donnelly P. Inferring coalescence times from DNA sequence data. Genetics. 1997 Feb;145(2):505–18.

81. Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population genetics. Genetics. 2002 Dec;162(4):2025–35.

1283

Supplementary Text

The admixture histories of Cabo Verde

Authors:
Romain Laurent[1], Zachary A. Szpiech[2,3], Sergio S. da Costa[1], Valentin Thouzeau[4,5], Cesar A. Fortes-Lima[6], Françoise Dessarps-Freichey[1], Laure Lémée[7], José Utgé[1], Noah A. Rosenberg[8], Marlyse Baptista[9,10], and Paul Verdu[1,*].

Authors' affiliations :
1. UMR7206 Eco-anthropologie, CNRS-MNHN-Université Paris Cité, Paris, France
2. Department of Biology, Pennsylvania State University, University Park, PA 16802, USA
3. Institute for Computational and Data Sciences, Pennsylvania State University, University Park, PA 16802, USA
4. UMR 7534 Centre de Recherche en Mathématiques de la Décision, CNRS-Université Paris-Dauphine-PSL University, Paris, France
5. Département d'Etudes Cognitives, Laboratoire de Sciences Cognitives et Psycholinguistique, ENS-PSL University-EHESS-CNRS, Paris, France
6. Department of Organismal Biology, Sub-department of Human Evolution, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden
7. Plate-forme Technologique Biomics–Centre de Ressources et Recherches Technologiques (C2RT), Institut Pasteur, Paris, France.
8. Department of Biology, Stanford University, Stanford, CA, USA
9. Department of Linguistics, University of Michigan, Ann Arbor, MI, USA
10. Department of Afroamerican and African Studies, University of Michigan, Ann Arbor, MI, USA

***Corresponding author:**
Paul Verdu, UMR7206 Eco-anthropologie, CNRS-MNHN-Université Paris Cité, Musée de l'Homme, 17, place du Trocadéro, 75016 Paris, France. Email : paul.verdu@mnhn.fr. Tel : +33 144057317

Authors' contributions:
PV, MB, NAR designed sampling strategy
PV, MB, SSC, VT, CFL conducted sampling field-work
FD, PV, LL, JU conducted molecular-genetics experiments
RL, PV conducted genetic data quality control and merging
SSC conducted linguistic transcriptions
MB, SSC conducted linguistic data categorization
RL, PV conducted linguistic data parsing
RL, PV, ZAS, MB, SSC, VT conducted genetic and linguistic statistical descriptions
RL, PV conducted genetic inferences
PV, RL, MB, ZAS wrote the paper with the help of all authors
PV, MB, and NAR supervised the research

Table of Contents

Supplementary Material and Methods **3**

Supplementary Results **8**

Supplementary results describing genetic diversity patterns with ASD-MDS at different geographical scales . 8

Supplementary results about genetic structure patterns in Cabo Verdean and other enslaved-African descending individuals obtained with ADMIXTURE 10

Supplementary description of local-ancestry inferences in Cabo Verde and other admixed populations related to the TAST 12

Supplementary results and discussion about admixture patterns across analyses and relations between individual admixture levels and birth-places geographical distances within Cabo Verde 13

Supplementary description of Runs of Homozygosity (ROH) patterns in Cabo Verde and other admixed populations related to the TAST 14

Detailed Results and Discussions of the genetic admixture histories of each Cabo Verdean island 15

Admixture history of Santiago (First official settlement 1460; Perennial peopling: 1470) 15

Admixture history of Fogo (First official settlement 1460; Perennial settlement: 1570) 17

Admixture history of Santo Antão (First official settlement 1570; Perennial settlement: 1580) 19

Admixture history of São Nicolau (First official settlement 1580; Perennial settlement: 1580) 20

Admixture history of Brava (First official settlement 1580; Perennial settlement: 1580) 22

Admixture history of Maio (First official settlement 1529; Perennial settlement: 1650) 23

Admixture history of Boa Vista (First official settlement 1566; Perennial settlement: 1650) 24

Admixture history of São Vicente (First official settlement 1570; Perennial settlement: 1780) 25

Admixture history of Sal (First official settlement 1529; Perennial settlement: 1841) 27

Final methodological discussion 28

REFERENCES **29**

Supplementary Figures and Tables **33**

Supplementary Material and Methods

Detailed *MetHis*-ABC (Fortes-Lima et al. 2021) procedures

MetHis-ABC Random Forest scenario-choice and Neural Network posterior parameter estimations

MetHis simulations for ABC inference

ABC inference relies on simulations conducted with scenario-parameter values drawn randomly within prior distributions set explicitly by the user. We used *MetHis* v1.0 (1) to simulate 60,000 autosomal independent SNP markers in the admixed population H under the four competing scenarios presented in **Figure F7**. In brief, at each generation, *MetHis* performs simple Wright-Fisher (2,3) forward-in-time discrete simulations, individual-centered, in a randomly-matting admixed population of N_g diploid individuals, following the (4) admixture scenario with admixture parameters specified by the user. We neglected mutation during the 21 generations of the admixture process.

Simulating the source and admixed populations

For simplicity, we considered that the African and European populations at the source of admixture in Cabo Verde were very large populations at the drift-mutation equilibrium during the entire admixture process. Furthermore, we considered that these source populations were accurately represented, respectively, by the Mandinka from Senegambia and the Iberian-IBS populations in our genome-wide SNP dataset (see **Results** section).

Therefore, to build the SNP set used in ABC inferences, we used the pruned 102,543 independent SNPs employed for the ADMIXTURE analyses, and then randomly drew 60,000 independent SNPs for ABC inferences.

We first built two separate datasets each comprising 20,000 haploid genomes of 60,000 independent SNPs. Each SNP was randomly drawn in the site frequency spectrum observed for the Mandinka and IBS pruned datasets respectively. We used both datasets as fixed gamete reservoirs respectively for the African and European source populations. At each generation, we built, from these gamete reservoirs, a parental effective individual by randomly pairing gametes avoiding selfing. These virtual source populations provided the African and European parents for simulating individuals in the admixed population, following the admixture scenario-parameters at each generation. Thus, while our gamete reservoirs were fixed over the 21 generations of the admixture processes here considered, the African or the European diploid parents possibly contributing to the gene-pool of the admixed population were randomly built anew at each generation.

At each generation, *MetHis* randomly draws parents for the admixed population at the following generation in proportions $s_{Afr,g}$, $s_{Eur,g}$, and h_g. Then, the software randomly builds gametes for each parent and randomly pair gametes, without selfing, to produce N_g diploid individuals in the admixed population.

Sampling simulated source and admixed populations

At the end of each simulation, we randomly drew individual samples from each source and the admixed population H matching observed sample sizes. We sampled 60 individuals in the simulated African source, 60 individuals in the European source, and the number of admixed individuals corresponding to the sample size of each Cabo Verde island of birth of individuals, in turn (**Table ST1**). We sampled

individuals without grand-parents in common, as allowed in *MetHis* by explicit genealogical flagging of individuals during the last two generations of the admixture process.

Number of simulations and scenario-parameter priors for ABC inference

We performed 10,000 such *MetHis* simulations under each four admixture scenarios for each nine birth-islands of Cabo Verde separately. Each simulation was performed under a vector of scenario parameters drawn from prior distributions described in **Material and Methods** and in **Table T2,** using *MetHis parameter generator* tools.

Separately for each island, we used this set of simulations to determine which scenario best explained the observed data using Random-Forest ABC (see below). Under the winning scenario for each birth-island, we then performed an additional 90,000 simulations to produce a total 100,000 simulations to be used for Neural-Network ABC posterior parameter estimation.

Summary statistics for ABC inference

ABC inferences use sets of summary statistics calculated on each simulated dataset, as well as the observed data. Therefore, in ABC, each unique vector of scenario-parameter values corresponds to a vector of calculated summary statistics. We used *MetHis* summary statistics calculation tools to produce, for each simulated dataset, 42 summary statistics listed with references in **Table ST3**, which we computed similarly on the observed datasets considering, in turn, all individuals born within each Cabo Verdean island, as Population H.

Within-Population H summary statistics

We computed the mean and variance of interindividual ASD (5) within Population H. Furthermore, we computed the mean and variance of average heterozygosities (averaged over independent SNPs for each individual and averaged over individuals within island) as in (6); and mean and variance of inbreeding coefficient F (7). Intuitively, we a priori expected these statistics to be particularly sensitive to genetic-drift and possibly informative specifically about reproductive population size *Ne* parameters in our scenarios.

Admixture-pattern summary statistics

We analytically showed previously that the distribution of admixture fractions across individuals within admixed populations carried identifiable information about the history of admixture (4,8). Furthermore, we showed with *MetHis* (1), that summary statistics describing this distribution could be successfully used in ABC inferences.

Therefore, we considered the 16 admixture-related summary statistics describing the mode, the first four-moments (as mean, variance, skewness, and kurtosis), and the minimum, maximum, and all 10% percentiles of the ASD-MDS admixture estimates of individual admixture fractions from the African source population (or one minus that from the European source in a two-source admixture scenario). ASD-MDS admixture estimates were obtained as described in **Material and Methods** represented schematically in **Figure SF12**, considering the 2D-MDS centroids, respectively, of the African and European sources, and the projection on the line joining these two points of each Population H individuals. Furthermore, to further capture the ASD-MDS admixture patterns, we calculated the distribution of the angles between population H's individuals and the source populations' centroids on the 2D ASD-MDS. We then considered as summary-statistics for ABC inference, the mode, the first

four-moments (as mean, variance, skewness, and kurtosis), the minimum and maximum, and the 10% percentiles of this distribution of angles in radian (**Figure SF12**).

Among-populations summary statistics

Using *MetHis* summary-statistics calculation tools, we computed F_{ST} values between either Source population (African or European) and Population H following (9). We also computed the mean ASD between either Source Population and Population H. Finally, we computed the f_3 statistic (10) with both African and European populations as the two sources, and Population H, as the targeted admixed population.

Prior-ABC checking

Before conducting any ABC inference, we needed to statistically evaluate whether the 42 summary-statistics calculated on the observed datasets were within the range of the sets of summary statistics computed on the simulated datasets with *MetHis*. To do so, we considered each Cabo Verdean birth-island, in turn as population H, and conducted the following three step procedured whose results are provided in **Figures SF13-22**.

First, (i) we conducted a goodness-of-fit test between the 40,000 vectors of summary statistics computed for each simulation under the four competing scenarios, and the vector of observed summary statistics with the *gfit* function of the R package *abc* (11), with 1000 random repetitions and a 1% tolerance level (**Figure SF13A**).

Second, (ii) we computed a Principal Component Analysis considering each simulation as an individual and all the summary statistics as observed variables. By projecting the observed summary statistics onto this PCA along the first three PCA axes of variation (with the *princomp* function in R), we visually evaluated whether the observed data fell into the range of simulated statistics along the major PCA axes (**Figure SF13B-C**).

Finally, (iii) we present the distribution of each summary statistic obtained for the 40,000 simulations and the observed value, for each summary statistic and each Cabo Verdean birth-island separately (**Figure SF14-SF22**).

Random-Forest ABC scenario-choice

We used the Random-Forest ABC algorithm implemented in the *abcrf* package in R (12,13) for scenario-choice, similarly as in the *MetHis*-ABC pipeline previously described (1). We considered the same prior probability for the four competing scenarios.

First, for each island separately, we used the *abcrf* function in this R package to conduct ABC scenario-choice cross-validation of each 40,000 simulations serving in-turn as pseudo-observed data and the remaining 39,999 simulations as training (**Figure SF23A**), using 1000 decision trees in the Random-Forest. We visually checked that prior error rates (the rates of erroneously assigned scenario in the cross-validation) was minimized for this number of decision trees, using the *err.abcrf* function in the same R package. Each summary-statistics importance to the cross-validation accurate decision was calculated and plotted using the *abcrf* function (**Figure SF23B**).

Second, for each birth-island separately, we used the *predict.abcrf* function on the trained Random Forest with all simulations in the reference table, to determine which competing scenario produced simulations whose summary statistics most resembled those from the observed data, and estimate the posterior probability of accurately finding such winning scenario in our framework (**Figure F7B**).

Neural Network ABC posterior parameter inference

While RF-ABC scenario-choice procedures are specifically designed to accurately work with a relatively small number of simulations compared to classical regression ABC (12,14), posterior joint-estimations of all scenario-parameters under the winning scenario remain difficult in Random-Forest (1,13). Furthermore, Neural Network ABC posterior parameter inferences have been shown in multiple previous investigations to require substantial amounts of information in the reference table to perform satisfactorily (1,11,15). Therefore, we conducted anew 90,000 simulations with *MetHis*, for each birth-island separately, under the winning scenario determined with RF-ABC, thus obtaining a total 100,000 simulations under the winning scenario, for each birth-island separately, for further Neural Network ABC posterior parameter estimation using the *abc* package in R (11).

NN-ABC parameter inference settings

We followed the same empirical approach as in *MetHis*-ABC (1) to determine the most suitable number of neurons in the hidden layer and associated number of simulations closest to the target data and retained for training the Neural Network (called "tolerance level") for further NN-ABC posterior parameter estimation for this target. Indeed, there is no rule-of-thumbs to determine these parameters a priori in order to obtain the most informative posterior parameter inference while minimizing overfitting using a Neural-Network approach under a given scenario (11,15).

Therefore, we empirically tested three different tolerance-level values to be used in NN-ABC posterior parameter estimation: 1%, 5%, or 10% of the total 100,000 simulations generated under the winning scenario determined by our RF-ABC procedure for each Cabo Verdean island respectively. Furthermore, for each one of the three tolerance-level values, we tested associated numbers of neurons in the hidden layer between 5 and 11, or 12, the number of free parameters minus one in the winning scenario for each Cabo Verdean island respectively.

Then, for each pair of tolerance-level value and number of neurons for each island separately, we performed 1,000 cross-validation NN-ABC posterior parameter inferences using the *cv4abc* function in the *abc* package (11). This procedure draws 1,000 simulations at random and considers them, in turn, as pseudo-observed data for posterior-parameter inference using the remaining 99,999 simulations as the reference table. For the 1,000 pseudo-observed cross-validations and each pair of tolerance level and number of hidden neurons, we calculated the error between the median posterior point-estimate of each parameter $(\hat{\theta}_i)$ and it's known true-value used for simulation (θ_i), as the mean-square error scaled by the parameter's variance across the 1,000 cross-validations: $\sum_1^{1000}(\hat{\theta}_i - \theta_i)^2 / (1000 \times Variance(\theta_i))$ (11).

Therefore, we conducted, in total, 198,000 NN-ABC cross-validation parameter inference procedures to identify the NN-ABC settings most suitable for further posterior parameter estimations using the observed data for each Cabo Verdean birth-island specifically. For each Cabo Verdean island separately under the corresponding winning scenario, we chose the pair of tolerance-level values and number of hidden neurons that minimized the average error across all estimated parameters, as the Neural-Network setting used in further ABC posterior parameter inferences based on the observed data (**Table ST4**).

NN-ABC posterior parameter estimation

For each Cabo Verdean island separately under the winning scenario identified with RF-ABC, we jointly estimated the posterior distributions of all parameters using NN-ABC "*neuralnet*" method option in the function *abc* of the R package *abc* (11), with logit-transformed parameters ("*logit*" transformation option) using the tolerance level and number of neurons in the hidden layer identified previously in **Table ST4**. For each birth-island separately, a synthetic schematic figure of the complex admixture processes identified using median point estimates of posterior scenario-parameter distribution is provided in **Figure F7**, and full posterior parameter distributions with 95% Credibility-Intervals (CI) are provided in **Figure SF24-26** and **Tables ST5-ST13**.

NN-ABC posterior parameter errors

We evaluated posterior parameter error rates and 95%-CI accuracies in the vicinity of the observed data, for each island separately and for each corresponding NN-ABC posterior parameter joint estimation.

First, we identified the 1,000 simulations closest to the observed "real" data for each island separately using a 1% tolerance level in the "*neuralnet*" option from the *abc* function. We then used each one of these 1,000 specific simulations, in turn, as pseudo-observed target data for cross-validation NN-ABC posterior parameter estimation using the same NN settings as previously, logit-transformed parameters, and the 99,999 remaining simulations in the reference table. We then compared the median posterior estimate of each scenario parameter $(\hat{\theta}_i)$, with the original parameter value used for the simulation (θ_i), respectively for each 1,000 cross-validation posterior parameter estimation; and calculated the cross-validation mean absolute error $\sum_1^{1000}|\hat{\theta}_i - \theta_i|/1000$ for each scenario parameter (**Tables ST5-ST13**).

Second, for each island and each scenario-parameter separately, we calculated how many times the true parameter values (θ_i), used for simulating the 1,000 simulations closest to the observed data, fell within the 95%-CI [2.5% quantile$(\hat{\theta}_i)$; 97.5% quantile$(\hat{\theta}_i)$] estimated using the observed data. Thus, if the 1,000 cross-validation true parameter values fell more than 95% of the time within the estimated 95%-CI, the length of the estimated 95%-CI was considered over-estimated and thus excessively conservative; and, alternatively, if it was the case less than 95% of the time, the length of the 95%-CI was considered under-estimated thus indicating a less conservative behavior of our CI estimation procedure (**Tables ST5-ST13**).

Supplementary Results

Supplementary results describing genetic diversity patterns with ASD-MDS at different geographical scales

At the worldwide scale (**Figure SF2**), the first MDS axis is mainly driven by genetic differentiation between African individuals and non-African individuals, the second MDS axis by differentiation between European and East Asian individuals, and the third axis by differentiation between South Asian and American individuals.

In this context, we find that Cabo Verdean individuals are located at intermediate genetic distances along an axis going from the European cluster to the African cluster, similarly to African-American ASW and Barbadian-ACB individuals. At this geographical scale, Cabo Verdean, ASW, and ACB individuals overall show no signs of genetic affinity with other Asian or American populations along the first three MDS axes, except for three ASW and one ACB individuals, consistently with previous reports (16). Notably, Puerto Rican-PUR individuals, also historically descending in part from enslaved-Africans forcibly deported to the Caribbean during the TAST, cluster separately at intermediate distances between the European, Sub-Saharan African, and American clusters along the first three MDS axes (**Figure SF2**), also consistently with previous observations of substantial three-way admixture between Europe, Africa and the Americas in this population (16,17).

We explored successive subsamples of this worldwide dataset in order to decompose apparent preferred relationships between each admixed population related to the TAST with respect to African and European populations only. We find (**Figure SF3**) that the first MDS axis differentiates African and European individuals; the second axis differentiates mainly Senegambian West Western African individuals from Congo Basin hunter-gatherer individuals (namely the Baka, Ba.Bongo, Ba.Koya, Ba.Mbuti, Ba.Twa and Bezan); and the third axis differentiates further these latter populations from Eastern African individuals. In this context, we find Cabo Verdean individuals clustering along an axis going from the European cluster to a cluster formed mainly by West Western African individuals, with the sole exception of two individuals with fathers born on Angola as per our familial anthropology questionnaires and consistently clustering closer to our Angolan samples. Instead, ASW, ACB and PUR individuals are found on a slightly different trajectory going from European individuals towards East Western Africa mainly (**Figure SF3**).

Further resampling our ASD matrix (**Figure SF4**), we find no signs of particular genetic affinity, along the first three MDS axes, between enslaved-African descendants and Eastern and Southern African individuals.

Notably, we find Cabo Verdean individuals clustering similarly to Senegambian Fulani individuals in between European and other Senegambian populations, the Fulani being known to be historically admixed between Senegambian and Northern African populations (18–21). In the absence of Northern African populations in our dataset, the South-Western European Iberian-IBS population represented the best such proxy-populations with known historical relationship with Northern Africa, hence resulting, at this geographical scale, in the similar clustering of Cabo Verdean and Fulani individuals. Note that Fulani and Cabo Verdean individuals substantially departed from each other along higher MDS axes, the Cabo Verdeans remaining on a European-African axis while the Fulani departed from this trajectory on the European side (**Figure SF4**).

Finally, considering only populations related to the TAST with respect to European and East, Central, and West Western African populations (excluding the Fulani), we find (**Figure SF5**) that Cabo

339 Verdeans cluster separately from other TAST populations not only with respect to African populations,
340 but also with respect to their relationships to European populations (**Figure F2**).
341
342

Supplementary results about genetic structure patterns in Cabo Verdean and other enslaved-African descending individuals obtained with ADMIXTURE

In **Figure F3A**, ADMIXTURE results at *K*=3 show that the new pink cluster is maximized in Chinese-CHB individuals, with a few African-American ASW and Barbadian-ACB individuals exhibiting relatively high membership to this cluster as visually identified in the ASD-MDS analysis (**Figure SF2**), and previously reported (16). While Finnish-FIN individuals all exhibit low membership to this cluster, such signal probably emerge from unresolved clustering for this population geographically between Western Europe and East Asia. This is further evidenced by the almost complete disappearing of this resemblance of FIN individuals with the Pink cluster at higher values of *K*. No other individuals or populations in our analysis show signals of strong resemblance with the East Asian Chinese-CHB for higher values of *K*.

The fourth gray cluster at *K*=4 is maximized in Southern African individuals from the Soto and Zulu populations. At this value of *K*, the original orange cluster is maximized in the West Western African Jola from Senegambia and all other African populations exhibit intermediate membership between the gray and the orange cluster. Interestingly, the average membership to the gray "Southern African" cluster decreases with increasing geographic distance from Southern Africa, along the Atlantic Ocean coast of the continent. This echoes the relatively continuous clustering observed in the ASD-MDS analysis for the Atlantic African populations in our dataset (**Figures SF3-5**).

Notably, the Cabo Verdean individuals born on the islands of Brava, Fogo, Santiago, and Maio (the four most Southern islands, **Figure F1**), exhibit membership to both the orange and the gray clusters, albeit with a majority of membership to the orange cluster rather than the gray cluster. All other Cabo Verdean individuals show virtually no membership to the gray cluster. While indicating the stronger resemblance between all Cabo Verdean individuals with genetic patterns observed mostly in West Western Africa and Senegambian populations, this provides a first indication of differentiated genetic structure across islands within Cabo Verde.

Conversely, the orange cluster previously identified in the ASW and ACB is, at *K*=4, divided into gray and orange with a slight excess of gray. A similar pattern is found among Central and East Western African populations such as the Yoruba-YRI, the Nigerian-ESN, the Igbo, and the Ga-Adangbe populations, hence providing a first indication of a closer resemblance of the individuals of these two TAST descending populations to this African region, rather than to West Western Africa.

At *K*=5, note that West Central and South-West African populations exhibit a distinctive unique pattern with high membership to both the red and gray clusters. However, note that ADMIXTURE clustering is largely unresolved at this value of *K* as no individuals exhibit close (<1%) to 100% membership to the red cluster only, and as only 27 out of the 30 ADMIXTURE iterations provide similar results (and an alternative minor mode is identified, **Figure SF8**).

At *K*=6, the novel green cluster is largely maximized in several Cabo Verdean islands (from Boa Vista, Sal, São Nicolau, São Vicente, Santo Antão, and, to a substantial extent, Maio), but seldom present in individuals from Brava, Fogo, and Santiago. Importantly, while the membership to the blue "European" cluster remain mostly unaltered in Brava, Fogo, Santiago, and Maio, only traces of this blue cluster remain for all other islands. The substantial membership to the green cluster in these birth-islands at this higher value of *K* show substantial genetic differentiation, possibly due to substantial drift having occurred specifically in the five islands. Notably, no other African or European population exhibit traces of membership to this novel green cluster, except for a marginal amount in the Senegambian Wolof individuals.

At K=7, the novel light-blue cluster is maximized only in Finnish-FIN individuals, while the blue cluster is now maximized only in Tuscan-TSI individuals. British-GBR and USA-CEU individuals exhibit intermediate membership between the two cluster while Iberian-IBS individuals strongly

resemble TSI individuals. This clustering patterns further echoes the observed ASD-MDS clustering of European populations identified in **Figure F2**. Importantly, the "European" membership of all Cabo Verdean islands resembles, relatively, the patterns exhibited in TSI and IBS individuals (with high proportions of blue and minimal proportion of light-blue), while African-American ASW and Barbadian-ACB individuals show relatively much higher proportions of light-blue, hereby resembling, for this genotype membership, GBR and CEU individuals rather than other Europeans. Nevertheless, variable results across runs at this value of K indicate unresolved clustering and these interpretations should be considered cautiously at this point (see alternative ADMIXTURE modes at *K*=7 in **Figure SF7**).

At K=8, the novel dark-red cluster is maximized in West Central African Tsogho individuals and at very high proportions in all other individuals from West-Central and South-West Africa, hence resolving the unresolved clustering identified for these populations at *K*=5. Interestingly, substantial traces of this cluster are found in YRI, Igbo and ESN East Western African populations as well as ASW and ACB individuals, but virtually absent from Cabo Verdeans or any other African population. This further indicates the closer resemblances of ASW and ACB African genetic component to East Western and West Central Africa than to other African populations in our dataset, as echoed in our ASD-MDS analyses (**Figure F2**, **Figures SF2-6**).

At *K*=9, the novel light-yellow cluster is maximized in Sierra Leone-MSL population from Central West Africa, albeit imperfectly. This further illustrates the intermediate clustering of Central Western African populations in between Senegambian and East Western African populations, largely overlapping on both sides, and observed in ASD-MDS analyses (**Figure F2**, **Figures SF2-6**). Nevertheless, patterns should be interpreted with caution at this value of *K*, as clustering is largely unresolved, also indicated by the several alternative ADMIXTURE solutions found (**Figure SF8**).

Supplementary description of local-ancestry inferences in Cabo Verde and other admixed populations related to the TAST

The ASD-MDS decomposition and ADMIXTURE analyses provided indications of differing admixture patterns among populations historically related to the TAST concerning possible source populations and patterns of admixture across individuals within populations, in particular since we know *a priori* that the Cabo Verdean and admixed American populations here investigated experienced historical admixture between European and African populations. However, as mentioned in the main text of this article, these descriptive population-genetics analyses do not formally test for admixture and ancestral origins of putative source populations, they rather disentangle genetic affinities among individuals, and results cannot, therefore, be directly interpreted as the product of admixture processes (22–26). Furthermore, they rely on SNPs considered independently, and do not explicitly account for haplotypic information known to massively inform the admixture history underlying observed genetic patterns (27,28).

Figure F3B presents the best (maximum likelihood across 20 iterations of each analysis) results for four and, respectively, six putative source populations for the shared haplotypic ancestry of either the ASW, ACB, and the nine Cabo Verdean birth islands, considering 24 possible sources in our analysis (**Figure F3A**, **Table ST1**).

In addition to **Results** explicated in the main text of the article, although every Cabo Verdean island almost exclusively share haplotypic ancestry with the Iberian-IBS, note that we found 6% shared ancestry with the Tuscan-TSI and 3% with the British-GBR, considering six possible source populations for Cabo Verdean individuals born on Brava; and a marginal 1% GBR possible shared ancestry in Santiago, also with six possible source populations only.

Furthermore, note that we find only reduced shared haplotypic ancestries between other non-Senegambian African populations and Cabo Verdeans. We found that the East Western African Igbo shared less than 6% haplotypic ancestry with the individuals born on Santo Antão and São Vicente (the latter being only found when considering four possible sources instead of six), and 3% with individuals from Boa Vista. Interestingly, we found that Angolan populations, also part of the Portuguese colonial empire during the TAST, had only trace-amounts (<3%) of ancestral haplotypes shared with Cabo Verdeans mainly from Maio, Sal, São Nicolau, and São Vicente, albeit with ambiguous calls depending on the number of putative source populations used in the analyses.

Supplementary results and discussion about admixture patterns across analyses and relations between individual admixture levels and birth-places geographical distances within Cabo Verde

We compared individual admixture estimates obtained between the three different approaches here conducted: the ASD-MDS individual estimates from **Figure F2A-C** (**Material and Methods**, **Figure SF12**), ADMIXTURE individual estimates at K=2 (**Figure F3A**), and SOURCEFIND population estimates (**Figure F3B**).

We find a strong positive Spearman correlation (ρ=0.99725; p-value < 2.2×10^{-16}) between the 225 Cabo Verde-born individual membership to the orange "African" cluster at K=2, and the ASD-MDS three-dimensional individual admixture estimates, considering the set of Senegambian individuals (GWD, Jola, Mandinka, and Wolof populations) and the set of Southern European (IBS and TSI) individuals used in the ADMIXTURE analysis, and visually identified as putative source populations from **Figure F2A-C**. Similarly strong correlations are found separately for the 50 ASW (ρ =0.99568; p-value < 2.2×10^{-16}) and 50 ACB individuals (ρ =0.99366; p-value < 2^{-16}), considering instead the British-GBR and USA-CEU individuals, and, respectively, the East Western African (Ga-Adangbe, Igbo, YRI, and ESN) and Central Western African Sierra Leone MSL individuals used in the ADMIXTURE analysis, as putative source populations visually identified from **Figure F2D-F**, and **Figure F2G-I**, respectively. This shows empirically that ASD-MDS projection and ADMIXTURE maximum-likelihood clustering capture very similar information with respect to inter-individual genetic differentiation of recently admixed populations, as previously reported in similar settings (e.g. Verdu et al. 2017; Fortes-Lima et al. 2021).

Altogether, our SOURCEFIND and ADMIXTURE results are highly consistent with one another (and with our ASD-MDS decompositions) as reported in **Figure F3** caption. Furthermore, we find substantially different admixture patterns across Cabo Verdean islands, with significant positive correlation between absolute individual pairwise differences in ADMIXTURE, or ASD-MDS, individual African admixture levels and absolute geographic distances between individual birth-places (see **Results**). In other words, on average across Cabo Verde, two individuals born further apart are also likely to exhibit more different estimated African admixture levels. This was visually somehow expected as we found the largest African genetic component in individuals from Santiago, Maio, and Boa Vista, and relatively more even African and European genetic components in all other islands (**Figure F3A-B**). While our results are consistent with previous results in Santiago and Fogo, further comparisons with previous studies remain difficult as individuals born on other islands were sometimes indiscriminately grouped or not investigated (29–32).

Supplementary description of Runs of Homozygosity (ROH) patterns in Cabo Verde and other admixed populations related to the TAST

Examining patterns of ROH less than 1cM long, we find generally low levels of short and medium ROH in continental Africa (**Figure SF9**), with the lowest levels in southern Africa. Conversely, we found high levels of such short and medium-size ROH among European and Asian populations, likely reflecting ancient migration bottlenecks during the Out-of-Africa expansion (33). We note that the Cabo Verdean individuals have slightly elevated levels of short and medium ROH compared to continental Africa, similarly to the ACB and ASW populations. This may stem from the known admixture with European populations that occurred recently during the TAST. Finally, overall total ROH levels reflected the generally higher levels of ROH in Cabo Verde compared to continental African and other admixed populations from the Americas (**Figure SF9C**), patterns likely related to the relative isolation of populations within the archipelago (**Figure F5**, **Table T1**).

Exploring how local ancestry intersects with ROH less than 1cM long, since short and medium ROH are comprised of old haplotypes that likely predate over 50 generations ago, ((34)) the colonization of the Cabo Verde islands and the subsequent admixture history of these people, we expected that the local ancestry content of these ROH classes should be correlated with total non-ROH local ancestry levels. **Figure SF10** illustrates the relationship between total ancestry in a given ROH size class versus the total ancestry not in that ROH size class for all Cabo Verdean individuals. We find strong correlations among the short ROH and local ancestry (**Figure SF10A**; African local ancestry: Pearson ρ=0.9372, $p < 2.2\text{x}10^{-16}$; European local ancestry: Pearson ρ=0.9496, $p < 2.2\text{x}10^{-16}$; Asian local ancestry: Pearson ρ=0.7516, $p < 2.2\text{x}10^{-16}$), and medium ROH and local ancestry (**Figure SF10B**; African local ancestry: Pearson ρ=0.9071, $p < 2.2\text{x}10^{-16}$; European local ancestry: Pearson ρ=0.9379, $p < 2.2\text{x}10^{-16}$; Asian local ancestry: Pearson ρ=0.3905, $p < 1.298\text{x}10^{-9}$).

Detailed Results and Discussions of the genetic admixture histories of each Cabo Verdean island

RF votes favor the Afr2Pulses-Eur2Pulses scenario with high posterior probabilities of identifying correctly this scenario for individuals born, respectively, in Santiago (RF-ABC posterior probability of choosing the Afr2Pulses-Eur2Pulses=0.6154), São Nicolau (post-prob.=0.5114), and Santo Antão (post-prob.=0.5388). While RF votes results also favor Afr2Pulses-Eur2Pulses scenarios for individuals born in Fogo, Brava and Maio, they do so less clearly with higher votes for alternative scenarios and weaker posterior probabilities (post- prob.=0.4632, 0.4152, and 0.4189, respectively). For Boa Vista and Sal, the Afr2Pulses-EurRecurring scenario is favored, while the AfrRecurring-Eur2Pulses scenario is favored for São Vicente. For São Vicente, the RF-ABC posterior probability is high (0.5151), while weaker for Boa Vista and Sal (0.3830 and 0.4910, respectively).

