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

Conflict of interest statement
The authors declare no conflict of interests for this work.

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 pro-

vided 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 CNRS 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).

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, Jorge M. Rocha, and Bruno Toupance, for useful comments and discussions about this work.

Keywords

Admixture; anthropology; population genetics; historical inference; machine-learning; approximate Bayesian computations; genetic diversity; linguistic diversity; 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 plantation economy, and after this era. Our results illustrate how shifting socio-cultural relationships between enslaved and non-enslaved communities shaped enslaved-African descendants' descent 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 genetic 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.

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 covariables 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 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**).

2. Genetic structure in Cabo Verde and other admixed populations related to the TAST

Based on these results, we further investigated patterns of individual genetic structure among Cabo Verdean-born individuals, ASW, and ACB populations with respect to European and Western, Central, South-Western, and Southern African populations (**Table ST1**), using ADMIXTURE [39]. Extended descriptions of the results are presented in **Supplementary Text**.

At K=2, the orange genetic-cluster in **Figure F3A** is maximized in African individuals while the blue alternative cluster is maximized in Europeans. Cabo Verdean, ASW and ACB individuals exhibit intermediate genotype-membership proportions between the two clusters, consistently with patterns expected for European-African admixed individuals. ACB individuals show, on average, the highest membership to the orange "African" cluster (88.23%, SD=7.33%), followed by the ASW (78.00%, SD=10.88%), and Cabo Verdeans (59.01%, SD=11.97%). Furthermore, membership proportions to this cluster are highly variable across Cabo Verdean islands, with highest average memberships for Santiago-born individuals (71.45%, SD=10.39%) and Maio (70.39%, SD=5.26%), and lowest for Fogo (48.10%, SD=6.89%) and Brava (50.61%, SD=5.80%). Furthermore, we found that inter-individual memberships' variation within Cabo Verde islands captured as F_{ST}/F_{ST}^{max} values [40], are significantly different across pairs of islands for 32 out of 36 comparisons (Wilcoxon rank-sum test $p<3.96\text{x}10^{-8}$), with variability across islands ranging from a lowest value of 0.010 in Santo Antao to a highest value of 0.0519 in Santiago (**Figure SF7**).

At *K*=5, the new red cluster is maximized in the YRI, Igbo and ESN populations, distinct from Western and Southern African orange and gray clusters, respectively. Note that the former orange cluster is almost completely replaced with red membership in the ACB and ASW populations. Conversely, Cabo Verdean individuals exhibit small to virtually no membership to this East Western African red cluster, but their African component, instead, resembled genetic patterns found among West Western Africans.

Ultimately, at *K*=10, the light-green cluster is maximized in Cabo Verdean individuals born on Maio, Boa Vista, Sal, and São Nicolau, distinct form the dark green cluster maximized in Santo Antão and São Vicente, and hence producing three distinct ADMIXTURE patterns among Cabo Verdean islands. Furthermore, an alternative mode at *K*=10 shows (**Figure SF8**), that Cabo Verdean individuals resemble more IBS and TSI patterns for their European-like membership than ASW and ACB individuals who, instead, resemble more CEU and GBR patterns, consistently with ASD-MDS results and previous studies [10,21].

3. Local-ancestry in Cabo Verde and other admixed populations related to the TAST

ASD-MDS and ADMIXTURE descriptive analyses do not formally test admixture and putative source populations of origins, they rather disentangle genetic-structure resemblances among individuals [39,41–44]. Therefore, we conducted admixture-LD haplotypic local-ancestry inferences with the SHAPEIT2-CHROMOPAINTER-SOURCEFIND pipeline [45–47], to identify the possible European and African populations at the sources of admixture patterns.

Figure F3B shows striking differences concerning both the European and the African source populations involved in the admixture history of ACB, ASW, and birth islands of Cabo Verdeans. We find that all Cabo Verdean islands of birth share almost all their European haplotypic ancestry with the Iberian-IBS population rather than other European populations. Santiago-born individuals present the smallest (27%) average haplotypic ancestry shared with IBS, and Fogo-born the highest (51%). Conversely, the ASW and ACB both share haplotypic ancestries only with the USA-CEU of North-Western European origin (20% and 12% respectively).

Furthermore, we find that all Cabo Verdeans almost exclusively share African haplotypic ancestries with two Senegambian populations (Mandinka and Wolof) and very reduced to no shared ancestries with other regions of Africa. More specifically, we find that the Mandinka from Senegal are virtually the sole African population to share haplotypic ancestry with Cabo Verdeans born in Brava, Fogo, Santiago, Maio, and Boa Vista, and the majority of the African shared ancestry for Sal, São Nicolau, São Vicente, and Santo Antão. In these four latter islands, we find shared haplotypic ancestry with the Wolof population ranging from 5%-4% for individuals born on Sal (considering six or four possible sources, respectively), up to 22%-16% for Santo Antão. Finally, we find limited (1%-6%) shared haplotypic ancestry with East Western (Igbo, YRI, or Ga-Adangbe) or South-West Central (Kimbundu, Kongo, or Ovimbundu) African populations in all Cabo Verdean islands, except Fogo and Brava, and the specific populations identified and their relative proportion of shared haplotypic ancestries varies across analyses. Conversely, we find that the ASW and ACB populations share African haplotypic ancestries in majority with East Western African populations (YRI, Ga-Adangbe, and Igbo), and substantial shared ancestries with Senegambian populations (5-10%), the MSL from Sierra Leone in Central Western Africa (5-12%), and South-West Central African populations (3-18%), albeit variable depending on the number of putative sources considered.

4. Genetic and linguistic diversity and admixture within Cabo Verde

Runs of Homozygosity (ROH) patterns within Cabo Verde

Runs of homozygosity (ROH) are Identical-By-Descent haplotypes resulting from recent parental relatedness and present in individuals as long stretches of homozygous genotypes. Their length and abundance can reflect demographic events, such as population bottlenecks and founder events, natural selection, and cultural preferences for endogamy [48–50].

We find higher levels of long ROH (≥1cM) in Cabo Verdeans compared to most other analyzed populations, including ASW and ACB (**Figures F4A, SF9**). We find the highest levels of long-ROH in individuals born on Maio, Boa Vista, Sal, São Nicolau, São Vincente, and Santo Antão, with a mean individual length of long-ROHs around 3cM (**Figure F4B**), and the lowest levels of long-ROH in Santiago and Brava. Among long ROH (**Figure SF10C**), we find little to no correlation with total non-ROH levels for African local-ancestry segments (Pearson ρ=-0.06689, p=0.3179), European (ρ=0.1551, p=0.01989), or Asian (ρ=0.06239, p=0.3516).

In admixed populations, we expected that some of the long ROH spanned local ancestry breakpoint switches (see **Materials and Methods**), indicating that the most recent common ancestor existed after the initial admixture event having generated local-ancestry patterns. Furthermore, we expected that these "spanning" ROH should be among the longest ROH observed if admixture occurred only in the past few generations. We find that (**Figure F4C**), almost uniformly across Cabo Verde, the longest ROH identified indeed spanned at least one ancestry breakpoint, excluding the very few Asian ancestry regions identified. Furthermore, correcting ancestry-specific long-ROH sizes (**Figure SF11**) for individual's total ancestry fraction of that ancestry, we find that individuals born in Brava and Fogo have, on average, an apparent overrepresentation of European ancestry in long-ROH (**Figure F4D**), and that individuals from Sal and São Vincente have, conversely, an apparent overrepresentation of African ancestry. Finally, we find that individuals from Santiago, Maio, Boa Vista, São Nicolau, and Santo Antão have relatively similar long-ROH levels in either African or European segments. This latter pattern may be consistent with these populations being founded by admixed individuals, while the former patterns could indicate, in addition to such founding effect, more recent or recurring contributions from either source, respectively (Mooney et al. 2018).

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 pair-wise 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 T2**, and **Table ST14**).

The first ASD-MDS axis mainly differentiates Brava and Fogo-born individuals 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 Boa Vista, São Nicolau, Sal, and Maio from all other islands. Furthermore, we find a significant positive correlation between ASD and actual individual birth-places across Cabo Verde (**Table T2**; Spearman ρ=0.2916, two-sided Mantel $p<2\text{x}10^{-4}$). This correlation increases within-islands excluding inter-island distances (Spearman ρ=0.3460, two-sided Mantel $p<2\text{x}10^{-4}$), 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 $p<2\text{x}10^{-4}$), as well as paternal and maternal birth-places respectively (**Table T2**). 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 (Verdu et al. 2017), 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 T2**). 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 (Categories A and B; **Material and Methods**) providing an estimate of individual African linguistic-admixture scores (Verdu et al. 2017). 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).

Finally, we find a significant positive correlation (Spearman ρ=0.2070, p=0.0018) between genetic and linguistic admixture in Cabo Verde (**Figure 5E**), indicating that individuals who frequently use African-related utterances in their manner of speaking Kriolu are more likely to exhibit higher levels of African genetic-admixture. This correlation remains, respectively, marginally significant and significant when considering utterances of strictly African-origin (Category A) or utterances with a dual European-African etymology (Category B) separately (Spearman ρ=0.1631, p=0.0143, and ρ=0.1829, p=0.0059, respectively).

