



**HAL**  
open science

## Nested admixture during and after the Trans-Atlantic Slave Trade on the island of São Tomé

Marta Ciccarella, Romain Laurent, Zachary Szpiech, Etienne Patin, Françoise Dessarps-Freichey, José Utgé, Laure Lémée, Armando Semo, Jorge Rocha, Paul Verdu

► **To cite this version:**

Marta Ciccarella, Romain Laurent, Zachary Szpiech, Etienne Patin, Françoise Dessarps-Freichey, et al.. Nested admixture during and after the Trans-Atlantic Slave Trade on the island of São Tomé. 2025. hal-04797739

**HAL Id: hal-04797739**

**<https://hal.science/hal-04797739v1>**

Preprint submitted on 14 Feb 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

1 Nested admixture during and after the Trans-Atlantic Slave Trade on the island of São  
2 Tomé.

3

4 Marta Ciccarella <sup>1,2,3,\*</sup>, Romain Laurent <sup>1</sup>, Zachary A. Szpiech <sup>4,5</sup>, Etienne Patin <sup>6</sup>, Françoise  
5 Dessarps-Freichy <sup>1</sup>, José Utgé <sup>1</sup>, Laure Lémée <sup>7</sup>, Armando Semo <sup>2,3</sup>, Jorge Rocha <sup>2,3,8,\*¶</sup>, and  
6 Paul Verdu <sup>1,\*¶</sup>

7

8 <sup>1</sup> UMR7206 Eco-anthropologie, CNRS, MNHN, Université Paris Cité, France;

9 <sup>2</sup> CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, *InBIO* Laboratório  
10 Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal

11 <sup>3</sup> BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de  
12 Vairão, 4485-661 Vairão, Portugal

13 <sup>4</sup> Department of Biology, Penn State University, United States;

14 <sup>5</sup> Institute for Computational and Data Sciences, Penn State University, United States;

15 <sup>6</sup> Human Evolutionary Genetics Unit, Institut Pasteur, Université Paris Cité, CNRS UMR2000,  
16 Paris, France;

17 <sup>7</sup> Plateforme Technologique Biomics, C2RT, Institut Pasteur, France.

18 <sup>8</sup> Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, 4099-002 Porto,  
19 Portugal

20

21 \* Authors for correspondence:

22 [marta.ciccarella@mnhn.fr](mailto:marta.ciccarella@mnhn.fr), [jrocha@cibio.up.pt](mailto:jrocha@cibio.up.pt), [paul.verdu@mnhn.fr](mailto:paul.verdu@mnhn.fr)

23 ¶ Equally supervised this work

- 24 MC – designed the study, conducted raw data QC and merging, conducted statistical and  
25 population genetics analyses, analyzed results, wrote the first draft of the article
- 26 RL – conducted raw data QC and merging
- 27 ZAS – helped design statistical and population genetics analyses
- 28 EP – contributed raw data
- 29 FDF – conducted molecular genetics raw data generation
- 30 JU – conducted molecular genetics raw data generation
- 31 LL – conducted molecular genetics raw data generation
- 32 AS – contributed raw data
- 33 JR – designed the study, contributed samples, analyzed results, participated to write the first  
34 draft of the article
- 35 PV – designed the study, contributed samples, conducted statistical and population genetics  
36 analyses, analyzed results, participated to write the first draft of the article
- 37 All authors – participated to write the article
- 38 The authors declare no competing interests.
- 39

## 40 Abstract

41 Human admixture history is rarely a simple process in which distinct populations, previously  
42 isolated for a long time, come into contact once to form an admixed population. In this study,  
43 we aim to reconstruct the complex admixture histories of the population of São Tomé, an  
44 island in the Gulf of Guinea that was the site of the first slave-based plantation economy, and  
45 experienced successive waves of forced and deliberate migration from Africa. We examined  
46 2.5 million SNPs newly genotyped in 96 São Toméans and found that geography alone cannot  
47 explain the observed patterns of genetic differentiation within the island. We defined five  
48 genetic groups in São Tomé based on the hypothesis that individuals sharing the most  
49 haplotypes are more likely to share similar genetic histories. Using Identical-by-Descent and  
50 different local ancestry inference methods, we inferred shared ancestries between 70 African  
51 and European populations and each São Toméan genetic group. We identified admixture  
52 events between admixed groups that were previously isolated on the island, showing how  
53 recently admixed populations can be themselves the sources of other admixture events. This  
54 study demonstrates how complex admixture and isolation histories during and after the  
55 Transatlantic Slave-Trade shaped extant individual genetic patterns at a local scale in Africa.

## 56 Keywords

57 Genetic admixture; Genetic structure; Demography; Trans-Atlantic Slave Trade

## 58 Introduction

59 The Trans-Atlantic Slave Trade (TAST) was one of the largest human migrations of historical  
60 times, involving the forced displacement of more than 12 million enslaved Africans according  
61 to historical sources [1]. Population genetic studies have extensively investigated the impact  
62 of these forced migrations on the genetic diversity of admixed populations on either sides of  
63 the Atlantic [2–6]. These studies provided important contributions for the understanding of



64 genetic admixture dynamics in the context of the TAST and European colonization, as well as  
65 for the general understanding of human genetic admixture processes [7–9].

66 Admixture histories are rarely simple, with distinct populations coming into contact only once  
67 over time. Numerous admixed populations descended from enslaved Africans experienced  
68 several periods of recurring introgressions from populations originating from different regions  
69 of Africa [3,8,10]. Moreover, often in the colonial context, the enslaved and non-enslaved  
70 communities were reproductively isolated and subject to strict demographic constraints. Over  
71 time, changes in the economic viability of plantations, shifts in colonial policies, and the  
72 abolition of slavery contributed to evolving socio-cultural contexts, resulting in new forms of  
73 community integration and segregation [11]. The aim of this study is to account for successive  
74 admixture and isolation histories on the island of São Tomé, in light of the documented history  
75 of settlement and social segregation during five centuries of Portuguese colonization.

76 The archipelago of São Tomé e Príncipe, in the Gulf of Guinea, was uninhabited at the time  
77 Portuguese sailors arrived on the island of São Tomé around 1470 and rapidly settled the  
78 island from 1493 onwards [12]. On this island, the colonizers established for the first time an  
79 economic system based on plantations and the exploitation of enslaved labor for the massive  
80 production of sugar cane, known as the Plantation Economic System, that was later deployed  
81 throughout European colonies in the Caribbeans and the Americas [11]. Furthermore, the  
82 archipelago served as an important hub for the Trans-Atlantic Slave Trade between the 16th  
83 century and the abolition of the TAST and slavery during the 19<sup>th</sup> century [13]. Historical  
84 records show that enslaved Africans were initially traded from the Niger delta in the kingdom  
85 of Benin, in a region that is now Nigeria. However, as the local demand for enslaved labor-  
86 force increased during the 16<sup>th</sup> century, together with the growing trade with colonies from the  
87 other side of the Atlantic, São Toméan merchants expanded their slave recruitment areas  
88 southwards, to the kingdom of Kongo and other regions of northern Angola [14].

89 In the 19<sup>th</sup> centuries, São Tomé e Príncipe became an important producer of coffee and cacao.  
90 Following the abolition of slavery in 1875, the plantation-based economy relied this time on

91 the exploitation of indentured servitude. Contractual workers, known as “*serviçais*”, were  
92 mostly recruited from other Portuguese colonies, including Angola, Mozambique and Cabo  
93 Verde. Remarkably, over the course of the 20th century, approximately 80,000 individuals  
94 from the archipelago of Cabo Verde in West-Western Africa migrated to São Tomé [15]. As  
95 São Tomé e Príncipe, the Cabo Verde archipelago was also colonized by the Portuguese  
96 crown in the 1460’s and underwent more than 400 years of history linked with the TAST [16].  
97 Nevertheless, the regions of origins of enslaved Africans and the histories of colonial  
98 exploitation and plantation economy differed significantly between the two archipelagos [17].