Results for the detailed admixture history of each island separately are henceforth presented in the chronological order of the first time in an islands' history when more than 100 individuals are reported in historical census information, and after which further census reports always exceeded this number until today; hereafter called "first perennial peopling" of an island (**Figure F7**, **Table ST2**). As in the main text of the article, note that this historical date can vastly differ from the date for the first official settlement of an island inferred via political, administrative and tax records, and witnessed by parishes foundations, illustrating the complex and versatile peopling history of each Cabo Verdean island after the initial establishment of the Portuguese dominion over the archipelago.

Admixture history of Santiago (First official settlement 1460; Perennial peopling: 1470)

MetHis-ABC inferences detailed results for Santiago

For individuals born on Santiago island, we find that the most recent pulses of admixture from Africa and Europe, respectively, occurred after the official abolition of the TAST in the Portuguese Empire in the 1830's, and after the effective abolition of slavery in the Portuguese empire in the 1860's (**Figure F7**). Both admixture pulses are characterized by posterior distributions with narrow 95%CIs clearly departing from their respective priors, both for the timing of each pulse and their respective intensities, in particular for the European admixture pulse (**Figure SF24-26**, **Table ST5**). For these recent events, observed genetic patterns are consistent with a relatively intense African admixture event (median=0.5913, mode=0.6837, 95%CI=[0.0642–0.9038]), occurring between generation sixteen and seventeen after the founding of Cabo Verde at generation 0 (generation median=16.6, mode=17.2, 95%CI=[10.0–18.9]). Furthermore, we find strong evidence for the most recent European admixture pulse being weak (median=0.0684, mode=0.0258, 95%CI=[0.0028–0.6021]), and having occurred between one and two generations before the sampled generation (generation median=18.7, mode =19.1, 95%CI=[11.7–19.6]).

NN-ABC posterior parameter estimations perform less accurately to infer the dates of the earlier admixture pulses and their associated intensities, both from Africa and Europe respectively, as shown by posterior distribution departing less clearly from their respective priors and overall large 95%CI. Nevertheless, results indicate that both the African and European pulses may have occurred in a more remote past; the African pulse was inferred at generation 9.2 (median) after founding at generation 0 (mode=13.9, 95%CI=[1.7–16.8]); the European pulse at generation 7.5 (median, mode=3.4, 95%CI=[2.3–15.9]), towards the beginning of the plantation economy era in the Americas and during the intensification of the TAST in the second half of the 17th century (35,36). Conversely, the initial founding admixture pulse, set to occur at generation 0 at the beginning of Cabo Verde settlement history

556 in the 1460's, is more accurately estimated *a posteriori*, as indicated by substantial departure from the
557 prior distribution and relatively narrow CI. Results here indicate a stronger contribution from Africa
558 than Europe (median=0.6365, mode=0.8816, 95%CI=[0.0538–0.9858]), at the root of the genetic make-
559 up of Santiago-born individuals in our sample set.

560 Finally, effective population sizes in Santiago indicate a strong, relatively linear (u_{Ne}
561 median=0.365, mode=0.440, 95%CI=[0.150–0.494]), increase since the 1460's up to 80,077 effective
562 individuals (median, mode=92,109, 95%CI=[30,408–98,975]), in the sampled generation born on
563 average between the 1960's and the 1980's.

564 *Discussion for the admixture history of Santiago*

565 Altogether, our results indicate that admixture in Santiago occurred mainly between Senegambian and
566 South Western European populations (**Figures F2-F3**), at the very beginning of the Portuguese
567 colonization of the archipelago in the 15th century. This early admixture history is likely due to an initial
568 peopling of Cabo Verde strongly biased towards solitary males (familial permanent migrations from
569 Portugal being a minority), who reproduced with African enslaved women, and whose admixed
570 offspring constituted the first Cabo Verdean generations, as shown by some recorded instances of
571 legitimizations of the admixed offspring to allow them to legally inherit dating back to the early 16th
572 century (37).

573 Our results further indicate that two substantial admixture events from Africa and Europe,
574 respectively, occur concomitantly to the early expansion of the TAST during the 17th-century, although
575 we are unable to precisely date both events. Nevertheless, our results support a more tenuous genetic
576 admixture history from Africa experienced by Cabo Verdeans from Santiago during the most intense
577 period of the TAST (18th-century) and the height of plantation economy in European colonial empires
578 (35,36). This apparent reduced influence of the most intense period of the TAST on African admixture
579 pattern in Santiago is also echoed in our ASD-MDS, ADMIXTURE, and SOURCEFIND results, where
580 Santiago Cabo Verdeans today share ancestry with Senegambian Mandinka almost exclusively (**Figures**
581 **F2-F3**). However, historical records unquestionably demonstrated that numerous other populations from
582 West Central and South Western Africa have been enslaved and forcibly deported to Cabo Verde during
583 this era (37). Thus, our results highlight that the strong and recurrent demographic migrations of
584 enslaved Africans brought to Santiago during the second half of the TAST do not seem to have left a
585 similar genetic admixture signature. This is consistent with several, mutually non-exclusive, historical
586 scenarios, where enslaved African populations brought to Santiago may have been mostly re-deported
587 to the Americas and Europe after the 17th century, without substantially contributing to the genetic
588 landscape of this island, whether due to rapid re-deportation, and/or to slave-owners and colonial society
589 strongly preventing marriages and reproduction between newly arrived enslaved individuals and pre-
590 existing enslaved or non-enslaved communities in Santiago (see **Discussion** in main text).

591 Conversely, our results indicate an intense introgression event from Africa in Santiago, occurring
592 shortly after the TAST and the abolition of slavery in the Portuguese empire in the 1860's. This may
593 have been caused by a relaxation of the socio-economic constraints that predominated during most of
594 the TAST against marriages between enslaved and non-enslaved communities as well as the control of
595 marriages among enslaved individuals imposed by slave-owners; concomitant with the strong
596 intensification of the illegal slave-trade which deported numerous enslaved Africans in a short amount
597 of time shortly after the end of the TAST (37–40).

598 Our results also clearly identify a European introgression event in Santiago having occurred at the
599 beginning of the 20th century, which may be explained by the European migrations towards African
600 colonies triggered by European empires' policies during the first half of this century. Interestingly, we
601 find that this recent admixture event was of very limited intensity, thus indicating that European

602 emigration to Santiago during the 20th century only had a quantitatively limited, albeit precisely
603 detectable, influence on genetic admixture patterns.
604 Finally, our results for the relatively constantly increasing reproductive population in Santiago over
605 the course of the peopling of Cabo Verde is consistent with this island being the administrative historical
606 capital, and principal entry to the archipelago throughout its' history (38,39).
607 Santiago-born individuals also show the lowest ROH levels and generally some of the shortest
608 ROH in Cabo Verde (**Figure F4** and **Figures SF9-SF10**). When examining ancestry patterns within
609 ROH in Santiago-born individuals, although proportions within ROH are on average consistent with an
610 individuals' overall genetic ancestry, there are some outlier individuals with much higher European
611 ancestry within ROH compared to the remainder of the genome. This could possibly reflect historical
612 marriage prohibitions and other social constraints against intercommunity reproductions (see
613 **Discussion** in main text). Our ROH results combined with the evidence for the relatively constantly
614 increasing reproductive population in Santiago over the course of the peopling of Cabo Verde, was
615 consistent with this island being the administrative historical capital, and principal entry to the
616 archipelago throughout its' history.
617

618 ***Admixture history of Fogo (First official settlement 1460; Perennial settlement: 1570)***

619 *MetHis-ABC inferences detailed results for Fogo*
620 For individuals born on Fogo island, we find that both European admixture pulses were precisely
621 inferred with NN-ABC, for the time of their occurrence and for their respective intensities, as posterior
622 parameter distributions and 95%CI substantially depart from their respective priors (**Figure F7**, **Figure**
623 **SF24-26**, **Table ST6**). The most recent European admixture pulse in Fogo occurred concomitantly to
624 the abolition of the TAST and of slavery in the Portuguese empire in the 1860's (generation
625 median=15.5 after the founding of Cabo Verde at generation 0, mode=16.7, 95%CI=[6.7–18.2]), and
626 was of substantial intensity (median=0.2428, mode=0.2101, 95%CI=[0.0157–0.8485]). The older
627 European pulse occurred much more remotely in the past (generation median=5.3, mode=2.5,
628 95%CI=[1.9 –13.9]), shortly after the first perennial census above 100 individuals in 1570 for this island,
629 and long before the increase in TAST intensity due to the expansion of Plantation Economy in the
630 Americas in the second half of the 17th century. Furthermore, note that our results indicate that this
631 European pulse is of relatively similar intensity compared to the most recent one (median=0.2234,
632 mode=0.1149, 95%CI=[0.0223–0.8383]).
633 Interestingly, we find that the African admixture history of individuals from Fogo occurred at very
634 different periods of time than the European ones, the timing for both African admixture pulses being
635 also well estimated with substantial departure from the prior and relatively narrow 95%CI. Indeed, we
636 find strong indications of a very recent pulse of African admixture of overall limited intensity
637 (median=0.2034, mode=0.1174, 95%CI=[0.0120–0.8225]), having occurred around the first part of the
638 20th century (generation median=19.3 after the founding of Cabo Verde at generation 0, mode=19.5,
639 95%CI=[15.6–19.8]). Furthermore, we find that the previous African admixture pulse likely occurred
640 between six and eight generations before the sampled one (generation median=12.0 after Cabo Verde
641 founding, mode=13.3, 95%CI=[4.7–16.9]), in the midst of the most intense period of the TAST and
642 Plantation Economy in the Americas, around the 1760's. However, we cannot recover a precise
643 estimation for the intensity of this older pulse, as its' posterior parameter distribution seldom departe
644 from its' prior. We also find that the founding admixture event set to occurred at the beginning of the
645 colonization history of the archipelago in the 1460's is substantially departing from its' prior distribution

646 and very similar to the founding admixture event found for individuals born on Santiago
647 (median=0.6831, mode=0.7407, 95%CI=[0.1679–0.9600]).
648 Finally, our posterior estimation of the demographic history of Fogo-born individuals substantially
649 differs from that of Santiago. Indeed, while we find similar effective sizes towards the present between
650 the two islands (median=72,023, mode=90,370, 95%CI=[18,011–98,846]), we find that the
651 demographic increase was much more recent and rapid in Fogo (u_{Ne} median=0.150, mode=0.115,
652 95%CI=[0.055–0.447]) than in Santiago.
653
654 *Discussion for the admixture history of Fogo*
655 Our results indicate a very similar founding admixture event for Fogo island compared to that of
656 Santiago, strongly suggesting that the admixture event at the root of Fogo peopling occurred early in
657 Cabo Verde history, likely in Santiago island. This scenario is consistent with historical data suggesting
658 an initial peopling of Fogo from Santiago at the end of the 15th century and the beginning of the 16th
659 century, rather than a separate founding peopling (38,39).
660 Interestingly, we find a European pulse of admixture concomitant with the perennial peopling of
661 the island at the end of the 16th century and the establishment of Fogo as an important center of
662 agricultural and free-range cattle herding in the economic triangle between Europe, Cabo Verde, and
663 continental Africa. Later on, we identify traces of an African admixture pulse having occurred during
664 the most intense period of the TAST in the mid-18th century, thus consistent with increased African
665 enslaved individuals' forced deportations to Fogo as showed by historical records, but we could not
666 infer its intensity, suggesting the difficult identifiability of such admixture event in the genetic landscape
667 of Fogo island today with the methods here deployed.
668 This history of admixture is likely related to the intense control on the island's economic expansion
669 and regulated immigration imposed after 1532 by the central political, administrative, and commercial
670 power of Santiago, with the notable exception of direct commercial relationships allowed between Fogo
671 and Europe under the rule of the Portuguese crown (38–40). This could thus explain the limited African
672 admixture during Fogo history and more intense European admixture on this island, identified with our
673 ASD-MDS/ADMIXTURE, and SOURCEFIND results (**Figures F2-F3**). Indeed, these results showed
674 a substantially more intense South-West European admixture and a more limited Senegambian shared
675 ancestry with Fogo-born individuals than in Santiago, despite reduced diversity in European and
676 continental African source populations in both islands. Furthermore, these socio-political factors likely
677 also explain the maintaining of a relatively low reproductive population size on Fogo until very recently,
678 as opposed to the more continuous population increase on Santiago. In addition, this demographic
679 history may be further explained by the recurrent eruptions of the Fogo volcano, which unquestionably
680 chronically disrupted the island's development, although historical records showed limited direct
681 mortality due to these events and that communities fleeing the island often migrated back after each
682 cataclysm (38).
683 Finally, we identify a much more recent pulse of European admixture, having occurred in the first
684 half of the 19th century, shortly after the abolition of the TAST in the Portuguese empire, possibly due
685 to the political power and commercial shifts induced by this major socio-historical change that reshaped
686 socio-cultural relationships between enslaved individuals and non-enslaved communities as well as
687 triggered European migrations throughout the colonial empires (see **Discussion** in main text). Finally,
688 we identify precisely a weak African admixture pulse having occurred in Fogo during the 20th century
689 possibly consistent with more recent work-related migrations from Africa to Cabo Verde (40), albeit
690 with limited impact on the global genetic landscape of Fogo island observed today.

ROH patterns in Fogo are similar to Santiago, in that ROH levels are relatively low and individual ROH are relatively short (**Figure F4**, **Figure SF9-SF10**). This is consistent with Fogo's history of economic control and influx of European individuals. Interestingly, when exploring ancestry proportions within ROH relative to the remainder of the genome, we see on average more European ancestry within ROH. This could possibly reflect the historical influx of Europeans onto the island and/or a legacy of preferential marriages between individuals of shared family origin.

Admixture history of Santo Antão (First official settlement 1570; Perennial settlement: 1580)

MetHis-ABC inferences detailed results for Santo Antão

For individuals born on Santo Antão island, we find that both European admixture pulses after the founding admixture event are substantially departing from their respective priors and have relatively narrow 95%CI, both for the timing of their occurrence and their respective intensities (**Figure F7**, **Figure SF23-25**, **Table ST7**). We find that Santo Antão-born individuals have experienced a recent European admixture pulse of weak intensity (median=0.0847, mode=0.0467, 95%CI=[0.0056–0.6382]), that occurred roughly three generations before the sampled generation, at the turn of the 20th century (generation median=17.2 after Cabo Verde founding at generation 0, mode=18.1, 95%CI=[7.6–19.1]). Furthermore, we find a much more ancient European admixture pulse of strong intensity (median=0.3592, mode=0.1542, 95%CI=[0.0216–0.9522]), occurring concomitantly (generation median=5.8, mode=3.2, 95%CI=[2.6–15.6]) with the first perennial census in this island dated in 1580, and before the massive increase of TAST during the 17th century.

Interestingly, we find that the first African admixture pulse after founding at generation 0 occurred at generation 4.9 (median, mode=2.4, 95%CI=[1.9–11.9]), almost synchronically to the oldest European admixture pulse and the first perennial settlement of the island recorded in 1580. Furthermore, this African admixture pulse was intense (median=0.6255, mode=0.7574; 95%CI=[0.0451–0.9749]), which summed close to 100% with the European admixture pulse having occurred at very similar times. This obliterated any older admixture events, consistently with the wide 95%CI obtained for the intensity of the founding admixture event for this island (median $s_{Afr,0}$=0.5681, mode=0.8047, 95%CI=[0.0260–0.9839]). Importantly, our NN-ABC posterior parameter estimation fail to infer the second African admixture pulse identified by our RF-ABC scenario-choice. Indeed, the posterior distributions of both the timing of this most recent pulse and its' intensity depart little from their respective priors.

Finally, despite the relatively poor performances of our inferences of demographic changes in Santo Antão, our results are indicative of a very reduced population in this island over the course of Cabo Verdean peopling history until a very recent increase up to 41,256 individuals (median, mode=12,056, 95%CI=[3570–96,855]).

Discussion for the admixture history of Santo Antão

Our results for the admixture history of Santo Antão indicate that individuals born today on this island experienced an African-European admixture event concomitant with the first perennial settlement of the island in the 1580's, at the root of the genetic makeup of this island.

Most interestingly, we find no identifiable genetic admixture events, from Europe or Africa, throughout the height of the TAST on Santo Antão, albeit this island experienced a strong agricultural expansion during the plantation economy era which triggered recurrent African enslaved individuals' deportations to this island, as shown by historical records (38–40). Thus, even more strikingly than observed in Santiago, our results indicate that the enslaved-African forced deportations to Santo Antão during most of the TAST did not leave a statistically significant genetic contribution to the genetic

landscape of the island. This could be due to a transient migration of enslaved individuals used temporarily as labor-force on Santo Antão, before deportation to the Americas and Europe during the TAST, and/or to sociocultural factors strongly disfavoring inter-marriages between enslaved and non-enslaved communities, and constraining reproductive integration between new enslaved individuals and pre-established enslaved communities, during the plantation economy era on the island (see **Discussion** in main text).

In this context, our ADMIXTURE and SOURCEFIND results identify two different Senegambian populations, the Mandinka and the Wolof, sharing significant African haplotypic ancestry with Santo Antão-born individuals, as well as a much more limited shared ancestry with the West Central Africa Igbo population (**Figure F3B**). Our ABC results would thus be consistent with a scenario where different enslaved-African populations from Senegambia and West Central Africa where brought together to settle Santo Antão in the early stages of the TAST in the late 16th century. Alternatively, it is possible that the Igbo shared ancestry identified in our analyses stem from another admixture event that we failed to identify. This is likely due to an overall limited shared ancestry between this population and Santo Antão identified only with haplotypic local ancestry inferences, as our *MetHis*-ABC procedure considered a single Senegambian population at the root of the admixture history of the island and a much more limited number of independent SNPs compared to the haplotypic local ancestry analysis (Fortes-Lima et al. 2021).

Finally, we find that a weak but significant European admixture pulse occurred at the turn between the 19th and the 20th centuries, substantially later than the end of the TAST and the abolition of slavery and before the 20th European colonial migrations, whose putative historical causes remain to be elucidated. Furthermore, we find that the reproductive population of Santo Antão remained overall small during its' entire history until a very recent expansion, although much more limited than the expansions observed in Santiago and Fogo. This is somewhat surprising with the census records showing that Santo Antão was the second most peopled island after Santiago during most of Cabo Verde history (**Table ST2**). Nevertheless, our results could be explained by the demographic collapse experienced by Santo Antão since the beginning of the 20th century, due to several intense famine episodes and rural-urban emigration from the island; a demographic history scenario thus much more complex than the relatively simple scenario here explored. Reflecting the historically small population size on the island, Santo Antão show high total ROH and generally quite long individual ROH among the islands (**Figures F4** and **SF9**).

Admixture history of São Nicolau (First official settlement 1580; Perennial settlement: 1580)

MetHis-ABC inferences detailed results for São Nicolau

For individuals born on São Nicolau island, we find that the two most recent admixture pulses from Europe and Africa, respectively, are precisely estimated with posterior distributions substantially departing from their priors and reduced 95%CI, for the pulses' timing and intensities (**Figure F7**, **Figure SF24-26**, **Table ST8**). We find that the most recent pulse of European admixture is weak in intensity (median=0.0731, mode=0.0297, 95%CI=[0.0038–0.6642]), and occurred very recently at the generation before that of the sampled individuals (generation median=19.0 after the founding at generation 0, mode=19.2, 95%CI=[10.3–19.8]). Furthermore, we find that the most recent African admixture pulse is very intense (median=0.6398, mode=0.7132, 95%CI=[0.0877–0.9380]), and occurred at generation 14.5 after founding (median, mode=14.6, 95%CI=[8.3–16.4]), during the abolition of the TAST in the first half of the 19th century.

Interestingly, for both the more ancient admixture pulses from either Africa or Europe, our NN-ABC posterior parameter estimation provide reasonably informative posterior distributions pointing towards relatively weak African and European admixture intensities (median=0.3133, mode=0.1695, 95%CI=[0.0171–0.9262]; and median=0.3240, mode=0.1193, 95%CI=[0.0145–0.9460], respectively). However, our results only provide indications that both pulses were ancient and occurred synchronically (generation median=5.2 after founding, mode=2.4, 95%CI=[1.3–15.5]; generation median=6.7, mode=2.7, 95%CI=[2.0–16.1], for the African and European older pulses respectively), as posterior distributions for the timing of these events are largely confounded with their respective priors.

Finally, we obtain a posterior distribution of the founding admixture intensity largely confounded with its prior, similarly as for the demographic parameters, thus showing a limit of our ABC inferences procedures.

Discussion for the admixture history of São Nicolau

Our results for the admixture history of São Nicolau indicate that two moderate admixture pulses from Africa and Europe occurred synchronically between the date for the first perennial settlement of this island in the 1580s and the onset of the intensification of the TAST at the beginning of the 17th century. Much later in the course of Cabo Verde peopling history at the beginning of the 19th century, we find an intense introgression event from Africa, consistent with massive number of enslaved-African forcible deportations during this era, before the abolition of the TAST in the Portuguese Empire. Interestingly, our haplotypic local ancestry inferences identify three African populations with shared ancestries with São Nicolau-born individuals (**Figure F3**); Senegambian Wolof and Mandinka shared-ancestry representing the majority, and South West Central African Angolan Ovimbundu a small minority. Given the large intensity of the most recent African admixture pulse and the majority of shared ancestry from Senegambian populations here identified, it is likely that both African admixture events in São Nicolau involved a majority of enslaved-Africans from this continental African region only 500 km away from Cabo Verde, also consistent with Senegambia being the major African commercial partner of the archipelago throughout history (37–39). However, and analogous to the Santo Antão inferences' limitations, our analyses cannot disentangle whether both African pulses in São Nicolau involved the limited South West Central African populations shared ancestry here identified, or whether these populations where brought in a single migration wave. Furthermore, note that the second African pulse appears to have arrived after the end of the agro-slavery system and plantation economy in Cabo Verde in the 1780's (40), which further questions the historical context and determinants of this major genetic event in the history of the island. Indeed, there is, to our knowledge, little historical information about this period in the island that may more precisely explain the intense African admixture pulse here identified, in particular bearing in mind that it occurred during a period of time when socio-economic and legal constraints against marriages and reproductions between enslaved and non-enslaved communities were very prevalent, and when slave-owners largely controlled reproduction and marriages between newly arrived enslaved individuals and pre-existing enslaved communities.

Finally, we identify, much later during the 20th century, a weak event of European introgression for the genetic admixture history of São Nicolau. This latter European admixture event may be accounted for by the European work-related immigrations during the first half of the 20th century similarly as for other islands, and in particular due to the expansion of fishing and cannery industry in São Nicolau at this time (40).

Altogether, our genetic data contain little identifiable information for the demographic history of this island. This may be due to the recurrent demographic collapses of the permanent settlement in this island due to intense starvation events, epidemic outbreaks, and destructive pirate raids that affected the

826 peopling of this island throughout history (**Table ST2**). However, one genetic pattern that was consistent
827 with these recurrent demographic collapses is the high total levels of ROH observed in the data (**Figure**
828 **F4** and **SF9-SF10**).
829

830 ***Admixture history of Brava (First official settlement 1580; Perennial settlement: 1580)***

831 *MetHis-ABC inferences detailed results for Brava*

832 For individuals born on Brava island, we find that the most recent pulse of European admixture is
833 relatively weak (median=0.1697, mode=0.1281, 95%CI=[0.0188–0.6504]), and occurred towards the
834 abolition of slavery in the Portuguese empire in the second half of the 19th century (generation
835 median=16.3 after the founding of Cabo Verde at generation 0, mode=17.9, 95%CI=[4.2–19.3]), as
836 indicated by posterior distributions of parameters for this event largely differing from their priors
837 (**Figure F7**, **Figure SF24-26**, **Table ST9**). We identify a similarly intense pulse of admixture from
838 Africa (median=0.2473, mode=0.1261, 95%CI=[0.0191–0.8266]), for which our inferences only bore
839 indications that it occurred during the second half of the 18th century during the TAST, as the 95%CI
840 for the timing of this event is wide (generation median=13.6 after founding at generation 0, mode=15.7,
841 95%CI=[2.4–19.2]).
842 More remotely in the past, we identify strong indications that a very weak European admixture
843 pulse (median=0.0352, mode=0.0157, 95%CI=[0.0007–0.7239]) may have occurred at the beginning of
844 the 17th century (generation median=7.3 after founding, mode=3.6, 95%CI=[2.8–16.1]). Furthermore,
845 we identify a strong signal for an intense African admixture pulse (median=0.6688, mode=0.7403,
846 95%CI=[0.1358–0.9685]), occurring at the beginning of Cabo Verdean peopling history, during the
847 early 16th century, at generation 2.8 after the founding of the archipelago at generation 0 (median,
848 mode=1.6, 95%CI=[1.1–12.2]), and thus largely before the first perennial settlement of this island.
849 Finally, our results strongly support a very limited effective size linear increase (u_{Ne} median=0.420,
850 mode=0.435, 95%CI=[0.243–0.494]) throughout the entire history of the island reaching 7337 effective
851 individuals (median, mode=3278, 95%CI=[647–73,114]) in the present.
852

853 *Discussion for the admixture history of Brava*

854 Our results for the admixture history of individuals born on Brava, in light of the historical record
855 indicating a first perennial peopling of this island in the 1580's (**Table ST2**), strongly suggest that Brava
856 was initially peopled by individuals originating from another Cabo Verdean island, where their ancestors
857 had already experienced an intense genetic introgression from Africa. Indeed, we find an intense first
858 African admixture pulse roughly three generations prior the first official and perennial peopling of the
859 island at this date (**Figure F7**). We hypothesize that Fogo admixed individuals were originally at the
860 root of this peopling, as the patterns of shared haplotypic ancestry in Brava-born individuals strongly
861 resemble that of Fogo, rather than Santiago, with shared ancestry exclusively with Senegambian
862 Mandinka and in similar proportion, as well as a significant (albeit very reduced) signal for a shared
863 ancestry with Tuscan TSI otherwise only found in Fogo (**Figure F3**).
864 Then, we find a limited introgression event from Europe during the late 17th century, possibly
865 consistent with a known immigration event from Madeira and the Açores, and from Fogo after a volcano
866 eruption, around this time (39). We also find a relatively weak African admixture pulse at some point
867 during the TAST for which we were unable to obtain a more precise timing, and which will require
868 further investigations.
869 Finally, we identify a more intense pulse of European admixture shortly after the abolition of
870 slavery in the Portuguese empire in the second half of the 19th century. This later event may be due to

871 the expansion of the whaling industry at this time, particularly in Brava due to favorable maritime
872 currents and routes for this industry, which brought European and European-North American sailors to
873 the island and increased locally contacts with Europe, and North America. Most interestingly, this
874 historical scenario is also consistent with the signal, virtually unique across Cabo Verde, of a significant
875 albeit reduced shared haplotypic ancestry between Brava-born individuals and the British GBR.
876 Notably, our results indicate that the effective population of Brava remained low during the entire
877 peopling history of Cabo Verde and until today, consistent with this island being the smallest inhabited
878 and the most south-west ward in the archipelago. Curiously, in spite of the historically low effective
879 population size, Brava shows low total ROH levels and relatively short ROH lengths (**Figure F4** and
880 **SF9-SF10**). As in Fogo, we see on average more European ancestry within ROH. This could possibly
881 reflect the historical influx of Europeans to the island and/or a legacy of preferential marriages between
882 individuals of shared family origin.
883

884 ***Admixture history of Maio (First official settlement 1529; Perennial settlement: 1650)***

885 *MetHis-ABC inferences detailed results for Maio*
886 For individuals born on Maio island, we find evidence that the most recent pulse of European admixture
887 was of relatively modest intensity (median=0.1472, mode=0.1134, 95%CI=[0.0233–0.5957]), and likely
888 occurred during the 18th century, in the midst of the TAST (generation median=12.7 after founding at
889 generation 0, mode=14.1, 95%CI=[3.7–17.9]), as both posterior distributions substantially departe from
890 their respective priors (**Figure F7**, **Figure SF24-26**, **Table ST10**). Although our RF-ABC scenario-
891 choice favors several pulses of admixture from both sources, our NN-ABC posterior parameter
892 estimation fails to capture sufficient information to infer satisfactorily the most recent pulse of admixture
893 from Africa, as both the posterior distribution of the timing of this event and its intensity seldom depart
894 from their respective priors and have wide 95%CI.
895 However, we capture substantial information strongly suggesting two, much older, admixture
896 pulses from Europe and Africa, respectively, occurring virtually synchronically in the middle 16th
897 century (generation median=4.1 after founding for the European pulse, mode=2.3, 95%CI=[1.7–14.2];
898 generation median=4.2 after founding for the African pulse, mode=2.4, 95%CI=[1.5–12.8]), thus much
899 earlier than the first perennial settlement of this island dated in 1650 by historians (**Table ST2**).
900 Furthermore, we find that the respective intensities of the two pulses (median=0.3260 for the European
901 pulse, mode=0.1075, 95%CI=[0.0117–0.9497]; median=0.5589 for the African pulse, mode=0.7703,
902 95%CI=[0.0346–0.9769]), sum close to 100%. This suggests a large replacement of the Maio population
903 at that time, with indications that these synchronic admixture events occur on the basis of a founding
904 admixed population largely of African origin (median $s_{Afr,0=}$0.6757, mode=0.9135, 95%CI=[0.017–
905 0.421]).
906 Finally, the demographic inference of reproductive size changes in Maio indicates, similarly as for
907 Brava, that the reproductive size of the island remained relatively small during its entire history, and
908 only very recently increased up to a median 21,454 effective individuals (mode=7159), albeit with a
909 large 95%CI ([914–93,656]).
910

911 *Discussion for the admixture history of Maio*
912 Our results for the admixture history of Maio indicate a relatively simple scenario with a single pulse of
913 European and African synchronic admixture having occurred in the mid-16th century in another Cabo
914 Verdean island, long before the first perennial settlement of the island in 1650. We hypothesize that this
915 event occurred on Santiago, as shared local haplotypic ancestry patterns are highly resembling between

islands, and also consistently with Maio being located very close to Santiago and historical records showing unequivocally the strong relationships between islands throughout history (**Figure F3**).

After the descendants of this initial admixture event settled Maio, they experienced only a single identifiable admixture pulse of weak intensity from Europe, which likely occurred at the turn between the 18th and the 19th century during the TAST, and no clearly identifiable further signal of African admixture in the island. As for some other Cabo Verde islands, we cannot conclude about the timing of the African admixture event that brought the reduced signal of West Central and South West Central Africa shared haplotypic ancestry with Maio, observed in our ADMIXTURE and SOURCEFIND analyses (**Figure F3**).

Finally, the reduced population size throughout the history of Maio is consistent with the island being located close to Santiago and with much less water resources, having made it difficult to maintain and expand the settlement on this island, despite its first peopling having occurred early in the history of Cabo Verde almost exclusively dedicated to free-range cattle herding (37–39).

ROH patterns are also consistent with this history, with high total ROH indicating long term isolation and population bottlenecks. Similar to Santiago there were some outlier individuals with much higher European ancestry within ROH compared to the remainder of the genome (**Figure F4** and **SF9-SF10**). This could possibly reflect historical marriage prohibitions and other social constraints against intercommunity reproductions (see **Discussion** in main text).

Admixture history of Boa Vista (First official settlement 1566; Perennial settlement: 1650)

MetHis-ABC inferences detailed results for Boa Vista

For the admixture history of individuals born on Boa Vista, we find an intense admixture pulse from Africa (median=0.6553, mode=0.7516, 95%CI=[0.1262–0.8949]), that occurred in the mid-19th century (generation median=15.0 after Cabo Verde founding at generation 0, mode=14.8, 95%CI=[11.0–17.4]), during the abolition of the TAST, with posterior distributions for both these parameters substantially departing from their respective priors and relatively reduced 95%CI (**Figure F7**, **Figure SF24-26**, **Table ST11**). Prior to that event, we fail to estimate the timing and intensity of the older African admixture pulse into Boa Vista, as posterior parameter distributions seldom depart from their respective priors for this event.