Altogether, our genetic and linguistic diversity results generalize our previous results obtained in Santiago only [21], and further suggest that genetic and linguistic admixture histories occurred in parallel in Cabo Verde. Together with the above genetic and linguistic isolation-by-distance signals and ROH results, this strongly suggests that different founding events and admixture histories for each Cabo Verde islands are at the root of genetic and linguistic diversity in the archipelago.

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

MetHis–ABC prior checking

We considered four competing genetic-admixture scenarios (**Figure F6**), tested separately for individuals born on each Cabo Verdean island with *MetHis*–ABC machine-learning scenario-choice and posterior parameter inferences [37,52,53]. 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.

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 more complex scenarios involving a period of recurring admixture from either source best explain the history of the remaining three islands. Conversely, periods of recurring admixture processes from both Africa and Europe ("AfrRecurring-EurRecurring") are least favored throughout Cabo Verde.

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).

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 a priori, increased assignation-errors in large parts of the parameter-space where scenarios are, by definition, biologically equivalent [37,54]. 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 [37,55].

MetHis–Neural Network-ABC posterior parameter estimations

For each Cabo Verdean individuals from each island of birth separately, we performed NN-ABC joint posterior parameter estimation based on 100,000 *MetHis* simulations under the winning scenario. Posterior parameter estimation errors were based on NN-ABC cross-validations for the 1,000 simulations closest to the observed data considered, in turn, as pseudo-observed [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.

DISCUSSION

Which African and European populations contributed genetically to Cabo Verde?

The genetic heritage of continental Africa to Cabo Verde

Numerous enslaved-African populations from Western, Central, and South-Western Central Africa have been forcibly deported during the TAST to both Cabo Verde and the Americas, as shown by historical demographic records [1,23]. There is still extensive debate about whether enslaved-Africans remained or more briefly transited in Cabo Verde during the most intense period of the TAST, in the 18th and 19th centuries, when the archipelago served as a slave-trade platform between continents [23,56,57]; a question also of major interest throughout the Americas [18,35].

In this context, our genetic results favor scenarios where mostly certain Senegambian populations (Mandinka and Wolof in our study) contribute to the genetic makeup of Cabo Verde (**Figures F2** and **F3**). Other Western, Central, and South-Western African populations historically deported during the TAST seem to have had only very limited contributions to the genomic diversity of Cabo Verde, and only in certain islands. This could be due to Cabo Verde being only a temporarily waypoint for these latter enslaved-African populations between Africa, the Americas, and Europe, but would also be consistent with additional socio-historical processes (see below). Interestingly, and further echoing these genetic results, the Cabo Verdean Kriolu language carries specific signatures mainly from the Mande language-family, and Wolof and Temne languages from Western Africa, and largely more limited signatures of Kikongo and Kimbundu Bantu languages from Central and South-Western Africa [15,28].

These results contrast with the admixture patterns identified in other enslaved-African descendant populations in the Americas in our dataset (African American and Barbadian, **Figures F2 and F3**), and in previous studies [3,4,10,36,58]. Indeed, the origins of African ancestries in numerous populations throughout the Caribbean and the Americas traced to varied continental African regions, from Western to South-Western Africa, thus qualitatively consistent with the known diversity of slave-trade routes used between continents and within the Americas after the Middle Passage.

The genetic heritage of continental Europe to Cabo Verde

After the initial settlement of Cabo Verde by Portuguese migrants, temporary changes of European dominion in certain islands, newly developing commercial opportunities, and intense piracy during the 16th and 17th centuries have triggered different European populations to migrate to the archipelago [24,25,34]. Nevertheless, these latter historical events do not seem to have left a major signature in the genetic landscape of Cabo Verde (**Figures F2 and F3**). Instead, we find that Cabo Verdean-born individuals in our dataset share virtually all their European genetic ancestry with Iberian populations, with extremely limited evidence of contributions from other European regions, consistent with previous studies [3,21]. Interestingly, the reduced diversity of European populations' contributions to the genomic landscape of Cabo Verde is also identified in other enslaved-African descendant populations in our study, as well as in previous studies in Caribbean and American populations [11,12,36]. Our results thus show that European admixture in enslaved-African descendant populations on both sides of the Atlantic mainly stem from the gene-pool of the European empires historically and socio-economically dominant locally, rather than subsequent European migrations [2].

Genetic and linguistic isolation-by-distance and difficult demographic expansion in Cabo Verde

A scenario of island peopling via series of founding events followed by slow-growing population sizes and relative isolation would consistently explain the increasing differentiation of island-specific patterns with increasing values of *K* found with ADMIXTURE, ASD-MDS and isolation-by-distance results (**Figures F2-F5**, **Table T2**), as well as *MetHis*-ABC demographic inferences (**Figure F7**).

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 T2**), 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 [59], 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 [60], 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 [61]. 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 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 specific 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 [1,19,62]. 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 history, which likely durably influenced genetic 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.335-362,63]. In addition, illegal slave-trade has flourished during this era, bringing numerous African deportees 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 [11,64], 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 sociocultural 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 genetic and cultural diversity 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 [65]. 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.

MATERIALS AND METHODS

1. Cabo Verde genetic and linguistic datasets

We conducted joint sampling of anthropological, genetic, and linguistic data in Cabo Verde with the only inclusion criteria that volunteer-participants would be healthy adults with Cabo Verdean citizenship and self-reporting speaking Kriolu [21]. Between 2010 and 2018, six interdisciplinary sampling-trips were conducted to interview 261 participants from more than thirty interview-locations throughout the archipelago (**Figure F1**).

1.a. Familial anthropology and geography data

The 261 Cabo Verdean individuals were each interviewed to record a classical familial anthropology questionnaire informing self-reported life-history mobility. In particular, we recorded primary residence location, birth location, parental and grand-parental birth and residence locations, and history of islands visited in Cabo Verde and foreign experiences [21]. Furthermore, we also recorded age, sex, marital status, and cumulated years of schooling and higher education (for 211 individuals only), and languages known and their contexts of use.

GPS coordinates for each reported location was acquired on site during field-work, supplemented by paper maps and Google Earth™ for foreign locations and islands that we did not visit (Sal and Maio). While participants' birth-locations were often precise, increasing levels of uncertainty arose for the reported parental and grand-parental locations. We arbitrarily assigned the GPS coordinates of the main population center of an island when only the island of birth or residence could be assessed with some certainty by participants. All other uncertain locations where recorded as missing data. Figure maps were designed with the software *QuantumGIS* v3.10 "Bucure□ti" and using Natural Earth free vector and raster map data (https://www.naturalearthdata.com).

1.b. Genome-wide genotyping data

The 261 participants each provided 2mL saliva samples collected with DNAGenotek™ OG-250 kits, and complete DNA was extracted following manufacturer's recommendation. DNA samples were genotyped using an Illumina HumanOmni2.5Million-BeadChip genotyping array following manufacturer's instructions. We followed the quality-control procedures for genotypic data curation using Illumina's GenomeStudio® Genotyping Module v2.0 described in [21]. Genotype-calling and population-level quality-controls, and merging procedures are detailed in **Figure SF1**.

In summary, we extracted a preliminary dataset of 259 Cabo Verdean Kriolu-speaking individuals, including relatives, genotyped at 2,118,835 polymorphic di-allelic autosomal SNPs genome-wide. We then merged this dataset with 2,504 worldwide samples from the 1000 Genomes Project Phase 3 [66]; with 1,307 continental African samples from the African Genome Variation Project [67] (EGA accession number EGAD00001000959); and with 1,235 African samples [7] (EGA accession number EGAS00001002078). We retained only autosomal SNPs common to all data sets, 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 [68], following previous procedure [21].

After merging all datasets (**Figure SF1**), we considered a final working-dataset of 5,157 worldwide unrelated samples, including 233 unrelated Cabo Verdean Kriolu-speaking individuals of which 225 were Cabo-Verde-born, genotyped at 455,705 autosomal bi-allelic SNPs (**Figure F1**; **Table ST1**).

1.c. Individual utterances of Kriolu

We collected linguistic data for each of the 261 Cabo Verdean individuals using anthropological linguistics questionnaires, and semi-directed interviews. Each participant was shown a brief (~6 min) speech-less movie, *"The Pear Story"* [69], 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 [70-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

We calculated pairwise Allele Sharing Dissimilarities [38] using the *asd* software (v1.1.0a; https://github.com/szpiech/asd), among the 5,157 individuals in our worldwide dataset (**Figure F1**; **Table ST1**), using 455,705 autosomal SNPs, considering, for a given pair of individuals, only the SNPs with no missing data. We then projected this dissimilarity matrix in three dimensions with metric Multidimensional Scaling using the *cmdscale* function in R [71]. We conducted successive MDS analyses on different individual subsets of the original ASD matrix, by excluding, in turn, groups of individuals and recomputing each MDS separately. Lists of populations included in each analysis can be found in **Table ST1**. 3D MDS animated plots in *gif* format (**Figures SF2-6**) were obtained with the *plot3d* and *movie3d* functions from the R packages *rgl* and *magick*.

Recently admixed individuals are intuitively expected to be at intermediate distances between the clusters formed by their putative proxy source populations on ASD-MDS two-dimensional plots. A putative estimate of individual admixture rates can then be obtained by projecting admixed individuals orthogonally on the line joining the respective centroids of each proxy-source populations and, then, calculating the distance between the projected points and either centroid, scaled by the distance between the two centroids (**Figure SF12**; [21,37]). We estimated such putative individual admixture rates in Cabo Verdean, ASW, and ACB individuals, considering sets of individuals for the putative proxy-source populations identified visually and resulting from our ASD-MDS decomposition (**Supplementary Text**).