99 The island of São Tomé provides unique opportunities to disentangle the impact of social  
100 segregation on genetic diversity in the colonial context of the TAST, throughout the expansion  
101 of the Plantation Economy system, and after the abolition of the TAST and that of slavery.  
102 Extant communities within São Tomé are associated with different histories related to  
103 marooning, slavery, and post-slavery migrations, which are also reflected in the linguistic  
104 landscape on the island. Since the beginning of Portuguese colonization, the São Tomé  
105 forested interior of the island served as a refuge for runaway slaves, who remained largely  
106 isolated in maroon communities until the 19<sup>th</sup> century, when one of these communities became  
107 known as the *Angolares* [18]. Today, *Angolares* communities in the north-eastern and south-  
108 western coasts of the island are known to speak Angolar, one of the two autochthonous creole  
109 languages of the island [19,20]. In parallel, freed enslaved-Africans became the largest social  
110 group in São Tomé during the 17<sup>th</sup> and the 18<sup>th</sup> century, a period of profound economic decline  
111 and relative abandon of the colony by Portuguese settlers and by the Portuguese crown [21].  
112 Their descendants are often associated with the Forro creole language, “Forro” literally  
113 meaning “freed slave” [20]. Finally, descendants of Cabo Verdean *serviçais* may be  
114 considered to form a third historical community associated with yet another creole language  
115 spoken on the island: the Cabo Verdean Kriolu [22].

116 In this complex context of forced and deliberate migrations to São Tomé and segregation  
117 within the island, previous genetic studies have identified substantial genetic structure with a

118 limited number of loci [23,24] or samples [5]. Coelho et al. suggested that the identified genetic  
119 structure within São Tomé was mainly caused by a strong signal of genetic differentiation  
120 between *Angolares* and non-*Angolares* individuals, with increased genetic drift having  
121 occurred in the former group. More recently, Almeida et al. generated exome sequence data  
122 in 25 individuals from the islands of São Tomé and Príncipe, and found similar genetic  
123 contributions from both the Gulf of Guinea and Angola in individuals speaking the Angolar and  
124 Forro creoles in São Tomé. The nested genetic structures and the mosaic of African genetic  
125 diversity found by these previous studies in São Tomé thus raised a series of fundamental  
126 questions about the histories of admixture in the island. What are the origins of the genetic  
127 ancestors of São Toméans? Are the diverse genetic ancestries evenly distributed among  
128 present-day São Toméans? Within São Tomé, and spanning existing linguistic-  
129 anthropological communities, when and how did genetic admixture and isolation processes  
130 occur?

131 In this study, we investigated 96 unrelated individuals sampled in 13 sampling sites from São  
132 Tomé, each genotyped at 2.5 million SNPs genome-wide. We identified five genetic groups of  
133 individuals in the sample based on haplotype-sharing patterns. We inferred the shared-  
134 haplotypic ancestries for each five São Toméan genetic groups using previously published  
135 genome-wide data from 70 African and European populations, including an extensive sample  
136 from the archipelago of Cabo Verde [6,25–28]. We reconstructed recent shared ancestries  
137 between admixed groups using sharing patterns of long tracts Identical-by-Descent. We finally  
138 propose a chronology of successive admixture and isolation events that may explain the  
139 genetic diversity observed in São Tomé, using model-based ancestry inference approaches.  
140 This study highlights the influence of the complex history of European colonization, including  
141 TAST, social segregation, and post-slavery migrations, on extant genetic patterns in São  
142 Tomé.

## 143 **Results**

### 144 **São Tomé genetic diversity in the global context**

145 To investigate the genetic diversity across the island of São Tomé in the worldwide context,  
146 we combined our genotype data on 96 unrelated individuals from São Tomé (Fig 1) with a  
147 reference dataset of population samples from Africa, Europe, and the Americas, including  
148 other populations descended from enslaved Africans in Cabo Verde, African-Americans in  
149 southwest USA (ASW) and African-Caribbeans in Barbados (ACB) (S1 Table) [6,25–28].

150 Fig 2A shows the sampling locations of the 70 populations included in the reference dataset,  
151 which are grouped into 10 distinct geographical regions. We consider a subset of population  
152 samples from previous studies, focusing particularly on regions in Africa that were used as  
153 key ports of embarkation during the Trans-Atlantic Slave Trade [1] (S4 Table). Fig 2B shows  
154 the first two dimensions of a Multi-Dimensional Scaling analysis calculated on Allele-Sharing  
155 Dissimilarities (ASD) [29], between each pair of individuals in our data set including 411,121  
156 autosomal SNPs. The first axis of variation is explained by major genetic differentiation  
157 between African and European populations, while the second axis captures genetic  
158 differentiation within Africa, revealing a north-south gradient ranging from West-Western  
159 African to South African populations.

160 The 96 unrelated individuals born in São Tomé are highly dispersed across the two first  
161 dimensions of the ASD-MDS, compared to the 70 population samples from Africa, Europe,  
162 and the Americas (Fig 2B). Other populations descended from enslaved Africans show a  
163 relatively simpler pattern, in spite of being also highly admixed: Cabo Verdean individuals fall  
164 along a trajectory going from European to West-Western African populations, reflecting the  
165 impact of slave recruitment from the neighbor areas of Senegambia and Guinea-Bissau in  
166 their current genetic profile [6,30]; African-Americans (ASW) and African-Barbadians (ACB)

167 cluster along a trajectory going from Central-Western African to European populations, in  
168 agreement with a different history of recruitment of slaves mainly in the Gulf of Guinea [6,31].

169 We further applied an unsupervised clustering algorithm, ADMIXTURE [32], to visualize  
170 several axes of inter-individual genetic variation and population structure (Fig 3). The results  
171 of the ADMIXTURE analysis reflect the patterns of genetic diversity observed in the ASD-  
172 MDS, while providing further information on genetic differentiation within Cabo Verde and São  
173 Tomé at higher values of  $K$ .

174 As for the first dimension of MDS, ADMIXTURE clustering at  $K=2$  is explained by differences  
175 between African and European populations, maximizing the red and blue clusters,  
176 respectively. At higher values of  $K$ , new clusters capture differences among African  
177 populations. At  $K=7$ , West-Western African populations maximize the green cluster, and East  
178 Western African populations maximize the brown cluster. Also, at this value of  $K$ , West Bantu-  
179 speaking populations from West Central and South Western Africa present the highest  
180 proportions of the red cluster, while East Bantu-speaking groups from South Eastern and  
181 South Africa maximize the yellow cluster.