However, our results strongly support a scenario with an intense period of recurring monotonically decreasing admixture from Europe (median admixture intensity at the period's start=0.8057, mode=0.9219, 95%CI=[0.2736–0.9907]; median admixture intensity at the period's end=0.340, mode=0.120, 95%CI=[0.021–0.880]), between the first official settlement of the island in 1566 (generation median at the start of the European period of admixture=3.9 after founding, mode=2.2, 95%CI=[1.7–11.8]), until shortly after its first perennial peopling dated in 1650 (generation median at the end of the European period of admixture=8.2, mode=8.6, 95%CI=[3.1–14.9]). Interestingly, the initial founding admixture event at the root of the genetic patterns of Boa Vista-born individuals supports a substantially lower amount of African admixture (median=0.3917, mode=0.1706, 95%CI=[0.1814–0.6678]), compared to the rest of Cabo Verde where this parameter could be reliably identified.

Finally, our results show that the demography of Boa Vista remained relatively constant throughout the history of Cabo Verde, and indicate a possible very recent increase in effective size, albeit posterior parameter estimation of the most recent effective size was ambiguous due to large 95%CI (median Ne_{20}=47,744, mode=13,128, 95%CI=[2657–98,587]).

Discussion for the admixture history of Boa Vista

Altogether, our results for the admixture history of Boa Vista born-individuals indicate a largely differing admixture history compared to all other islands of the archipelago. First, we find that the admixture event at the root of the peopling of the island is largely biased towards European admixture, as opposed to all other islands in the archipelago. We find that this initial admixture event occurred before the mid-16th, concomitantly with the first official settlement of the island. We then find a recurring gene-flow from Europe into Boa Vista until its first perennial settlement in the 1650s. Interestingly, this corresponds to a period when Boa Vista seems to have been used by Cabo Verde as a penitentiary island as well as grounds for free-range herding, which may explain the continuous European admixture process at the root of the genetic landscape of this island (37–39).

Later-on towards the end of the TAST in the first half of the 19th century, Boa Vista experienced a strong pulse of African admixture, while no more recent African admixture pulse is identifiable in our inferences. This massive event likely reshaped the genetic admixture patterns of Boa Vista-born individuals, strongly increasing the African gene-pool from Senegambia. The timing inferred for this most recent admixture event in Boa Vista corresponds to the very end of the TAST in the Portuguese empire, a period characterized by intense African legal and illegal enslaved individuals' deportations, and corresponding to the maximal census of enslaved individuals in Boa Vista from the overall limited historical records specific to this island (**Table ST2**).

Finally, the effective size of Boa Vista remained constant and low throughout the history of this island, which echoes its' small historical census sizes, mainly due to limited water resources on the island, up until today. ROH patterns are also consistent with this history, with the highest total ROH among the Cabo Verdean islands indicating long term isolation and population bottlenecks (**Figures F4** and **SF9-SF10**).

Admixture history of São Vicente (First official settlement 1570; Perennial settlement: 1780)

MetHis-ABC inferences detailed results for São Vicente

For the admixture history of individuals born on São Vicente, we find a recent European admixture pulse of substantial intensity (median=0.2840, mode=0.2596, 95%CI=[0.0397–0.8044]), having likely occurred towards the end of the 19th century (generation median=17.4 after founding of Cabo Verde at generation 0, mode=17.7, 95%CI=[10.6–18.9]); both posterior distributions have relatively narrow 95%CI and clearly depart from their respective priors (**Figure F7**, **Figure SF24-26**, **Table ST12**).

This most recent admixture event is preceded by a period of recurrent decreasing African admixture likely spanning the second half of the most intense period of the TAST starting in the early 18th century (generation median at the start of the African period of admixture=9.7, mode=12.6, 95%CI=[2.5–16.9]), and ending during the end of the TAST and the abolition of slavery in the Portuguese Empire in the mid-19th century (generation median at the end of the African period of admixture=15.8, mode=17.5, 95%CI=[7.4–19.1]), albeit our results are ambiguous for the estimate of the start of this admixture period as indicated by the large 95%CI. Furthermore, our results are also ambiguous to estimate the intensity of the African introgression during the period of recurring admixture, as the posterior distributions for the onset and offset intensities, as well as the shape of the monotonic recurring admixture do not clearly depart from their respective priors and had large 95%CIs.

We find traces of an older European admixture pulse of substantial intensity (median=0.3403, mode=0.1772, 95%CI=[0.0149–0.9133]), which likely occurred at the beginning of the 17th century (generation median =6.4 after founding, mode=2.8, 95%CI=[1.9–15.1]). Our results indicate that this first admixture event occurred on a root-population that exhibited a large African admixture proportion

(median $s_{Afr,0}$=0.5983, mode=0.8596, 95%CI=[0.0529–0.9835]), albeit this result should be considered with cautious as CI is overall wide.

Finally, our results clearly indicate a strong effective population expansion since the perennial peopling of the island in the late 18th century, up to 66,715 effective individuals today (median, mode=85,838, 95%CI=[12,159–98,375]).

Discussion for the admixture history of São Vicente

Our results for the admixture history of individuals born on São Vicente indicate a strong European admixture event concomitant with the first official peopling of the island occurring at the same time as the nearby island of Santo Antão, consistently with historical records (**Table ST2**). However, the perennial peopling of this island was difficult due to the very limited water resources, in particular compared to the much more hospitable, larger, and nearby Santo Antão. Thus, São Vicente essentially served as a coal and salt harbor depot during most of its history with a minimal settlement (40), thus possibly consistent with the initial European admixture pulse identified in these analyses (**Figure F7**), and our local haplotypic shared ancestry showing a single European source of shared genetic ancestry with São Vicente (**Figure F3**).

Later-on and unique in Cabo Verde, our results support a long period of recurring African admixture occurring during the entire second half of the TAST and ending with the abolition of slavery in the Portuguese empire. We thus obtained informative results concerning a substantial African admixture pulse occurring during the mid-19th century shortly after the end of the TAST in the Portuguese empire. The small haplotypic ancestry shared between São Vicente-born individuals and West Central and/or South West Central African population may stem from work-related forced or voluntary African migrations having occurred during this period when illegal slave-trade is known to have surged. However, we could not satisfactorily estimate the onset timing and associated admixture intensity for this recurrent admixture event.

Interestingly, São Vicente census only very recently started to increase during this period, with newly available water sources and a protected deep-water harbor in the bay of Mindelo, unique in Cabo Verde, which favored the massive economic expansion of the island in the late 19th century and throughout the 20th century (**Table ST2**). Furthermore, the end of the agro-slavery system and end of plantation economy in Cabo Verde at the end of the 18th century and beginning of the 19th century are known historically to have triggered emigration from Santo Antão and rural-urban migrations during the 19th and 20th century, essential for São Vicente perennial peopling (40,41). This is further consistent with extensive familial grand-parental birth-places in Santo Antão and extensive familial relationships reported by São Vicente-born individuals in our familial anthropology interviews.

Altogether, these historical, economic, and demographic processes may well explain the onset of the large reproductive population increase identified in our analyses, as well as haplotypic local ancestry patterns largely resembling those obtained for Santo Antão (**Figure F3**). Furthermore, they are consistent with the substantial European admixture pulse identified at the turn of the 19th and 20th century with our analyses. However, the detailed history of African admixture prior to the early 19th century largely remains to be assessed for this island.

Reflecting the historical small population size, São Vicente had high total levels of ROH (**Figure F4**). Interestingly, individuals from this island, on average, show more African ancestry in ROH compared to the remainder of the genome. This could possibly reflect historical marriage prohibitions or the historically high levels of recurrent African admixture, or both.

Admixture history of Sal (First official settlement 1529; Perennial settlement: 1841)

MetHis-ABC inferences detailed results for Sal

Finally, the admixture history of individuals born on Sal inferred with NN-ABC is overall ambiguous *a posteriori* (**Figure F7**, **Figure SF24-26**, **Table ST13**). Indeed, while our RF-ABC scenario-choice leaned towards a complex admixture history with a period of recurring admixture from Europe and two African admixture pulses after the initial founding admixture event for the population of Sal, the duration and intensity of the European admixture period, as well as the first African admixture pulse, all showed posterior parameter distributions seldom departing from their priors. Nevertheless, we clearly identify a recent African admixture pulse of moderate intensity (median=0.2401, mode=0.1780, 95%CI=[0.0197–0.6590]), which likely occurred at the end of the 19th century (generation median=17.6 after the founding of Cabo Verde at generation 0, mode=17.9, 95%CI=[12.5–19.0]). Furthermore, our results indicate a strong and very recent effective size expansion in Sal, up to 63,289 individuals (median, mode =88,248, 95%CI=[5689–98,181]).

Discussion for the admixture history of Sal

Sal was among the early Cabo Verdean islands to be officially settled and exploited, in 1529 (**Table ST2**), as a source of sea-salt both exported and employed locally in the free-range herding meat-production and tannery industry at the historical root of Cabo Verde economy with Europe and Africa (38,41). However, its perennial peopling dated only from 1841, the last perennially peopled island of Cabo Verde, as water sources and agricultural surfaces are scarce on this island. This slow demographic expansion is well captured by our inference showing a reduced population size during most Cabo Verde history on Sal, followed very recently by a strong reproductive size increase. Furthermore, in this context, our results clearly identifying only a recent admixture pulse of African origin at the beginning of the 20th century could reflect the recent economic migrations of Western African populations to this island, during the development of its salt mining activities and its travel and tourism activities since then. Nevertheless, our scenario-choice results suggested the occurrence of a period of recurrent European admixture in Sal, which our posterior parameter inference largely failed to identify. This is possibly due to the complex recent peopling history of Sal that we fail to capture with genetic data, which will need to be clarified in the future.

Nevertheless, reflecting the historical small population size, Sal has high total levels of ROH (**Figure F4**). Interestingly, individuals on this island, on average, show more African ancestry in ROH compared to the remainder of the genome. As in São Vicente, this could possibly reflect historical marriage prohibitions or the recent pulse of African admixture, or both.

1085 **Final methodological discussion**

1086 Overall, our NN-ABC posterior parameter inferences results show that European admixture
1087 parameters are often more accurately identified than African ones, as also shown by cross-validation
1088 posterior parameter errors obtained using simulated data (**Tables ST5-12**). It is not clear how this
1089 relative asymmetry in parameter identifiability may emerge. It is plausible that the low diversity of
1090 European populations of origin of the Cabo Verde gene-pool better fitted a priori our two source-
1091 population scenarios with respect to the European source compared to the African ones. Indeed, while
1092 we also found a relatively homogeneous African origin of the Cabo Verde gene-pool from Senegambia,
1093 our haplotypic local-ancestry results nevertheless indicated different, genetically very close, putative
1094 source African populations for several Cabo Verdean islands (**Figure F3**). As our *MetHis*-ABC
1095 procedures relied exclusively on independent markers, they were a priori not able to capture this
1096 haplotypic-based signal, but may nevertheless have been affected. Other scenario misspecifications
1097 specific to the African admixture process may further explain this identifiability asymmetry, for instance
1098 with more overlapping generations or more assortative mating in the African admixture process than in
1099 the European one. Finally, it is plausible that these observed limits of our methods arise from the
1100 European-ascertained nature of our genotyping-chip SNP set, despite the LD-pruning here performed,
1101 which may alter summary-statistics accurate estimations differently for the African and European
1102 sources.
1103
1104
1105

REFERENCES

1. Fortes-Lima CA, Laurent R, Thouzeau V, Toupance B, Verdu P. Complex genetic admixture histories reconstructed with Approximate Bayesian Computation. Mol Ecol Resour [Internet]. 2021 May [cited 2021 Jun 4];21(4):1098–117. Available from: https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13325

2. Fisher RA. Darwinian evolution of mutations. Eugen Rev. 1922 Apr;14(1):31–4.

3. Wright S. Evolution in Mendelian Populations. Genetics. 1931 Mar;16(2):97–159.

4. Verdu P, Rosenberg NA. A General Mechanistic Model for Admixture Histories of Hybrid Populations. Genetics [Internet]. 2011 Dec 1 [cited 2021 Jun 4];189(4):1413–26. Available from: https://academic.oup.com/genetics/article/189/4/1413/6089061

5. Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL. High resolution of human evolutionary trees with polymorphic microsatellites. Nature [Internet]. 1994 Mar [cited 2022 Feb 2];368(6470):455–7. Available from: http://www.nature.com/articles/368455a0

6. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics [Internet]. 1978 Jul 20 [cited 2022 Feb 2];89(3):583–90. Available from: https://academic.oup.com/genetics/article/89/3/583/5992737

7. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics [Internet]. 2011 Aug 1 [cited 2022 Feb 2];27(15):2156–8. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btr330

8. Buzbas EO, Verdu P. Inference on admixture fractions in a mechanistic model of recurrent admixture. Theoretical Population Biology [Internet]. 2018 Jul [cited 2022 Feb 2];122:149–57. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0040580918300662

9. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population-structure. Evolution [Internet]. 1984 Nov [cited 2022 Feb 2];38(6):1358–70. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1558-5646.1984.tb05657.x

10. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient Admixture in Human History. Genetics [Internet]. 2012 Nov 1 [cited 2022 Feb 2];192(3):1065–93. Available from: https://academic.oup.com/genetics/article/192/3/1065/5935193

11. Csilléry K, François O, Blum MGB. abc: an R package for approximate Bayesian computation (ABC): *R package: abc*. Methods in Ecology and Evolution [Internet]. 2012 Jun [cited 2021 Jun 4];3(3):475–9. Available from: http://doi.wiley.com/10.1111/j.2041-210X.2011.00179.x

12. Pudlo P, Marin JM, Estoup A, Cornuet JM, Gautier M, Robert CP. Reliable ABC model choice via random forests. Bioinformatics [Internet]. 2016 Mar 15 [cited 2021 Jun 4];32(6):859–66. Available from: https://academic.oup.com/bioinformatics/article/32/6/859/1744513

13. Raynal L, Marin JM, Pudlo P, Ribatet M, Robert CP, Estoup A. ABC random forests for Bayesian parameter inference. Stegle O, editor. Bioinformatics [Internet]. 2019 May 15 [cited 2022 Feb 2];35(10):1720–8. Available from: https://academic.oup.com/bioinformatics/article/35/10/1720/5132692

14. Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population genetics. Genetics. 2002 Dec;162(4):2025–35.

15. Jay F, Boitard S, Austerlitz F. An ABC Method for Whole-Genome Sequence Data: Inferring Paleolithic and Neolithic Human Expansions. Hernandez R, editor. Molecular Biology and Evolution [Internet]. 2019 Jul 1 [cited 2022 Feb 2];36(7):1565–79. Available from: https://academic.oup.com/mbe/article/36/7/1565/5345566

16. Mathias RA, Taub MA, Gignoux CR, Fu W, Musharoff S, O'Connor TD, et al. A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. Nat Commun [Internet]. 2016 Nov [cited 2021 Jun 4];7(1):12522. Available from: http://www.nature.com/articles/ncomms12522

17. Moreno-Estrada A, Gravel S, Zakharia F, McCauley JL, Byrnes JK, Gignoux CR, et al. Reconstructing the Population Genetic History of the Caribbean. Tarazona-Santos E, editor. PLoS Genet [Internet]. 2013 Nov 14 [cited 2021 Jun 4];9(11):e1003925. Available from: https://dx.plos.org/10.1371/journal.pgen.1003925

18. Pereira L, Černý V, Cerezo M, Silva NM, Hájek M, Vašíková A, et al. Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of the Tuareg nomads from the African Sahel. Eur J Hum Genet. 2010 Aug;18(8):915–23.

19. Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, Byrnes JK, et al. Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. Schierup MH, editor. PLoS Genet [Internet]. 2012 Jan 12 [cited 2022 Feb 2];8(1):e1002397. Available from: https://dx.plos.org/10.1371/journal.pgen.1002397

20. Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K, et al. The African Genome Variation Project shapes medical genetics in Africa. Nature [Internet]. 2015 Jan 15 [cited 2021 Jun 4];517(7534):327–32. Available from: http://www.nature.com/articles/nature13997

21. Busby GB, Band G, Si Le Q, Jallow M, Bougama E, Mangano VD, et al. Admixture into and within sub-Saharan Africa. eLife [Internet]. 2016 Jun 21 [cited 2022 Feb 2];5:e15266. Available from: https://elifesciences.org/articles/15266

22. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000 Jun;155(2):945–59.

23. Rosenberg NA. Genetic Structure of Human Populations. Science [Internet]. 2002 Dec 20 [cited 2021 Jun 4];298(5602):2381–5. Available from: https://www.sciencemag.org/lookup/doi/10.1126/science.1078311

24. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 2003 Aug;164(4):1567–87.

25. Lawson DJ, van Dorp L, Falush D. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nat Commun [Internet]. 2018 Dec [cited 2021 Jun 4];9(1):3258. Available from: http://www.nature.com/articles/s41467-018-05257-7

26. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Research [Internet]. 2009 Sep 1 [cited 2021 Jun 4];19(9):1655–64. Available from: http://genome.cshlp.org/cgi/doi/10.1101/gr.094052.109

27. Gravel S. Population Genetics Models of Local Ancestry. Genetics [Internet]. 2012 Jun 1 [cited 2021 Jun 4];191(2):607–19. Available from: https://academic.oup.com/genetics/article/191/2/607/5935164

28. Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, et al. A Genetic Atlas of Human Admixture History. Science [Internet]. 2014 Feb 14 [cited 2021 Jun 4];343(6172):747–51. Available from: https://www.sciencemag.org/lookup/doi/10.1126/science.1243518

29. Gonçalves R, Rosa A, Freitas A, Fernandes A, Kivisild T, Villems R, et al. Y-chromosome lineages in Cabo Verde Islands witness the diverse geographic origin of its first male settlers. Human Genetics [Internet]. 2003 Nov 1 [cited 2021 Jun 4];113(6):467–72. Available from: http://link.springer.com/10.1007/s00439-003-1007-4

30. Beleza S, Johnson NA, Candille SI, Absher DM, Coram MA, Lopes J, et al. Genetic Architecture of Skin and Eye Color in an African-European Admixed Population. Spritz RA, editor. PLoS Genet [Internet]. 2013 Mar 21 [cited 2021 Jun 4];9(3):e1003372. Available from: https://dx.plos.org/10.1371/journal.pgen.1003372

31. Beleza S, Campos J, Lopes J, Araújo II, Hoppfer Almada A, e Silva AC, et al. The Admixture Structure and Genetic Variation of the Archipelago of Cape Verde and Its Implications for Admixture Mapping Studies. Relethford JH, editor. PLoS ONE [Internet]. 2012 Nov 30 [cited 2021 Jun 4];7(11):e51103. Available from: https://dx.plos.org/10.1371/journal.pone.0051103

32. Hamid I, Korunes KL, Beleza S, Goldberg A. Rapid adaptation to malaria facilitated by admixture in the human population of Cabo Verde. eLife [Internet]. 2021 Jan 4 [cited 2021 Jun 4];10:e63177. Available from: https://elifesciences.org/articles/63177

33. Szpiech ZA, Mak ACY, White MJ, Hu D, Eng C, Burchard EG, et al. Ancestry-Dependent Enrichment of Deleterious Homozygotes in Runs of Homozygosity. The American Journal of Human Genetics [Internet]. 2019 Oct [cited 2021 Jun 4];105(4):747–62. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0002929719303374

34. Thompson EA. Identity by Descent: Variation in Meiosis, Across Genomes, and in Populations. Genetics [Internet]. 2013 Jun 1 [cited 2021 Jun 4];194(2):301–26. Available from: https://academic.oup.com/genetics/article/194/2/301/6065409

35. Eltis D, Richardson D. Atlas of the transatlantic slave trade. 2015. 336 p.

36. Fortes-Lima C, Verdu P. Anthropological genetics perspectives on the transatlantic slave trade. Human Molecular Genetics [Internet]. 2021 Apr 26 [cited 2021 Jun 4];30(R1):R79–87. Available from: https://academic.oup.com/hmg/article/30/R1/R79/6040749

37. Carreira A. Cabo Verde: formação e extinção de uma sociedade escravocrata (1460-1878). 3rd ed. Praia, Cabo Verde: IPC; 2000. 522 p. (Estudos e ensaios).

38. Albuquerque L de, Santos MEM, editors. História geral de Cabo Verde T1. Vol. 1. Lisboa : Praia [Cape Verde]: Centro de Estudos de História e Cartografia Antiga, Instituto de Investigação Científica Tropical ; Instituto Nacional da Cultura de Cabo Verde; 1991. 523 p.

39. Albuquerque L de, Santos MEM, editors. História geral de Cabo Verde T2. 2. ed. Vol. 2. Lisboa: Centro de Estudos de História e Cartogr. Antiga; 1995. 596 p.

40. Albuquerque L de, Santos MEM, editors. História concisa de Cabo Verde (Resumo da História geral de Cabo Verde T1-3). Lisboa: Inst. de Investigação Científica Tropical; 2007. 427 p.

1228 41. Patterson KD. Epidemics, Famines, and Population in the Cape Verde Islands, 1580-1900. The
1229 International Journal of African Historical Studies [Internet]. 1988 [cited 2021 Jun 4];21(2):291.
1230 Available from: https://www.jstor.org/stable/219938?origin=crossref

1231

1232

Supplementary Figures and Tables

Figure SF1:

Quality-Control and datasets merging procedures.

Quality controls at the genotyping call level (Stage 1) have been conducted with Illumina's GenomeStudio® software Genotyping Module. Cabo Verde original DNA samples have been collected during six separate field-trips between 2010 and 2018, genotyped in four batches using four different versions of the Illumina Omni2.5Million Beadchip genotyping array. The resulting dataset was merged with 2,504 worldwide samples from the 1000 Genomes Project Phase 3 (1000 Genomes Project Consortium, 2015); with 1,307 continental African samples from the African Genome Variation Project (EGA accession number EGAD00001000959, Gurdasani et al. 2015); and with 1,235 African samples from Patin et al. 2017 (EGA accession number EGAS00001002078). We retained only autosomal SNPs common to all datasets, and excluded one individual for each pair of individuals related at the 2nd degree (at least one grand-parent in common) as inferred with KING (Manichaikul et al. 2010), following previous procedure (Verdu et al. 2017).

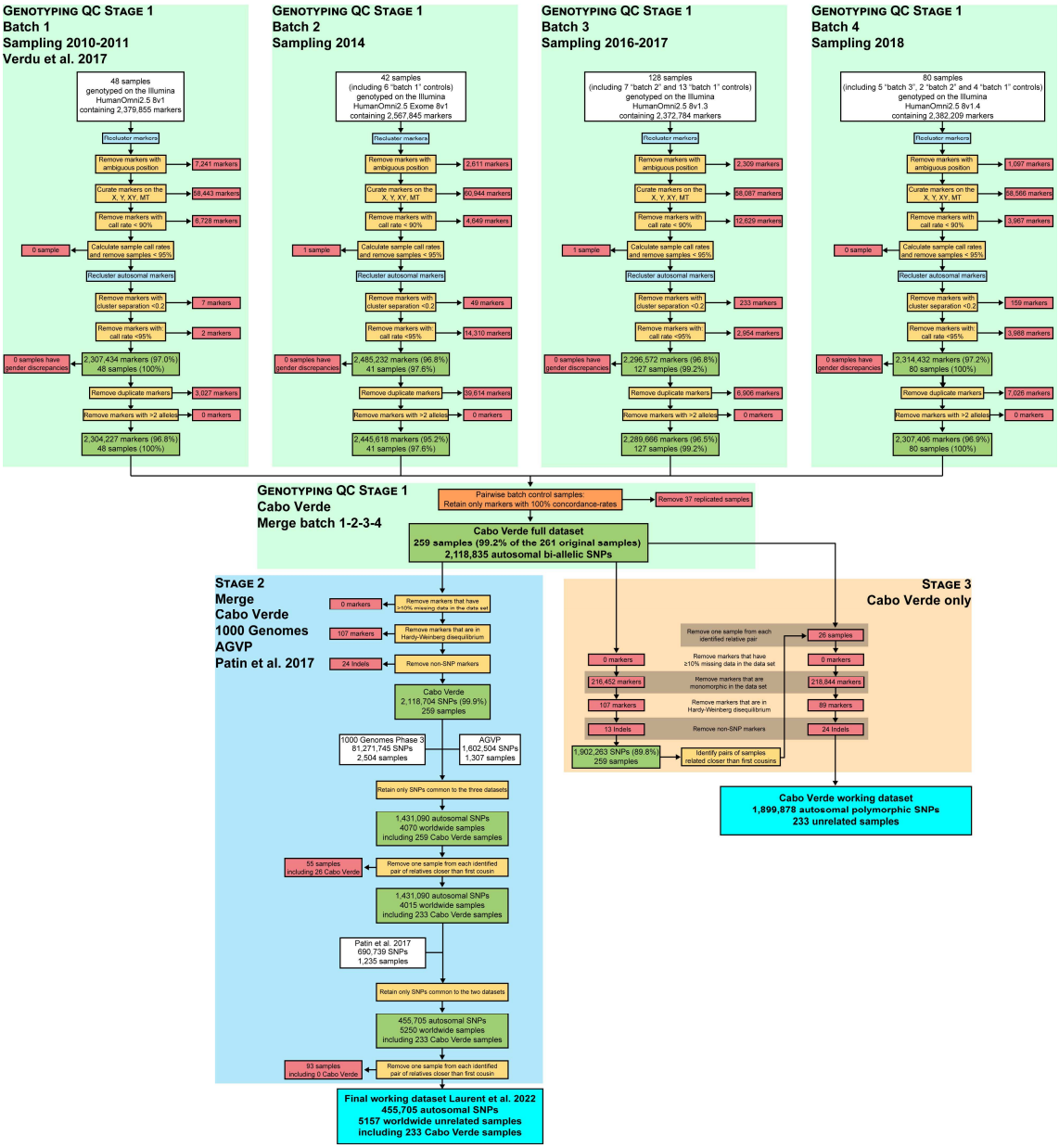


Figure SF2:

Multidimensional Scaling three-dimensional projection of Allele Sharing Dissimilarity pairwise dissimilarities, for all worldwide populations in our dataset.

All populations and individuals from our working-dataset are included in this analysis: see **Table ST1** for the population list used and **Figure F1** for sample locations and symbols.

Each individual is represented by a single point.

A) Axis 1 and 2; B) Axis 1 and 3. 3D animated plot is provided in *.gif* format.

We evaluate the Spearman correlation between 3D MDS projections and the original ASD pairwise matrix to evaluate the precision of the dimensionality reduction along the first three axes of the MDS. We find significant ($p<2.2x10^{-16}$) Spearman ρ of 0.9796.

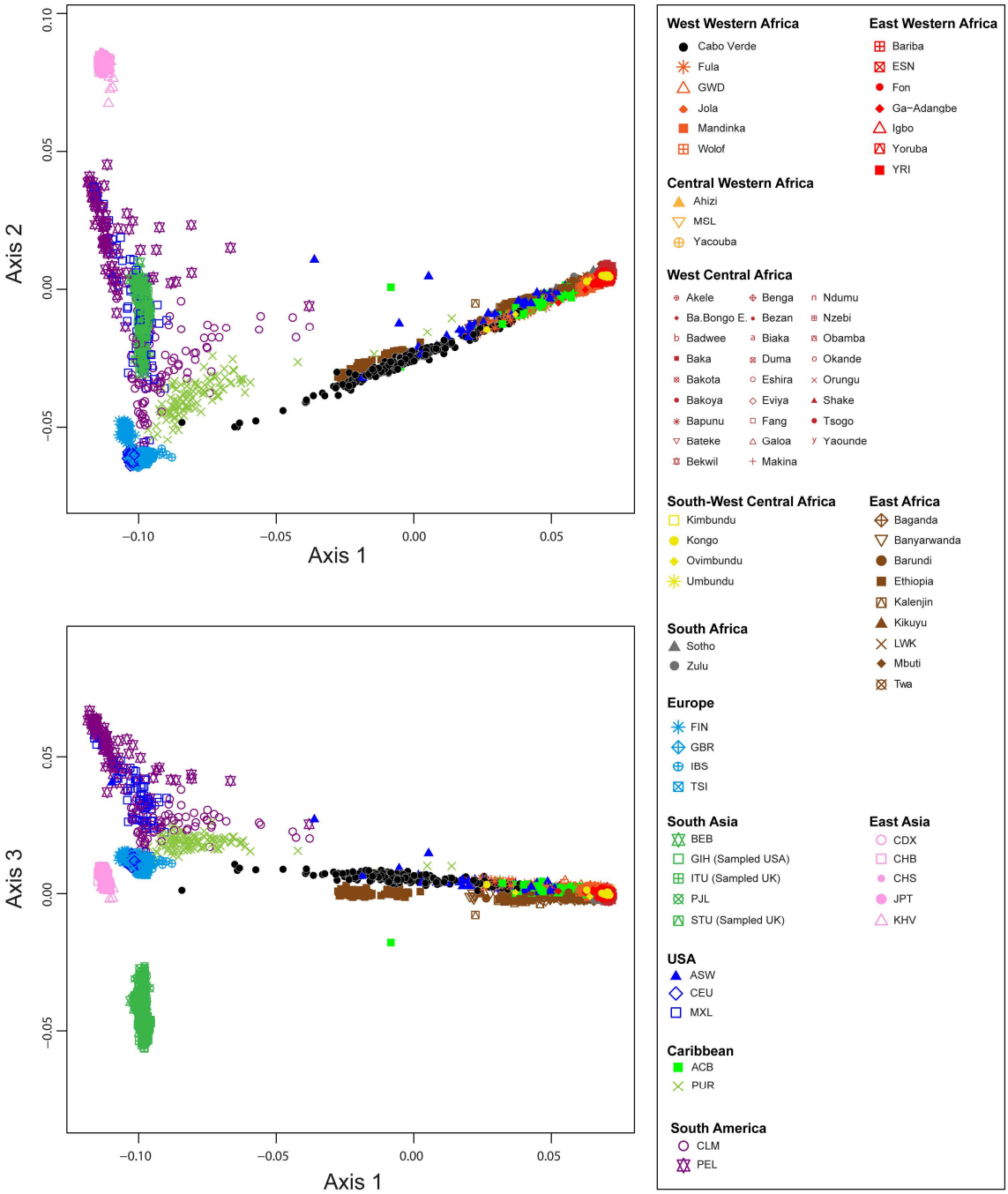
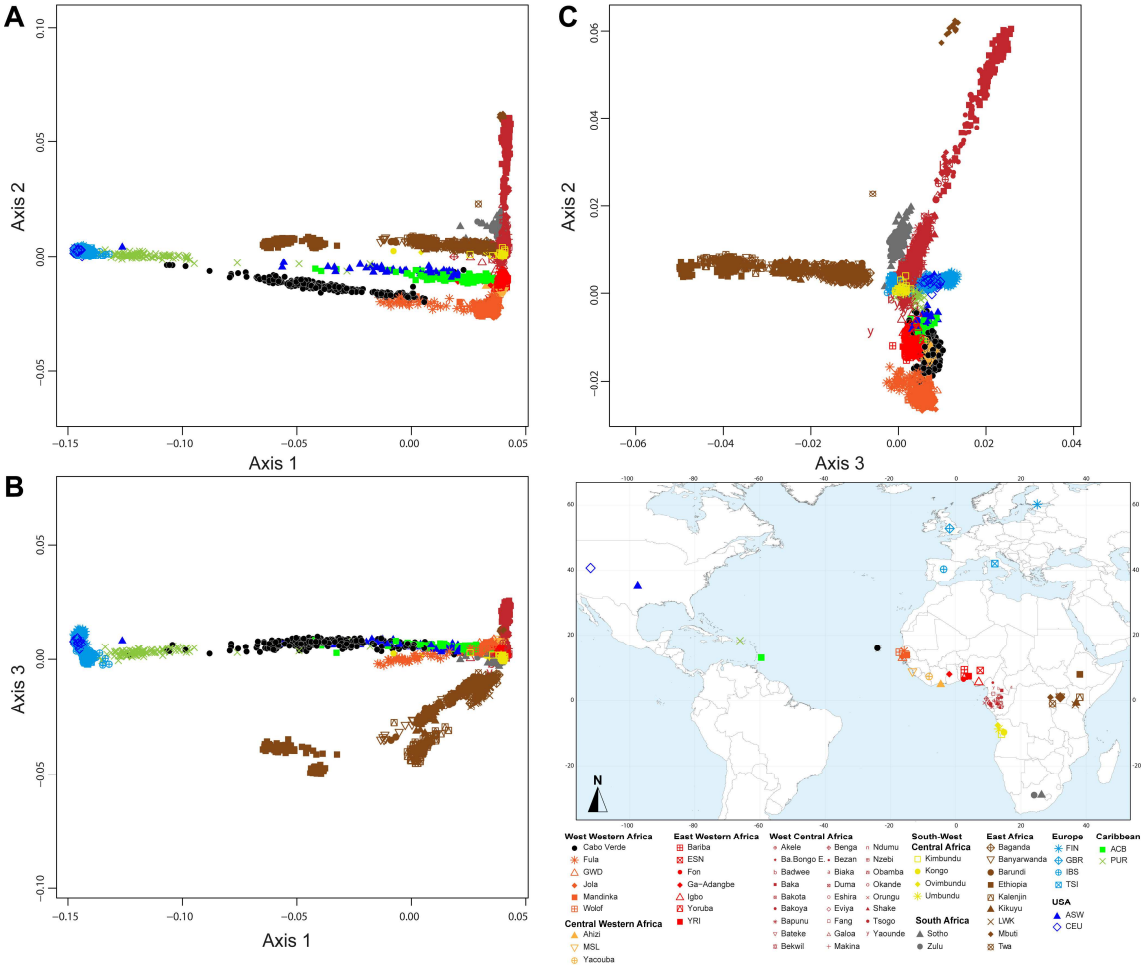


Figure SF3:

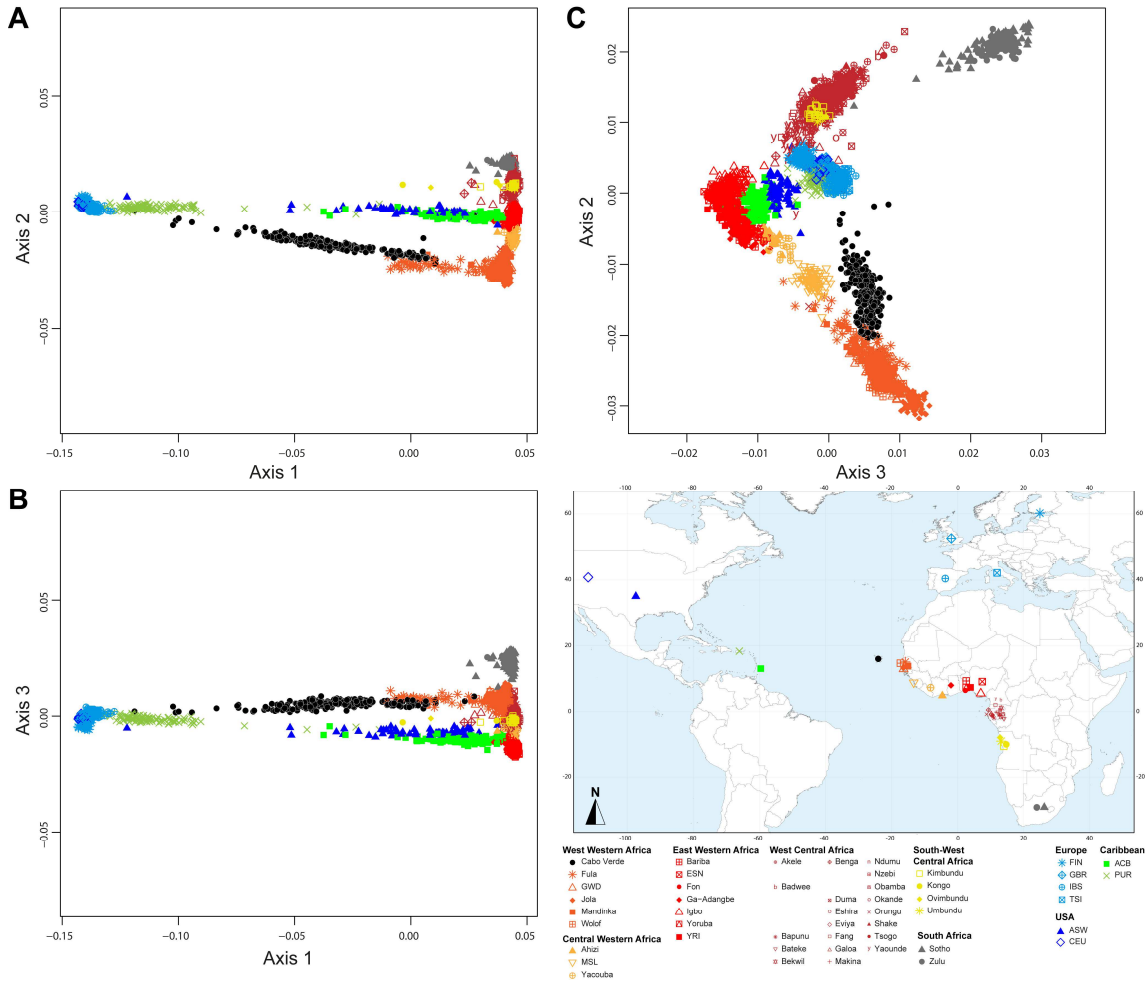
Multidimensional Scaling three-dimensional projection of Allele Sharing Dissimilarity pairwise dissimilarities, for African, European and other admixed populations related to the TAST.