2.b. ADMIXTURE-CLUMPP-DISTRUCT analyses

Based on ASD-MDS explorations, we focused on the genetic structure of individuals born in Cabo Verde compared to that of other admixed populations related to TAST migrations. Therefore, we conducted ADMIXTURE analyses [39] using 1,100 individuals from 22 populations: four from Europe, 14 from Africa, the USA-CEU, the African-American ASW, the Barbadian-ACB populations, and the North Chinese-CHB population as an outgroup (**Table ST1**). To limit clustering biases due to uneven sampling, we randomly resampled 50 individuals, once, for each one of these 22 populations. Furthermore, we also included all 44 Angolan individuals from four populations in the analyses, as no other samples from the region were available in our dataset. In addition to the 1,100 individuals hence obtained, we included the 225 Cabo Verde-born individuals.

Following constructor recommendations, we pruned the initial 455,705 SNPs set for low LD using *plink* [72] function *--indep-pairwise* for a 50 SNPs-window moving in increments of 10 SNPs and a r^2 threshold of 0.1. We thus conducted all subsequent ADMIXTURE analyses considering 1,369 individuals genotyped at 102,543 independent autosomal SNPs.

We performed 30 independent runs of ADMIXTURE for values of *K* ranging from 2 to 10. For each value of *K* separately, we identified groups of runs providing highly similar results (ADMIXTURE "modes"), with Symmetric Similarity Coefficient strictly above 99.9% for all pairs of runs within a mode, using CLUMPP [73]. We plotted each modal result comprising two ADMIXTURE runs or more, for each value of *K* separately, using DISTRUCT [74]. Finally, we evaluated within-population variance of individual membership proportions as F_{ST}/F_{ST}^{max} values using FSTruct [40] with 1000 Bootstrap replicates per population, for the ADMIXTURE mode result at $K=2$ (**Figure SF7**).

2.c. Local-ancestry inferences

To identify all populations sharing a likely common ancestry with the Cabo Verdean, ASW or ACB individuals in our dataset using local-ancestry haplotypic inferences, we considered the same sample-set as for the ADMIXTURE analysis (**Table ST1**), including all 455,705 SNPs from the merged dataset.

Phasing with ShapeIT2

We first phased individual genotypes using SHAPEIT2 [46] for the 22 autosomal chromosomes separately using the joint Hap Map Phase 3 Build GRCh38 genetic recombination map [75]. We considered default parameters using phasing windows of 2Mb and 100 states per SNP. We considered by default 7 burn-in iterations, 8 pruning iterations, 20 main iterations, and missing SNPs were imputed as monomorphic. Finally, we considered a "Ne" parameter of 30,000, and all individuals were considered unrelated.

Chromosome painting with ChromoPainter2

We determined the possible origins of each Cabo Verdean, ASW, and ACB individual phased haplotypes among European, African, USA-CEU, and Chinese-CHB populations using CHROMOPAINTER v2 [45] with the same recombination map used for phasing. Following authors' recommendations, we conducted a first set of 10 replicated analyses on a random subset of 10% of the individuals for chromosomes 2, 5, 12, and 19, which provided a posteriori Ne=233.933 and θ=0.0004755376, on average by chromosome weighted by chromosome sizes, to be used in the subsequent analysis. Then, we conducted a full CHROMOPAINTER analysis using these parameters to paint all individuals in our dataset, in turn set as Donor and Recipient. Finally, we combined painted chromosomes for each individual in the Cabo Verdean, ASW, and ACB population, separately.

Estimating possible source populations for the Cabo Verde gene-pool using SOURCEFIND

We used CHROMOPAINTER results aggregated for each Cabo Verdean, ACB, and ASW individual separately and conducted two SOURCEFIND [47] analyses using all other populations in the dataset as a possible source, separately for four or six possible source populations ("surrogates"). We considered 400,000 MCMC iteration steps (including 100,000 burn-in) and kept only one MCMC step every 10,000 steps for final likelihood estimation. Each individual genome was divided in 100 slots with possibly different ancestry, with an expected number of surrogates equal to 0.5 times the number of surrogates, for each SOURCEFIND analysis. We aggregated results obtained for all individuals in the ACB, ASW, and each Cabo Verdean birth-island, separately. We present the highest likelihood results across 20 separate iterations of the SOURCEFIND analysis in **Figure F3B.** The second-best results were highly consistent and thus not shown.

3. Genetic and linguistic diversities, and admixture within Cabo Verde

3.a. Runs of Homozygosity (ROH)

Considering the same sample and SNP set as in the above local-ancestry analyses (**Table ST1**), we called ROH with GARLIC [76]. For each population separately, we ran GARLIC using the weighted logarithm of the odds (wLOD) score-computation scheme, with a genotyping-error rate of $1e^{-3}$ (a likely overestimate), and using the same recombination map as for phasing, window sizes ranging from 30 to 90 SNPs in increment of 10 SNPs, 100 resampling to estimate allele frequencies, and all other GARLIC parameters set to default values. We only considered results obtained with a window size of 40 SNPs, which was the largest windows size associated with a bimodal wLOD score distribution and a wLOD score cutoff between the two modes, for all populations.

For each populations and Cabo Verdean islands separately, we considered three classes of ROH that correspond to the approximate time to the most recent common ancestor of the IBD haplotypes, which can be estimated from the equation g=100/2L, where L is the ROH length in cM and g is the number of generations to the most recent common ancestor of the haplotypes [48]. Short ROH are less than 0.25cM reflecting homozygosity of haplotypes from more than 200 generations ago; medium ROH are between 0.25cM and 1cM reflecting demographic events having occurred between approximately 200 and 50 generations ago; and long ROH are longer than 1cM reflecting haplotypes with a recent common ancestor less than 50 generations ago. In **Figure F4A-B** and **Figure SF9**, we plot the distribution of the summed

length of ROH per individual per size-classes.

3.b. Intersecting ROH and local ancestry painting

Using the same phasing results as described above, we conducted 10 EM iterations of the RFMIX2 [77] algorithm to assign, for each Cabo Verdean individual and each SNP on each chromosome, separately, it's putative source population of origin among the 24 African, European, Chinese-CHB, and USA-CEU populations. We collapsed the local ancestry assignments for each SNP in each Cabo Verdean individuals hence obtained into three continental regions, representing broadly African, broadly European, and broadly Asian ancestries respectively. We considered an approach similar to previous work [50], and intersected local ancestry calls with ROH calls (**Figure 4C-D**).

RFMIX2 only makes local ancestry calls at loci that are present in the dataset. Therefore, a gap of unclassified ancestry exists where inferred ancestry switches between two successive genotyped loci as called by RFMIX2. These gaps necessarily each contain an odd number of ancestry switch points ($\geq$1) absent from our marker set. Therefore, when computing the total ancestry content of an ROH that overlaps one of these ancestry breakpoints, we assign half of the length of this gap to each ancestry classification, effectively extending each local ancestry segment to meet at the midpoint.

We then plotted the length distribution of long ROH for each island and we breakout the distributions by ancestral haplotype background: those with only African ancestry, those with only European ancestry, and those that span at least one ancestry breakpoint (**Figure F4C**). We excluded long ROH in Asian ancestry segments from this and the following analyses, as we found overall very limited such ancestry in the samples (**Figure SF10**).

Finally, we explored how total African/European ancestry in long ROH varies between islands. For each individual we summed the total amount of each ancestry in long-ROH and plot the distributions across islands (**Figure SF11**). High levels of a given ancestry in long ROH could stem from an overall high level of that ancestry in that individual. Therefore, for each individual, we computed their global ancestry proportions by summing up the total length of the genome inferred as a given ancestry and dividing by the length of the genome. We then plotted (**Figure F4D**), the difference of an individual's long-ROH ancestry proportion and their global ancestry proportion, for African and European ancestries separately. Values above zero indicated that a given ancestry is overrepresented in long-ROH relative to genome-wide proportions of that ancestry.

3.c. Genetic and linguistic diversity within Cabo Verde

We explored genetic pairwise levels of differentiation calculated with ASD as previously, considering the 1,899,878 non-monomorphic SNPs within Cabo Verde obtained after QC Stage 3 (**Figure SF1**). We projected the matrix for the 225 unrelated Cabo Verde-born individuals using metric MDS as above. Note that pruning this data sets for very low LD using *plink* function *--indep-pairwise* for a 50 SNPs window moving in increments of 10 SNPs and a r^2 threshold of 0.025 resulted in 85,425 SNPs for which the ASD matrix was highly correlated with the one using all SNPs (Spearman ρ=0.8745, $p<2.2\text{x}10^{-16}$). In parallel, we described the diversity of Kriolu idiolects ("individuals' manners of speaking Kriolu") among the 225 genetically unrelated Cabo Verde-born individuals by projecting, with metric MDS, the matrix of pairwise Euclidean distances between vectors of individuals' utterance-frequencies (**Material and Methods 1.d**), considering the 4831 unique uttered items.