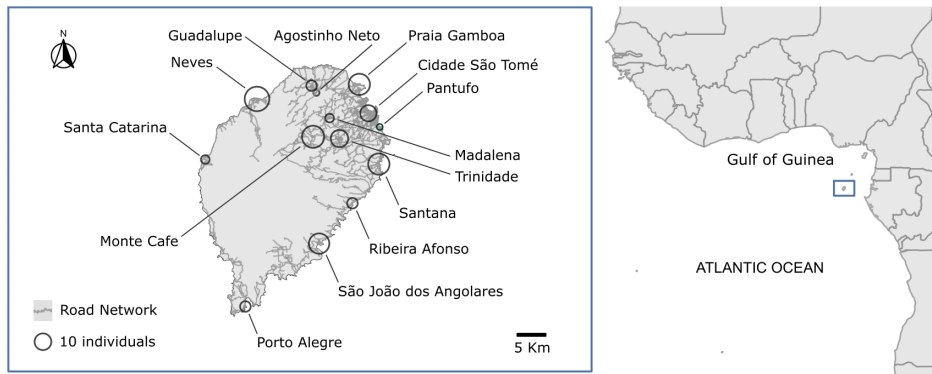
182 In general, genetic clusters at  $K=7$  discriminate among different areas of slave recruitment in  
183 Africa, revealing genetic similarities between populations currently living in these areas and  
184 populations descended from enslaved Africans. However, while African-Barbadians (ACB)  
185 and African-Americans (ASW) display relatively simple patterns with resemblance mainly to  
186 Central and East Western Africa, Cabo Verde and São Tomé exhibit more complex genetic  
187 structures that cannot be simply explained by the diverse African origins of their enslaved  
188 settlers.

189 In Cabo Verde, while the southern islands of Fogo, Santiago, and Brava show high  
190 proportions of the green cluster found in West Western Africa, individuals from the northern  
191 islands (Santo Antão, São Vicente and São Nicolau) maximize the orange cluster, which is  
192 found predominantly there, and might have resulted from *in situ* differentiation due to high  
193 genetic drift [6,30,33].

194 In São Tomé, some individual profiles are close to that of Cabo Verdeans, while others present  
195 varying proportions of clusters found in East-Western, West-Central and South-Western  
196 Africans. Finally, another group of individuals has a unique genetic pattern, represented by  
197 the violet cluster, which is not observed outside the island.

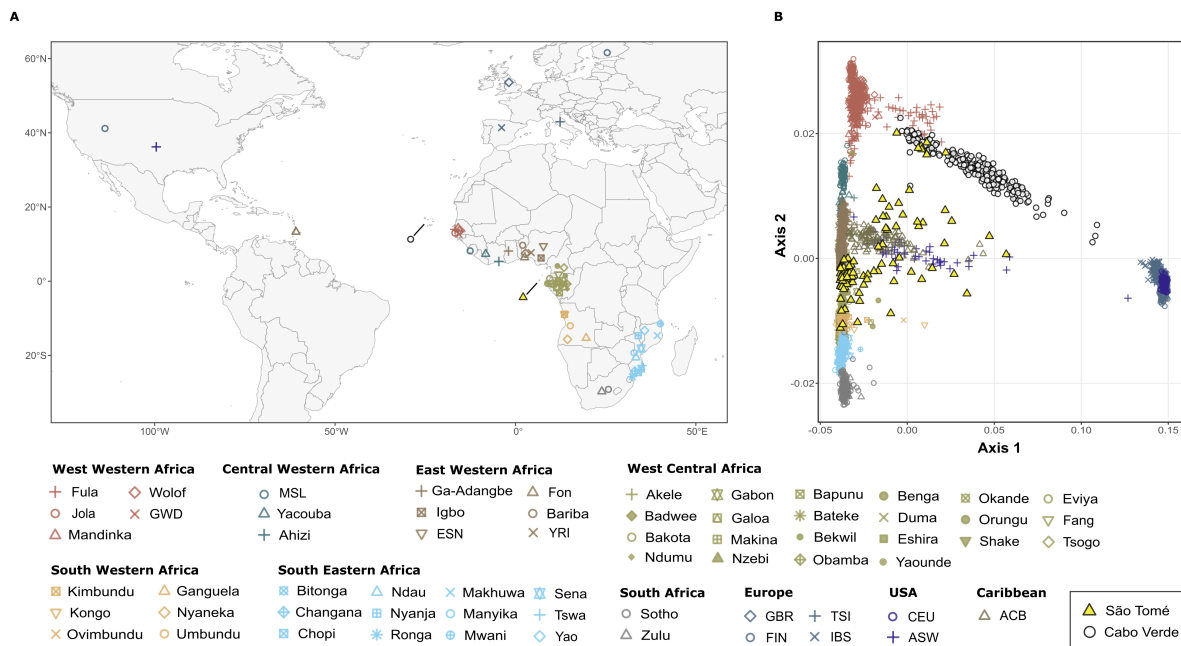
198

199



200

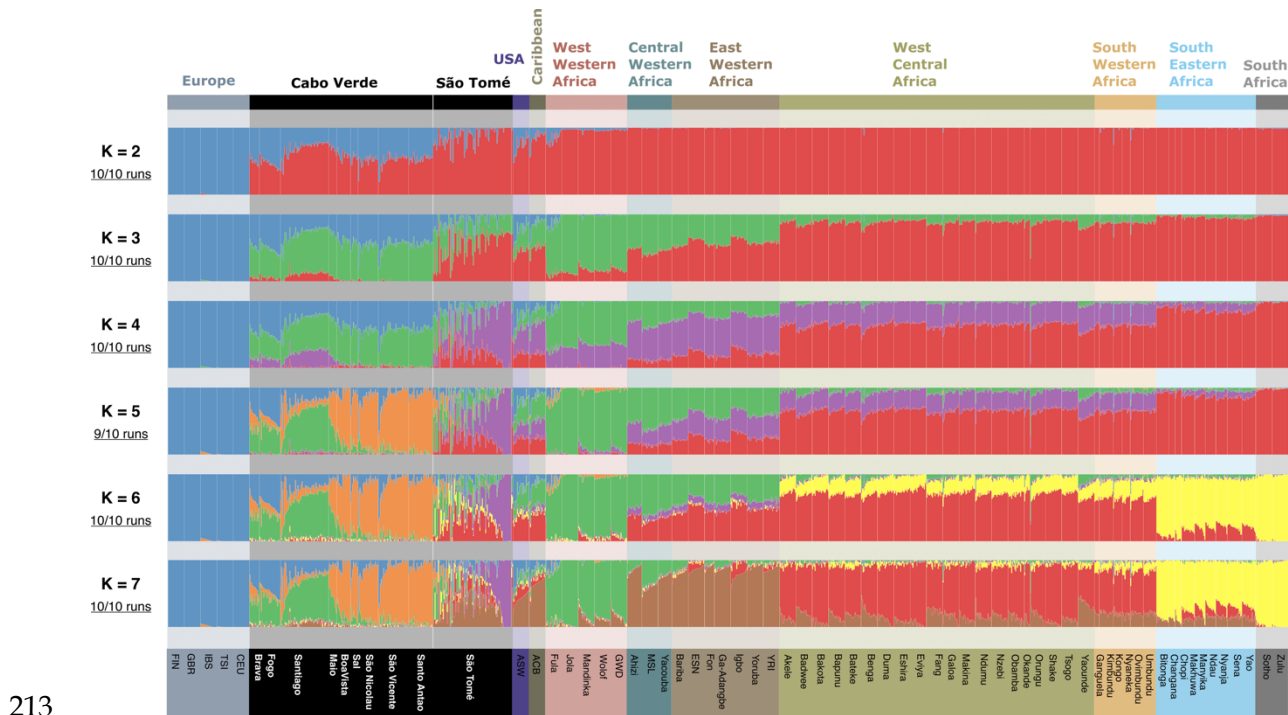
201 **Fig 1. The island of São Tomé in the Gulf of Guinea.** Distribution of individual samples  
 202 collected from 13 locations across the island of São Tomé. The road network of the island is  
 203 represented in grey. Each circle represents a location, with the size of the circle proportional  
 204 to the sample size. On the right, the map shows the position of São Tomé in the Gulf of Guinea.