We removed all Asian and South American samples compared to the sample set employed in **Figure SF2**, and recomputed the MDS based on this reduced sample set. See **Table ST1** for the population list used in these analyses. Sample map used in these analyses is extracted from **Figure F1** and provided in the bottom right corner. Each individual is represented by a single point. **A) Axis 1 and 2; B) Axis 2 and 3; C) Axis 1 and 3.** 3D animated plot is provided in *.gif* format.

We evaluate the Spearman correlation between 3D MDS projections and the original ASD pairwise matrix to evaluate the precision of the dimensionality reduction along the first three axes of the MDS. We find significant (p<2.2x10^{-16}) Spearman ρ of 0.9392.

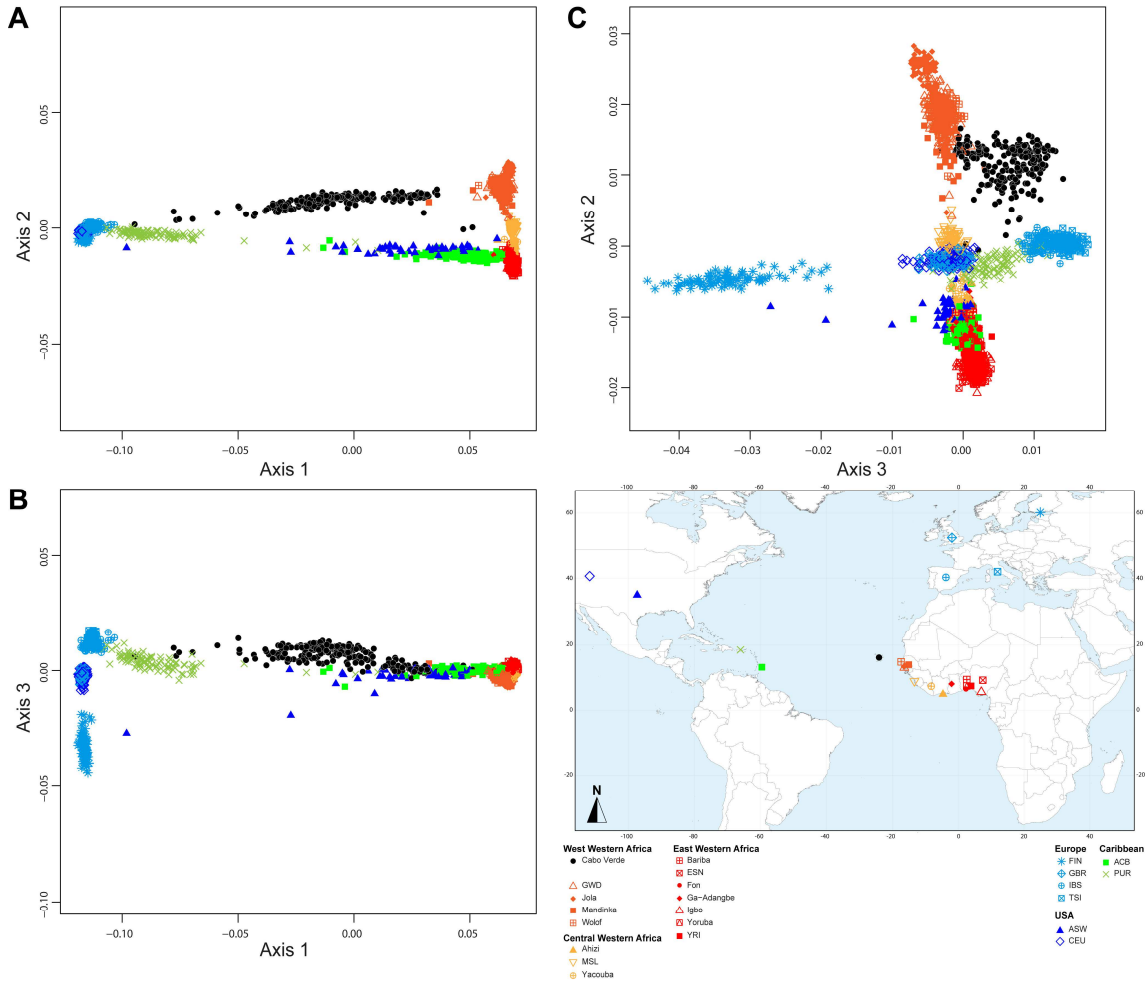


1274 **Figure SF4:**
1275 **Multidimensional Scaling three-dimensional projection of Allele Sharing Dissimilarity pairwise dissimilarities, for a**
1276 **subset of African, European and other admixed populations related to the TAST.**
1277 We removed all East African and Central African hunter-gatherer samples (Baka, Bezan, Ba.Bongo, Ba.Koya, Ba.Twa, Bi.Aka,
1278 Mbuti) compared to the sample set employed in **Figure SF3**, and recomputed the MDS based on this reduced sample set. See
1279 **Table ST1** for the population list used in these analyses. Sample map used in these analyses is extracted from **Figure F1** and
1280 provided in the bottom right corner. Each individual is represented by a single point.
1281 **A) Axis 1 and 2; B) Axis 2 and 3; C) Axis 1 and 3.** 3D animated plot is provided in *.gif* format.
1282 We evaluate the Spearman correlation between 3D MDS projections and the original ASD pairwise matrix to evaluate the
1283 precision of the dimensionality reduction along the first three axes of the MDS. We find significant (p<2.2x10^{-16}) Spearman ρ
1284 of 0.9450.
1285



1286
1287

1288 **Figure SF5:**
1289 **Multidimensional Scaling three-dimensional projection of Allele Sharing Dissimilarity pairwise dissimilarities, for a**
1290 **subset of African, European and and other admixed populations related to the TAST.**
1291 We removed all West and South-West Central African samples, as well as South African samples, compared to the sample set
1292 employed in **Figure SF4**, and recomputed the MDS based on this reduced sample set. See **Table ST1** for the population list
1293 used in these analyses. Sample map used in these analyses is extracted from **Figure F1** and provided in the bottom right corner.
1294 Each individual is represented by a single point.
1295 **A) Axis 1 and 2; B) Axis 2 and 3; C) Axis 1 and 3.** 3D animated plot is provided in *.gif* format.
1296 We evaluate the Spearman correlation between 3D MDS projections and the original ASD pairwise matrix to evaluate the
1297 precision of the dimensionality reduction along the first three axes of the MDS. We find significant (p<2.2x10^{-16}) Spearman ρ
1298 of 0.9636.
1299



1300
1301
1302

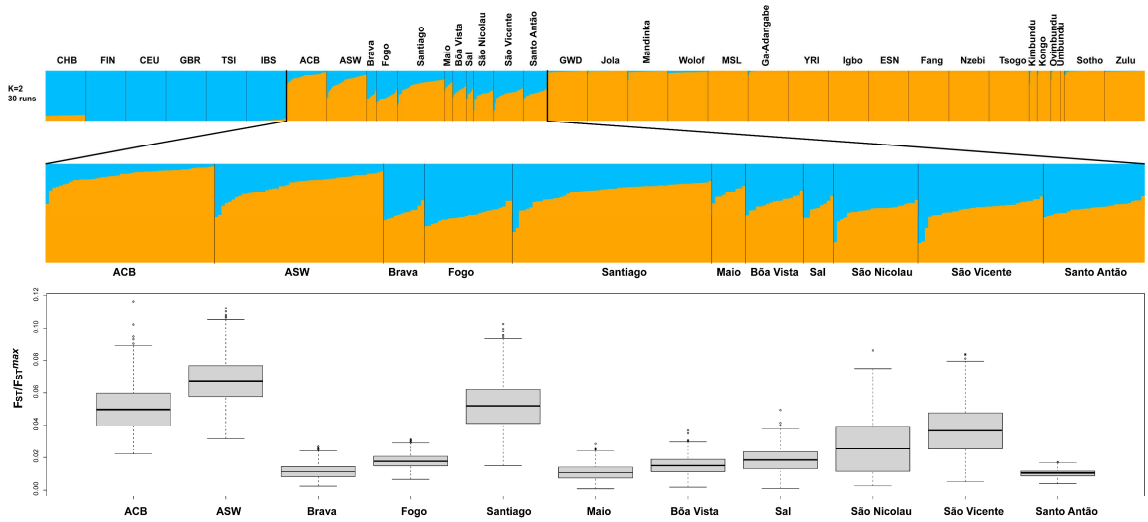
1303 **Figures SF6A-D:**
1304 **Four separate animated Multidimensional Scaling three-dimensional projections of Allele Sharing Dissimilarity**
1305 **pairwise dissimilarities, for a subset of African, European and other admixed populations related to the TAST.**
1306 Each four animated plot is provided in *.gif* format. Each individual is represented by a single point. See **Table ST1** for the
1307 population list used in these analyses, they correspond to each separate sample set provided in two-dimensional plots in **Figure**
1308 **F2**. Spearman correlations are provided in **Figure F2** caption.
1309 **Figure SF6A correspond to Figure F2 panels A-C.**
1310 **Figure SF6B correspond to Figure F2 panels D-F.**
1311 **Figure SF6C correspond to Figure F2 panels G-I.**
1312 **Figure SF6D correspond to Figure F2 panels J-L.**
1313
1314
1315

1316 **Figure SF7:**

1317 **Population *FST/FST^max^* values for the ASW, ACB, and each Cabo Verdean birth-island separately considering the**
1318 **ADMIXTURE mode result at *K=2*.**

1319 *FST/FST^max^* values were computed using FSTruct (Morrisson et al. 2021) with 1000 bootstrap replicates per population. All
1320 population pairwise distributions of bootstrap *FST/FST^max^* values were significantly different from one another after Bonferroni
1321 correction (Wilcoxon two-sided rank sum test $p<3.96\times10^{-8}$), except the following pairwise comparisons: ACB-Santiago,
1322 Brava-Maio, Fogo-Sal, and Maio-Santo Antão.

1323



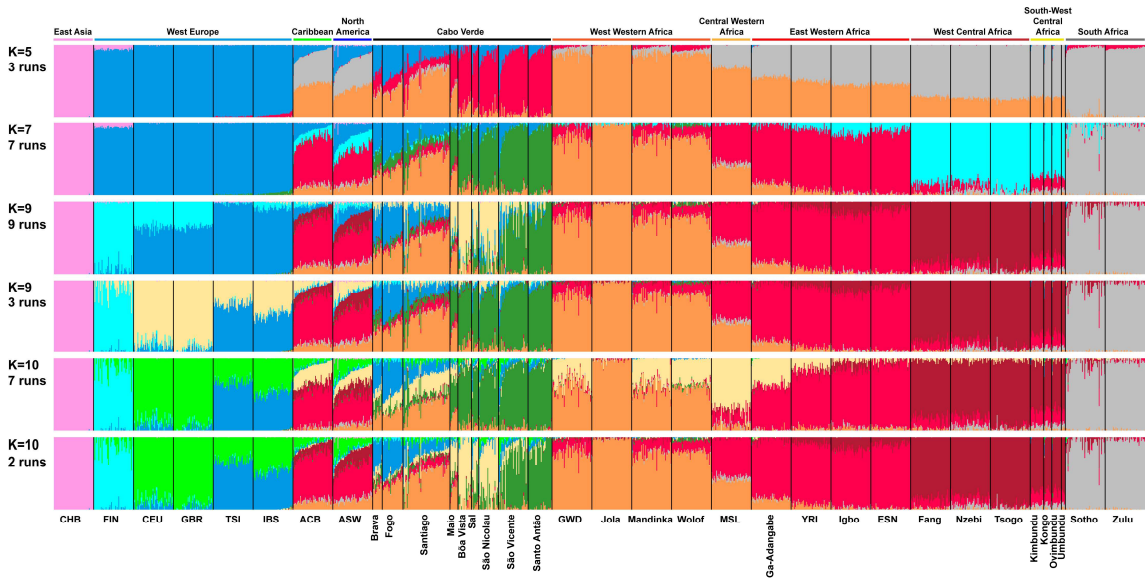
1324

1325

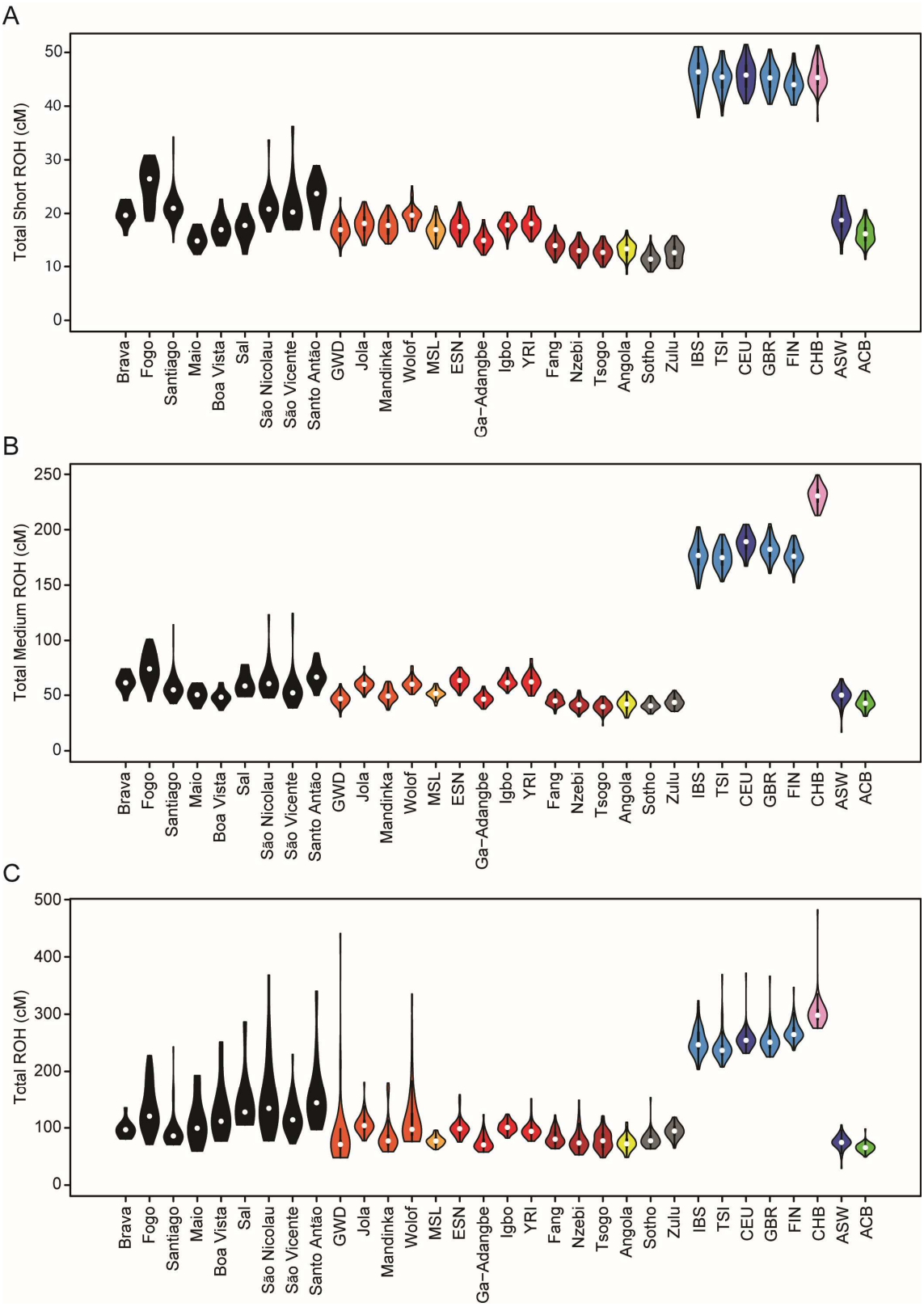
Figure SF8:

Alternative ADMIXTURE mode results for the individual genetic structure among Cabo Verdean, Barbadian-ACB and African-American ASW populations related to the TAST.

ADMIXTURE analyses using resampled individual sets for the population sets originally containing more than 50 individuals (**Table ST1**). 225 unrelated Cabo Verdean-born individuals are comprised in the analysis and grouped by island of birth (**Table ST1**). All analyses comprised 102,543 independent autosomal SNPs, for values of *K* between 2 and 10. 30 independent ADMIXTURE runs were performed for each value of *K*, and groups of runs (>2) providing similar results (all pairwise SSC > 99.9%) were averaged in a single "mode" result using CLUMPP, and plotted with DISTRUCT. Number of runs in the presented modes are indicated below the value of *K*. All other mode results are presented in **Figure F3**.

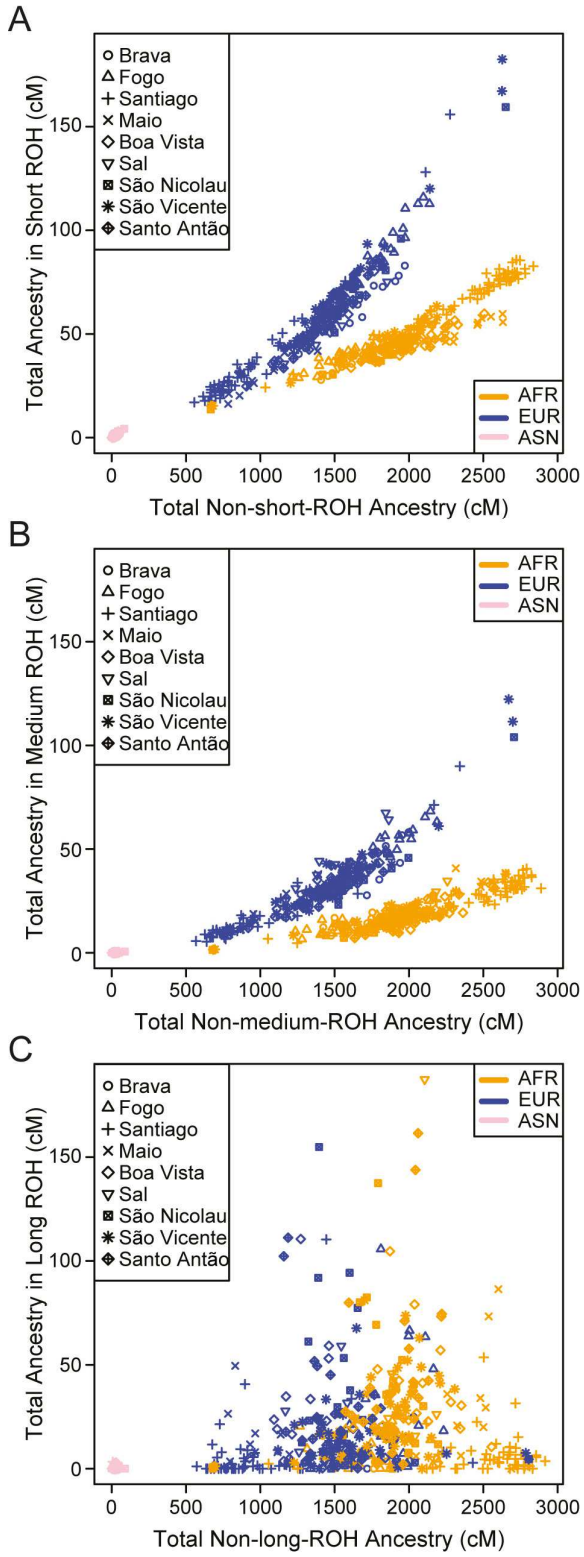


1338 **Figure SF9:**
1339 **The distribution of (A) short, (B) medium, and (C) all ROH per individual, for each Cabo Verdean birth-island and**
1340 **other worldwide populations.**
1341



1342

1343 **Figure SF10:**
1344 **Individual total ancestry contained in (A) short, (B) medium, and (C) long ROH versus total ancestry not in that class**
1345 **ROH, for each Cabo Verdean birth-island and other worldwide populations.**
1346



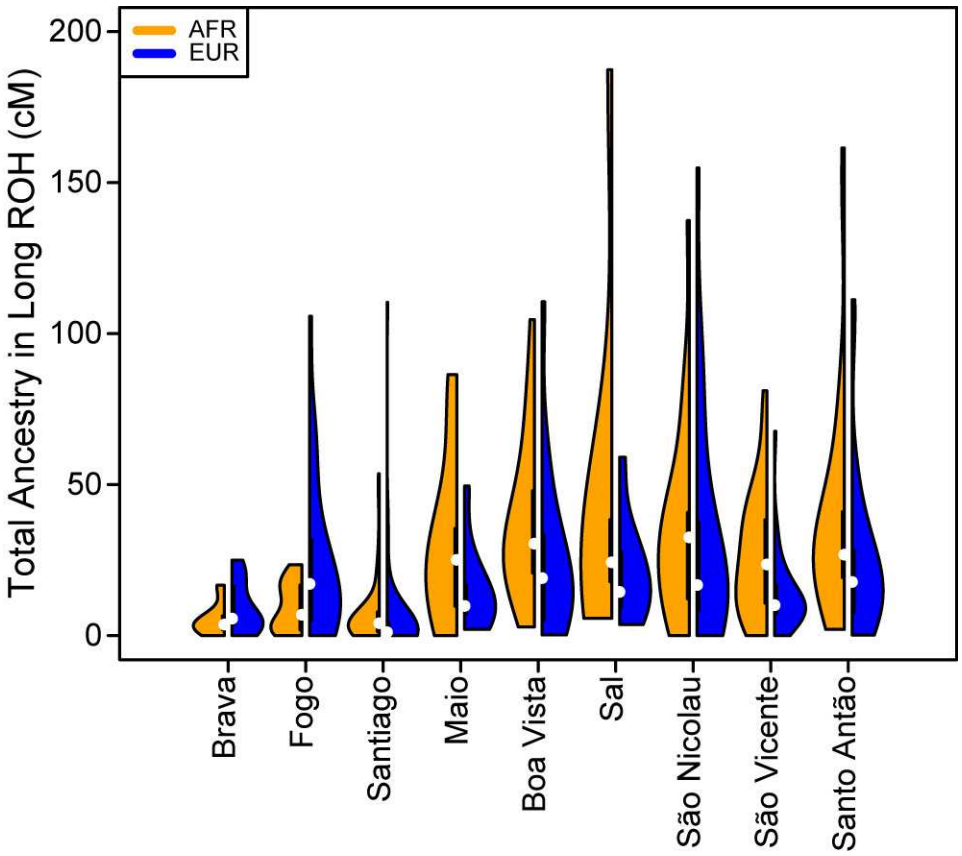
1347
1348

1349 **Figure SF11:**

1350 **The distribution of total ancestry in long ROH per individual for each Cabo Verdean birth-island.**

1351

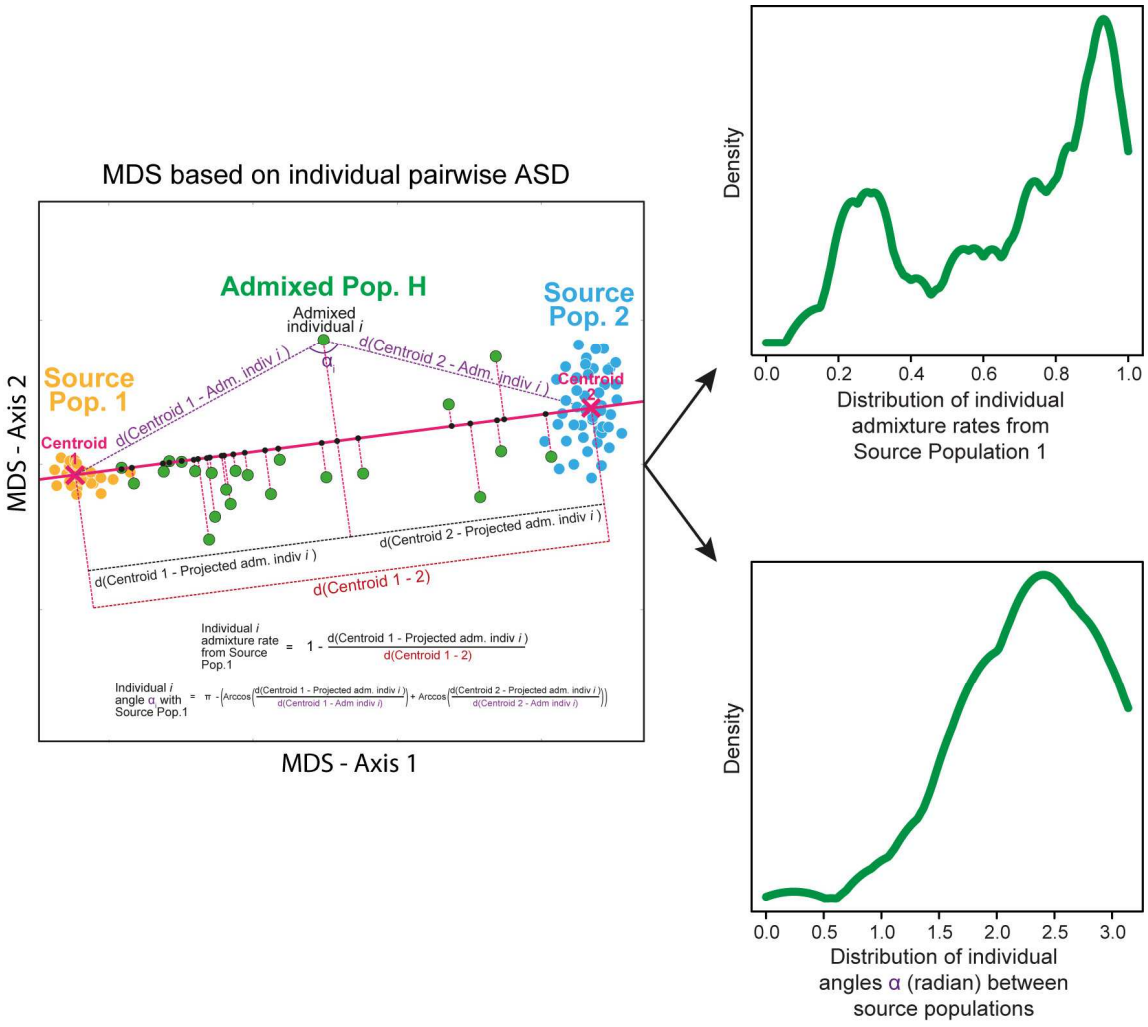
1352



1353

1354

1355 **Figure SF12:**
1356 **Schematic representation of the 2-Dimensional ASD-MDS estimates of individual admixture proportions and angles,**
1357 **also used as summary statistics for ABC inference and implemented in *MetHis* summary-statistics calculation tools.**
1358



1359
1360

Figure SF13:

ABC Prior checking for *MetHis* simulations for each island separately.

10,000 simulations are conducted under each four competing scenarios of historical admixture considered in Random-Forest ABC scenario-choice (see **Figure F6**). **A)** Goodness-of-fit histograms obtained considering the 42 summary-statistics calculated from each simulation and 1000 permutations. Observed statistic is indicated by a red vertical line**. B)** First two axes of a Principal Component Analysis performed on the 42 summary-statistics obtained for 40,000 simulations per island (10,000 simulations for each four competing-scenario), each color corresponding to a given scenario. The red white-circled dot corresponds to the vector of summary-statistics from the observed dataset. **C)** Axes 1 and 3 of the same PCA projection as in panels **B**).