Then, we conducted series of Mantel and partial Mantel correlation tests (Legendre and Legendre 1998), using the *partial.mantel.test* function in the R package *ncf*, with Spearman correlation and 10,000 Mantel permutations, to explore possible isolation-by-distance [51] patterns as well as correlations with other covariables of interest. We conducted correlation tests between either the pairwise Euclidean distances between vectors of individuals' utterance-frequencies or genetic ASD separately, and individual pairwise matrices of *i)* absolute age differences, *ii)* absolute differences in academic education duration, *iii)* geographic distances

between residence-locations (logarithmic scale), *iv)* between birth-locations, *v)* between mothers' birth-locations, *vi)* between fathers' birth-locations (**Table T2**, **Table ST14**). All pairwise geographic distances were calculated with the Haversine great circle distance formulation [78], 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* [53], followed by Neural-Network posterior parameter estimation with the R package *abc* [52], 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 [55] 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 T1**), 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 $t_{Source,p1}$ and $t_{Source,p2}$ in [1,20], with $t_{Source,p1} \neq t_{Source,p2}$; and with intensity $s_{Source,tSource,p1}$ and $s_{Source,tSource,p2}$ 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 [79,80]. To do so, we conducted simulations using *MetHis* [37] forward-in-time simulator, followed by Random-Forest ABC scenario-choice [53], and Neural Network ABC posterior parameter inferences [52], 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 T1**. 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 [37,55]. 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.

FIGURE AND TABLE LEGENDS

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**.

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 evaluate 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 ρ=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.

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 populations (surrogates) 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, $p<8\text{x}10^{-12}$, and 0.98861, $p<2\text{x}10^{-8}$, 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; $p<2.2\text{x}10^{-16}$ for Cabo Verde; ρ =0.99568; $p<2.2\text{x}10^{-16}$ for ASW; ρ =0.99366; $p<2.2\text{x}10^{-16}$ for ACB, **Supplementary Text**).

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.

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; $p<2.2\text{x}10^{-16}$). **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; $p<2.2\text{x}10^{-16}$). **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** (Wang et al. 2010).

Figure F6:

Four competing scenarios for the genetic admixture history 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.

After the initial founding pulse of admixture, *Scenario 1 Afr2Pulses-Eur2Pulses*: the admixed population receives two separate introgression pulses from the African and European sources, respectively; *Scenario 2 Afr2Pulses-EurRecurring*: 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*: 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*: 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 **Materials and Methods** for detailed descriptions of model-parameters.

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. 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 dashed-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 dashed 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.

Table T1:

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**.

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

Table T2:

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**.

				Utterance-Frequency Euclidean Distance - 4831 uttered items		Genetic ASD - 1,899,978 SNPs	
Mantel variable	**Partial-Mantel control**	**n**	**Geographic scale**	**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*

REFERENCES

1. Eltis, D. & Richardson, D. *Atlas of the transatlantic slave trade*. (2015).
2. Fortes-Lima, C. & Verdu, P. Anthropological genetics perspectives on the transatlantic slave trade. *Human Molecular Genetics* **30**, R79–R87 (2021).
3. Micheletti, S. J. *et al.* Genetic Consequences of the Transatlantic Slave Trade in the Americas. *The American Journal of Human Genetics* **107**, 265–277 (2020).
4. Ongaro, L. *et al.* The Genomic Impact of European Colonization of the Americas. *Current Biology* **29**, 3974-3986.e4 (2019).
5. Winkler, C. A., Nelson, G. W. & Smith, M. W. Admixture Mapping Comes of Age. *Annu. Rev. Genom. Hum. Genet.* **11**, 65–89 (2010).
6. Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019).
7. Patin, E. *et al.* Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. *Science* **356**, 543–546 (2017).
8. Gravel, S. Population Genetics Models of Local Ancestry. *Genetics* **191**, 607–619 (2012).
9. Hellenthal, G. *et al.* A Genetic Atlas of Human Admixture History. *Science* **343**, 747–751 (2014).
10. Mathias, R. A. *et al.* A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. *Nat Commun* **7**, 12522 (2016).
11. Moreno-Estrada, A. *et al.* Reconstructing the Population Genetic History of the Caribbean. *PLoS Genet* **9**, e1003925 (2013).
12. Baharian, S. *et al.* The Great Migration and African-American Genomic Diversity. *PLoS Genet* **12**, e1006059 (2016).
13. Holm, J. A. *An introduction to pidgin and creoles*. (Cambridge University Press, 2000).
14. Escure, G. & Schwegler, A. *Creoles, contact, and language change: linguistics and social implications*. (John Benjamins Pub, 2004).
15. Quint, N. *Le cap-verdien: origines et devenir d'une langue métisse*. (L'Harmattan, 2000).
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* **129**, 1–8 (2013).
17. Baptista, M. Continuum and variation in Creoles: Out of many voices, one language. *JPCL* **30**, 225–264 (2015).
18. Eltis, D. *Coerced and free migration: global perspectives*. (Stanford University Press, 2002).
19. Berlin, I. *Many Thousands Gone: the First Two Centuries of Slavery in North America*. (Harvard University Press, 2009).
20. Ansari-Pour, N. *et al.* Palenque de San Basilio in Colombia: genetic data support an oral history of a paternal ancestry in Congo. *Proc. R. Soc. B.* **283**, 20152980 (2016).
21. Verdu, P., Jewett, E. M., Pemberton, T. J., Rosenberg, N. A. & Baptista, M. Parallel Trajectories of Genetic and Linguistic Admixture in a Genetically Admixed Creole Population. *Current Biology* **27**, 2529-2535.e3 (2017).

22. Hagemeijer, T. & Rocha, J. Creole languages and genes: the case of São Tomé and Príncipe. *Faits Lang.* **49**, 167–182 (2019).
23. Carreira, A. *Cabo Verde: formação e extinção de uma sociedade escravocrata (1460-1878)*. (IPC, 2000).
24. Albuquerque, L. de, Madeira Santos, M. E. *História geral de Cabo Verde T1*. vol. 1 (Centro de Estudos de História e Cartografia Antiga, Instituto de Investigação Científica Tropical□; Instituto Nacional da Cultura de Cabo Verde, 1991).
25. Albuquerque, L. de, Madeira Santos, M. E. *História geral de Cabo Verde T2*. vol. 2 (Centro de Estudos de História e Cartografia Antiga, Instituto de Investigação Científica Tropical□; Instituto Nacional da Cultura de Cabo Verde, 1995).
26. Albuquerque, L. de, Madeira Santos, M. E. *História concisa de Cabo Verde (Resumo da História geral de Cabo Verde T1-3)*. (Centro de Estudos de História e Cartografia Antiga, Instituto de Investigação Científica Tropical□; Instituto Nacional da Cultura de Cabo Verde, 2007).
27. Baptista, M. *The syntax of Cape Verdean Creole: the Sotavento varieties*. (John Benjamins Pub, 2002).
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*. (Narr, 2009).
29. Brehm, A., Pereira, L., Bandelt, H.-J., Prata, M. J. & Amorim, A. Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. *Ann. Hum. Genet.* **66**, 49–60 (2002).
30. Gonçalves, R. *et al.* Y-chromosome lineages in Cabo Verde Islands witness the diverse geographic origin of its first male settlers. *Human Genetics* **113**, 467–472 (2003).
31. Beleza, S. *et al.* The admixture structure and genetic variation of the archipelago of Cape Verde and its implications for admixture mapping studies. *PLoS ONE* **7**, e51103 (2012).
32. Beleza, S. *et al.* Genetic architecture of skin and eye color in an African-European admixed population. *PLoS Genet* **9**, e1003372 (2013).
33. Hamid, I., Korunes, K. L., Beleza, S. & Goldberg, A. Rapid adaptation to malaria facilitated by admixture in the human population of Cabo Verde. *eLife* **10**, e63177 (2021).
34. Soares, M. J. The British Presence on the Cape Verdean Archipelago (Sixteenth to Eighteenth centuries). *African Economic History* **39**, 129–146 (2011).
35. Berlin, I. *The making of African America: the four great migrations*. (Penguin Books, 2010).
36. Gouveia, M. H. *et al.* Origins, Admixture Dynamics, and Homogenization of the African Gene Pool in the Americas. *Molecular Biology and Evolution* **37**, 1647–1656 (2020).
37. Fortes□Lima, C. A., Laurent, R., Thouzeau, V., Toupance, B. & Verdu, P. Complex genetic admixture histories reconstructed with Approximate Bayesian Computation. *Mol Ecol Resour* **21**, 1098–1117 (2021).
38. Bowcock, A. M. *et al.* High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455–457 (1994).
39. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655–1664 (2009).