205

206 **Fig 2. The genetic differentiation of the worldwide population samples.** (A) Geographical  
 207 locations of the 70 population samples considered in the present study. (B) MDS projection of  
 208 pairwise allele sharing dissimilarities (ASD) among São Toméans, Cabo Verdeans and other  
 209 African, American and European populations. The first two axes of variation in the MDS are

210 calculated among 3203 individuals using 411,121 autosomal SNPs. On the bottom, the color-  
 211 coded legend sorts the populations into 10 distinct geographical regions in continental Africa,  
 212 Europe and the Americas.



214 **Fig 3. Unsupervised ADMIXTURE analysis.** Unsupervised ADMIXTURE analysis using  
 215 110,499 LD-pruned ( $r^2 < 0.1$ ) autosomal SNPs from 1347 unrelated individuals, including 96  
 216 individuals that reported to be born in São Tomé, 2 individuals born in Príncipe, 225 individuals  
 217 born on seven different islands in Cabo Verde, and a random sample of 20 individuals for each  
 218 of the remaining population samples (S4 Table). The proportion of independent ADMIXTURE  
 219 runs closely resembling one-another is indicated on the left of each panel below each  $K$  value.  
 220 Population samples are ordered by geographical regions according to the color code indicated  
 221 in Fig 2.

222



## 223 **Haplotype-based population structure within São Tomé**

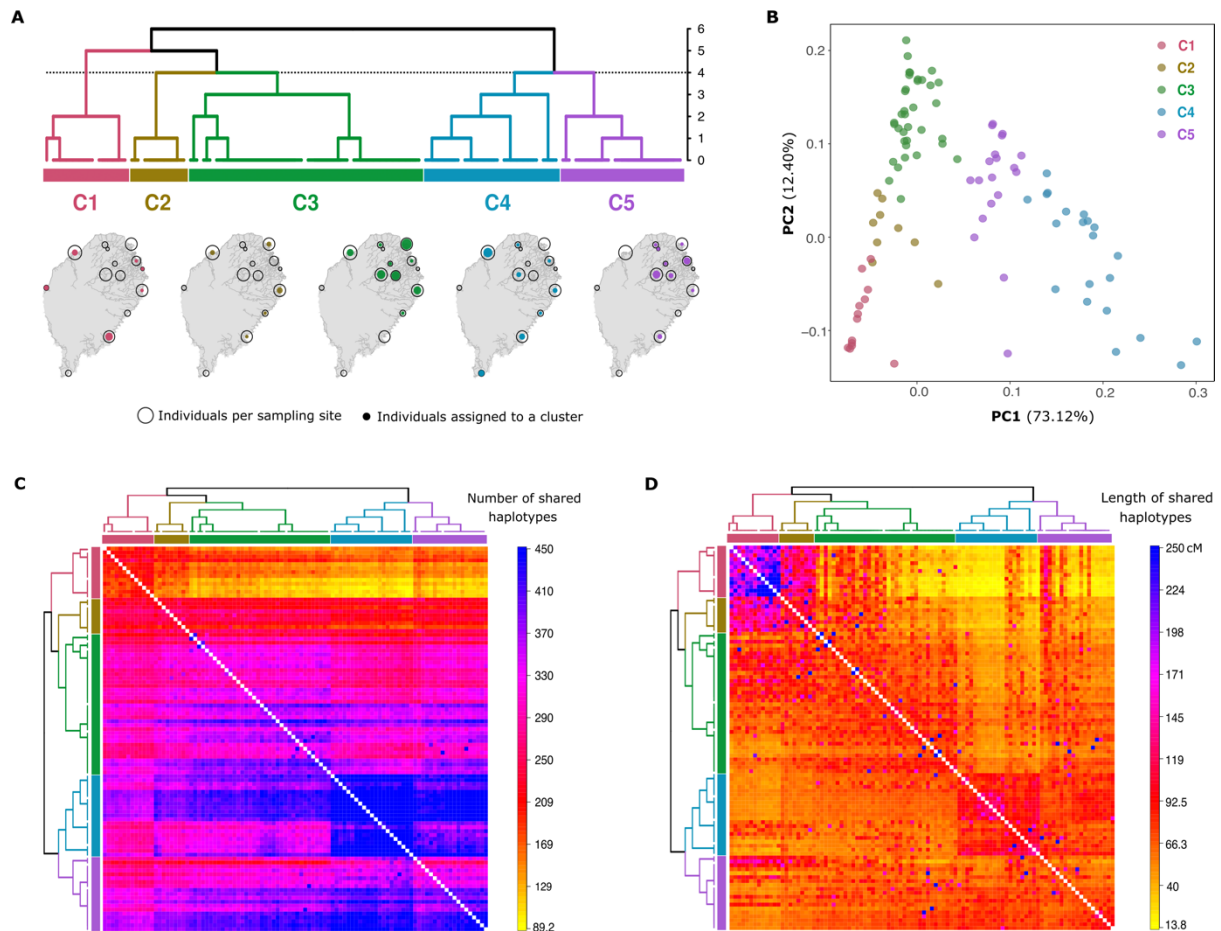
224 We analyzed patterns of shared haplotypes between all sampled individuals from São Tomé  
225 using ChromoPainter2 and fineSTRUCTURE [34]. By looking at the patterns of the number  
226 and length of shared haplotypes, we identified five different clusters of individuals (C1-C5) with  
227 high levels of inter-individual haplotypic resemblance (Fig 4).

228 Fig 4A presents a dendrogram displaying the phylogenetic relationships between the five  
229 clusters, as well as their spatial distribution across São Tomé. The majority of individuals  
230 sampled from the same locations are found to belong to more than one genetic cluster. This  
231 indicates that the geographical location of sampled individuals does not fully account for the  
232 observed genetic clustering patterns. Instead, it suggests that these patterns may be  
233 influenced by variation in admixture proportions. In a PCA calculated on the same co-ancestry  
234 matrix used to infer the fineSTRUCTURE dendrogram, individuals grouped in clusters C1 and  
235 C4 are separated along the first PC, while the second PC separates individuals belonging to  
236 the C3 cluster (Fig 4B). Within this triangle, C2 individuals lay in between C1 and C3 along  
237 the second PC, and C5 individuals lay in between C4 and the rest of São Tomé individuals  
238 along the first PC.

239 Moreover, we found that C1 individuals mostly share haplotypes with each other - both in  
240 terms of total number (Fig 4C) and cumulative length (Fig 4D) - and copy relatively few  
241 haplotypes from other individuals on the island. On the contrary, C2 individuals copy relatively  
242 long haplotypes from C1 individuals, as well as several haplotypes from the rest of São  
243 Toméan individuals with varying cumulative lengths. Individuals of the C3 cluster copy  
244 relatively more haplotypes from themselves than from the rest of the island, while individuals  
245 from clusters C4 and C5 differ markedly from other clusters, sharing the highest number of  
246 haplotypes among one another.

247

248



249

250 **Fig 4. Genetic clustering and population structure in São Tomé.** (A) Clustering of 96  
251 individuals from São Tomé into five genetic clusters using the fineSTRUCTURE algorithm  
252 based on the number of haplotype segments shared between each pair of individuals. The  
253 vertical axis of the dendrogram represents the distance or dissimilarity between clusters. The  
254 geographical distribution of individuals within each cluster is illustrated below the  
255 dendrogram. The size of empty black circles indicates the number of individuals sampled at  
256 each location, while the size of colored dots represents the number of individuals per cluster  
257 at each location. (B) Principal Component Analysis (PCA) based on the co-ancestry matrix  
258 represented in panel (C). Individual labels are color-coded according to the genetic clusters  
259 identified in the fineSTRUCTURE dendrogram in Panel (A). (C) Co-ancestry matrix based on  
260 the total number of haplotype chunks that each individual (row) copies from any other  
261 individual (column). (D) Co-ancestry matrix based on the cumulated length (in cM) of  
262 haplotype segments that each individual (row) copies from any other individual (column).