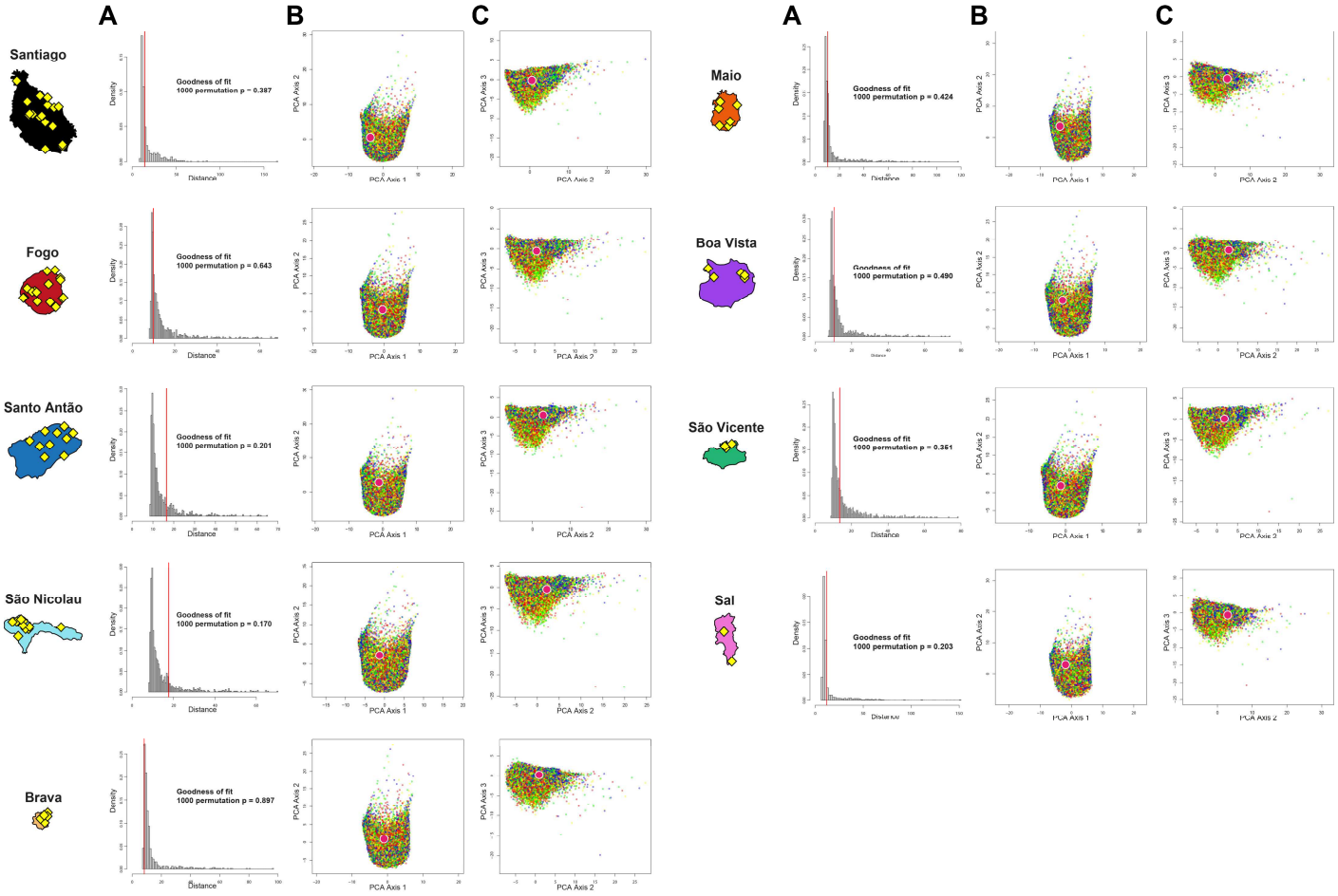


Figure SF14:

Density distributions (black line) of each 42 summary-statistics obtained from 40,000 simulations, 10,000 under each four competing scenarios, for Santiago birth-island, compared to the observed statistic obtained in this island (red vertical line).

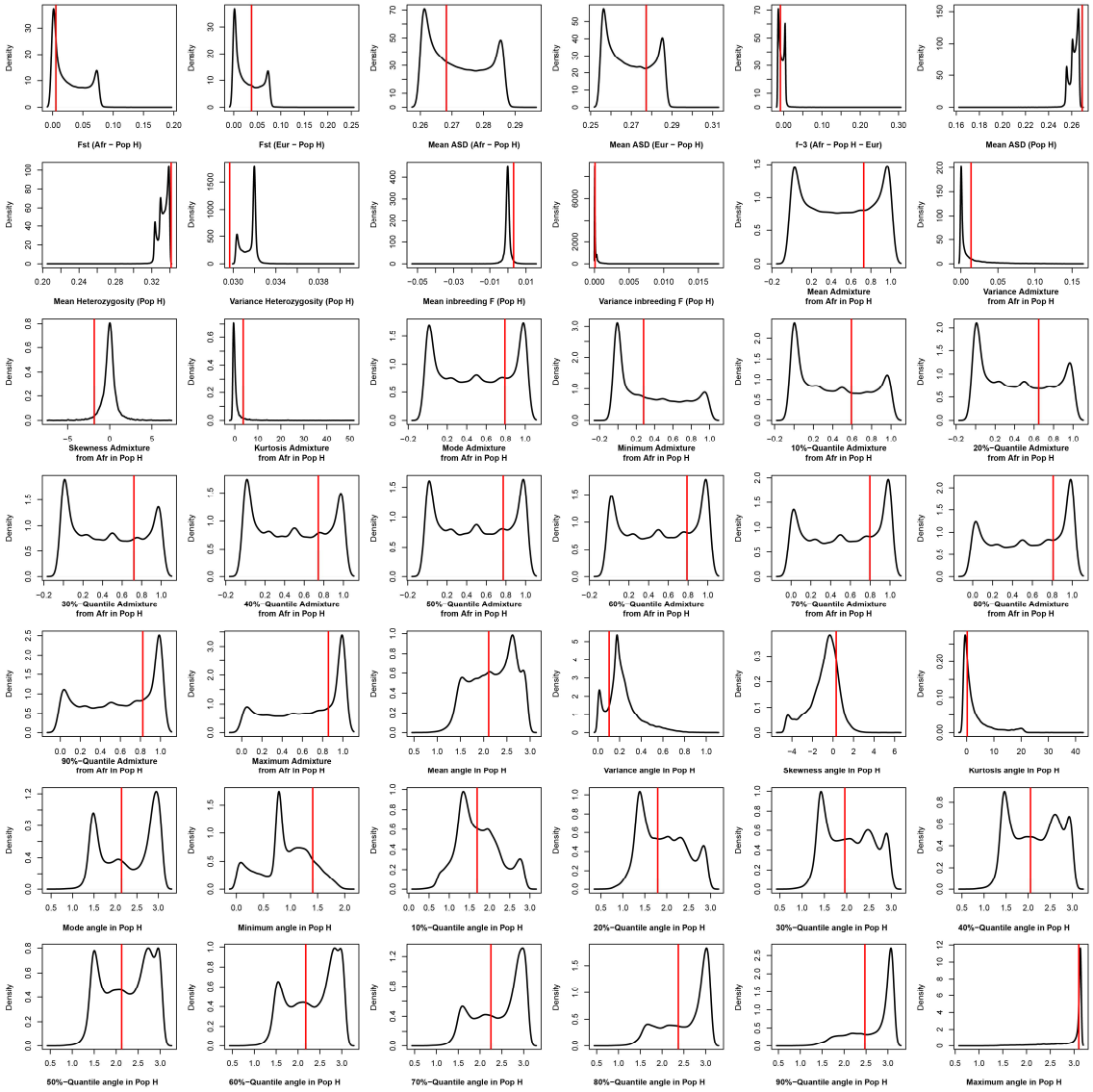


Figure SF15:

Density distributions (black line) of each 42 summary-statistics obtained from 40,000 simulations, 10,000 under each four competing scenarios, for Fogo birth-island, compared to the observed statistic obtained in this island (red vertical line).

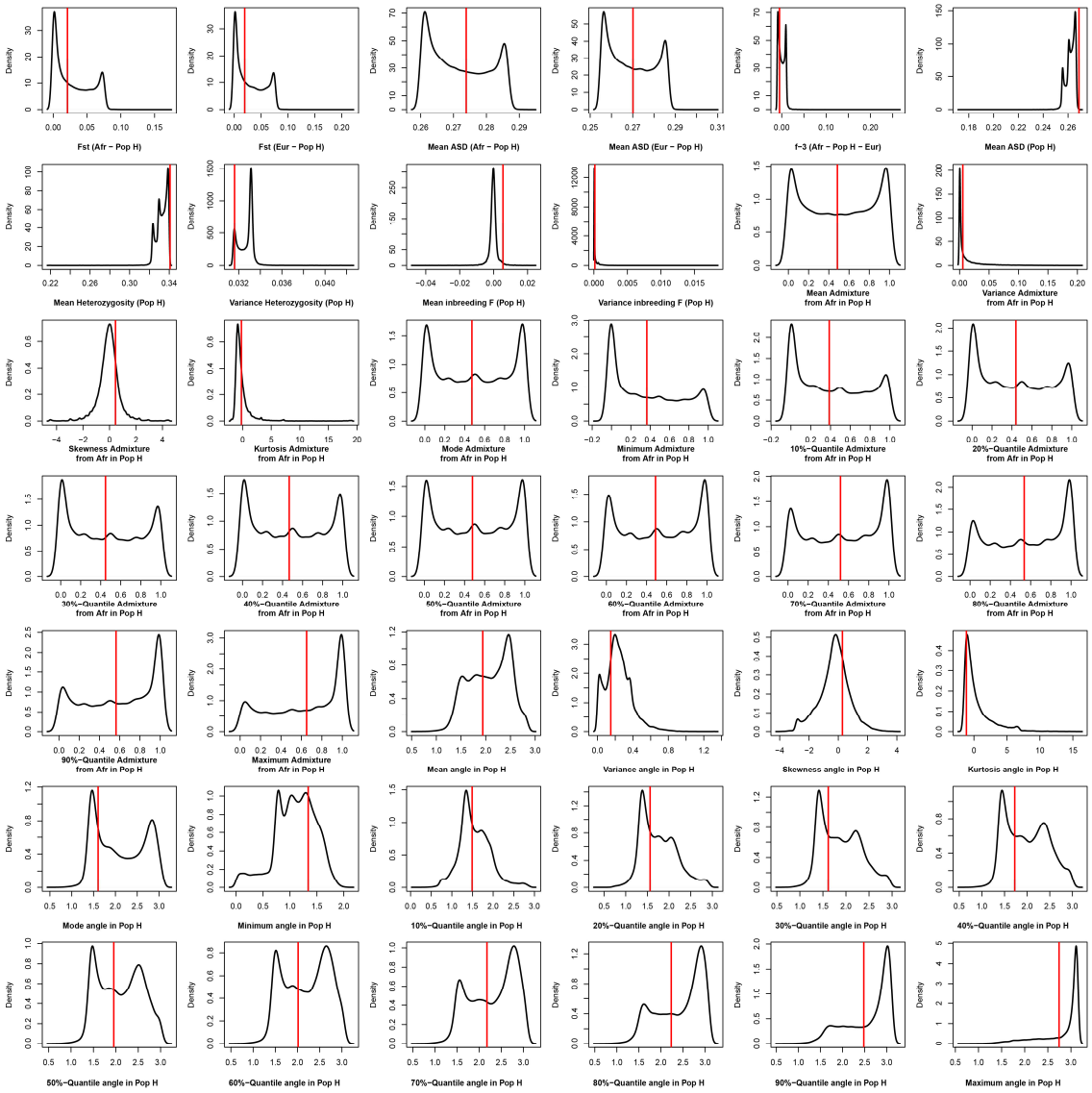


Figure SF16:

Density distributions (black line) of each 42 summary-statistics obtained from 40,000 simulations, 10,000 under each four competing scenarios, for Santo Antão birth-island, compared to the observed statistic obtained in this island (red vertical line).

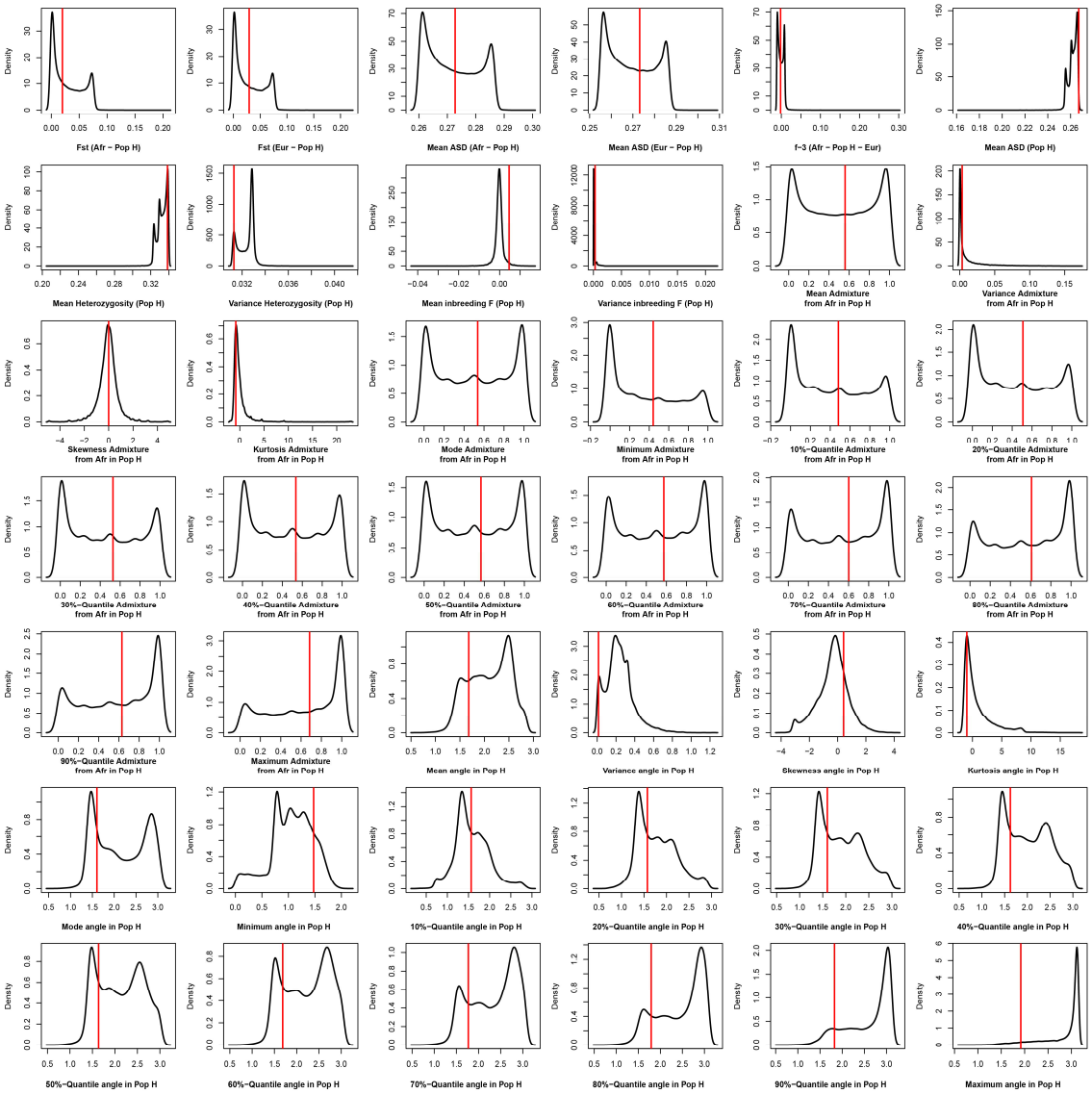


Figure SF17:

Density distributions (black line) of each 42 summary-statistics obtained from 40,000 simulations, 10,000 under each four competing scenarios, for São Nicolau birth-island, compared to the observed statistic obtained in this island (red vertical line).

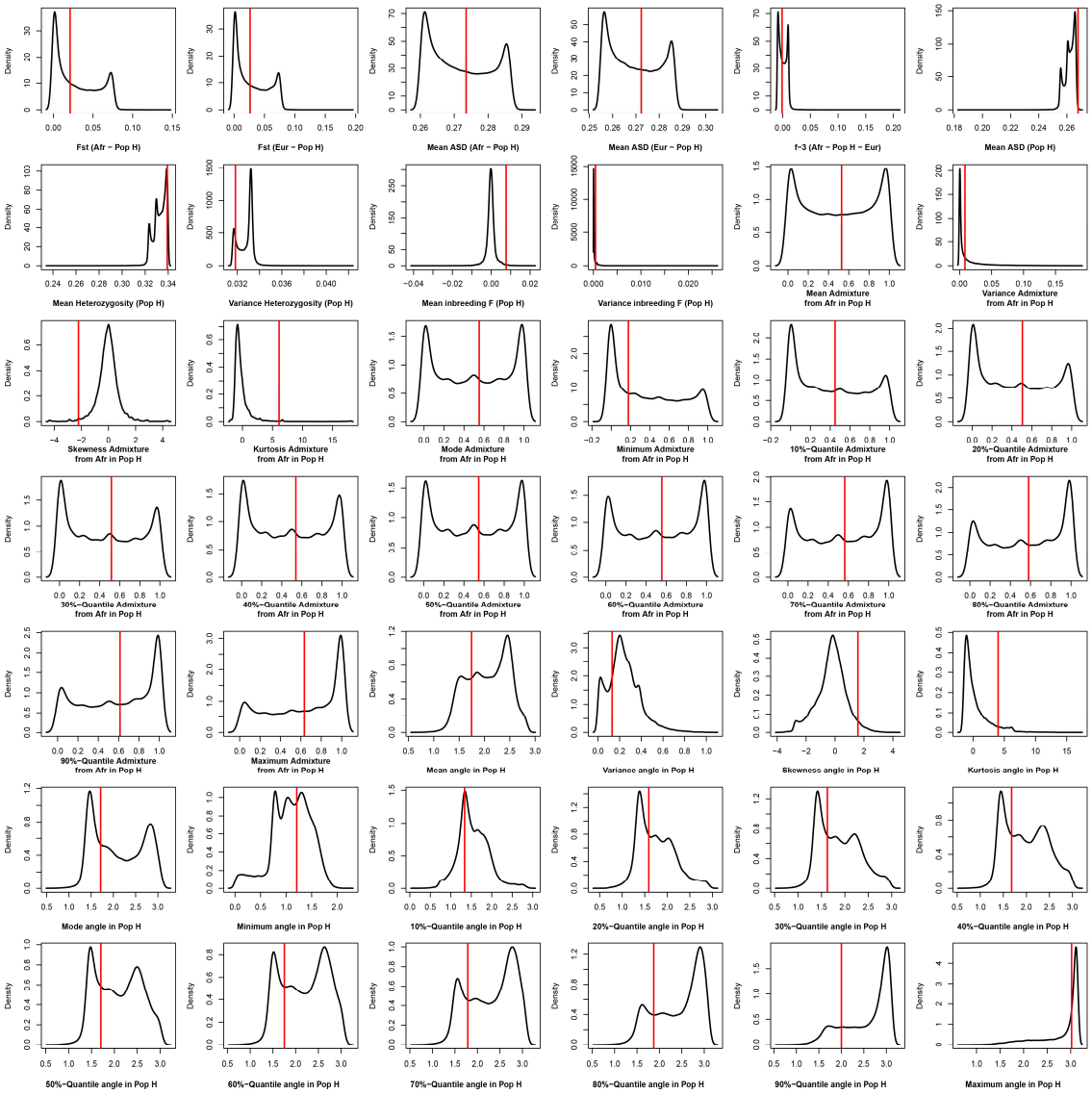


Figure SF18:

Density distributions (black line) of each 42 summary-statistics obtained from 40,000 simulations, 10,000 under each four competing scenarios, for Brava birth-island, compared to the observed statistic obtained in this island (red vertical line).

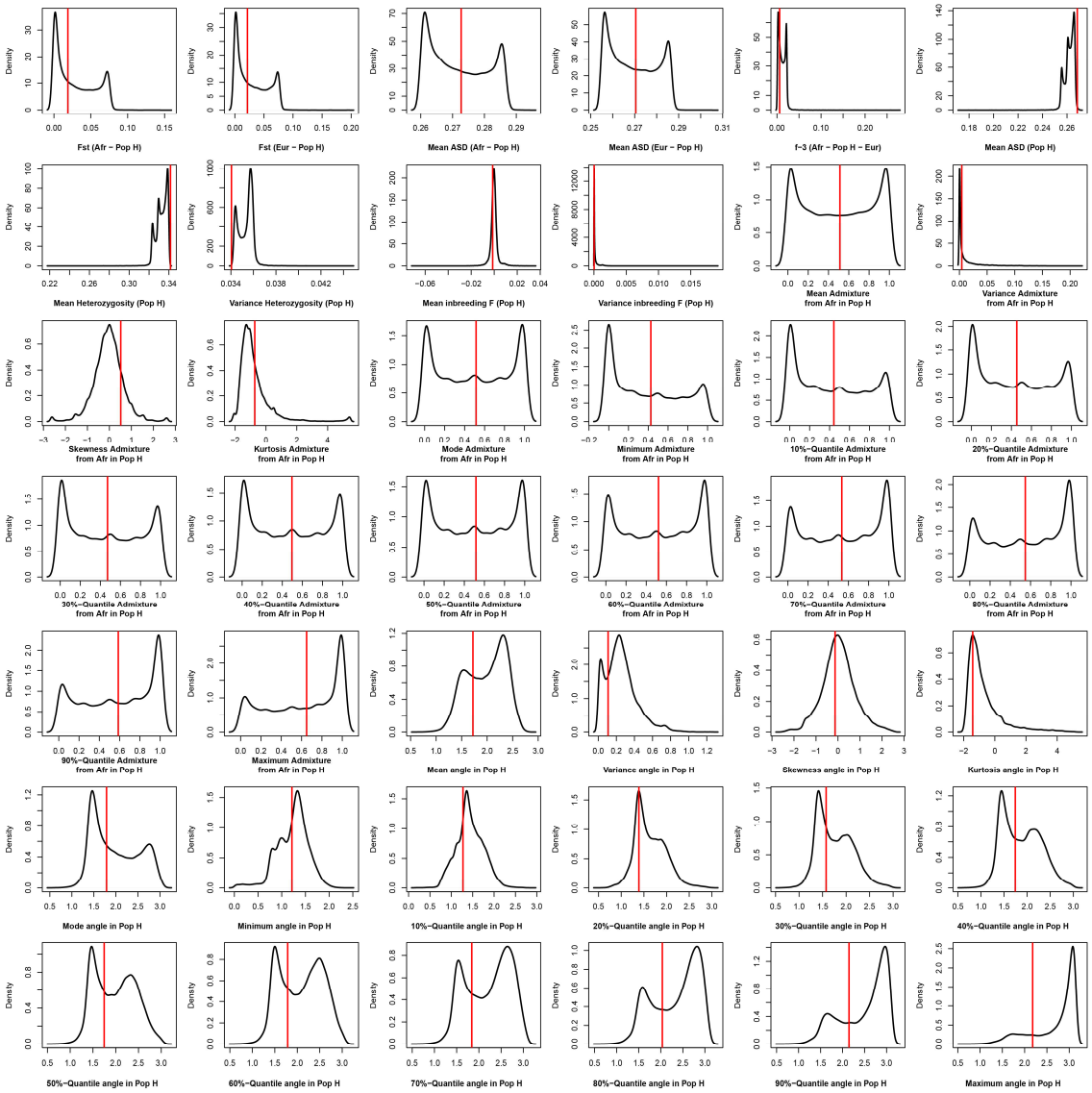


Figure SF19:

Density distributions (black line) of each 42 summary-statistics obtained from 40,000 simulations, 10,000 under each four competing scenarios, for Maio birth-island, compared to the observed statistic obtained in this island (red vertical line).

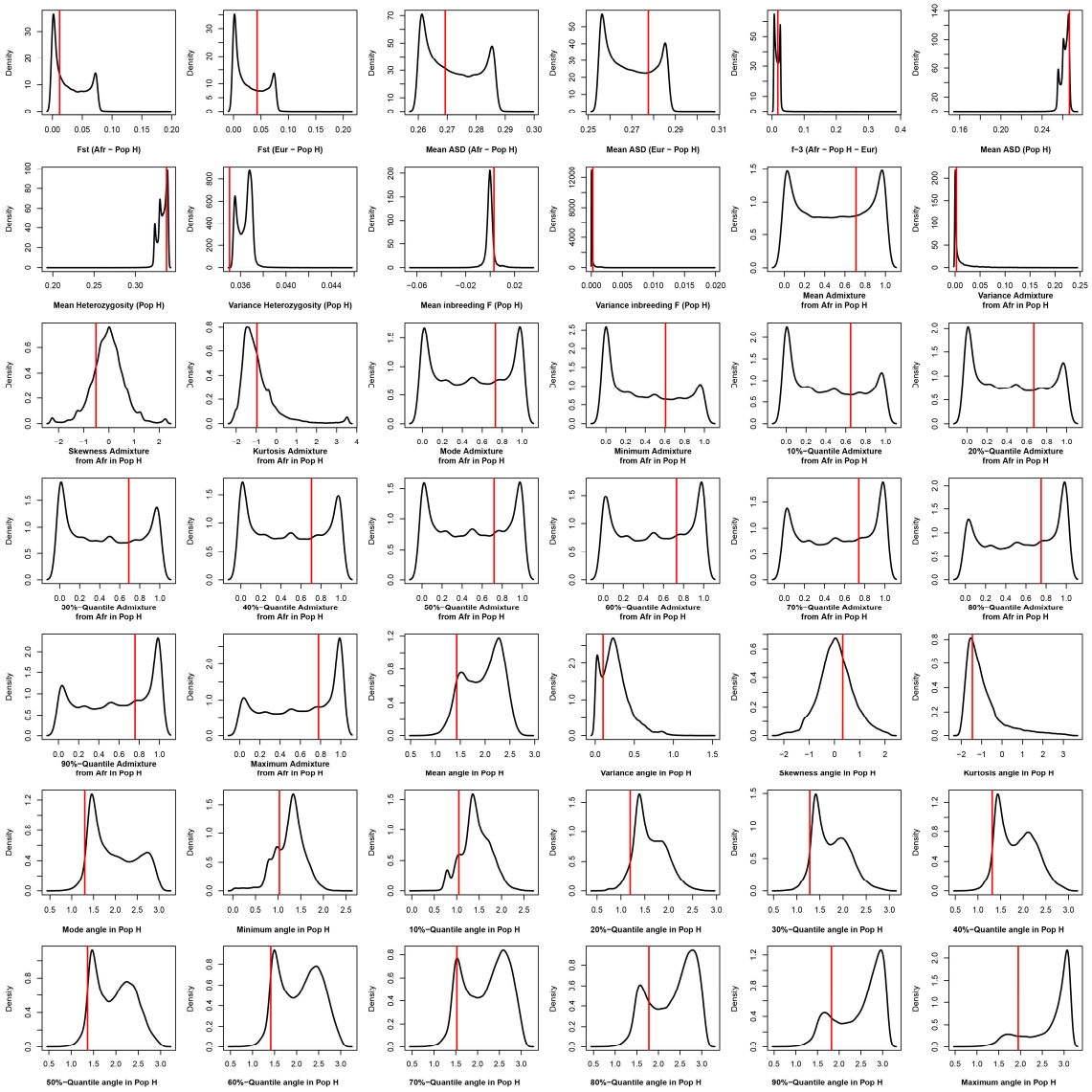


Figure SF20:

Density distributions (black line) of each 42 summary-statistics obtained from 40,000 simulations, 10,000 under each four competing scenarios, for Boa Vista birth-island, compared to the observed statistic obtained in this island (red vertical line).

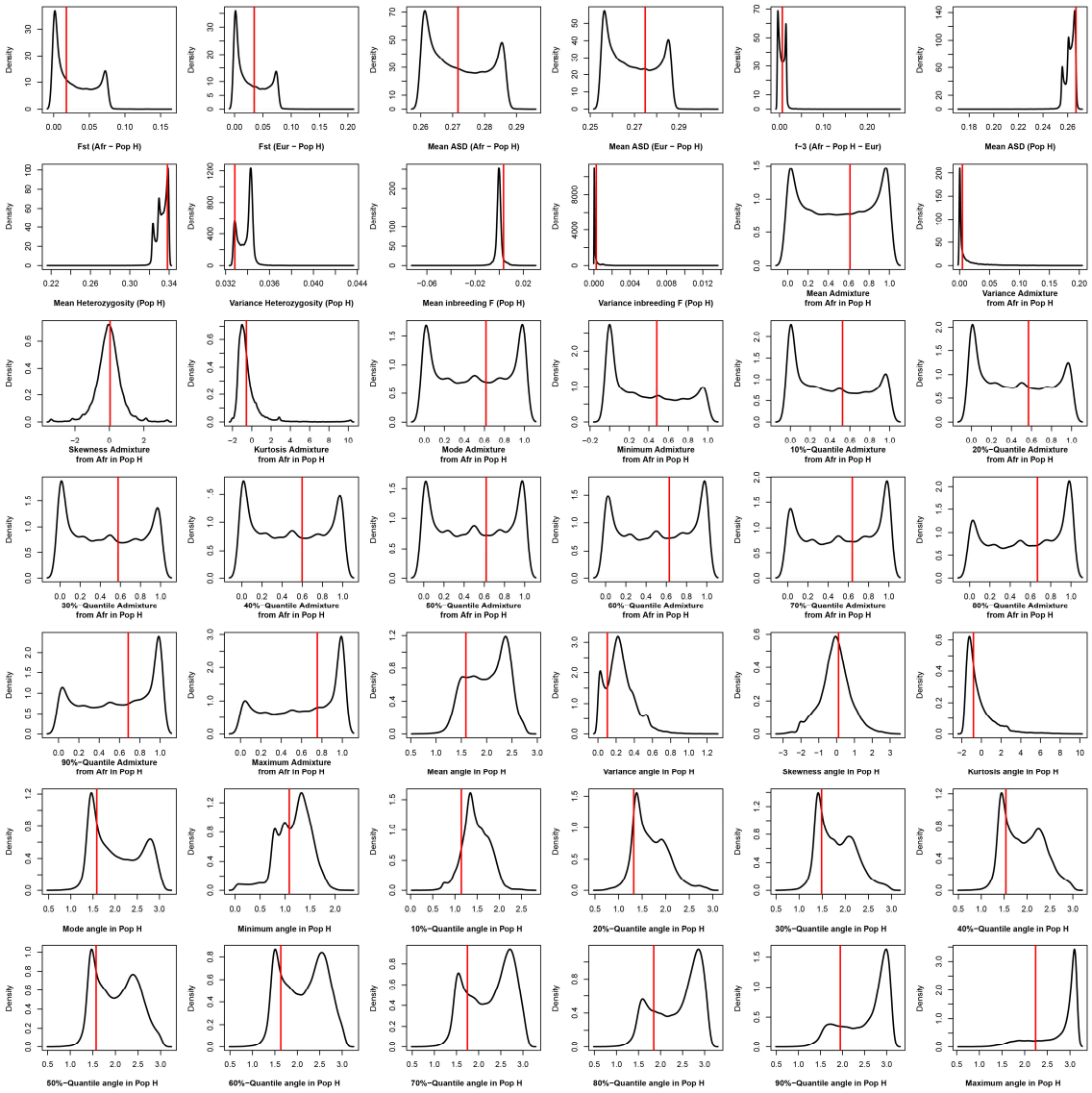


Figure SF21:

Density distributions (black line) of each 42 summary-statistics obtained from 40,000 simulations, 10,000 under each four competing scenarios, for São Vicente birth-island, compared to the observed statistic obtained in this island (red vertical line).