40. Morrison, M. L., Alcala, N. & Rosenberg, N. A. *FSTruct: an FST-based tool for measuring ancestry variation in inference of population structure*. http://biorxiv.org/lookup/doi/10.1101/2021.09.24.461741 (2021) doi:10.1101/2021.09.24.461741.
41. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
42. Rosenberg, N. A. Genetic Structure of Human Populations. *Science* **298**, 2381–2385 (2002).
43. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).
44. Lawson, D. J., van Dorp, L. & Falush, D. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nat Commun* **9**, 3258 (2018).
45. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of Population Structure using Dense Haplotype Data. *PLoS Genet* **8**, e1002453 (2012).
46. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179–181 (2012).
47. Chacón-Duque, J.-C. *et al.* Latin Americans show wide-spread Converso ancestry and imprint of local Native ancestry on physical appearance. *Nat Commun* **9**, 5388 (2018).
48. Thompson, E. A. Identity by Descent: Variation in Meiosis, Across Genomes, and in Populations. *Genetics* **194**, 301–326 (2013).
49. Mooney, J. A. *et al.* Understanding the Hidden Complexity of Latin American Population Isolates. *The American Journal of Human Genetics* **103**, 707–726 (2018).
50. Szpiech, Z. A. *et al.* Ancestry-Dependent Enrichment of Deleterious Homozygotes in Runs of Homozygosity. *The American Journal of Human Genetics* **105**, 747–762 (2019).
51. Rousset, F. Genetic Differentiation and Estimation of Gene Flow from *F* -Statistics Under Isolation by Distance. *Genetics* **145**, 1219–1228 (1997).
52. Csilléry, K., François, O. & Blum, M. G. B. abc: an R package for approximate Bayesian computation (ABC): *R package: abc*. *Methods in Ecology and Evolution* **3**, 475–479 (2012).
53. Pudlo, P. *et al.* Reliable ABC model choice via random forests. *Bioinformatics* **32**, 859–866 (2016).
54. Robert, C. P., Mengersen, K. & Chen, C. Model choice versus model criticism. *Proceedings of the National Academy of Sciences* **107**, E5–E5 (2010).
55. Verdu, P. & Rosenberg, N. A. A General Mechanistic Model for Admixture Histories of Hybrid Populations. *Genetics* **189**, 1413–1426 (2011).
56. Patterson, K. D. Epidemics, Famines, and Population in the Cape Verde Islands, 1580-1900. *The International Journal of African Historical Studies* **21**, 291 (1988).
57. Brooks, G. E. Cabo Verde: Gulag of the South Atlantic: Racism, Fishing Prohibitions, and Famines. *Hist. Afr.* **33**, 101–135 (2006).
58. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *The American Journal of Human Genetics* **100**, 635–649 (2017).
59. Korunes, K. L. *et al. Sex-biased admixture and assortative mating shape genetic variation and influence demographic inference in admixed Cabo Verdeans*.

http://biorxiv.org/lookup/doi/10.1101/2020.12.14.422766 (2020) doi:10.1101/2020.12.14.422766.

60. Zaitlen, N. *et al.* The Effects of Migration and Assortative Mating on Admixture Linkage Disequilibrium. *Genetics* **205**, 375–383 (2017).
61. Versluys, T. M. M., Mas-Sandoval, A., Flintham, E. O. & Savolainen, V. Why do we pick similar mates, or do we? *Biol. Lett.* **17**, 20210463 (2021).
62. Chaudenson, R. & Mufwene, S. S. *Creolization of language and culture*. (Routledge, 2001).
63. Cooper, F., Holt, T. C. & Scott, R. J. *Beyond slavery: explorations of race, labor, and citizenship in postemancipation societies*. (University of North Carolina Press, 2000).
64. Fortes-Lima, C. *et al.* Exploring Cuba's population structure and demographic history using genome-wide data. *Sci Rep* **8**, 11422 (2018).
65. Goldberg, A., Verdu, P. & Rosenberg, N. A. Autosomal Admixture Levels Are Informative About Sex Bias in Admixed Populations. *Genetics* **198**, 1209–1229 (2014).
66. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
67. Gurdasani, D. *et al.* The African Genome Variation Project shapes medical genetics in Africa. *Nature* **517**, 327–332 (2015).
68. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
69. Chafe, W. L. *The Pear stories: cognitive, cultural, and linguistic aspects of narrative production*. (Ablex Pub. Corp, 1980).
70. Croft, W. Linguistic Selection: An Utterance-based Evolutionary Theory of Language Change. *Nordic Journal of Linguistic* **19**, 99–139 (1996).
71. R Development Core Team. *R: A language and environment for statistical computing.* (R Foundation for Statistical Computing, 2020).
72. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).
73. Jakobsson, M. & Rosenberg, N. A. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801–1806 (2007).
74. Rosenberg, N. A. DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137–138 (2003).
75. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
76. Szpiech, Z. A., Blant, A. & Pemberton, T. J. GARLIC: Genomic Autozygosity Regions Likelihood-based Inference and Classification. *Bioinformatics* **33**, 2059–2062 (2017).
77. Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. *The American Journal of Human Genetics* **93**, 278–288 (2013).
78. Snyder, J. P. *Map projections--a working manual*. (U.S. Government Printing Office 1987).
79. Tavaré, S., Balding, D. J., Griffiths, R. C. & Donnelly, P. Inferring coalescence times from

DNA sequence data. *Genetics* **145**, 505–518 (1997).

80. Beaumont, M. A., Zhang, W. & Balding, D. J. Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025–2035 (2002).

SUPPLEMENTARY FIGURES AND TABLES LEGENDS

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.

Figure SF4:

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

We removed all East African and Central African hunter-gatherer samples (Baka, Bezan, Ba.Bongo, Ba.Koya, Ba.Twa, Bi.Aka, Mbuti) compared to the sample set employed in **Figure SF3**, 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 ($p<2.2\text{x}10^{-16}$) Spearman ρ of 0.9450.

Figure SF5:
Multidimensional Scaling three-dimensional projection of Allele Sharing Dissimilarity pairwise dissimilarities, for a subset of African, European and and other admixed populations related to the TAST.
We removed all West and South-West Central African samples, as well as South African samples, compared to the sample set employed in **Figure SF4**, 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 ($p<2.2\text{x}10^{-16}$) Spearman ρ of 0.9636.

Figures SF6A-D:
Four separate animated Multidimensional Scaling three-dimensional projections of Allele Sharing Dissimilarity pairwise dissimilarities, for a subset of African, European and other admixed populations related to the TAST.
Each four animated plot is provided in *.gif* format. Each individual is represented by a single point. See **Table ST1** for the population list used in these analyses, they correspond to each separate sample set provided in two-dimensional plots in **Figure F2**. Spearman correlations are provided in **Figure F2** caption.

Figure SF7:
Population F_{ST}/F_{ST}^{max} values for the ASW, ACB, and each Cabo Verdean birth-island separately considering the ADMIXTURE mode result at *K=2*.
F_{ST}/F_{ST}^{max} values were computed using FSTruct (Morrisson et al. 2021) with 1000 bootstrap replicates per population. All population pairwise distributions of bootstrap F_{ST}/F_{ST}^{max} values were significantly different from one another after Bonferroni correction (Wilcoxon two-sided rank sum test $p<3.96\text{x}10^{-8}$), except the following pairwise comparisons: ACB-Santiago, Brava-Maio, Fogo-Sal, and Maio-Santo Antão.

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 (Alexander et al. 2009) 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 per-

formed 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 (Jakobsson and Rosenberg 2009), and plotted with DISTRUCT (Rosenberg et al. 2004). Number of runs in the presented modes are indicated below the value of *K*. All other mode results are presented in **Figure F3**.

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

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

Figure SF11:
The distribution of total ancestry in long ROH per individual for each Cabo Verdean birth-island.

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

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. 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* (Pudlo et al. 2016).

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.

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.

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

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).

Table ST5-Santiago:
NN-ABC posterior parameter estimations for Santiago 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 6 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**.

Table ST6-Fogo:

NN-ABC posterior parameter estimations for Fogo 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 10 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**.

Table ST7-Santo Antão:

NN-ABC posterior parameter estimations for Santo Antão 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 11 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**.

Table ST8-São Nicolau:

NN-ABC posterior parameter estimations for São Nicolau 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 11 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**.

Table ST9-Brava:

NN-ABC posterior parameter estimations for Brava 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 10 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**.

Table ST10-Maio:
NN-ABC posterior parameter estimations for Maio 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 10 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**.

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**.

Table ST12-São Vicente:
NN-ABC posterior parameter estimations for São Vicente 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 11 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**.

Table ST13-Sal:

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

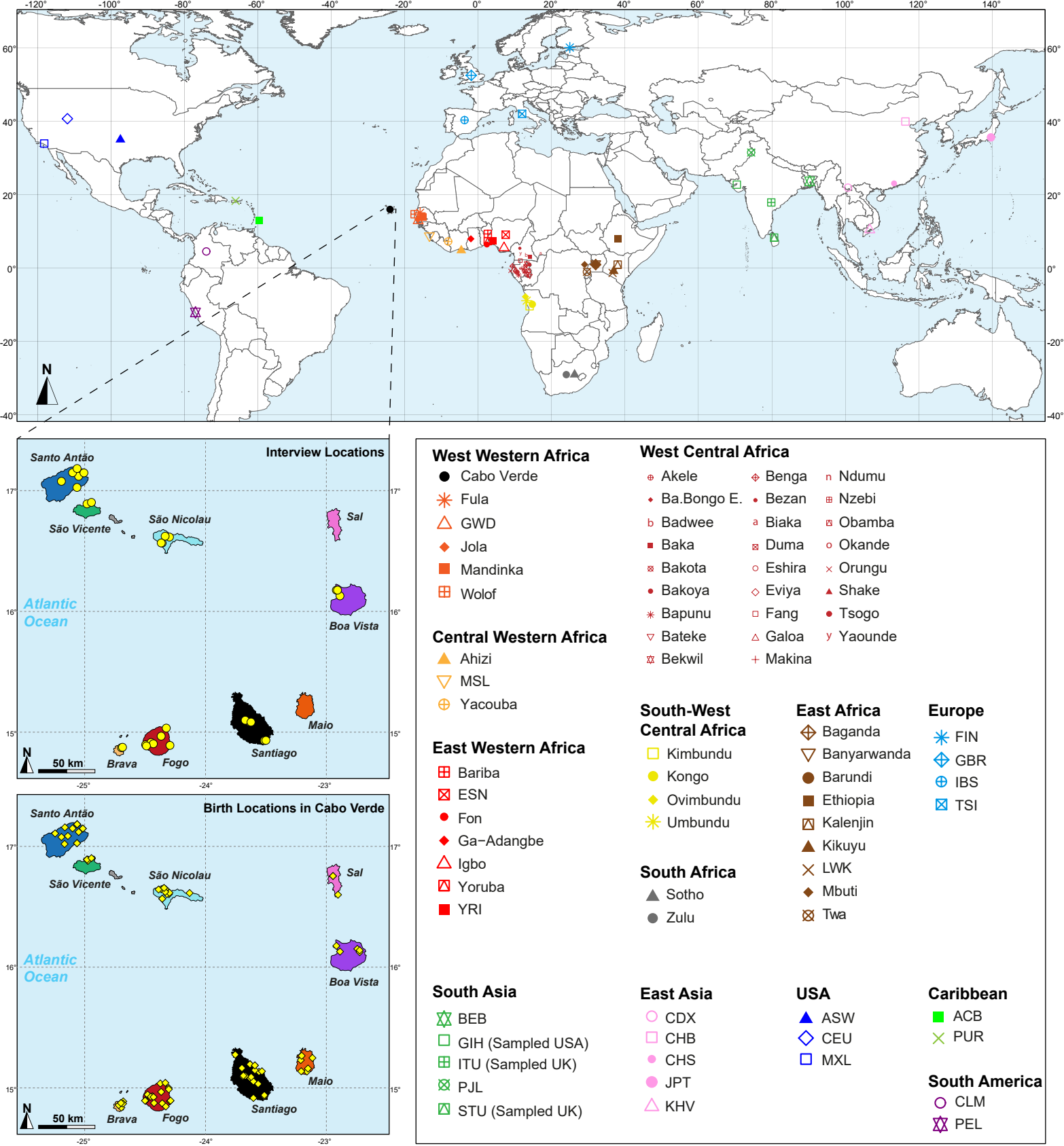
Cross-validation errors and 95% CI accuracies are based on 1,000 NN-ABC posterior parameter inferences using, in turn, each one of the 1,000 simulations closest to the observed data used as pseudo-observed data, and the remaining 99,999 simulations as the reference table.

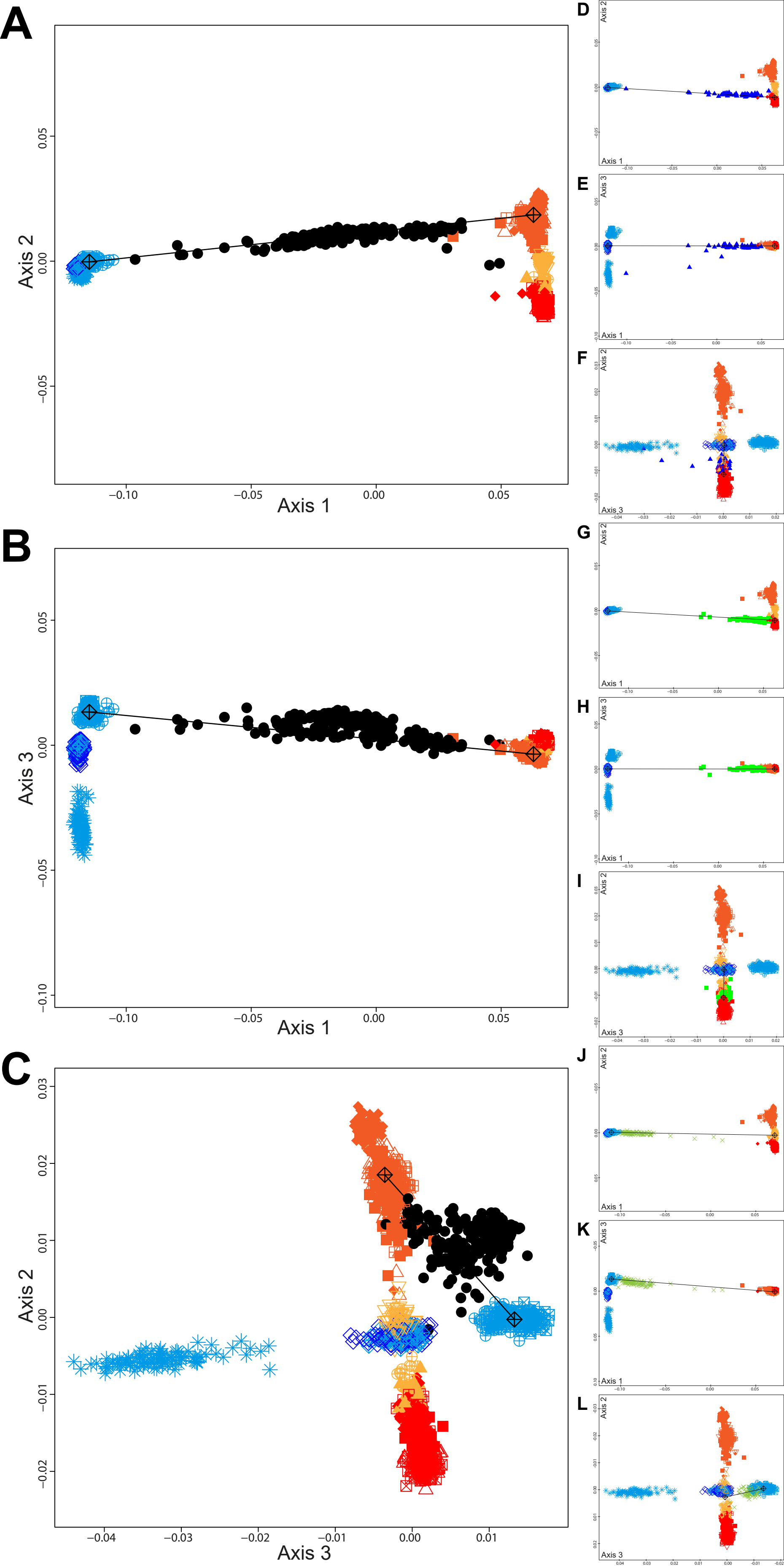
Error calculations are based on the mode and, separately, the median point estimate for each 1,000 pseudo-observed simulations posterior parameter estimation, compared to the known parameter used for simulation. We considered 11 neurons in the hidden layer and a tolerance level of 0.01 (1000 simulations) 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**.

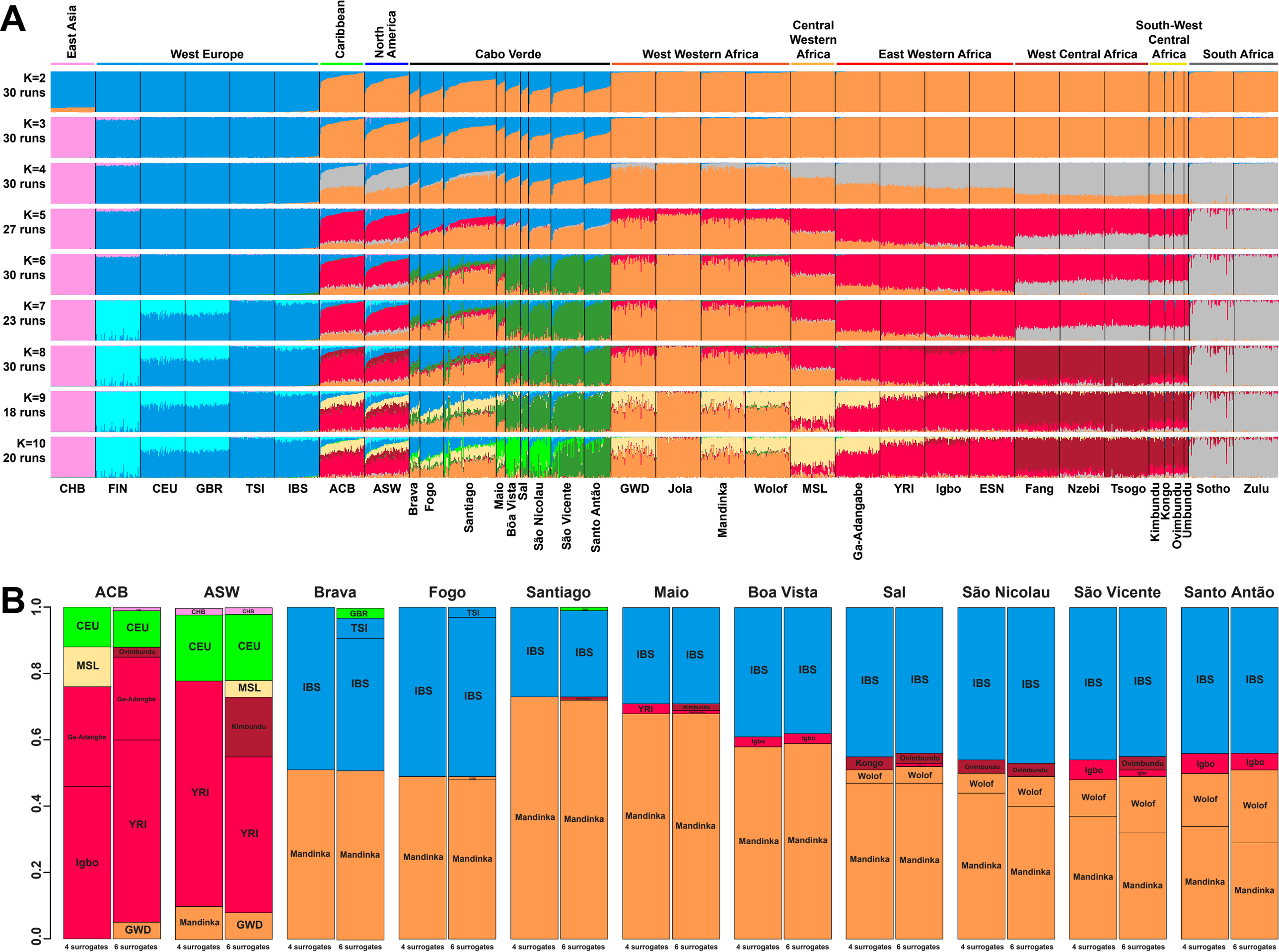
Table ST14:

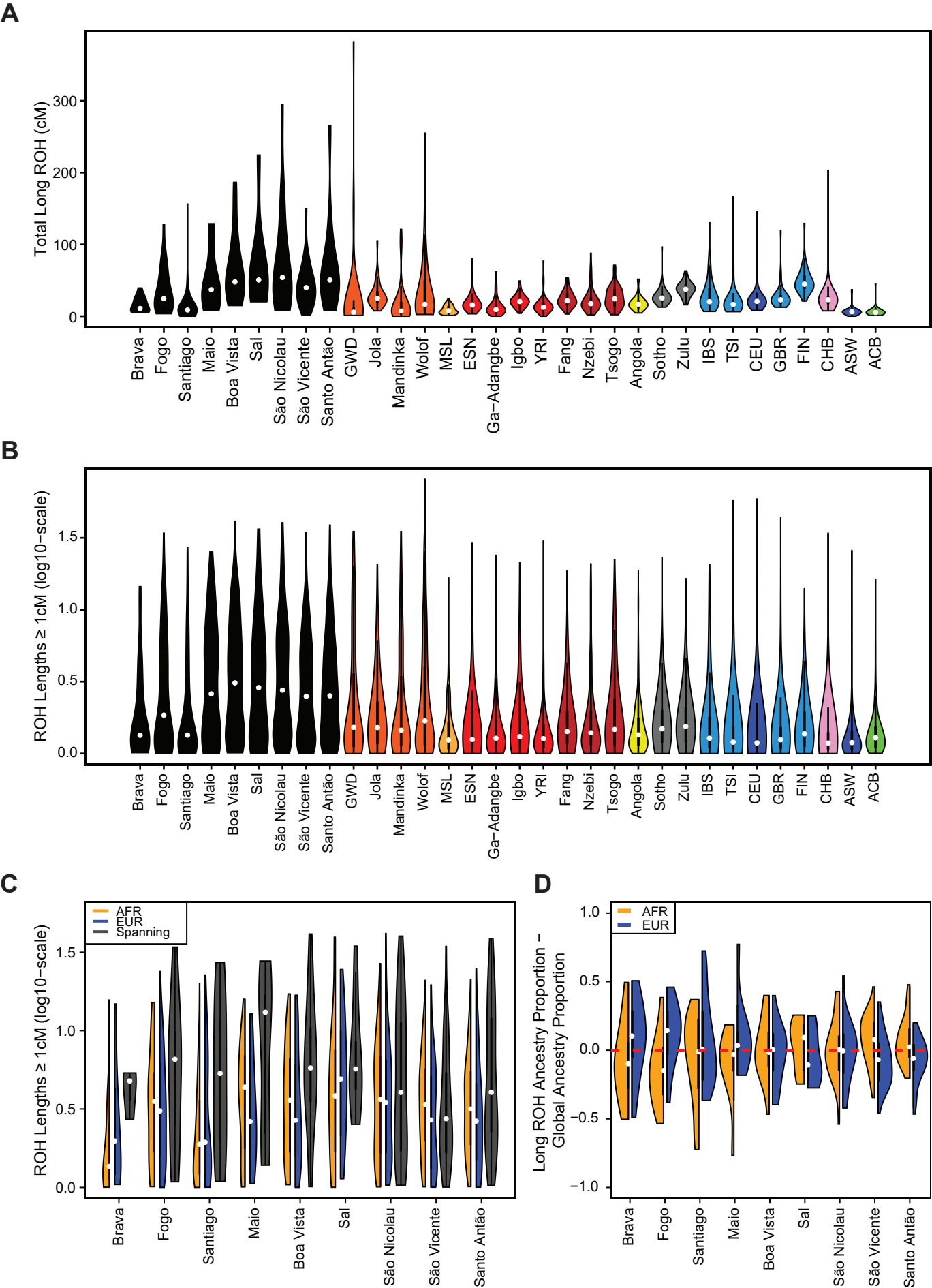
Mantel correlations among individual birth-places, residence-places, maternal and paternal birth places, age, and academic education duration.

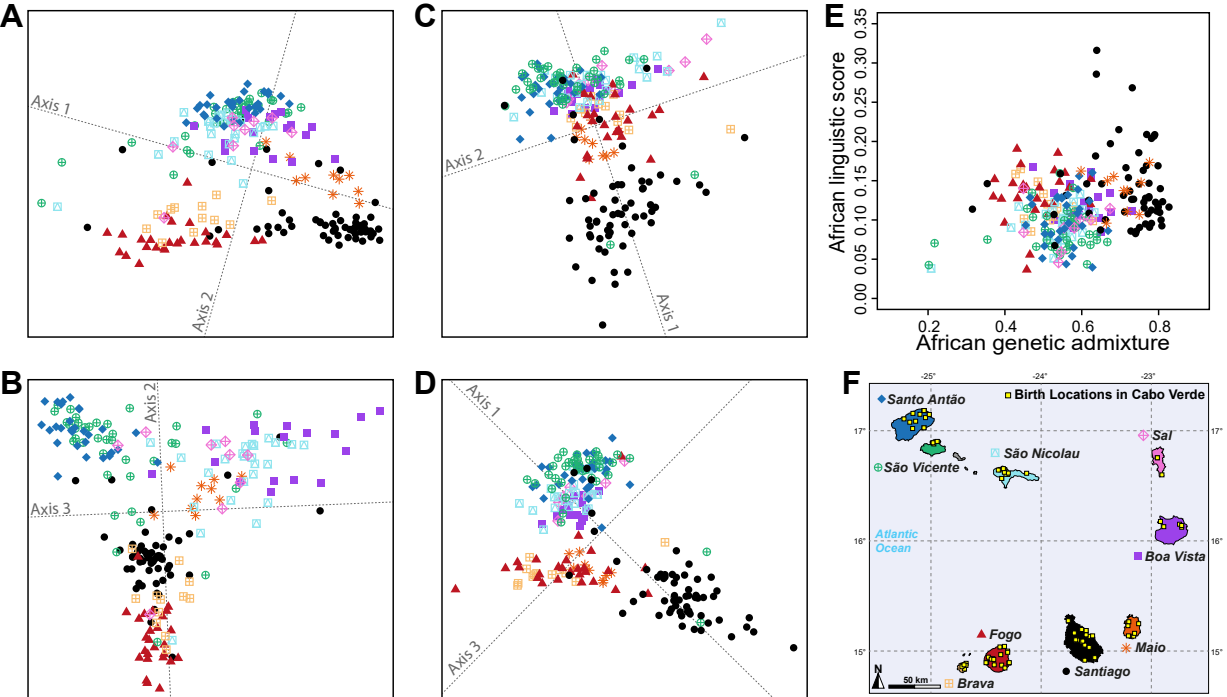
We considered only the 225 genetically unrelated Cabo Verde-born individuals in these analyses.