## 263 Haplotypic ancestry of genetic clusters

264 The haplotypic ancestry of each São Toméan cluster was modelled as a mixture of the  
265 possible source populations included in the reference dataset using the Chromopainter2-  
266 SOURCEFIND pipeline [35].

267 Based on the co-ancestry matrix computed with Chromopainter2, we inferred a maximum of  
268 eight source populations per “Target” population sample using SOURCEFIND. First, we  
269 considered each one of the admixed populations descended from enslaved-Africans as a  
270 separate “Target” population, and all other populations from continental Africa and Europe as  
271 “Donors” (Fig 5A). In agreement with previous studies [3,4,6,27,30], we found that the genetic  
272 composition of island populations from Cabo Verde essentially resulted from admixture  
273 between Europeans (from 27% in Santiago to 51% in Fogo) and African individuals from  
274 West-Western Africa (from 46% in Fogo to 69% in Santiago), while African-Barbadians (ACB),  
275 and African-Americans (ASW) resulted from admixture between Europeans (15% in ACB,  
276 24% in ASW) and Africans originating in Central and East-Western Africa (49% in ASW, 82%  
277 in ACB), South-Western Africa (3% in ACB, 18% in ASW), and West-Western Africa (9% in  
278 ASW only). São Tomé as a whole displays a more complex ancestry profile, including  
279 contributions from Europe (9%), West-Western Africa (16%), East-Western Africa (37%),  
280 South-Western Africa (36%), as well as South-Eastern Africa (2%). Importantly, when Cabo  
281 Verdean populations are considered as possible “Donors” (Fig 5B), the haplotypic ancestry  
282 from the island of Santiago in Cabo Verde (15%) replaces the Mandinka GWD haplotypic  
283 ancestry and part of the Iberian IBS component that were identified previously in São Tomé  
284 (Fig 5A).

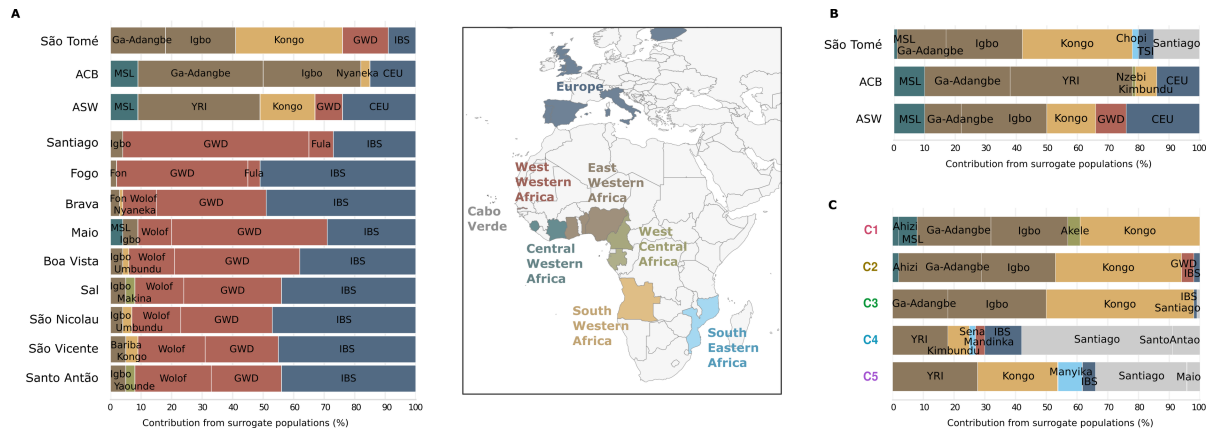
285 In Fig 5C, each of the five São Toméan genetic clusters (C1 to C5), is considered as a  
286 separate “Target” population and Cabo Verdean individuals from seven different islands are  
287 also included as seven separate “Donor” populations. Interestingly, clusters corresponding to  
288 the most basal bifurcation of the fineSTRUCTURE dendrogram (C1-C3 vs. C4-C5, Fig 4A)

289 have different ancestry profiles: the C1, C2 and C3 clusters share haplotypic ancestries with  
290 Kongo-speakers from Angola, in South-Western Africa (from 39% in C1 to 48% in C3), and  
291 the Ga-Adangbe from Ghana and the Igbo from Nigeria, in East-Western Africa (from 49% in  
292 C1 to 51% in C2); the C4 and C5 clusters share haplotypic ancestry with Cabo Verdeans,  
293 mostly from the island of Santiago (30% in C5, 49% in C4). Interestingly, we found evidence  
294 for small amounts of shared haplotypic ancestry between the C4 and C5 cluster and the Sena  
295 and Manyika population from Mozambique, in South-Eastern Africa (2% and 8%,  
296 respectively).

297 Taken together, these observations suggests that the major division between the São Toméan  
298 clusters (C1-C3 vs. C4-C5) is due to variable contributions from different source populations,  
299 while further splits occurring within each major division may be related to subsequent episodes  
300 of *in situ* isolation and gene flow.

301

302



303

304 **Fig 5. Haplotype ancestry of São Toméan and other admixed populations descended**

305 **from enslaved-Africans. (A) SOURCEFIND results showing the shared haplotypic ancestry**

306 **between each São Toméan (STP), Barbadian (ACB), Afro-American (ASW), and Cabo**

307 **Verdean population samples, set as Target, and 57 populations from continental Africa and**

308 **Europe, set as Donors. (B) SOURCEFIND results showing the shared haplotypic ancestry**

309 **between each STP, ACB, and ASW population samples, set as Target, with 66 populations**

310 **from Africa and Europe, including Cabo Verdean populations from nine islands, set as Donors.**

311 **(C) SOURCEFIND results showing the shared haplotypic ancestry between each São**

312 **Toméan cluster, set as Target, with 66 populations from Africa and Europe, including Cabo**

313 **Verdean populations from nine islands, set as Donors. The Donor populations are color-coded**

314 **by geographic regions in Africa and Europe, as indicated in the legend at the center of the**

315 **figure.**

316

## 317 **Long IBD-tracts shared among São Toméans and Cabo** 318 **Verdeans**

319 We explored the sharing of identical by descent (IBD) tracts of the five genetic clusters in São  
320 Tomé with each other, and with each of the island populations of Cabo Verde. Fig 6A  
321 summarizes sharing patterns of IBD tracts longer than 18 cM, representing approximately the  
322 last eight generations of recombination between individuals [2]. We observe differences in IBD  
323 sharing patterns between and within each São Toméan group. Notably, individuals of the C1  
324 cluster share the highest total length of long-IBD tracts between each other, and with  
325 individuals of the C3 cluster. Individuals of the C2 cluster share relatively low levels of long  
326 IBD tracts with each other, while sharing long IBD tracts with both the C1 and C3 clusters.  
327 Individuals of the C4 and C5 clusters are the only ones to share long IBD tracts with Cabo  
328 Verdean populations, in accordance with the SOURCEFIND results (Fig 5), although the C5  
329 cluster shows higher values of total length of long IBD tracts shared with the C1 and C3  
330 clusters in São Tomé, when compared to the C4 cluster.