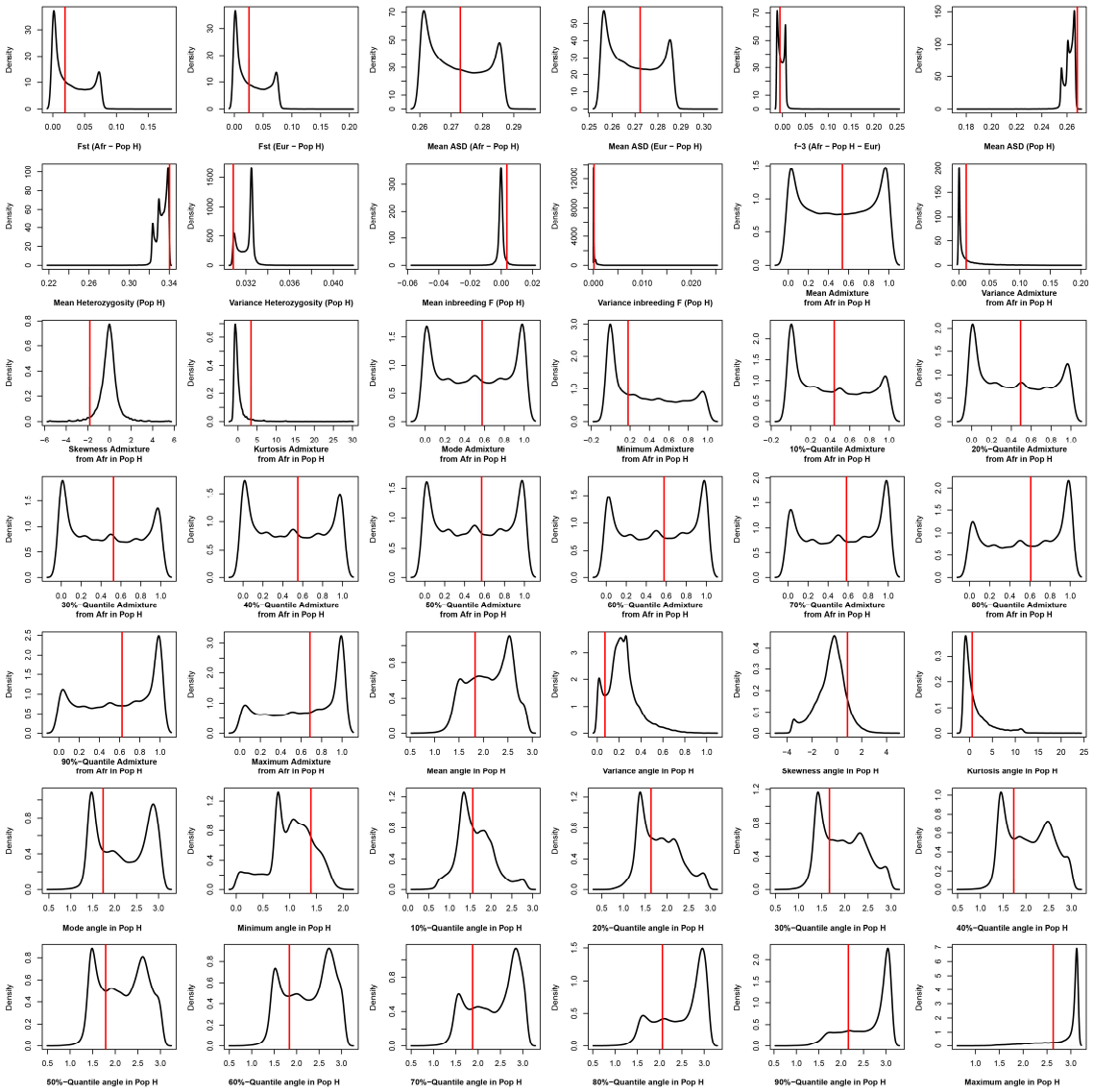


Figure SF22:

Density distributions (black line) for each 42 summary-statistics obtained from 40,000 simulations, 10,000 under each four competing scenarios, for Sal birth-island, compared to the observed statistic obtained in this island (red vertical line).

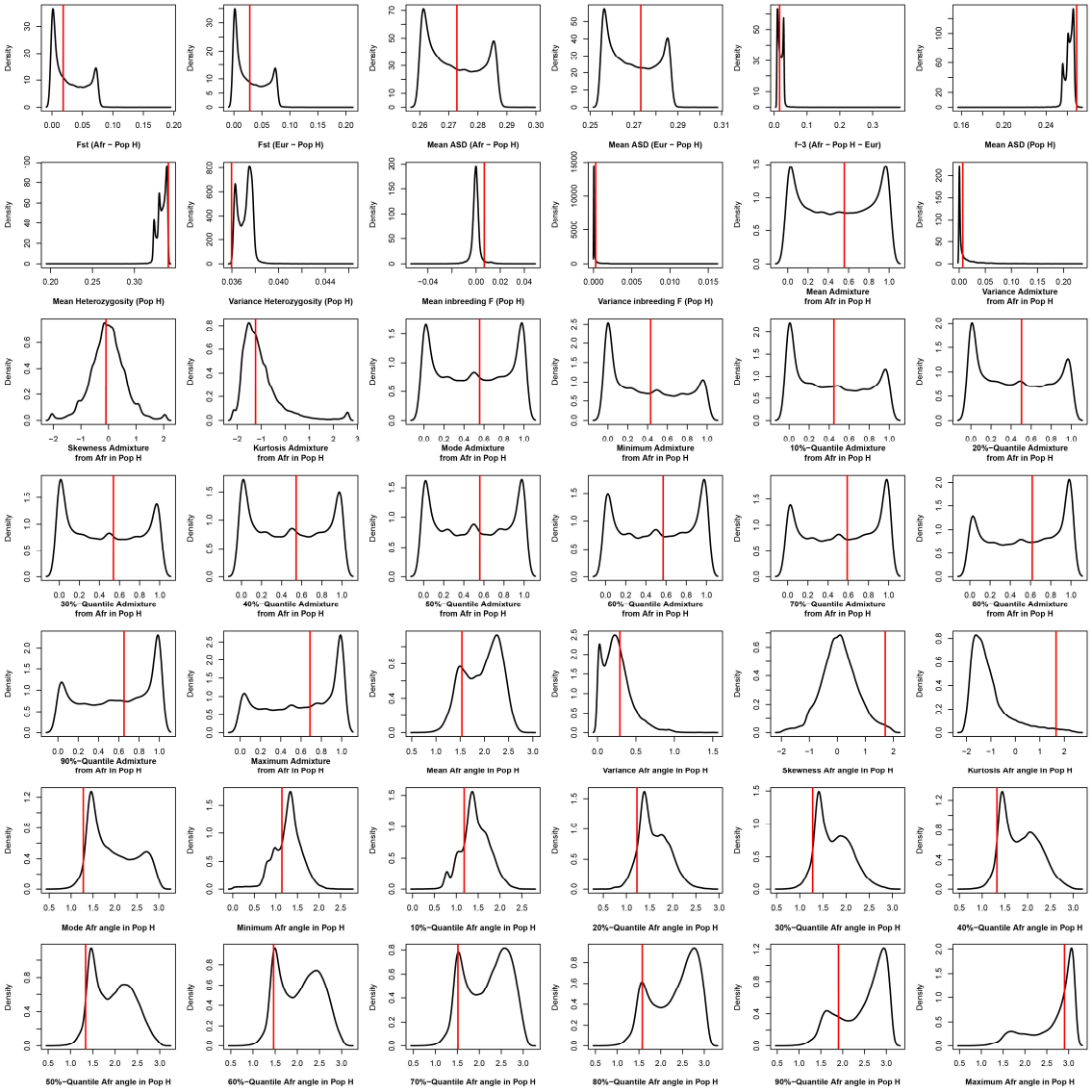


Figure SF23:

Island per island Random-Forest ABC cross validation results and variable importance.

We considered four competing scenarios as described in **Material and Methods** and in **Figure F6**, with associated prior parameter distributions in **Table T2**. Cross-validation results are obtained by conducting RF scenario-choices using in-turn each 40,000 simulations (10,000 per competing scenario) as pseudo-observed data and the remaining 39,999 simulations in the reference table. Note that no real data is used in this analysis. We considered 1,000 trees in the random-forest for each analysis. Cross-validation results and associated summary-statistics' importance for the RF decisions are obtained from the same function *abcrf* in the R package *abcrf*.

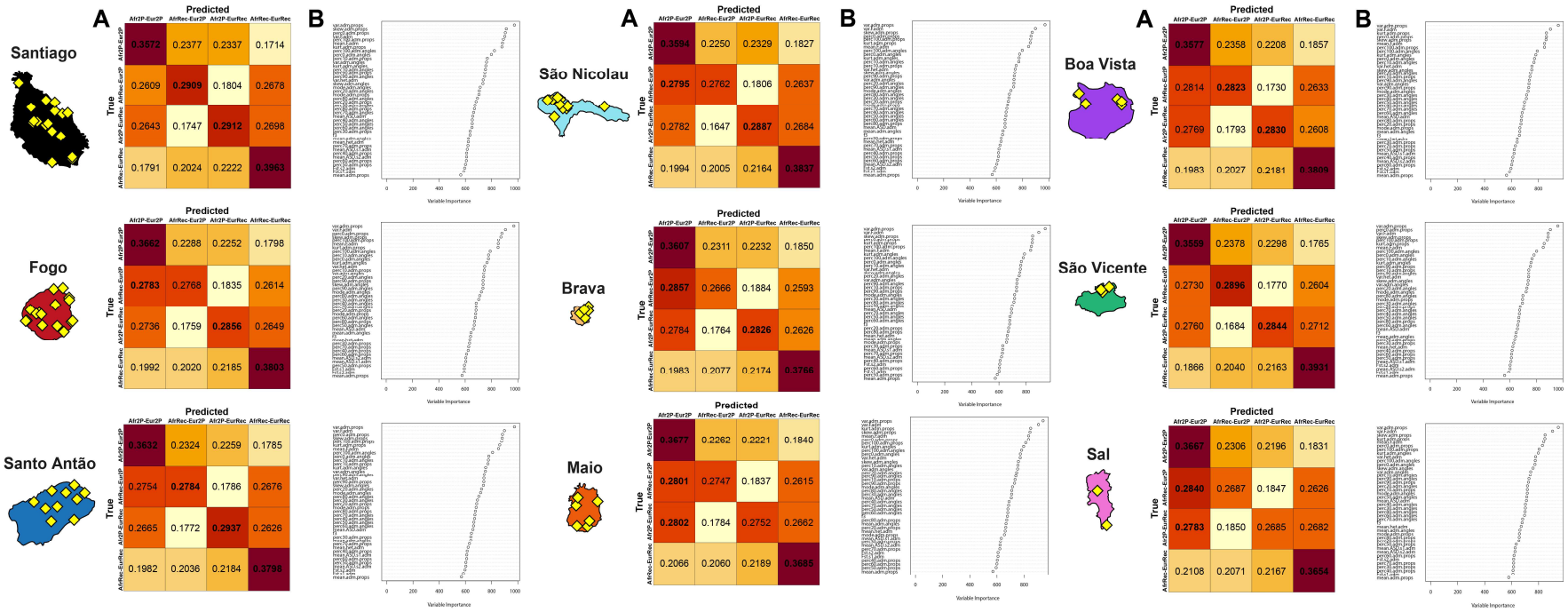


Figure SF24:

Reproductive-size posterior parameter distributions and associated priors obtained with Neural Network ABC inference for each island separately.

We considered, for each island separately, 100,000 simulations computed under the winning scenario obtained with RF-ABC scenario-choice procedures and provided on the left of the figure. Neural-Network tolerance levels and number of neurons in the hidden layer, for each birth-island separately, are chosen based on posterior parameter cross-validation error minimization procedures conducted on 1,000 random simulations used in-turn, as pseudo observed data (see **Supplementary Text**, and **Table ST4**). For each island, posterior parameter distributions correspond to the solid lines and corresponding priors correspond to the doted black lines. Density distributions are based on the logit transformation of parameter values (see **Supplementary Text**) using an Epanechnikov kernel between the corresponding prior bounds. See **Results** and **Supplementary Text** for description and discussion of the results. Results are synthesized in **Figure F7**.

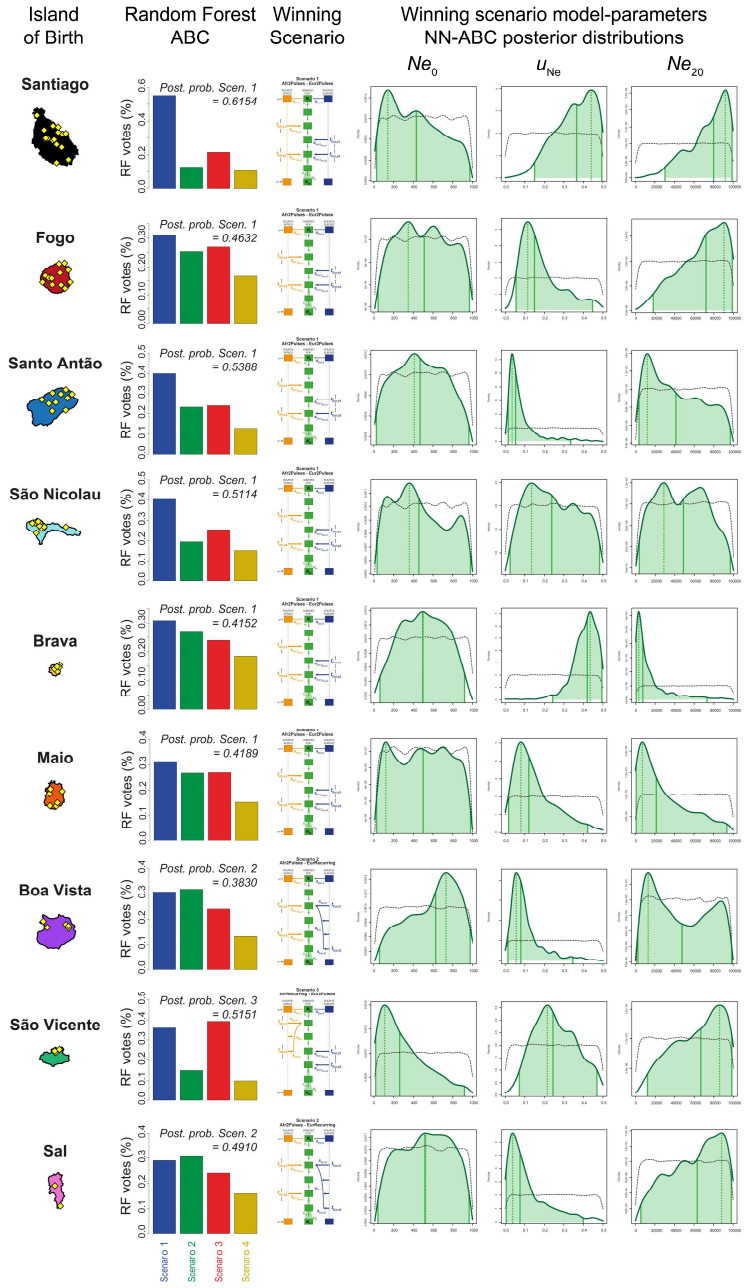


Figure SF25:

African admixture histories posterior parameter distributions and associated priors obtained with Neural Network ABC inference for each island separately.

We considered, for each island separately, 100,000 simulations computed under the winning scenario obtained with RF-ABC scenario-choice procedures and provided on the left of the figure. Neural Network tolerance levels and number of neurons in the hidden layer, for each island separately, are chosen based on posterior parameter cross-validation error minimization procedures conducted on 1,000 random simulations used in-turn, as pseudo observed data (see **Supplementary Text**, and **Table ST4**). For each island, posterior parameter distributions correspond to the solid lines and corresponding priors correspond to the doted black lines. Density distributions are based on the logit transformation of parameter values (see **Supplementary Text**) using an Epanechnikov kernel between the corresponding prior bounds. See **Results** and **Supplementary Text** for description and discussion of the results. Results are synthesized in **Figure F7**.

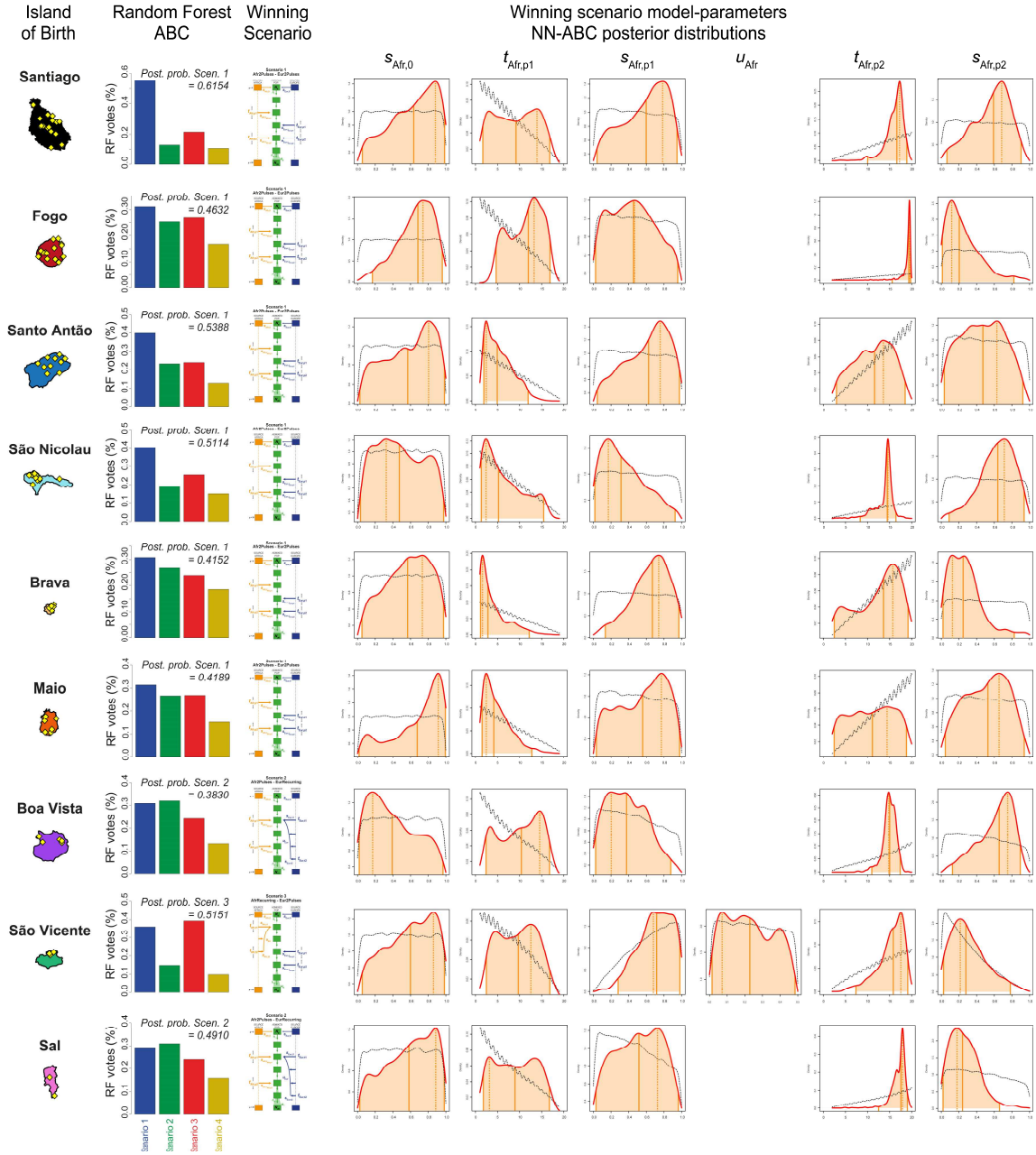


Figure SF26:

European admixture histories posterior parameter distributions and associated priors obtained with Neural Network ABC inference for each island separately.

We considered, for each island separately, 100,000 simulations computed under the winning scenario obtained with RF-ABC scenario-choice procedures and provided on the left of the figure. Neural Network tolerance levels and number of neurons in the hidden layer, for each island separately, are chosen based on posterior parameter cross-validation error minimization procedures conducted on 1,000 random simulations used in-turn, as pseudo observed data (see **Supplementary Text**, and **Table ST4**). For each island, posterior parameter distributions correspond to the solid lines and corresponding priors correspond to the doted black lines. Density distributions are based on the logit transformation of parameter values (see **Supplementary Text**) using an Epanechnikov kernel between the corresponding prior bounds. See **Results** and **Supplementary Text** for description and discussion of the results. Results are synthesized in **Figure F7**.

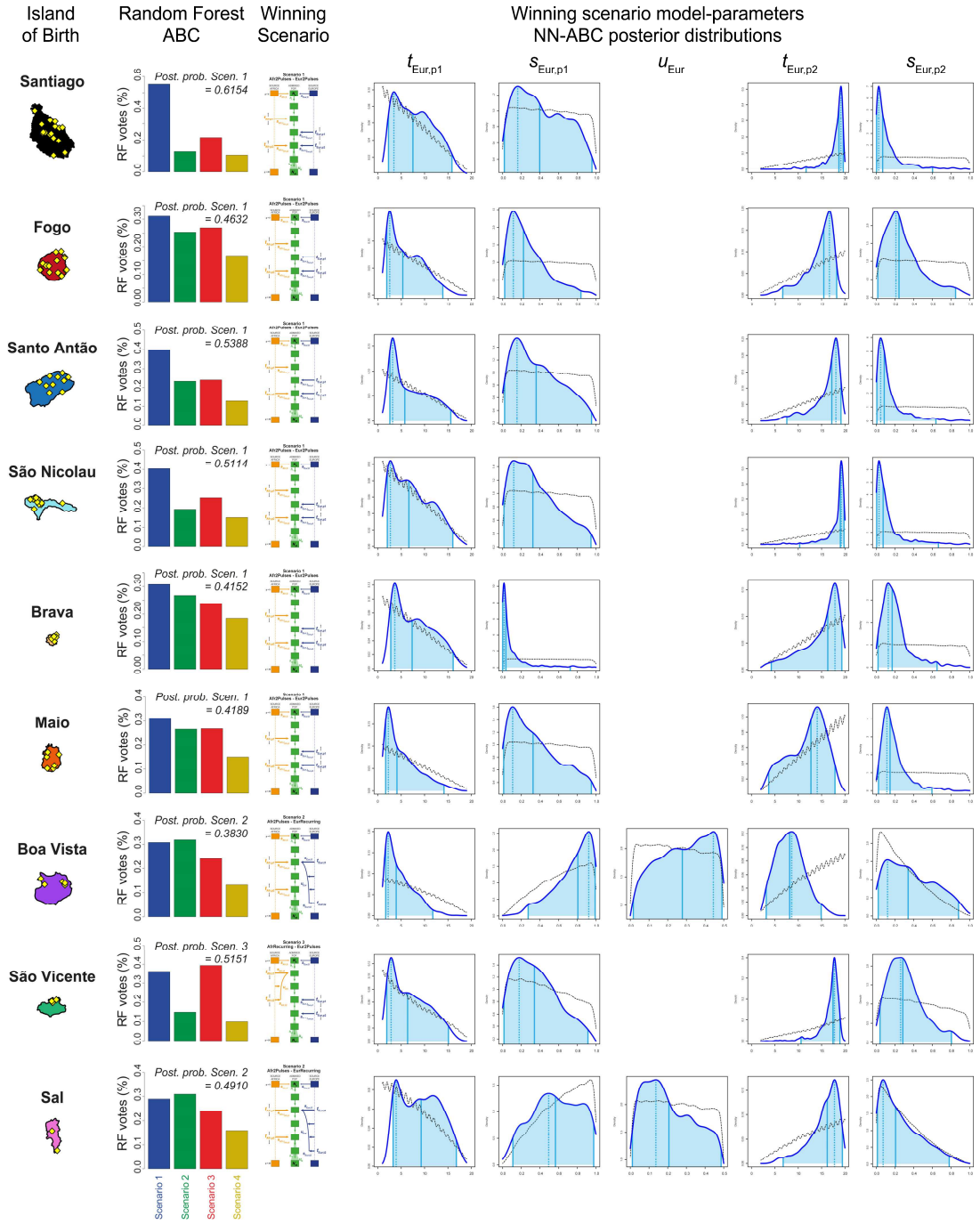


Table ST1:

Population table corresponding to Figure F1 map and sample inclusion in all analyses, respectively.

See .xls file

Table ST1: Part – 1/3

Geographic Region	Population Code	Population description	N	Latitude	Longitude	Data Reference	F 1	SF 2	SF 3	SF 4	SF 5	F2;SF6	F3;SF7-8	F4; SF9-11	F5	F7;SF13-26
West Western Africa	Cabo Verde Kriolu speakers	Born on Santiago	59	15.12	-23.64	This Study & Verdu et al. 2017	x	x	x	x	x	x	x	x	x	x
	Cabo Verde Kriolu speakers	Born on Fogo	26	15.12	-23.64	This Study	x	x	x	x	x	x	x	x	x	x
	Cabo Verde Kriolu speakers	Born on Santo Antão	30	15.12	-23.64	This Study	x	x	x	x	x	x	x	x	x	x
	Cabo Verde Kriolu speakers	Born on São Nicolau	25	15.12	-23.64	This Study	x	x	x	x	x	x	x	x	x	x
	Cabo Verde Kriolu speakers	Born on Maio	10	15.12	-23.64	This Study	x	x	x	x	x	x	x	x	x	x
	Cabo Verde Kriolu speakers	Born on Brava	12	15.12	-23.64	This Study	x	x	x	x	x	x	x	x	x	x
	Cabo Verde Kriolu speakers	Born on Bõa Vista	17	15.12	-23.64	This Study	x	x	x	x	x	x	x	x	x	x
	Cabo Verde Kriolu speakers	Born on São Vicente	37	15.12	-23.64	This Study	x	x	x	x	x	x	x	x	x	x
	Cabo Verde Kriolu speakers	Born on Sal	9	15.12	-23.64	This Study	x	x	x	x	x	x	x	x	x	x
	Cabo Verde Kriolu speakers	Foreign born (Portugal, Mozambic, Angola, France, USA)	8	15.12	-23.64	This Study	x	x	x	x	x	x	-	-	-	-
	Fula	Gambian in Western Division, The Gambia - Fula	74	13.45	-16.58	Gambian Genome Variation Project; 1000 Genomes Phase 3	x	x	x	x	-	-	-	-	-	-
	GWD	Gambian in Western Division, The Gambia - Mandinka	113	13.45	-16.58	Gambian Genome Variation Project; 1000 Genomes Phase 3	x	x	x	x	x	x	**Resample 50**	**Resample 50**	-	-
	Jola	Gambian in Western Division, The Gambia - Jola	79	13.45	-16.58	Gambian Genome Variation Project; 1000 Genomes Phase 3	x	x	x	x	x	x	**Resample 50**	**Resample 50**	-	-
	Mandinka	Gambia	87	13.47	-16.57	AGVP ; Gurdasani et al. 2014	x	x	x	x	x	x	**Resample 50**	**Resample 50**	-	**Resample 60**
	Wolof	Gambian in Western Division, The Gambia - Wolof	78	13.45	-16.58	Gambian Genome Variation Project; 1000 Genomes Phase 3	x	x	x	x	x	x	**Resample 50**	**Resample 50**	-	-
Central Western Africa	Ahizi	Ivory Coast	18	5.25	-4.61	Patin et al. 2017	x	x	x	x	x	x	-	-	-	-
	MSL	Mende in Sierra Leone	85	8.48	-13.23	1000 Genomes Phase 3	x	x	x	x	x	x	**Resample 50**	**Resample 50**	-	-
	Yacouba	Ivory Coast	17	7.27	-8.17	Perry et al. 2014	x	x	x	x	x	x	-	-	-	-
East Western Africa	Bariba	Benin	20	9.35	2.62	Patin et al. 2017	x	x	x	x	x	x	-	-	-	-
	ESN	Esan in Nigeria	98	9.07	7.48	1000 Genomes Phase 3	x	x	x	x	x	x	**Resample 50**	**Resample 50**	-	-
	Fon	Benin	12	6.35	2.41	Patin et al. Science 2017	x	x	x	x	x	x	-	-	-	-
	Ga-Adangbe	Ghana	100	8.00	-2.00	AGVP ; Gurdasani et al. 2014	x	x	x	x	x	x	**Resample 50**	**Resample 50**	-	-
	Igbo	Nigeria	99	6.00	7.00	AGVP ; Gurdasani et al. 2014	x	x	x	x	x	x	**Resample 50**	**Resample 50**	-	-
	Yoruba	Benin	20	7.36	2.60	Patin et al. 2017	x	x	x	x	x	x	-	-	-	-
	YRI	Yoruba in Ibadan, Nigeria	108	7.40	3.92	1000 Genomes Phase 3	x	x	x	x	x	x	**Resample 50**	**Resample 50**	-	-

1490 **Table ST1: Part – 2/3**

Geographic Region	Population Code	Population description	N	Latitude	Longitude	Data Reference	F1	SF2	SF3	SF4	SF5	F2;SF6	F3;SF7-8	F4; SF9-11	F5	F7;SF13-26
West Central Africa	Akele	Gabon	37	-0.70	10.22	Patin et al.2017	**x**	**x**	**x**	**x**	-	-	-	-	-	-
	Ba.Bongo East	Gabon	27	-2.27	13.58	Patin et al. 2017	**x**	**x**	**x**	-	-	-	-	-	-	-
	Badwee	Cameroon	38	3.56	13.12	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	-	-	-	-
	Baka	Cameroon	97	3.08	14.07	Patin et al. 2017	**x**	**x**	**x**	-	-	-	-	-	-	-
	Bakota	Gabon	46	0.57	12.87	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	-	-	-	-
	Bakoya	Gabon	20	1.00	14.00	Patin et al. 2017	**x**	**x**	**x**	-	-	-	-	-	-	-
	Bapunu	Gabon	47	-1.87	11.02	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	-	-	-	-
	Bateke	Gabon	41	-0.82	12.70	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	-	-	-	-
	Bekwil	Gabon	5	1.18	13.22	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	-	-	-	-
	Benga	Gabon	38	0.58	9.33	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	-	-	-	-
	Bezan	Cameroon	20	5.45	11.36	Patin et al. 2017	**x**	**x**	**x**	-	-	-	-	-	-	-
	Biaka	Aka from CAR	15	4.00	17.00	Patin et al. 2017	**x**	**x**	**x**	-	-	-	-	-	-	-
	Duma	Gabon	39	-0.82	12.70	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	-	-	-	-
	Eshira	Gabon	40	-1.22	10.60	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	-	-	-	-
	Eviya	Gabon	23	-1.21	10.60	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	-	-	-	-
	Fang	Gabon	65	2.08	11.48	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	**Resample 50**	**Resample 50**	-	-
	Galoa	Gabon	46	-0.70	10.22	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	-	-	-	-
	Makina	Gabon	39	-0.10	11.93	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	-	-	-	-
	Ndumu	Gabon	37	-1.63	13.58	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	-	-	-	-
	Nzebi	Gabon	61	-1.57	13.20	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	**Resample 50**	**Resample 50**	-	-
	Obamba	Gabon	45	-0.68	13.78	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	-	-	-	-
	Okande	Gabon	7	-0.05	11.62	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	-	-	-	-
	Orungu	Gabon	19	-0.72	8.78	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	-	-	-	-
	Shake	Gabon	47	-0.82	12.70	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	-	-	-	-
	Tsogo	Gabon	60	-1.03	10.67	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	**Resample 50**	**Resample 50**	-	-
	Yaounde	Cameroon	39	3.84	11.50	AGVP ; Gurdasani et al. 2014	**x**	**x**	**x**	**x**	-	-	-	-	-	-
Sout-West Central Africa	Kimbundu	Angola	17	-8.83	13.22	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	**x**	**x**	-	-
	Kongo	Angola	10	-8.83	13.22	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	**x**	**x**	-	-
	Ovimbundu	Angola	12	-8.83	13.22	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	**x**	**x**	-	-
	Umbundo	Angola	5	-8.83	13.22	Patin et al. 2017	**x**	**x**	**x**	**x**	-	-	**x**	**x**	-	-
South Africa	Sotho	South Africa	86	-29.13	26.26	AGVP ; Gurdasani et al. 2014	**x**	**x**	**x**	**x**	-	-	**Resample 50**	**Resample 50**	-	-
	Zulu	South Africa	100	-29.00	24.00	AGVP ; Gurdasani et al. 2014	**x**	**x**	**x**	**x**	-	-	**Resample 50**	**Resample 50**	-	-

1491
1492

1493 **Table ST1: Part 3/3**

Geographic Region	Population Code	Population description	N	Latitude	Longitude	Data Reference	F1	SF2	SF3	SF4	SF5	F2;SF6	F3;SF7-8	F4; SF9-11	F5	F7;SF13-26
East Africa	Baganda	Uganda	99	1.00	32.00	AGVP ; Gurdasani et al. 2014	**x**	**x**	**x**	-	-	-	-	-	-	-
	Banyarwanda	Uganda	96	1.00	32.00	AGVP ; Gurdasani et al. 2014	**x**	**x**	**x**	-	-	-	-	-	-	-
	Barundi	Uganda	93	1.00	32.00	AGVP ; Gurdasani et al. 2014	**x**	**x**	**x**	-	-	-	-	-	-	-
	Ethiopia	Ethiopia	107	8.00	38.00	AGVP ; Gurdasani et al. 2014	**x**	**x**	**x**	-	-	-	-	-	-	-
	Kalenjin	Kenya	100	1.00	38.00	AGVP ; Gurdasani et al. 2014	**x**	**x**	**x**	-	-	-	-	-	-	-
	Kikuyu	Kenya	99	-0.40	36.90	AGVP ; Gurdasani et al. 2014	**x**	**x**	**x**	-	-	-	-	-	-	-
	LWK	Luhya in Webuye, Kenya	97	-1.27	36.61	1000 Genomes Phase 3	**x**	**x**	**x**	-	-	-	-	-	-	-
	Mbuti	DRC	13	1.00	29.00	Patin et al. 2017	**x**	**x**	**x**	-	-	-	-	-	-	-
	Twa	Uganda	2	-0.98	29.62	Perry et al. 2014	**x**	**x**	**x**	-	-	-	-	-	-	-
Europe	FIN	Finnish in Finland	99	60.17	24.93	1000 Genomes Phase 3	**x**	**x**	**x**	**x**	**x**	**x**	**Resample 50**	**Resample 50**	-	-
	GBR	British in England and Scotland	90	52.49	-1.89	1000 Genomes Phase 3	**x**	**x**	**x**	**x**	**x**	**x**	**Resample 50**	**Resample 50**	-	-
	IBS	Iberian populations in Spain	107	40.38	-3.72	1000 Genomes Phase 3	**x**	**x**	**x**	**x**	**x**	**x**	**Resample 50**	**Resample 50**	-	**Resample 60**
	TSI	Toscani in Italy	107	42.10	12.00	1000 Genomes Phase 3	**x**	**x**	**x**	**x**	**x**	**x**	**Resample 50**	**Resample 50**	-	-
South Asia	BEB	Bengali in Bangladesh	86	23.70	90.35	1000 Genomes Phase 3	**x**	**x**	-	-	-	-	-	-	-	-
	GIH	Gujarati Indians in Houston, TX	101	22.62	70.69	1000 Genomes Phase 3	**x**	**x**	-	-	-	-	-	-	-	-
	ITU	Indian Telugu in the UK	101	17.18	79.79	1000 Genomes Phase 3	**x**	**x**	-	-	-	-	-	-	-	-
	PJL	Punjabi in Lahore, Pakistan	94	31.55	74.36	1000 Genomes Phase 3	**x**	**x**	-	-	-	-	-	-	-	-
	STU	Sri Lankan Tamil in the UK	99	8.38	80.50	1000 Genomes Phase 3	**x**	**x**	-	-	-	-	-	-	-	-
East Asia	CDX	Chinese Dai in Xishuangbanna, China	92	22.00	100.78	1000 Genomes Phase 3	**x**	**x**	-	-	-	-	-	-	-	-
	CHB	Han Chinese in Beijing, China	103	39.92	116.38	1000 Genomes Phase 3	**x**	**x**	-	-	-	-	**Resample 50**	**Resample 50**	-	-
	CHS	Han Chinese South	104	23.13	113.27	1000 Genomes Phase 3	**x**	**x**	-	-	-	-	-	-	-	-
	JPT	Japanese in Tokyo, Japan	104	35.68	139.68	1000 Genomes Phase 3	**x**	**x**	-	-	-	-	-	-	-	-
	KHV	Kinh in Ho Chi Minh City, Vietnam	99	10.78	106.68	1000 Genomes Phase 3	**x**	**x**	-	-	-	-	-	-	-	-
USA	ASW	African Ancestry in Southwest US	55	35.48	-97.53	1000 Genomes Phase 3	**x**	**x**	x	x	x	x	**Resample 50**	**Resample 50**	-	-
	CEU	US Utah residents (CEPH) with Northern and Western European ancestry	99	40.77	-111.89	1000 Genomes Phase 3	**x**	**x**	x	x	x	x	**Resample 50**	**Resample 50**	-	-
	MXL	Mexican Ancestry in Los Angeles, California	64	34.05	-118.24	1000 Genomes Phase 3	**x**	**x**	-	-	-	-	-	-	-	-
Carribbean	ACB	African Caribbean in Barbados	95	13.10	-59.62	1000 Genomes Phase 3	**x**	**x**	**x**	**x**	**x**	**x**	**Resample 50**	**Resample 50**	-	-
	PUR	Puerto Rican in Puerto Rico	104	18.40	-66.10	1000 Genomes Phase 3	**x**	**x**	**x**	**x**	**x**	**x**	-	-	-	-
South America	CLM	Colombian in Medellin, Colombia	94	4.58	-74.07	1000 Genomes Phase 3	**x**	**x**	-	-	-	-	-	-	-	-
	PEL	Peruvian in Lima, Peru	85	-12.04	-77.03	1000 Genomes Phase 3	**x**	**x**	-	-	-	-	-	-	-	-

1494

1495

Table ST2:

Historical landmark chronology for the peopling history of Cabo Verde as provided by previous historical work, respectively for each island.