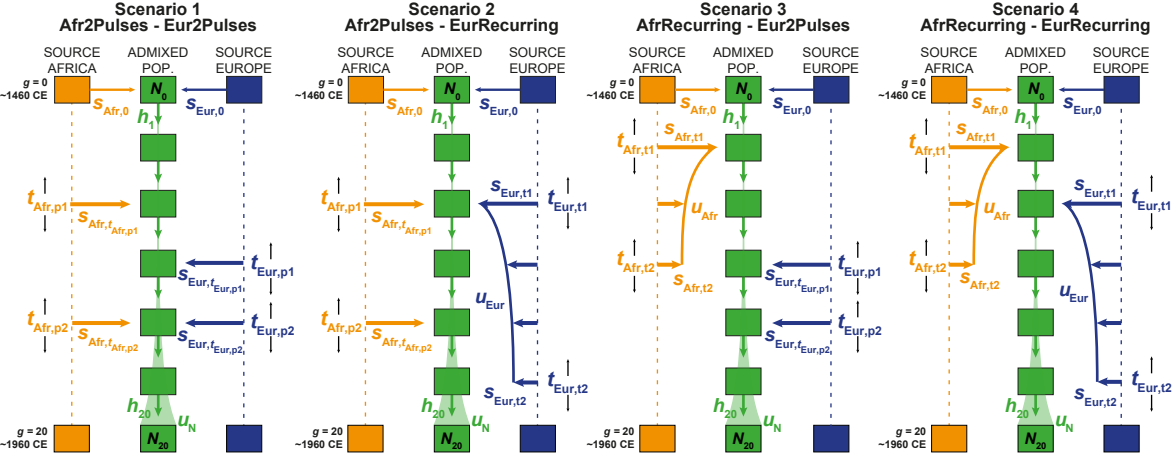

Interview Locations
Birth Locations in Cabo Verde
Santo Antão
São Vicente
São Nicolau
Sal
Boa Vista
Atlantic Ocean
Maio
Santiago
Brava
Fogo
50 km
N
West Western Africa
Cabo Verde
Fula
GWD
Jola
Mandinka
Wolof
Central Western Africa
Ahizi
MSL
Yacouba
East Western Africa
Bariba
ESN
Fon
Ga-Adangbe
Igbo
Yoruba
YRI
West Central Africa
Akele
Ba.Bongo E.
Badwee
Baka
Bakota
Bakoya
Bapunu
Bateke
Bekwil
Benga
Bezan
Biaka
Duma
Eshira
Eviya
Fang
Galoa
Makina
Ndumu
Nzebi
Obamba
Okande
Orungu
Shake
Tsogo
Yaounde
South-West Central Africa
Kimbundu
Kongo
Ovimbundu
Umbundu
South Africa
Sotho
Zulu
East Africa
Baganda
Banyarwanda
Barundi
Ethiopia
Kalenjin
Kikuyu
LWK
Mbuti
Twa
Europe
FIN
GBR
IBS
TSI
South Asia
BEB
GIH (Sampled USA)
ITU (Sampled UK)
PJL
STU (Sampled UK)
East Asia
CDX
CHB
CHS
JPT
KHV
USA
ASW
CEU
MXL
Caribbean
ACB
PUR
South America
CLM
PEL


A
Axis 2
Axis 1
B
Axis 3
Axis 1
C
Axis 2
Axis 3
D
E
F
G
H
I
J
K
L

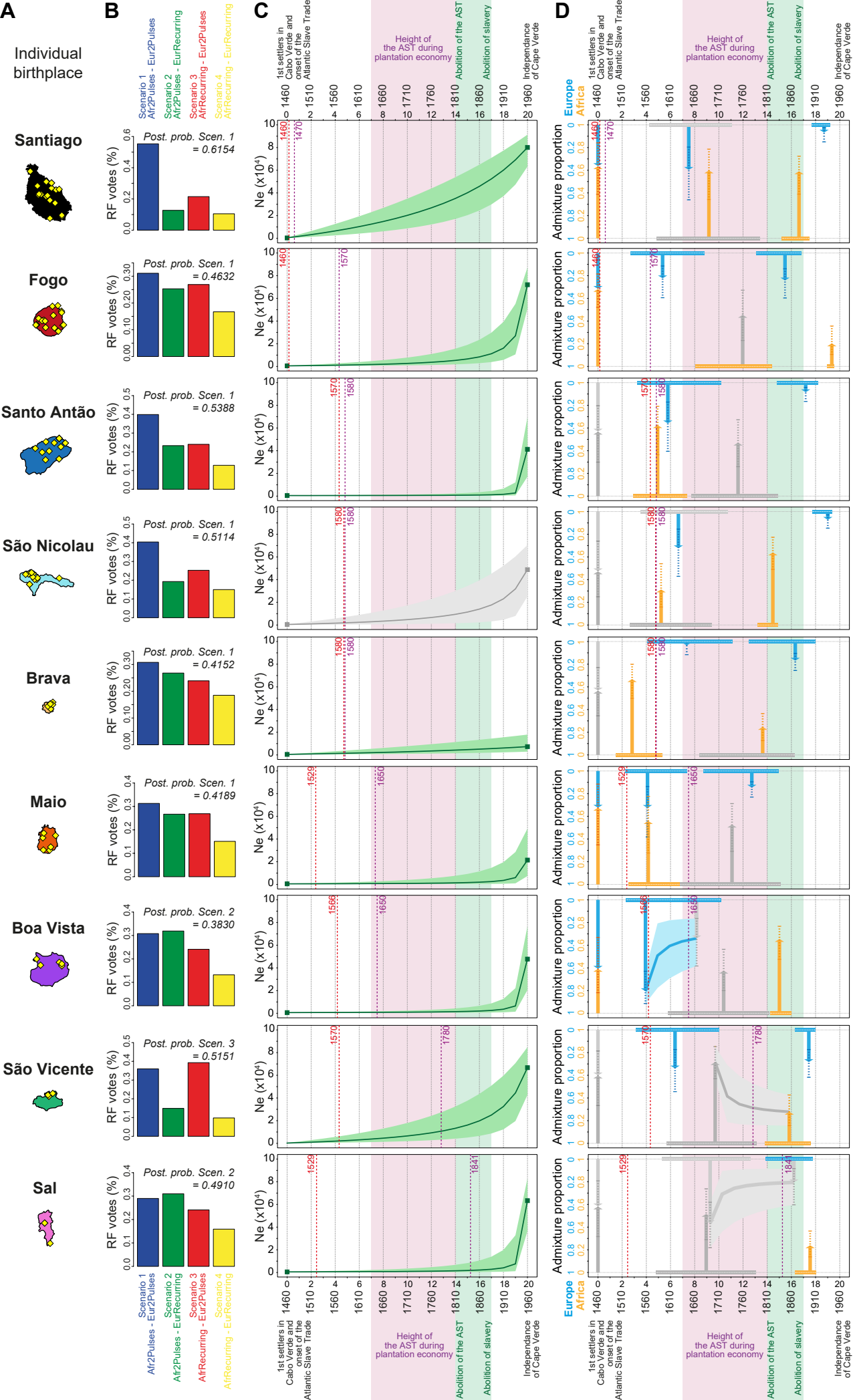

A
East Asia
West Europe
Caribbean
North America
Cabo Verde
West Western Africa
Central Western Africa
East Western Africa
West Central Africa
South-West Central Africa
South Africa
K=2 30 runs
K=3 30 runs
K=4 30 runs
K=5 27 runs
K=6 30 runs
K=7 23 runs
K=8 30 runs
K=9 18 runs
K=10 20 runs
CHB FIN CEU GBR TSI IBS ACB ASW Brava Fogo Santiago Maio Bõa Vista Sal São Nicolau São Vicente Santo Antão GWD Jola Mandinka Wolof MSL Ga-Adangbe YRI Igbo ESN Fang Nzebi Tsogo Kimbundu Kongo Ovimbundu Umbundu Sotho Zulu
B
ACB
ASW
Brava
Fogo
Santiago
Maio
Boa Vista
Sal
São Nicolau
São Vicente
Santo Antão
1.0
0.8
0.6
0.4
0.2
0.0
CEU
MSL
Ga-Adangbe
Igbo
Ovimbundu
YRI
GWD
CHB
Kimbundu
GBR
TSI
IBS
Mandinka
Kongo
Wolof
4 surrogates
6 surrogates


A
Total Long ROH (cM)
0
100
200
300
Brava
Fogo
Santiago
Maio
Boa Vista
Sal
São Nicolau
São Vicente
Santo Antão
GWD
Jola
Mandinka
Wolof
MSL
ESN
Ga−Adangbe
Igbo
YRI
Fang
Nzebi
Tsogo
Angola
Sotho
Zulu
IBS
TSI
CEU
GBR
FIN
CHB
ASW
ACB
B
ROH Lengths ≥ 1cM (log10−scale)
0.0
0.5
1.0
1.5
C
AFR
EUR
Spanning
D
AFR
EUR
Long ROH Ancestry Proportion − Global Ancestry Proportion
−1.0
−0.5


A
Axis 1
Axis 2
B
Axis 2
Axis 3
C
Axis 2
Axis 1
D
Axis 1
Axis 3
E
African linguistic score
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.2
0.4
0.6
0.8
African genetic admixture
F
Birth Locations in Cabo Verde
Santo Antão
São Vicente
São Nicolau
Sal
Atlantic Ocean
Boa Vista
Fogo
Maio
Santiago
Brava
50 km
N
-25°
-24°
-23°
17°
16°
15°



Scenario 1
Afr2Pulses - Eur2Pulses
Scenario 2
Afr2Pulses - EurRecurring
Scenario 3
AfrRecurring - Eur2Pulses
Scenario 4
AfrRecurring - EurRecurring
SOURCE AFRICA
ADMIXED POP.
SOURCE EUROPE
g = 0
~1460 CE
g = 20
~1960 CE
N0
N20
sAfr,0
sEur,0
h1
tAfr,p1
sAfr,tAfr,p1
sEur,tEur,p1
tEur,p1
tAfr,p2
sAfr,tAfr,p2
sEur,tEur,p2
tEur,p2
h20
uN
sEur,t1
tEur,t1
uEur
sEur,t2
tEur,t2
sAfr,t1
tAfr,t1
uAfr
tAfr,t2
sAfr,t2


A
B
C
D
Individual birthplace
Scenario 1 Afr2Pulses - Eur2Pulses
Scenario 2 Afr2Pulses - EurRecurring
Scenario 3 AfrRecurring - Eur2Pulses
Scenario 4 AfrRecurring - EurRecurring
1st settlers in Cabo Verde and onset of the Atlantic Slave Trade
Height of the AST during plantation economy
Abolition of the AST
Abolition of slavery
Independance of Cape Verde
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
Europe
Africa