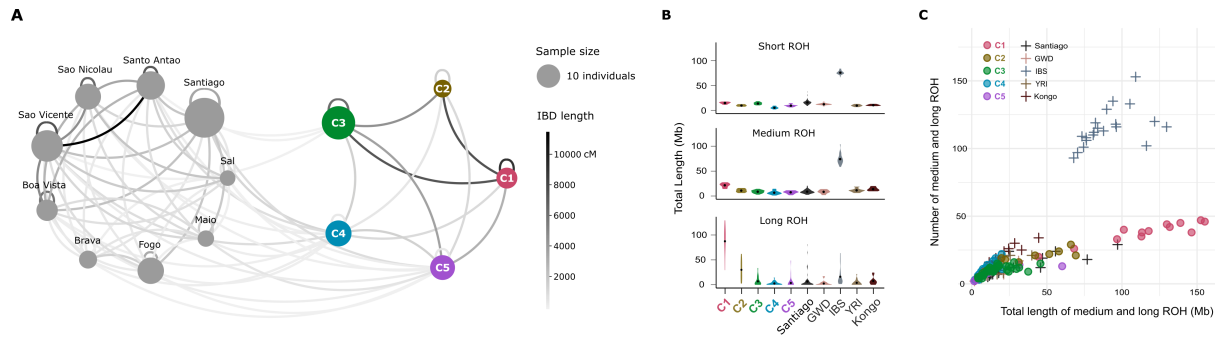
## 331 **Runs of Homozygosity**

332 We further investigated the presence of Runs of Homozygosity (ROH), arising when an  
333 individual inherits IBD segments from a recent common ancestor. ROH of different lengths are  
334 usually generated by different demographic processes: medium ROH can be generated by  
335 recent demographic events such as bottlenecks, while long ROH are usually due to recent  
336 parental relatedness [36–38]. Fig 6B reports the total length of long, medium and short ROH  
337 for each population separately, and Fig 6C reports the total length of long and medium ROH  
338 versus their total number for all populations, calculated with GARLIC [37]. C1 individuals  
339 present a total length of long ROH higher than any other population considered. The  
340 relationship between the number and size of ROH highlights that overall C1 has few, very long  
341 ROH, suggesting that it underwent inbreeding [39]. The sum of total length of long ROH in C2

342 individuals is intermediate between the C1 and C3 clusters in São Tomé. The C4 and C5  
343 clusters present similar average levels of medium and long ROH as the Santiago population  
344 in Cabo Verde.

345

346



347

348 **Fig 6. IBD tracts sharing and ROH analysis. (A)** Network displaying the cumulative length  
349 of Identical by Descent (IBD) tracts longer than 18 cM shared within and between population  
350 samples from São Tomé (right) and Cabo Verde (left). The size of the points is proportional to  
351 the size of the population sample. The color of the lines is proportional to the total cumulative  
352 length of long IBD tracts (in Mb). **(B)** Violin plots illustrating the total length of Runs of  
353 Homozygosity (ROH) within each cluster of São Tomé, the Santiago population in Cabo  
354 Verde, and other African and European populations. **(C)** Scatter plot where each point  
355 represents an individual sample, depicting the number of medium and long ROH plotted  
356 against their total length in Megabases (Mb).

357



## 358 **Dating admixture events for each São Toméan cluster** 359 **separately**

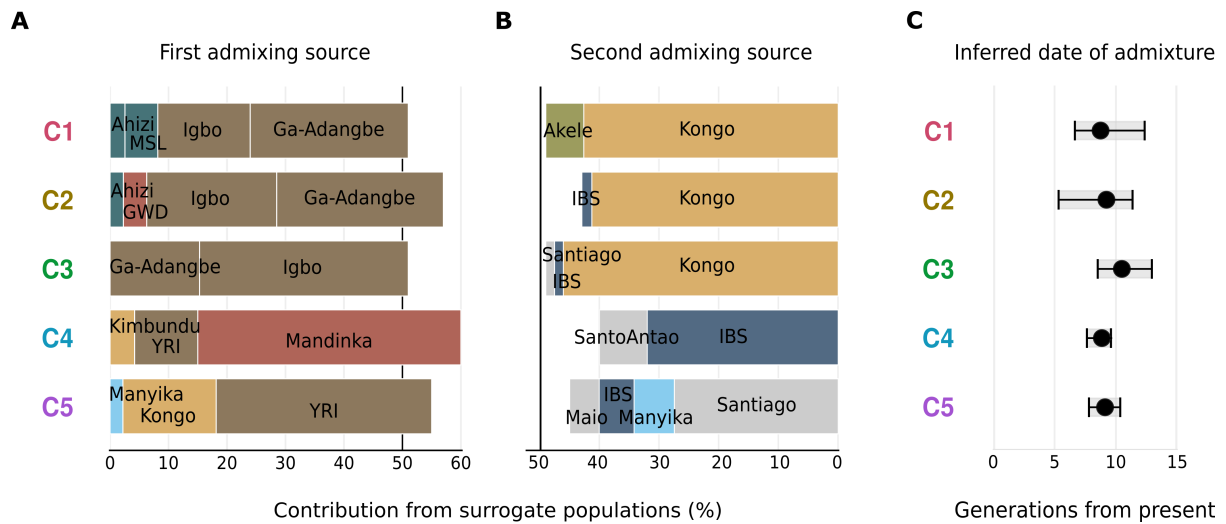
360 We inferred the timings of major admixture events that occurred in each of the five São  
361 Toméan genetic clusters, based on recombination distances between ancestry tracts using  
362 the algorithm implemented in fastGLOBETROTTER [40]. In this approach, we specified a  
363 separate set of “Surrogate” populations for each São Toméan genetic cluster, which were  
364 identified by SOURCEFIND as the “Donor” populations most likely involved in the admixture  
365 events of each cluster (Fig 5C). The fastGLOBETROTTER algorithm utilizes a mixture model  
366 for the “Surrogate” populations, allowing it to infer the relative contributions of each “Surrogate”  
367 population to two alternative mixtures, representing the first and second admixing sources for  
368 each admixture event. Overall, fastGLOBETROTTER identifies strong signals of admixture in  
369 all the five São Toméan clusters and attempts to date admixture events by considering a  
370 simplified scenario that assumes one major pulse of admixture, an important caveat that need  
371 to be kept in mind for the interpretation of the following results (See Materials and Methods  
372 and Discussion).

373 For the C1, C2, and C3 clusters, the major admixture event primarily involves populations from  
374 East-Western Africa, specifically the Igbo and Ga-Adangbe, on one side (Fig 7A), and the  
375 Kongo population from South-Western Africa on the other (Fig 7B). The dates of admixture  
376 events, considering a single pulse of admixture, range from 9 to 11 generations ago, and their  
377 relative Confidence Intervals are largely overlapping (Fig 7C). For the C3 cluster, the single  
378 pulse of admixture inferred 11 generations before present would correspond to the year 1694  
379 (CI: 1636-1769) when considering a generation time of 30 years. Moreover, admixture is  
380 estimated to approximately 9 generations before present for the C1 cluster, corresponding to  
381 the year 1763 (CI: 1654-1824), and similarly for the C2 cluster, around 1748 (CI: 1683-1864).  
382 The C4 and C5 clusters likely result from varying levels of admixture between Cabo Verdeans  
383 and São Toméans. In both clusters, the contribution from admixed São Toméans is

384 represented as a mixture of East-Western Africans and South-Western Africans (Fig 7A), and  
385 admixture events are inferred at approximately 9 generations before present, around 1760  
386 (CI: 1736-1795) for the C4 cluster, and 1751 (CI: 1714-1790) for the C5 cluster (Fig 7C).