See detailed historical references in the table. **See .xls file**

Part 1/6

- First significant indication of permanent settlement
- Administrative information
- Church establishment
- Census

Year	Cabo Verde Unspecified island	SANTIAGO	FOGO	SANTO ANTAO	SAO NICOLAU	BRAVA	MAIO	BOA VISTA	SAO VICENTE	SAL
1460-1466	Discovery of the archipelago .	First settlers from Portugal **all**	First settlers from Portugal **all**							
1472	**Portuguese crown restricting slave-trade out of Cabo Verde. First trade and population movements regulation (All references)**									
1480-1493			Remote administration of the island from Santiago, São Filipe church foundation **1. p143**							
1495			Royal donation changes island ownership and administration **1. p53**							
1504				Royal donation changes island ownership and administration **1. p56**			Royal donation changes island ownership and administration **1. p55**			
1505							Census* = 225 **7. p294**			
1507				Extensive livestock farming without permanent peopling, from tax records **3. p135**	Extensive livestock farming without permanent peopling, from tax records **3. p135**		Extensive livestock farming without permanent peopling, from tax records **3. p135**	Extensive livestock farming without permanent peopling, from tax records **3. p135**	Extensive livestock farming without permanent peopling, from tax records **3. p135**	
1509						Royal donation changes island ownership and administration **1. p57**				Royal donation changes island ownership and administration **1. p57**
1513	Census* = 3500 **1. p147-148**	382 slaves imported from "Guinea" **1. p264** **4. p135**								
1514		1354 slaves imported from "Guinea" **1. p264** **4. p135**								
1515		1404 slaves imported from "Guinea" **1. p264** **4. p135**								
1516		26 slaves imported from "Guinea" **4. p135**								
1526 1529		Commerce records indicate ~300 slaves sent to Portugal per year during 3 years **1. p97**								
1527				Military troops sent from Santiago against Spanish marauders **3. p329**	Military troops sent from Santiago against Spanish marauders **3. p329**				Military troops sent from Santiago against Spanish marauders **3. p329**	

Table ST2: Part 2/6

Year	Cabo Verde Unspecified island	SANTIAGO	FOGO	SANTO ANTAO	SAO NICOLAU	BRAVA	MAIO	BOA VISTA	SAO VICENTE	SAL
1528		1404 slaves imported from "Guinea" **1. p267**			Fishing watters and boat halt **1. p225**				Fishing watters and boat halt **1. p225**	
1529							Permanent harbour and salt depot, possible small-size permanent settlement **1. p223-224**			Permanent harbour and salt depot, possible small-size permanent settlement **1. p223-224**
1532			All trade, administration, and population movements under control by Santiago, except direct trade with Portuguese crown **1. p101-105**							
1540						Fogo fishing watters, no mention of permanent settlement **1. p225**				
1546 1557	Instances of donations/inheritance of both goods/money and land to the benefit of admixed offsprings **1. p162-164**									
1548					Request to Santiago for additional workforce, no trace of the outcome **1. p228**				Request to Santiago for additional workforce, no trace of the outcome **1. p228**	
1566								Census* = 30, prisonners kept on the island, extensive livestock farming and goat-skin production **5. p275**		
1570				Reports of small-size permanent settlement, no census **1. p146**		Reports of illegal slave trade **3. p329**			Reports of small-size permanent settlement, no census **1. p146**	
1572		Census* = 8900, only town residents, no children, no marginalized people **2. p279 1. p231, note**	Census* = 1000, in São Filipe **1. p146, note**							
1573				Church foundation **1. p146 note**						
1580	Census* = 9940 **7. p297**	Census* = 8000 **7. p294**	Census* = 1200 **7. p294**	Census* = 400, Permanent harbor **1. p213 7. p294**	Census* = 140, Permanent harbor **1. p213 7. p294**	Census* = 100 **7. p294**	Census* = 50 **7. p294**	Illegal slave trade of 50 individuals **3. p329 7. p294**	Census* = 0 **7. p294**	Census* = 0 **7. p294**
1580-1614	Admixed individuals own and/or manage commercial firms **2. p515**									
1580-1582			Famine reports (thus implying prior settlement ?) **1. p12-14 6. p106**			Famine reports (thus implying prior settlement ?) **1. p12-14 6. p106**	Famine reports (thus implying prior settlement ?) **1. p12-14 6. p106**			
1582-1731	**Second migration wave into Santo Antão and São Nicolau and in the interior of Fogo and Santiago (1. p230; 3. p302; 4. p387; 7. p293)**									

1507
1508

Table ST2: Part 3/6

Year	Cabo Verde Unspecified island	SANTIAGO	FOGO	SANTO ANTAO	SAO NICOLAU	BRAVA	MAIO	BOA VISTA	SAO VICENTE	SAL
1582		Census* = 2600 only town residents, no children, no marginalized people, 11700 slaves. Population movements from Ribeira Grande and Praia into the interior of the island until 1731 **1. p230 3. p302 4. p387 7. p293**	Census* = 260 only town residents, no children, no marginalized people, 2000 slaves. **1. p230 4. p387**	Reports of increasing census, no numbers **1. p230**	Reports of increasing census, no numbers **1. p230**					
1599					Census* = 6 **7. p294**					
1606							Census* = 10-12, extensive livestock farming **1. p229**			
1608							Census* = 10 **7. p.294**			
1613-1639	Admixed individuals own and/or manage commercial firms **2. p515**									
1620	Portuguese crown sends boats of women for the peopling of the colony **1. p11**									
1623									Census* = 19 **7. p294**	
1628							Census* = 50 **7. p294**			
1639									Census* = 0 **7. p294**	
1650	Census* = 13980 **7. p297**	Census* = 9500 **7. p294**	Census* = 2500 **7. p294**	Census* = 1000 **7. p294**	Census* = 300 **7. p294**	Census* = 400 **7. p294**	Census* = 120 **7. p294**	Census* = 150 **7. p294**	Census* = 0 **7. p294**	Census* = 10 **7. p294**
1677					Census* = 1000 ?, Permanent clergy **3. p326-327**		Permanent clergy **3. p326-327**	Permanent clergy **3. p326-327**		
1680	Weavery exportations reserved to Portuguese citizen only, Cabo Verdean firms go bankrupt **3. p310**		Earthquake and volcano eruption, migration to Brava **1. p12-14**	Census* = 500 **7. p294**		Migration from Fogo **1. p12-14**			Census* = 0 **7. p294**	
1683					Census* = 900 **7. p294**			Census* = 6 **7. p294**		
1699							Census* = 230 **7. p294**			
1700						Census* =100 **7. p294**				
1708		400 soldiers sent to repress marooning in the interior regions. Failure of the expedition **3. p306**								
1710		Reports of over 600 maroon slaves **3. p306**								
1719		Famine **1. p12-14**								
1720	Census* = 23130 **7. p297**	Census* = 12000 **7. p294**	Census* = 5000 **7. p294**	Census* = 3000 **7. p294**	Census* = 650 **7. p294**	Census* = 1200 **7. p294**	Census* = 250 **7. p294**	Census* = 1000 **7. p294**	Census* = 0 **7. p294**	Census* = 30 **7. p294**
1723			Census* = 700 **7. p294**	Census* = 2500 **7. p294**	Census* = 1350 **7. p294**	Census* = 200 **7. p294**	Census* = 200 **7. p294**		Census* = 0 **7. p294**	Census* = 0 **7. p294**
1727				Census* = 4000, including 502 slaves **3. p295**						
1730		Census* = 25000 **7. p294**	Census* = 13000 **7. p294**							

1509

1510 **Table ST2: Part 4/6**

1511

Year	Cabo Verde Unspecified island	SANTIAGO	FOGO	SANTO ANTAO	SAO NICOLAU	BRAVA	MAIO	BOA VISTA	SAO VICENTE	SAL
1731		Census* = 17464, ~1800 RibeiraGrande/Praia + ~15664 interior, 16,28% slaves on the whole island **3. p300-302**	No census, 25% slaves and highest "white" people proportion among islands **3. p300**	No census, 15% slaves ~1600 individuals **3. p296 7. p294**	Census* = 360 housholds ~2600 individuals (7/household), 11% slaves **3. p296**	No census, 6% slaves **3. p296**	Census* = less than 400 **3. p299**	Census* = Less than 2% of the total population of Cabo Verde **3. p299**		
1750-1780	**Third migration wave into all remaining islands: commercial-monopoly shifts on all trades dictate population settlements and movements among Cabo Verde islands. End of the agro-slavery system in Cabo Verde (3. p304)**									
1770	Census* = 26000 **1. p11**									
1772				Census* = 11000 **7. p294**						
1773			Census* = 5900 **1. p12-14**							
1774	Census* = 50639 **3. p355**	Census* = 24358 **3. p355**	Census* = 5728 **3. p355**	Census* = 10215 **3. p355**	Census* = 5000 **3. p355**	Census* = 3190 **3. p355**	Census* = 708 **3. p355**	Census* = 1440 **3. p355**	Census* = 0 **3. p355 7. p294**	Census* = 0 **3. p355 7. p294**
1773-1775	Famine kills 22666 individuals. People sell themsleves into slavery to avoid death **1. p12-14 3. p200**									
1775	Census* = 28368 **3. p355**	Census* = 11580 **3. p355**	Census* = 4225 **1. p12-14 3. p355**	Census* = 5668 **3. p355**	Census* = 2920 **3. p355**	Census* = 2115 **3. p355**	Census* = 604 **3. p355**	Census* = 1256 **3. p355**	Census* = 0 **3. p355 7. p294**	Census* = 0 **3. p355 7. p294**
1788									Census* = 0 **7. p294**	Census* = 0 **7. p294**
1790	Famine **1. p12-14**									
1793										Census* = 0 **7. p294**
1794									Census* = 100, first settlers **7. p294**	
1800	Census* = 56050 **7. p294**	Census* = 26000 **7. p294**	Census* = 8000 **7. p294**	Census* = 12000 **7. p294**	Census* = 4000 **7. p294**	Census* = 3000 **7. p294**	Census* = 700 **7. p294**	Census* = 2200 **7. p294**	Census* = 100 **7. p294**	Census* = 50 **7. p294**
1803					Census* = 4500 **7. p294**					
1807	Census* = 58401 **7. p294**	Census* = 14200 **7. p294**	Census* = 13150 **7. p294**	Census* = 13650 **7. p294**	Census* = 8300 **7. p294**	Census* = 6950 **7. p294**	Census* = 451 **7. p294**	Census* = 1500 **7. p294**	Census* = 200 **7. p294**	Census* = 0 **7. p294**
1814								Famine -> Mig BV -> F et SN **1. Vol1. p12-14**		
1819									Census* = 120 **7. p294**	
1820									Census* = 300 **7. p294**	
1827	Census* = 74307 **7. p297**	Census* = 2505 slaves **4. p387**	Census* = 1212 slaves **4. p387**	Census* = 207 slaves **4. p387**	Census* = 171 slaves **4. p387**	Census* = 213 slaves **4. p387**	Census* = 240 slaves **4. p387**	Census* = 489 slaves **4. p387**	Census* = 14 slaves **4. p387**	Census* = 72 slaves **4. p387**
1831	Census* = 88368 **7. p294**	Census* = 26220 **7. p294**	Census* = 17000 **1. p12-14 7. p294**	Census* = 21670 **7. p294**	Census* = 8530 **7. p294**	Census* = 9320 **7. p294**	Census* = 1648 **7. p294**	Census* = 3860 **7. p294**	Census* = 250 **7. p294**	Census* = 0 **7. p294**
1832	Census* = 60000 **7. p294**									
1831-1833	Famine kills 30000 **1. p12-14**		Famine kills 17000 **1. p12-14**	Famine kills 13000 **1. p12-14**						
1831	**End of the AST in the Portuguese Empire with the official abolition of the AST in Brazil, progressive termination of the AST since the 1815 and AST-abolition in the British Empire (9)**									

1512

Table ST2: Part 5/6

Year	Cabo Verde Unspecified island		SANTIAGO		FOGO		SANTO ANTAO		SAO NICOLAU		BRAVA		MAIO		BOA VISTA		SAO VICENTE		SAL	
1834	Census* = 55833	**7. p297**	Census* = 21696	**7. p294**	Census* = 5615	**1.p12-14 7. p294**	Census* = 13587	**7. p294**	Census* = 5418	**7. p294**	Census* = 3990	**7. p294**	Census* = 1905	**7. p294**	Census* = 3331	**7. p294**	Census* = 341	**7. p294**	Census* = 0	**7. p294**
1841	Census* = 63000	**7. p294**	Census* = 19000	**7. p294**	Census* = 12000	**7. p294**	Census* = 15000	**7. p294**	Census* = 7000	**7. p294**	Census* = 4000	**7. p294**	Census* = 2000	**7. p294**	Census* = 3000	**7. p294**	Census* = 350	**7. p294**	Census* = 500	**7. p294**
1844	Census* = 69700	**7. p297**	Census* = 25000	**7. p294**	Census* = 7000	**7. p294**	Census* = 17500	**7. p294**	Census* = 7200	**7. p294**	Census* = 4600	**7. p294**	Census* = 2200	**7. p294**	Census* = 5200	**7. p294**	Census* = 400	**7. p294**	Census* = 600	**7. p294**
1848																	Census* = 553	**7. p294**		
1850	Census* = 86700	**7. p295**																		
1851	Census* = 101700	**7. p295**	Census* = 27800	**7. p295**	Census* = 7500	**7. p295**	Census* = 30200	**7. p295**	Census* = 20200	**7. p295**	Census* = 4300	**7. p295**	Census* = 4100	**7. p295**	Census* = 4500	**7. p295**	Census* = 2100	**7. p295**	Census* = 1000	**7. p295**
1854-1855					Cholera kills 800	**1. p12-14**														
1855	Census* = 120000	**7. p295**							Census* = 11000	**7. p295**					Census* = 4000	**7. p295**				
1856-1878	**Abolition of Slavery in the Portugese Empire**																			
1856							Census* = 16907	**7. p295**												
1857	Census* = 100000	**7. p295**															Census* = 1400	**7. p295**		
1858	Census* = 85393	**7. p295**																		
1860	Census* = 90000	**1. p12-14**																		
1861	Census* = 89310	**7. p297**	Census* = 40852	**7. p295**	Census* = 14341	**7. p295**	Census* = 14643	**7. p295**	Census* = 6372	**7. p295**	Census* = 6557	**7. p295**	Census* = 1863	**7. p295**	Census* = 2647	**7. p295**	Census* = 1141	**7. p295**	Census* = 894	**7. p295**
1862	Census* = 97009	**7. p295**	Census* = 44200	**7. p295**	Census* = 14426	**1. p12-14 7. p295**	Census* = 17965	**7. p295**	Census* = 6731	**7. p295**	Census* = 6824	**7. p295**	Census* = 2067	**7. p295**	Census* = 2621	**7. p295**	Census* = 1337	**7. p295**	Census* = 838	**7. p295**
1864	Census* = 97009	**7. p297**																		
1863-1866	Famine kills 29845	**1. p12-14**	Famine kills 18000	**1. p12-14**	Famine forces 7000 individuals into emigration, unknown destinations	**1. p12-14**											Census increase by 353 individuals (exception)	**1. p12-14**		
1867	Census* = 67517	**7. p297**	Census* = 26428	**7. p295**	Census* = 7431	**1. p12-14 7. p295**	Census* = 17403	**7. p295**	Census* = 5522	**7. p295**	Census* = 5874	**7. p295**	Census* = 955	**7. p295**	Census* = 1400	**7. p295**	Census* = 1690	**7. p295**	Census* = 814	**7. p295**
1870	Census* = 76003	**1. p12-14 7. p295**																		
1871	Census* = 76053	**7. p297**																		
1872					Census* = 10000	**7. p295**					Census* = 8673	**7. p295**					Census* = 2000	**7. p295**		
1874	Census* = 90710	**7. p297**																		
1878	Census* = 99311	**7. p297**	Census* = 41077	**7. p295**	Census* = 12221	**7. p295**	Census* = 20507	**7. p295**	Census* = 8733	**7. p295**	Census* = 8151	**7. p295**	Census* = 1600	**7. p295**	Census* = 2643	**7. p295**	Census* = 3297	**7. p295**	Census* = 1082	**7. p295**
1882	Census* = 103761	**7. p295**																		
1885	Census* = 115461	**7. p297**	Census* = 45488	**7. p295**	Census* = 16000	**7. p295**	Census* = 22000	**7. p295**	Census* = 11000	**7. p295**	Census* = 9013	**7. p295**	Census* = 1000	**7. p295**	Census* = 4000	**7. p295**	Census* = 6560	**7. p295**	Census* = 400	**7. p295**
1888	Census* = 121127	**7. p295**																		

Table ST2: Part 6/6

Year	Cabo Verde Unspecified island		SANTIAGO		FOGO		SANTO ANTAO		SAO NICOLAU		BRAVA		MAIO		BOA VISTA		SAO VICENTE		SAL	
1890	Census* = 127390	**1. p12-14 7. p297**																		
1898	Census* = 142537	**7. p295**																		
1899-1900					Famine and smallpox	**1. p12-14**			Famine and smallpox	**1. p12-14**										
1900	Census* = 147424	**1. p12-14 7. p297**																		
1901-1902	Famine	**1. p12-14**			Great poverty	**1. p12-14**	Great poverty	**1. p12-14**	Diseases and fevers	**1. p12-14**	Diseases and fevers	**1. p12-14**	Great population reduction, no census	**1. p12-14**						
1903-1904			Famine kills 15000	**1. p12-14**																
1905	Census* = 135000	**1. p12-14**																		
1910	Census* = 142252	**1. p12-14 7. p295**											Census* =							
1913	Census* = 147754	**7. p295**	Census* = 59222	**7. p295**	Census* = 17800	**7. p295**	Census* = 33724	**7. p295**	Census* = 12041	**7. p295**	Census* = 9207	**7. p295**	Census* = 1867	**7. p295**	Census* = 2823	**7. p295**	Census* = 10491	**7. p295**	Census* = 579	**7. p295**
1920	Census* = 160000	**1. p12-14**																		
1930	Census* = 146299	**1. p15**																		
1940	Census* = 181740	**1. p16**	Census* = 77382	**1. p16**	Census* = 23022	**1. p16**	Census* = 35977	**1. p16**	Census* = 14846	**1. p16**	Census* = 8528	**1. p16**	Census* = 2237	**1. p16**	Census* = 2779	**1. p16**	Census* = 15848	**1. p16**	Census* = 1121	**1. p16**
1942					Census* = 25000	**1. p12-14**														
1941-1943	Famine	**1. p12-14**			Famine kills 7500 (31%)	**1. p12-14**			Famine kills 28%, no census	**1. p12-14**										
1944					Census* = 17500	**1. p12-14**														
1946-1948	Great famine kills 65% of the total population	**1. p12-14**																		
1949																				
1950	Census* = 149989-148331	**1. p16 7. p295**	Census* = 59384	**1. p16**	Census* = 17582	**1. p16**	Census* = 28379	**1. p16**	Census* = 10366	**1. p16**	Census* = 7937	**1. p16**	Census* = 1942-1044	**1. p16 7. p297**	Census* = 2985	**1. p16**	Census* = 19576	**1. p16**	Census* = 1838	**1. p16**
1960	Census* = 199902	**1. p16**	Census* = 88587	**1. p16**	Census* = 25615	**1. p16**	Census* = 33953	**1. p16**	Census* = 13866	**1. p16**	Census* = 8625	**1. p16**	Census* = 2680	**1. p16**	Census* = 3263	**1. p16**	Census* = 20705	**1. p16**	Census* = 2608	**1. p16**
1970	Census* = 270999-272071	**1. p16 7. p295**	Census* = 128782-129358	**1. p16 7. p295**	Census* = 29412	**1. p16**	Census* = 44623	**1. p16**	Census* = 16308	**1. p16**	Census* = 7756	**1. p16**	Census* = 3466	**1. p16**	Census* = 3569	**1. p16**	Census* = 31578	**1. p16**	Census* = 5505	**1. p16**
1975	**Independence of Cabo Verde**																			
1980	Census* = 295703-320000	**1. p16 7. p295**	Census* = 145957	**1. p16**	Census* = 30978	**1. p16**	Census* = 43321	**1. p16**	Census* = 13572	**1. p16**	Census* = 6985	**1. p16**	Census* = 4098	**1. p16**	Census* = 3372	**1. p16**	Census* = 41594	**1. p16**	Census* = 5826	**1. p16**
1980-1985	Emigration -2552/year + Diaspora = 405-420K -> 250K USA ; 83-88K West Europe ; 68-76K continental Africa	**1. p16**																		
2015	Census* = 524833	**8. p.36**	Census* = 294135	**8. p.36**	Census* = 35837	**8. p.36**	Census* = 40547	**8. p.36**	Census* = 12424	**8. p.36**	Census* = 5698	**8. p.36**	Census* = 6980	**8. p.36**	Census* = 14451	**8. p.36**	Census* = 81014	**8. p.36**	Census* = 33747	**8. p.36**

REFERENCES

1. Albuquerque, Luís de, and Maria Emília Madeira Santos, eds. 1991. *História geral de Cabo Verde T1*. Vol. 1. 3 vols. Lisboa : Praia [Cape Verde]: Centro de Estudos de História e Cartografia Antiga, Instituto de Investigação Científica Tropical ; Instituto Nacional da Cultura de Cabo Verde.

2. Albuquerque, Luís de, and Maria Emília Madeira Santos, eds. 1995. *História geral de Cabo Verde T2*. 2. ed. Vol. 2. 3 vols. Lisboa: Centro de Estudos de História e Cartogr. Antiga.

3. Albuquerque, Luís de, and Maria Emília Madeira Santos, eds. 2007. *História concisa de Cabo Verde (Resumo da História geral de Cabo Verde T1-3)*. 1 vols. Lisboa: Inst. de Investigação Científica Tropical.

4. Carreira, A. 2000. *Cabo Verde: Formação e Extinção de Uma Sociedade Escravocrata (1460-1878)*. 3rd ed. Estudos e Ensaios. Praia, Cabo Verde: IPC.

5. *The voyage of M. George Fenner to Guinea and to the Isles of Capo Verde, An. 1566.* In Hakluyt, Richard. 2014. The Principal Navigations, Voyages, Traffiques and Discoveries of the English Nation. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107286306.

6. Brooks, George E. 2006. "Cabo Verde: Gulag of the South Atlantic: Racism, Fishing Prohibitions, and Famines." *History in Africa* 33: 101–35. https://doi.org/10.1353/hia.2006.0008.

7. Patterson, K. David. 1988. "Epidemics, Famines, and Population in the Cape Verde Islands, 1580-1900." *The International Journal of African Historical Studies* 21 (2): 291. https://doi.org/10.2307/219938.

8. Cabo Verde, Statistical Yearbook 2015, Instituto Nacional de Estatística, Praia, Cabo Verde. https://ine.cv/wp-content/uploads/2017/02/statistical-yearbook-cv-2015_en.pdf.

9. Conrad, Robert. 1969. "The Contraband Slave Trade to Brazil, 1831-1845." *Hispanic American Historical Review* 49 (4): 617–38. https://doi.org/10.1215/00182168-49.4.617.

1543 **Table ST3:**
1544 **Summary-statistics used for *MetHis*-machine-learning ABC inferences.**
1545

Summary Statistic for ABC		n	ref	software
within population	Mean ASD within population H	1	Bowcock et al. 1994	MetHis SumStat tools
	Mean Heterozygosity (SNP by SNP) within population H	1	Nei 1978	MetHis SumStat tools
	Variance Heterozygosity (SNP by SNP) within population H	1	Nei 1978	MetHis SumStat tools
	Mean inbreeding F within population H	1	Danecek et al. 2011	MetHis SumStat tools
	Variance inbreeding F within population H	1	Danecek et al. 2011	MetHis SumStat tools
admixture pattern	Mode ASD-MDS African admixture proportions in population H	1	Verdu and Rosenberg 2011; Fortes-Lima et al. 2021	MetHis SumStat tools
	Mean ASD-MDS African admixture proportions in population H	1	Verdu and Rosenberg 2011; Fortes-Lima et al. 2021	MetHis SumStat tools
	Variance ASD-MDS African admixture proportions in population H	1	Verdu and Rosenberg 2011; Fortes-Lima et al. 2021	MetHis SumStat tools
	Skewness ASD-MDS African admixture proportions in population H	1	Verdu and Rosenberg 2011; Fortes-Lima et al. 2021	MetHis SumStat tools
	Kurtosis ASD-MDS African admixture proportions in population H	1	Verdu and Rosenberg 2011; Fortes-Lima et al. 2021	MetHis SumStat tools
	Min ASD-MDS African admixture proportions in population H	1	Verdu and Rosenberg 2011; Fortes-Lima et al. 2021	MetHis SumStat tools
	Max ASD-MDS African admixture proportions in population H	1	Verdu and Rosenberg 2011; Fortes-Lima et al. 2021	MetHis SumStat tools
	10%-percentiles of ASD-MDS African admixture proportions in population H	9	Verdu and Rosenberg 2011; Fortes-Lima et al. 2021	MetHis SumStat tools
	Mode ASD-MDS "African-European angles" in population H	1	This study ; Supplementary Figure SF12	MetHis SumStat tools
	Mean ASD-MDS "African-European angles" in population H	1	This study ; Supplementary Figure SF12	MetHis SumStat tools
	Variance ASD-MDS "African-European angles" in population H	1	This study ; Supplementary Figure SF12	MetHis SumStat tools
	Skewness ASD-MDS "African-European angles" in population H	1	This study ; Supplementary Figure SF12	MetHis SumStat tools
	Kurtosis ASD-MDS "African-European angles" in population H	1	This study ; Supplementary Figure SF12	MetHis SumStat tools
	Min ASD-MDS "African-European angles" in population H	1	This study ; Supplementary Figure SF12	MetHis SumStat tools
	Max ASD-MDS "African-European angles" in population H	1	This study ; Supplementary Figure SF12	MetHis SumStat tools
	10%-percentiles of ASD-MDS"African-European angles" in population H	9	This study ; Supplementary Figure SF12	MetHis SumStat tools
between populations	FST(African Source - Population H)	1	Weir and Cockerham 1984	MetHis SumStat tools
	FST(European Source - Population H)	1	Weir and Cockerham 1984	MetHis SumStat tools
	Mean ASD (African Source - Population H)	1	Bowcock et al. 1994	MetHis SumStat tools
	Mean ASD (European Source - Population H)	1	Bowcock et al. 1994	MetHis SumStat tools
	f3 (African Source, European Source ; Population H)	1	Patterson et al. 2012	MetHis SumStat tools

1546

1547 **REFERENCES**

1548 Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J. R. Kidd, and L. L. Cavalli-Sforza. 1994. "High
1549 Resolution of Human Evolutionary Trees with Polymorphic Microsatellites." Nature 368 (6470): 455–57.
1550 https://doi.org/10.1038/368455a0.

1551 Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, et al. 2011. "The
1552 Variant Call Format and VCFtools." Bioinformatics 27 (15): 2156–58.
1553 https://doi.org/10.1093/bioinformatics/btr330.

1554 Fortes-Lima, C.A., Laurent, R., Thouzeau, V., Toupance, B., and Paul Verdu. 2021. "Complex Genetic Admixture
1555 Histories Reconstructed with Approximate Bayesian Computation." *Molecular Ecology Resources* 21 (4):
1556 1098–1117. https://doi.org/10.1111/1755-0998.13325.

1557 Nei, Masatoshi. 1978. "Estimation of Average Heterozygosity and Genetic Distance from a Small Number of
1558 Individuals." Genetics 89 (3): 583–90. https://doi.org/10.1093/genetics/89.3.583.

1559 Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and David
1560 Reich. 2012. "Ancient Admixture in Human History." Genetics 192 (3): 1065–93.
1561 https://doi.org/10.1534/genetics.112.145037.
1562 Verdu, Paul, and Noah A Rosenberg. 2011. "A General Mechanistic Model for Admixture Histories of Hybrid
1563 Populations." Genetics 189 (4): 1413–26. https://doi.org/10.1534/genetics.111.132787.
1564 Weir, B. S., and C. Clark Cockerham. 1984. "Estimating F-Statistics for the Analysis of Population-Structure."
1565 Evolution 38 (6): 1358–70. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x.
1566
1567

Table ST4:

Average parameter posterior random cross-validation error across all model parameters as a function of the number of neurons in the hidden layer and the rejection tolerance level (number of simulations retained for training the Neural-Network) under the winning scenario for each island respectively.

For each island separately, we considered, 1,000 random simulations in-turn as pseudo-observed data to estimate posterior parameter distributions and 100,000 total simulations in the reference table. For each cross-validation procedure, we considered between 5 and 11 neurons in the hidden layer ("NN-HL") for the winning scenarios with 12 original parameters, and between 5 and 12 NN-HL for the winning scenarios with 13 original parameters. We considered three different tolerance levels of 0.01, 0.03, and 0.1 ("Tol.") corresponding, respectively, to 1,000, 3,000, and 10,000 simulations, in turn closest to each one of the 1,000 cross-validation pseudo-observed simulation retained for training the NN. The median values of posterior parameter distributions were used as point estimates for the calculation of the error of each parameter (Csilléry et al. 2012).

For each birth-island and corresponding winning scenario separately, we considered, for further posterior parameter estimation using the observed "real" data, only the pair of tolerance level and number of hidden neurons that minimized the average error on posterior parameter estimations across all parameters (indicated in bold in the table).

Island	Winning scenario	Number of scenario parameters	NN-HL = 5 Tol. = 0.01	NN-HL = 5 Tol. = 0.03	NN-HL = 5 Tol. = 0.1	NN-HL = 6 Tol. = 0.01	NN-HL = 6 Tol. = 0.03	NN-HL = 6 Tol. = 0.1	NN-HL = 7 Tol. = 0.01	NN-HL = 7 Tol. = 0.03	NN-HL = 7 Tol. = 0.1	NN-HL = 8 Tol. = 0.01	NN-HL = 8 Tol. = 0.03	NN-HL = 8 Tol. = 0.1	NN-HL = 9 Tol. = 0.01	NN-HL = 9 Tol. = 0.03	NN-HL = 9 Tol. = 0.1	NN-HL = 10 Tol. = 0.01	NN-HL = 10 Tol. = 0.03	NN-HL = 10 Tol. = 0.1	NN-HL = 11 Tol. = 0.01	NN-HL = 11 Tol. = 0.03	NN-HL = 11 Tol. = 0.1	NN-HL = 12 Tol. = 0.01	NN-HL = 12 Tol. = 0.03	NN-HL = 12 Tol. = 0.1
SANTIAGO	Scenario 1 Afr2pulses-Eur2pulses	12	0.79264	0.80866	0.84002	**0.77139**	0.79501	0.84684	0.78525	0.80899	0.85600	0.78512	0.80595	0.83938	0.77910	0.81192	0.84379	0.78593	0.80159	0.85245	0.79994	0.82611	0.84914	*na*	*na*	*na*
FOGO	Scenario 1 Afr2pulses-Eur2pulses	12	0.79943	0.81252	0.84001	0.80186	0.81397	0.84072	0.79836	0.81151	0.84844	0.79805	0.81077	0.84904	0.79899	0.81046	0.84721	**0.79615**	0.80818	0.84242	0.81073	0.80802	0.83930	*na*	*na*	*na*
SANTO ANTAO	Scenario 1 Afr2pulses-Eur2pulses	12	0.80170	0.81257	0.84662	0.79800	0.80533	0.84028	0.79075	0.81000	0.83704	0.79127	0.81353	0.83681	0.81027	0.81474	0.83676	0.80275	0.81143	0.84222	**0.78479**	0.80825	0.84011	*na*	*na*	*na*
SÃO NICOLAU	Scenario 1 Afr2pulses-Eur2pulses	12	0.79380	0.80711	0.84970	0.80765	0.80483	0.84484	0.81595	0.81035	0.84737	0.80886	0.82455	0.85231	0.80103	0.81642	0.84813	0.79444	0.80949	0.83444	**0.78709**	0.81653	0.84941	*na*	*na*	*na*
BRAVA	Scenario 1 Afr2pulses-Eur2pulses	12	0.81251	0.81703	0.84526	0.81085	0.83318	0.85753	0.82657	0.83069	0.85071	0.82241	0.82334	0.84982	0.81898	0.82593	0.85218	**0.80692**	0.82496	0.85410	0.82527	0.82511	0.85471	*na*	*na*	*na*
MAIO	Scenario 1 Afr2pulses-Eur2pulses	12	0.83232	0.83470	0.85366	0.82516	0.83673	0.85428	0.82662	0.83474	0.85582	0.83397	0.83718	0.84614	0.82483	0.82631	0.85223	**0.82095**	0.82787	0.85424	0.84091	0.82334	0.85756	*na*	*na*	*na*
BOA VISTA	Scenario 2 Afr2pulses-EurReccurring	13	0.79972	0.79708	0.83374	0.78239	0.80492	0.82136	0.79304	0.80060	0.81083	**0.77998**	0.79490	0.83499	0.79744	0.80830	0.83254	0.80121	0.79994	0.83049	0.79414	0.80234	0.82120	0.79915	0.79959	0.81731
SÃO VICENTE	Scenario 3 AfrReccurring-Eur2pulses	13	0.77795	0.80208	0.82312	0.77743	0.80127	0.81617	0.77898	0.79491	0.81619	0.78077	0.78639	0.81671	0.78619	0.78640	0.81696	0.77644	0.78673	0.81742	**0.77365**	0.80333	0.81919	0.78138	0.79306	0.82090
SAL	Scenario 2 Afr2pulses-EurReccurring	13	0.82888	0.81783	0.83264	0.81601	0.81325	0.83773	0.80415	0.81855	0.83779	0.80992	0.80666	0.82696	0.81770	0.82212	0.81723	0.81801	0.81565	0.84950	**0.79984**	0.81750	0.83044	0.81418	0.81551	0.82284

1585 **Table ST5-Santiago:**
1586 **NN-ABC posterior parameter estimations for Santiago island, cross-validation posterior parameter errors, and cross-**
1587 **validation 95% Credibility Interval accuracies.**
1588 Cross-validation errors and 95% CI accuracies are based on 1,000 NN-ABC posterior parameter inferences using, in turn, each
1589 one of the 1,000 simulations closest to the observed data used as pseudo-observed data, and the remaining 99,999 simulations
1590 as the reference table.
1591 Error calculations are based on the mode and, separately, the median point estimate for each 1,000 pseudo-observed simulations
1592 posterior parameter estimation, compared to the known parameter used for simulation. We considered 6 neurons in the hidden
1593 layer and a tolerance level of 0.01 (1000 simulations) for all posterior parameter estimation in this analysis as identified in
1594 **Table ST4**. Plotted distributions for posterior parameter estimations can be found in **Figure SF24-26**.
1595
1596

SANTIAGO	**Parameter posterior estimation**				**1,000 cross-validation errors**		
Afr2P-Eur2P scenario parameters	**Mode**	**Median**	**50% Credibility Interval**	**95% Credibility Interval**	**Mode mean absolute error**	**Median mean absolute error**	**95% CI length accuracy**
Ne.0	142	431	[194 - 692]	[28 - 974]	305	255	0.059
Ne.20	92109	80077	[63220 - 91349]	[30408 - 98975]	25426	21558	0.045
u.Ne	0.440	0.365	[0.284 - 0.438]	[0.150 - 0.494]	0.113	0.096	0.042
sAfr,0	0.8816	0.6365	[0.3884 - 0.8415]	[0.0538 - 0.9858]	0.3147	0.2507	0.056
tAfr,p1	13.9	9.2	[4.9 - 13.4]	[1.7 - 16.8]	5.3	4.2	0.101
sAfr,tAfr,p1	0.7844	0.5986	[0.3421 - 0.7861]	[0.0420 - 0.9479]	0.2618	0.2204	0.054
tAfr,p2	17.2	16.6	[15.2 - 17.5]	[10.0 - 18.9]	1.8	1.6	0.199
sAfr,tAfr,p2	0.6837	0.5913	[0.3894 - 0.7241]	[0.0642 - 0.9038]	0.1597	0.1458	0.052
tEur,p1	3.4	7.5	[4.3 - 11.1]	[2.3 - 15.9]	4.0	3.1	0.161
sEur,tEur,p1	0.1617	0.3969	[0.1957 - 0.6606]	[0.0185 - 0.9564]	0.3042	0.2439	0.056
tEur,p2	19.1	18.7	[17.7 - 19.2]	[11.7 - 19.6]	3.0	2.9	0.115
sEur,tEur,p2	0.0258	0.0684	[0.0250 - 0.1501]	[0.0028 - 0.6021]	0.2749	0.2254	0.057

1597
1598
1599

1600 **Table ST6-Fogo:**
1601 **NN-ABC posterior parameter estimations for Fogo island, cross-validation posterior parameter errors, and cross-**
1602 **validation 95% Credibility Interval accuracies.**
1603 Cross-validation errors and 95% CI accuracies are based on 1,000 NN-ABC posterior parameter inferences using, in turn, each
1604 one of the 1,000 simulations closest to the observed data used as pseudo-observed data, and the remaining 99,999 simulations
1605 as the reference table.
1606 Error calculations are based on the mode and, separately, the median point estimate for each 1,000 pseudo-observed simulations
1607 posterior parameter estimation, compared to the known parameter used for simulation. We considered 10 neurons in the hidden
1608 layer and a tolerance level of 0.01 (1000 simulations) for all posterior parameter estimation in this analysis as identified in
1609 **Table ST4**. Plotted distributions for posterior parameter estimations can be found in **Figure SF24-26**.
1610
1611

FOGO	**Parameter posterior estimation**				**1,000 cross-validation errors**		
Afr2P-Eur2P scenario parameters	**Mode**	**Median**	**50% Credibility Interval**	**95% Credibility Interval**	**Mode mean absolute error**	**Median mean absolute error**	**95% CI length accuracy**
Ne.0	349	511	[274 - 737]	[44 - 969]	295	253	0.063
Ne.20	90370	72023	[50948 - 87248]	[18011 - 98846]	27142	22749	0.057
u.Ne	0.115	0.150	[0.104 - 0.230]	[0.055 - 0.447]	0.085	0.081	0.074
sAfr,0	0.7407	0.6831	[0.5218 - 0.8152]	[0.1679 - 0.9600]	0.2936	0.2432	0.069
tAfr,p1	13.3	12.0	[8.1 - 14.4]	[4.7 - 16.9]	4.2	3.3	0.157
sAfr,tAfr,p1	0.4667	0.4549	[0.2286 - 0.6721]	[0.0251 - 0.9555]	0.2498	0.2315	0.056
tAfr,p2	19.5	19.3	[18.9 - 19.5]	[15.6 - 19.8]	2.6	2.5	0.100
sAfr,tAfr,p2	0.1174	0.2034	[0.1051 - 0.3560]	[0.0120 - 0.8225]	0.2032	0.1891	0.060
tEur,p1	2.5	5.3	[2.7 - 8.8]	[1.9 - 13.9]	3.6	3.0	0.216
sEur,tEur,p1	0.1149	0.2234	[0.1146 - 0.3943]	[0.0223 - 0.8383]	0.2804	0.2371	0.062
tEur,p2	16.7	15.5	[13.1 - 16.8]	[6.7 - 18.2]	3.2	3.0	0.070
sEur,tEur,p2	0.2101	0.2428	[0.1393 - 0.3966]	[0.0157 - 0.8485]	0.1751	0.1686	0.060

1612
1613
1614

1615 **Table ST7-Santo Antão:**
1616 **NN-ABC posterior parameter estimations for Santo Antão island, cross-validation posterior parameter errors, and**
1617 **cross-validation 95% Credibility Interval accuracies.**
1618 Cross-validation errors and 95% CI accuracies are based on 1,000 NN-ABC posterior parameter inferences using, in turn, each
1619 one of the 1,000 simulations closest to the observed data used as pseudo-observed data, and the remaining 99,999 simulations
1620 as the reference table.
1621 Error calculations are based on the mode and, separately, the median point estimate for each 1,000 pseudo-observed simulations
1622 posterior parameter estimation, compared to the known parameter used for simulation. We considered 11 neurons in the hidden
1623 layer and a tolerance level of 0.01 (1000 simulations) for all posterior parameter estimation in this analysis as identified in
1624 **Table ST4**. Plotted distributions for posterior parameter estimations can be found in **Figure SF24-26**.
1625
1626

SANTO ANTAO	**Parameter posterior estimation**				**1,000 cross-validation errors**		
Afr2P-Eur2P scenario parameters	**Mode**	**Median**	**50% Credibility Interval**	**95% Credibility Interval**	**Mode mean absolute error**	**Median mean absolute error**	**95% CI length accuracy**
Ne.0	411	469	[254 - 706]	[26 - 964]	254	228	0.071
Ne.20	12056	41256	[17145 - 69559]	[3570 - 96855]	32106	25308	0.051
u.Ne	0.035	0.053	[0.033 - 0.093]	[0.015 - 0.334]	0.072	0.067	0.056
sAfr,0	0.8047	0.5681	[0.2995 - 0.7964]	[0.0260 - 0.9839]	0.3029	0.2586	0.068
tAfr,p1	2.4	4.9	[2.9 - 7.4]	[1.9 - 11.9]	3.3	2.6	0.164
sAfr,tAfr,p1	0.7574	0.6255	[0.3644 - 0.7916]	[0.0451 - 0.9749]	0.2512	0.2283	0.055
tAfr,p2	13.6	11.6	[7.7 - 14.9]	[3.0 - 18.5]	2.8	2.7	0.072
sAfr,tAfr,p2	0.6294	0.4716	[0.2596 - 0.6733]	[0.0338 - 0.9233]	0.2435	0.2150	0.049
tEur,p1	3.2	5.8	[3.2 - 10.2]	[2.6 - 15.6]	2.6	2.4	0.265
sEur,tEur,p1	0.1542	0.3592	[0.1709 - 0.6051]	[0.0216 - 0.9522]	0.2767	0.2347	0.059
tEur,p2	18.1	17.2	[14.8 - 18.2]	[7.6 - 19.1]	3.2	3.0	0.074
sEur,tEur,p2	0.0467	0.0847	[0.0416 - 0.1656]	[0.0056 - 0.6382]	0.1868	0.1763	0.059

1627
1628
1629
1630

1631 **Table ST8-São Nicolau:**
1632 **NN-ABC posterior parameter estimations for São Nicolau island, cross-validation posterior parameter errors, and**
1633 **cross-validation 95% Credibility Interval accuracies.**
1634 Cross-validation errors and 95% CI accuracies are based on 1,000 NN-ABC posterior parameter inferences using, in turn, each
1635 one of the 1,000 simulations closest to the observed data used as pseudo-observed data, and the remaining 99,999 simulations
1636 as the reference table.
1637 Error calculations are based on the mode and, separately, the median point estimate for each 1,000 pseudo-observed simulations
1638 posterior parameter estimation, compared to the known parameter used for simulation. We considered 11 neurons in the hidden
1639 layer and a tolerance level of 0.01 (1000 simulations) for all posterior parameter estimation in this analysis as identified in
1640 **Table ST4**. Plotted distributions for posterior parameter estimations can be found in **Figure SF24-26**.
1641
1642

SAO NICOLAU	**Parameter posterior estimation**				**1,000 cross-validation errors**		
Afr2P-Eur2P scenario parameters	**Mode**	**Median**	**50% Credibility Interval**	**95% Credibility Interval**	**Mode mean absolute error**	**Median mean absolute error**	**95% CI length accuracy**
Ne.0	360	457	[238 - 734]	[34 - 979]	309	255	0.059
Ne.20	28767	48793	[25155 - 70475]	[3686 - 96753]	27502	23452	0.055
u.Ne	0.135	0.238	[0.133 - 0.358]	[0.025 - 0.483]	0.132	0.113	0.057
sAfr,0	0.3249	0.4730	[0.2491 - 0.7393]	[0.0327 - 0.9752]	0.3017	0.2461	0.051
tAfr,p1	2.4	5.2	[2.7 - 9.5]	[1.3 - 15.5]	5.9	4.5	0.145
sAfr,tAfr,p1	0.1695	0.3133	[0.1584 - 0.5404]	[0.0171 - 0.9262]	0.2897	0.2363	0.069
tAfr,p2	14.6	14.5	[13.2 - 14.9]	[8.3 - 16.4]	1.5	1.4	0.522
sAfr,tAfr,p2	0.7132	0.6398	[0.4900 - 0.7736]	[0.0877 - 0.9380]	0.1044	0.0981	0.055
tEur,p1	2.7	6.7	[3.5 - 10.8]	[2.0 - 16.1]	4.6	3.5	0.139
sEur,tEur,p1	0.1193	0.3240	[0.1547 - 0.5711]	[0.0145 - 0.9460]	0.2806	0.2394	0.048
tEur,p2	19.2	19.0	[17.7 - 19.4]	[10.3 - 19.8]	3.3	2.9	0.199
sEur,tEur,p2	0.0297	0.0731	[0.0295 - 0.1465]	[0.0038 - 0.6642]	0.2419	0.2067	0.065

1643
1644
1645

1646 **Table ST9-Brava:**
1647 **NN-ABC posterior parameter estimations for Brava island, cross-validation posterior parameter errors, and cross-**
1648 **validation 95% Credibility Interval accuracies.**
1649 Cross-validation errors and 95% CI accuracies are based on 1,000 NN-ABC posterior parameter inferences using, in turn, each
1650 one of the 1,000 simulations closest to the observed data used as pseudo-observed data, and the remaining 99,999 simulations
1651 as the reference table.
1652 Error calculations are based on the mode and, separately, the median point estimate for each 1,000 pseudo-observed simulations
1653 posterior parameter estimation, compared to the known parameter used for simulation. We considered 10 neurons in the hidden
1654 layer and a tolerance level of 0.01 (1000 simulations) for all posterior parameter estimation in this analysis as identified in
1655 **Table ST4**. Plotted distributions for posterior parameter estimations can be found in **Figure SF24-26**.
1656
1657

BRAVA	**Parameter posterior estimation**				**1,000 cross-validation errors**		
Afr2P-Eur2P scenario parameters	**Mode**	**Median**	**50% Credibility Interval**	**95% Credibility Interval**	**Mode mean absolute error**	**Median mean absolute error**	**95% CI length accuracy**
Ne.0	495	501	[309 - 696]	[63 - 919]	308	254	0.068
Ne.20	3278	7337	[2908 - 17935]	[647 - 73114]	31523	24989	0.056
u.Ne	0.435	0.420	[0.374 - 0.455]	[0.243 - 0.494]	0.077	0.073	0.050
sAfr,0	0.7337	0.5666	[0.3452 - 0.7698]	[0.0497 - 0.9737]	0.2836	0.2405	0.059
tAfr,p1	1.6	2.8	[1.5 - 5.3]	[1.1 - 12.2]	3.9	3.1	0.159
sAfr,tAfr,p1	0.7403	0.6688	[0.4987 - 0.7995]	[0.1353 - 0.9685]	0.2418	0.2232	0.060
tAfr,p2	15.7	13.6	[8.4 - 16.3]	[2.4 - 19.2]	2.8	2.7	0.069
sAfr,tAfr,p2	0.1261	0.2473	[0.1272 - 0.3647]	[0.0191 - 0.8266]	0.2118	0.1905	0.058
tEur,p1	3.6	7.3	[4.1 - 11.1]	[2.8 - 16.1]	3.4	2.9	0.204
sEur,tEur,p1	0.0157	0.0352	[0.0116 - 0.1161]	[0.0007 - 0.7239]	0.2658	0.2338	0.064
tEur,p2	17.9	16.3	[12.5 - 18.0]	[4.2 - 19.3]	3.3	3.1	0.054
sEur,tEur,p2	0.1281	0.1697	[0.1054 - 0.2554]	[0.0188 - 0.6504]	0.1844	0.1716	0.057

1658
1659
1660

1661 **Table ST10-Maio:**
1662 **NN-ABC posterior parameter estimations for Maio island, cross-validation posterior parameter errors, and cross-**
1663 **validation 95% Credibility Interval accuracies.**
1664 Cross-validation errors and 95% CI accuracies are based on 1,000 NN-ABC posterior parameter inferences using, in turn, each
1665 one of the 1,000 simulations closest to the observed data used as pseudo-observed data, and the remaining 99,999 simulations
1666 as the reference table.
1667 Error calculations are based on the mode and, separately, the median point estimate for each 1,000 pseudo-observed simulations
1668 posterior parameter estimation, compared to the known parameter used for simulation. We considered 10 neurons in the hidden
1669 layer and a tolerance level of 0.01 (1000 simulations) for all posterior parameter estimation in this analysis as identified in
1670 **Table ST4**. Plotted distributions for posterior parameter estimations can be found in **Figure SF24-26**.
1671
1672

MAIO	**Parameter posterior estimation**				**1,000 cross-validation errors**		
Afr2P-Eur2P scenario parameters	**Mode**	**Median**	**50% Credibility Interval**	**95% Credibility Interval**	**Mode mean absolute error**	**Median mean absolute error**	**95% CI length accuracy**
Ne.0	122	500	[241 - 746]	[29 - 980]	276	240	0.059
Ne.20	7159	21454	[7388 - 49176]	[914 - 93656]	30301	25378	0.072
u.Ne	0.080	0.121	[0.071 - 0.210]	[0.017 - 0.421]	0.082	0.078	0.064
sAfr,0	0.9135	0.6757	[0.3486 - 0.8841]	[0.0177 - 0.9941]	0.3150	0.2509	0.061
tAfr,p1	2.4	4.2	[2.5 - 6.8]	[1.5 - 12.8]	4.1	3.2	0.126
sAfr,tAfr,p1	0.7703	0.5589	[0.2768 - 0.7758]	[0.0346 - 0.9769]	0.2630	0.2350	0.051
tAfr,p2	14.4	11.1	[6.6 - 15.1]	[2.4 - 18.8]	2.5	2.4	0.112
sAfr,tAfr,p2	0.6543	0.5276	[0.2918 - 0.7126]	[0.0452 - 0.9393]	0.2168	0.1937	0.052
tEur,p1	2.3	4.1	[2.2 - 7.4]	[1.7 - 14.2]	3.2	2.8	0.219
sEur,tEur,p1	0.1075	0.3260	[0.1376 - 0.5818]	[0.0117 - 0.9497]	0.3014	0.2415	0.068
tEur,p2	14.1	12.7	[8.7 - 14.9]	[3.7 - 17.9]	3.3	3.1	0.048
sEur,tEur,p2	0.1134	0.1472	[0.0957 - 0.2301]	[0.0233 - 0.5957]	0.2145	0.1899	0.057

1673
1674
1675

Table ST11-Boa Vista:

NN-ABC posterior parameter estimations for Boa Vista island, cross-validation posterior parameter errors, and cross-validation 95% Credibility Interval accuracies.

Cross-validation errors and 95% CI accuracies are based on 1,000 NN-ABC posterior parameter inferences using, in turn, each one of the 1,000 simulations closest to the observed data used as pseudo-observed data, and the remaining 99,999 simulations as the reference table.

Error calculations are based on the mode and, separately, the median point estimate for each 1,000 pseudo-observed simulations posterior parameter estimation, compared to the known parameter used for simulation. We considered 8 neurons in the hidden layer and a tolerance level of 0.01 (1000 simulations) for all posterior parameter estimation in this analysis as identified in **Table ST4**. Plotted distributions for posterior parameter estimations can be found in **Figure SF24-26**.

BOA VISTA	**Parameter posterior estimation**				**1,000 cross-validation errors**		
Afr2P-Eur Rec. scenario parameters	**Mode**	**Median**	**50% Credibility Interval**	**95% Credibility Interval**	**Mode mean absolute error**	**Median mean absolute error**	**95% CI length accuracy**
Ne.0	731	627	[357 - 793]	[57 - 977]	263	230	0.049
Ne.20	13128	47744	[20486 - 76777]	[2657 - 98587]	30386	25219	0.074
u.Ne	0.056	0.079	[0.051 - 0.125]	[0.014 - 0.346]	0.067	0.066	0.059
sAfr,0	0.1706	0.3917	[0.1814 - 0.6678]	[0.0188 - 0.9677]	0.3186	0.2517	0.061
tAfr,p1	14.6	10.4	[5.8 - 14.2]	[2.4 - 16.8]	4.6	3.7	0.137
sAfr,tAfr,p1	0.2027	0.3760	[0.1998 - 0.5640]	[0.0261 - 0.8771]	0.2494	0.2213	0.056
tAfr,p2	14.8	15.0	[14.2 - 16.0]	[11.0 - 17.4]	1.7	1.7	0.113
sAfr,tAfr,p2	0.7516	0.6553	[0.4758 - 0.7671]	[0.1262 - 0.8949]	0.1962	0.1801	0.048
tEur,p1	2.2	3.9	[2.3 - 6.6]	[1.7 - 11.8]	3.7	3.1	0.192
tEur,p2	8.6	8.2	[6.0 - 10.2]	[3.1 - 14.9]	3.3	3.2	0.090
sEur,tEur,p1	0.9219	0.8057	[0.6300 - 0.9141]	[0.2736 - 0.9907]	0.2723	0.2319	0.076
sEur,tEur,p2	0.120	0.340	[0.1643 - 0.5832]	[0.021 - 0.880]	0.1721	0.1579	0.071
u.sEur	0.4444	0.2783	[0.151 - 0.398]	[0.0160 - 0.4907]	0.151	0.128	0.065

1691 **Table ST12-São Vicente:**
1692 **NN-ABC posterior parameter estimations for São Vicente island, cross-validation posterior parameter errors, and**
1693 **cross-validation 95% Credibility Interval accuracies.**
1694 Cross-validation errors and 95% CI accuracies are based on 1,000 NN-ABC posterior parameter inferences using, in turn, each
1695 one of the 1,000 simulations closest to the observed data used as pseudo-observed data, and the remaining 99,999 simulations
1696 as the reference table.
1697 Error calculations are based on the mode and, separately, the median point estimate for each 1,000 pseudo-observed simulations
1698 posterior parameter estimation, compared to the known parameter used for simulation. We considered 11 neurons in the hidden
1699 layer and a tolerance level of 0.01 (1000 simulations) for all posterior parameter estimation in this analysis as identified in
1700 **Table ST4**. Plotted distributions for posterior parameter estimations can be found in **Figure SF24-26**.
1701
1702

SAO VICENTE	**Parameter posterior estimation**				**1,000 cross-validation errors**		
Afr Rec.-Eur2P scenario parameters	**Mode**	**Median**	**50% Credibility Interval**	**95% Credibility Interval**	**Mode mean absolute error**	**Median mean absolute error**	**95% CI length accuracy**
Ne.0	109	262	[115 - 486]	[19 - 918]	288	243	0.061
Ne.20	85838	66715	[42741 - 84818]	[12159 - 98375]	29947	23721	0.048
u.Ne	0.215	0.244	[0.174 - 0.338]	[0.072 - 0.469]	0.094	0.084	0.049
sAfr,0	0.8596	0.5983	[0.3334 - 0.8142]	[0.0529 - 0.9835]	0.3231	0.2499	0.063
tAfr,p1	12.6	9.7	[5.7 - 13.1]	[2.5 - 16.9]	3.9	3.4	0.153
tAfr,p2	17.5	15.8	[13.8 - 17.6]	[7.4 - 19.1]	2.8	2.7	0.235
sAfr,tAfr,p1	0.6796	0.7168	[0.5698 - 0.8525]	[0.2806 - 0.9795]	0.2031	0.1774	0.079
sAfr,tAfr,p2	0.2149	0.2786	[0.1561 - 0.4266]	[0.0256 - 0.7820]	0.1538	0.1457	0.063
u.sAfr	0.072	0.228	[0.111 - 0.360]	[0.015 - 0.484]	0.161	0.126	0.056
tEur,p1	2.8	6.4	[3.1 - 10.0]	[1.9 - 15.1]	4.1	3.3	0.214
sEur,tEur,p1	0.1772	0.3403	[0.1733 - 0.5442]	[0.0149 - 0.9133]	0.2448	0.2126	0.055
tEur,p2	17.7	17.4	[16.3 - 18.0]	[10.6 - 18.9]	2.0	2.1	0.143
sEur,tEur,p2	0.2596	0.2840	[0.1732 - 0.4177]	[0.0397 - 0.8044]	0.1585	0.1498	0.059

1703
1704
1705
1706

1707 **Table ST13-Sal:**
1708 **NN-ABC posterior parameter estimations for Sal island, cross-validation posterior parameter errors, and cross-**
1709 **validation 95% Credibility Interval accuracies.**
1710 Cross-validation errors and 95% CI accuracies are based on 1,000 NN-ABC posterior parameter inferences using, in turn, each
1711 one of the 1,000 simulations closest to the observed data used as pseudo-observed data, and the remaining 99,999 simulations
1712 as the reference table.
1713 Error calculations are based on the mode and, separately, the median point estimate for each 1,000 pseudo-observed simulations
1714 posterior parameter estimation, compared to the known parameter used for simulation. We considered 11 neurons in the hidden
1715 layer and a tolerance level of 0.01 (1000 simulations) for all posterior parameter estimation in this analysis as identified in
1716 **Table ST4**. Plotted distributions for posterior parameter estimations can be found in **Figure SF24-26**.
1717
1718

SAL	**Parameter posterior estimation**				**1,000 cross-validation errors**		
Afr2P-EurRec. scenario parameters	**Mode**	**Median**	**50% Credibility Interval**	**95% Credibility Interval**	**Mode mean absolute error**	**Median mean absolute error**	**95% CI length accuracy**
Ne.0	525	518	[285 - 745]	[41 - 966]	306	250	0.061
Ne.20	88248	63289	[36457 - 82778]	[5689 - 98181]	30007	24554	0.058
u.Ne	0.038	0.076	[0.035 - 0.155]	[0.005 - 0.399]	0.112	0.102	0.056
sAfr,0	0.8853	0.5820	[0.3164 - 0.8077]	[0.0247 - 0.9820]	0.3039	0.2508	0.060
tAfr,p1	3.2	9.0	[4.8 - 13.1]	[1.8 - 16.8]	4.9	4.1	0.122
sAfr,tAfr,p1	0.7261	0.5166	[0.2893 - 0.7374]	[0.0302 - 0.9626]	0.2667	0.2275	0.063
tAfr,p2	17.9	17.6	[16.3 - 18.0]	[12.5 - 19.0]	1.3	1.2	0.212
sAfr,tAfr,p2	0.1780	0.2401	[0.1374 - 0.3655]	[0.0197 - 0.6590]	0.1588	0.1462	0.071
tEur,p1	3.9	9.3	[5.3 - 12.6]	[3.3 - 16.8]	4.2	3.4	0.180
tEur,p2	17.8	16.2	[13.9 - 17.8]	[6.8 - 19.4]	3.7	3.5	0.102
sEur,tEur,p1	0.4931	0.5657	[0.3810 - 0.7810]	[0.1100 - 0.9754]	0.2585	0.2183	0.072
sEur,tEur,p2	0.0730	0.2061	[0.0842 - 0.4034]	[0.0113 - 0.7804]	0.1926	0.1734	0.076
u.sEur	0.137	0.207	[0.102 - 0.346]	[0.011 - 0.486]	0.156	0.128	0.061

1719
1720
1721

Table ST14:

Mantel correlations among individual birth-places, residence-places, maternal and paternal birth places, age, and academic education duration.

We considered only the 225 genetically unrelated Cabo Verde-born individuals in these analyses.

Target	Mantel variable	n	Geographic scale	Spearman rho	10000 Mantel two-sided permutation p
abs(Age difference)	log(Birth-loc dist.)	225	within and between islands	*0.0046*	*0.3524*
abs(Education duration difference)	log(Birth-loc dist.)	186	within and between islands	**0.0553**	**0.0007**
abs(Education duration difference)	log(Residence dist.)	185	within and between islands	**0.0809**	**< 2.10-4**
log(Residence dist.)	log(Birth-loc dist.)	224	within and between islands	**0.6910**	**< 2.10-4**
log(Father Birth-loc dist.)	log(Birth-loc dist.)	222	within and between islands	**0.5967**	**< 2.10-4**
log(Mother Birth-loc dist.)	log(Birth-loc dist.)	224	within and between islands	**0.7326**	**< 2.10-4**
abs(Age difference)	log(Birth-loc dist.)	225	within islands only	*0.1737*	*0.0047*
abs(Education duration difference)	log(Birth-loc dist.)	186	within islands only	*0.0889*	*0.1310*
abs(Education duration difference)	log(Residence dist.)	185	within islands only	*0.1632*	*0.0147*
log(Residence dist.)	log(Birth-loc dist.)	224	within islands only	*-0.0439*	*0.2752*
log(Father Birth-loc dist.)	log(Birth-loc dist.)	222	within islands only	*0.1865*	*0.0095*
log(Mother Birth-loc dist.)	log(Birth-loc dist.)	224	within islands only	**0.2356**	**0.0005**