387

388



389

Contribution from surrogate populations (%)

Generations from present

390 **Fig 7. Admixing sources and timing of admixture events.** Results of fastGLOBETROTTER  
391 analysis reporting the inferred composition of each admixing source for the most supported  
392 event, when assuming only a single date of admixture, for each São Toméan cluster. The  
393 surrogate populations are the ones inferred by SOURCEFIND in Fig 5C. For each São  
394 Toméan cluster, the relative contribution of surrogate populations to (A) the first and (B) the  
395 second admixing sources is depicted. The total contribution of both admixing sources sums to  
396 100%. (C) Admixture event dates are reported in generations from present, with grey bars  
397 indicating the 95% Confidence Interval calculated with 100 bootstraps.

398

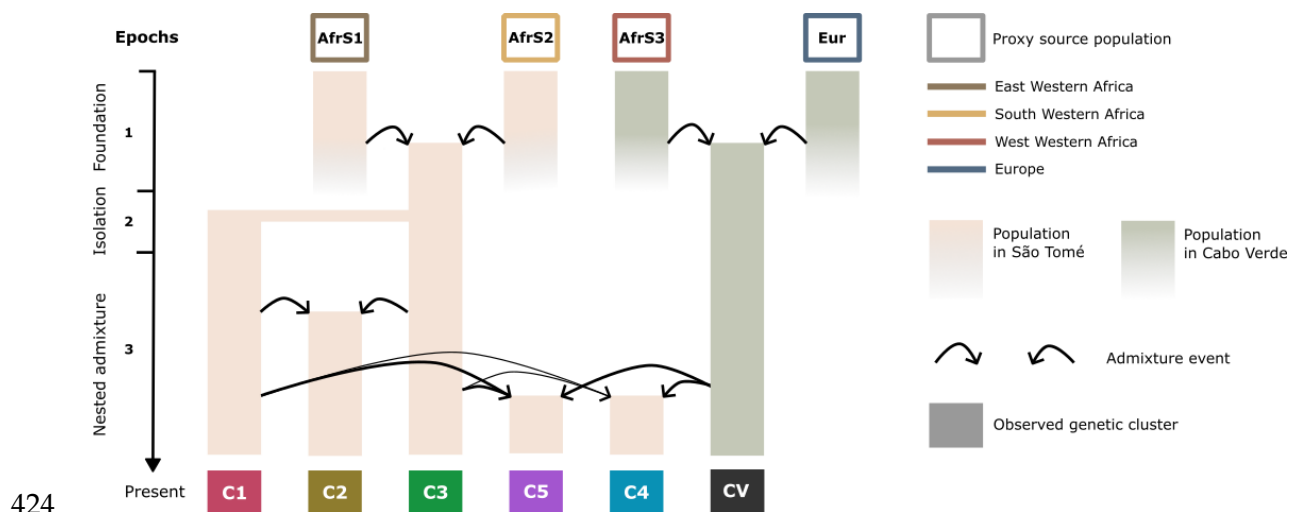
## 399 Discussion

400 In this work, we analyzed the detailed structure of the genome-wide diversity of São Tomé -  
401 the earliest European colony in Equatorial Africa, which became a model for the Plantation  
402 Economic system that was later deployed in numerous other European colonies across the  
403 Americas [11]. Our investigation reveals that the genetic structure of São Tomé is more  
404 complex than that of several other populations descended from enslaved Africans, due to an  
405 intricate pattern of admixture and isolation events during and after the TAST.

406 We identified five genetic groups based on the hypothesis that individuals sharing the most  
407 haplotypes across the island may result from similar genetic histories. Notably, individuals  
408 from the same sampling site on São Tomé were often assigned to more than one genetic  
409 group, showing how social segregation in the colonial context operated at a micro-scale on  
410 the island.

411 A major division separates groups sharing similar proportions of haplotypic ancestries with  
412 East Western and South Western African populations (C1-C3), from groups with a clear  
413 contribution from Cabo Verde (C4 and C5) (Fig 4, 5). This differentiation is in line with the  
414 historically known recruitment of enslaved Africans from the Gulf of Guinea and Congo-Angola  
415 for the peopling of São Tomé from the 16<sup>th</sup> to the 18<sup>th</sup> centuries, which was followed by the  
416 arrival of Cabo Verdean immigrants for indentured labor in the 20<sup>th</sup> century.

417 In addition to differences in the African origins of their genetic ancestors, the present gene  
418 pool of São Tomé's inhabitants was also shaped by demographic events occurring within the  
419 island. In Fig 8, we propose an interpretation of the demographic history of São Tomé based  
420 on different types of genetic evidence, which explores the role of external population sources  
421 as well as *in situ* admixture and isolation events in shaping the genetic landscape of the island.  
422 In the following paragraphs, we discuss the relationships between the genetic data and the  
423 available historical and linguistic information that underlie our demographic model.



425 **Fig 8. Nested admixture in São Tomé.** Schematic representation of the demographic  
426 histories that may have given rise to the five genetic cluster here identified in São Tomé (Fig  
427 4A). The timeline is divided into three epochs, spanning from the foundation of the São  
428 Toméan population to the present. Each bar represents a population evolving over time,  
429 colored according to the location of the admixture event. Proxies for admixture sources are  
430 shown at the top of the graph and are color-coded based on their origins in Africa and Europe.  
431 The colored rectangles at the bottom represents the five São Toméan genetic clusters (C1,  
432 C2, C3, C4, C5), and the Cabo Verdean population sample (CV). Each arrow illustrates the  
433 contribution of a source population at a certain epoch. In most cases, to explain the genetic  
434 diversity of a group of individuals in São Tomé, we need to consider that at least one of the  
435 source populations is itself recently admixed. We call this event “nested admixture”. Future  
436 studies should test this and similar models, including nested admixture, in their ensemble.

437

## 438 **A population descended from freed slaves**

439 Towards the end of the 16th century, competition from Brazilian sugar production, coupled  
440 with significant slave revolts, led to São Tomé's rapid economic decline and diminishing  
441 Portuguese interest in the colony [12]. Throughout the 17<sup>th</sup> and 18<sup>th</sup> centuries, a growing  
442 population of *Forros*, literally "freed slaves", eventually became the dominant social group in  
443 the island [21]. Linguistically, the *Forros* speak a creole language with a Nigerian Edoid-related  
444 syntax and a predominantly Portuguese-based lexicon, which also includes various lexical  
445 items derived from Edo and Kikongo, a Bantu language from Congo-Angola [41]. Although the  
446 number of Forro creole speakers has declined in recent decades, it remains the most widely  
447 spoken creole on the island [42].

448 Forro-speaking individuals are currently found throughout São Tomé, with a notable  
449 concentration in the northeastern areas where the capital city of São Tomé is located [22].  
450 Since most individuals belonging to the C3 cluster were sampled in the northeast of São Tomé  
451 (Fig 4A), it is likely that they are the descendants of the *Forros*. As noted before, individuals  
452 of the C3 cluster derive almost equal proportions of their genomic ancestry from populations  
453 currently living in Nigeria in East Western Africa, as a consequence of the slave trade in the  
454 Gulf of Guinea, and Angola in South Western Africa, reflecting the Bantu-speaking trading  
455 regions in Congo-Angola (Fig 5C). While these contributions result from extensive admixture  
456 between enslaved Africans of different origins, it is less clear when this admixture occurred.

457 According to the available historical and linguistic information, the features of the Forro creole  
458 that have a Nigerian origin can be traced back to the early phases of the peopling of São Tomé  
459 in the beginning of the 16<sup>th</sup> century, when the slave trade was mostly limited to the Gulf of  
460 Guinea. Bantu-derived linguistic features would have entered the Forro creole in a later phase,  
461 still in the 16<sup>th</sup> century, during the expansion of the plantation system, when Bantu-speaking  
462 regions became more important for slave recruitment [41]. According to our dating approach,  
463 using a single pulse model, the mixing of East Western and South Western African genetic

464 contributions in the C3 cluster would have occurred around 1694 (CI: 1636-1769) (Fig 7C),  
465 substantially later than the mixing of linguistic features and the timing of arrival of slaves from  
466 the African mainland would suggest.

467 However, the dates of genetic admixture must be considered in the light of methodological  
468 limitations. In particular, fastGLOBETROTTER cannot always distinguish between discrete  
469 pulses and continuous admixture: if the period of continuous admixture is short, the model  
470 may infer a single admixture event, and the estimated date may fall within the time frame of  
471 the actual recurring admixture period [43]. Therefore, genetic admixture between East  
472 Western and South Western Africans may have occurred recurrently, both before and after  
473 the late-17<sup>th</sup> century date inferred here.

## 474 **One of the first communities of runaway slaves**

475 According to historical records, enslaved Africans escaped plantations from the onset of  
476 Portuguese colonization in São Tomé, and managed to control the mountainous interior of the  
477 island until occupation by colonial authorities in the late 19<sup>th</sup> century, when one of the  
478 communities formed by runaway slaves became known as the *Angolares* [44]. Today, the  
479 descendants of the *Angolares* are mostly found in fishing communities scattered along the  
480 coast, with major settlements in the villages of São João dos Angolares and Santa Catarina  
481 where most individuals of the C1 cluster were collected (Figs 1 and 4A). It is therefore likely  
482 that the C1 cluster represents *Angolares* communities in São Tomé.

483 While colonial narratives have suggested that the *Angolares* descended from survivors of a  
484 shipwreck with enslaved Africans from Angola who remained almost completely isolated [44],  
485 historical, linguistic, and genetic studies have proved this hypothesis unrealistic [5,41]. The  
486 Angolar creole spoken by the *Angolares* shares many features with the Forro creole, although  
487 having a stronger lexical influence from the Bantu language Kimbundu, spoken in northern  
488 Angola. Thus, it seems likely that the two languages descend from the same proto-creole of

489 the Gulf of Guinea and subsequently accumulated differences in the amount and type of Bantu  
490 features that became incorporated in the ancestral creole [20].

491 Consistent with previous genetic studies [5], we found that the *Forros* and *Angolares* of the  
492 C3 and C1 clusters, respectively, derived the most of their genetic ancestry from the Gulf of  
493 Guinea (East Western Africa, 50% and 49%) and Bantu-speaking regions in Congo-Angola  
494 (South Western Africa, 48% and 39%). However, despite this similarity, C1 individuals  
495 exhibited a markedly distinct genetic pattern, maximizing the unique ADMIXTURE violet  
496 cluster (Fig 3), and sharing very long haplotypes (Fig 4D), long IBD tracts (Fig 6A), and ROH  
497 (Fig 6B). Together, these characteristics are typical of isolated populations with small effective  
498 population sizes who experienced high levels of genetic drift and inbreeding (Ceballos et al.,  
499 2018). Similar features have also been observed in admixed population isolates from South  
500 America [45], and in other populations descended from maroon communities, such as the Noir  
501 Marron from Suriname and French Guyana [46], and another sample of *Angolares* individuals  
502 from São Tomé [5].

503 The strong signal of genetic drift can significantly affect recombination distances between  
504 ancestry tracts [7,47]. Nevertheless, the single pulse of admixture is estimated approximately  
505 9 generations before present for the C1 cluster, corresponding to the year 1763 (CI: 1654-  
506 1824), which is only slightly earlier than the date inferred for the C3 cluster, and their  
507 confidence intervals largely overlap. Both clusters likely share a similar history of founding  
508 admixture, with the genetic signal of C1 more heavily impacted by subsequent genetic drift  
509 resulting from isolation and inbreeding.

510 It is therefore likely that the *Angolares*, sharing a major part of their genetic ancestry with the  
511 *Forros*, became differentiated through isolation after running away from the plantations (Fig  
512 8). According to the interpretation of previous studies [5], additional contributions from  
513 runaway slaves from Angola would explain the important impact of the Kimbundu language  
514 that distinguishes the Angolar and Forro creole languages.



## 515 Cabo Verdeans in São Tomé

516 The second period of Portuguese colonization in São Tomé, corresponding to the  
517 establishment of coffee and cacao plantations since the 19<sup>th</sup> century, marks a profound  
518 demographic change on the island. The number of indentured laborers working on the  
519 archipelago of São Tomé e Príncipe represented approximately half of the population during  
520 the first half of the 20<sup>th</sup> century [48]. According to the last census, 8% of the population in São  
521 Tomé currently speaks the Cabo Verdean Kriolu [22].

522 Individuals from the C4 cluster in São Tomé are more similar to Cabo Verdeans than to the  
523 rest of the São Toméan samples in terms of haplotypic diversity (S9 Fig). They share over half  
524 of their haplotypic ancestry with the islands of Santiago and Santo Antão, and present the  
525 lowest proportions of East-Western African and South-Western African contributions among  
526 all São Toméan clusters (Fig 5C). The individuals in the C4 cluster have mostly been sampled  
527 in sites close to historical *roça* plantations, such as Agostinho Neto, Monte Café, São João  
528 dos Angolares, and Porto Alegre, which hosted numerous Cabo Verdean *serviçais* to work in  
529 plantations. Moreover, they include the five São Toméan individuals that reported in our  
530 interviews that one or both their parents were born in Cabo Verde.

531 Our findings suggests that the Cabo Verdean *serviçais* who stayed in São Tomé after their  
532 contracts ended, and left descendants on the island, most probably came from the islands of  
533 Santiago and Santo Antão, in the south and the north of the Cabo Verde archipelago,  
534 respectively (Fig 5D). However, looking at sharing patterns of long-IBD tracts, the C4 cluster  
535 shows recent common ancestry with admixed populations from different islands in Cabo Verde  
536 (Fig 6A). Island populations of Cabo Verde share many long-IBD tracts with each other,  
537 particularly with Santiago, the first island to be settled in the 15<sup>th</sup> century, which played a key  
538 role in the founder effects that shaped the genetic diversity of the other islands [6].  
539 SOURCEFIND potentially underestimated contributions from other islands, and likely

540 identified Santiago as a major contributor to Cabo Verdean genetic contribution in São Tomé,  
541 partly because the algorithm was limited to using a maximum of eight surrogate populations.

542 The main admixture event inferred for individuals of the C5 cluster, approximately 9  
543 generations before present (circa 1760, CI: 1736-1795, assuming 30 years per generation;  
544 Fig 7C) between West-Western African and European populations, likely reflects the founding  
545 of the admixed gene pool in Cabo Verde. However, this date is more recent than expected for  
546 the settlement of the archipelago [6].

547 The accuracy of admixture dating depends strongly on how well the model captures the  
548 underlying history of admixture and how well the surrogate populations represent the actual  
549 sources. The simplified admixture model used here, which assumes a single admixture pulse,  
550 does not account for subsequent, albeit minor, admixture between Cabo Verdean and São  
551 Toméan individuals. In this model, a mixture of East Western and South Western African  
552 contributions is included in the first admixture source (Fig 7A), which consists mainly of West  
553 Western African contribution, while admixed Cabo Verdean genomes contribute to the second  
554 source (Fig 7B), which consists mainly of European contribution.

## 555 **Limited European genetic contribution**

556 Overall, São Toméan individuals exhibit limited levels of haplotypic ancestries shared with  
557 European populations, in particular compared to numerous other enslaved-African  
558 descendant populations in the Americas and in Africa [8], and even specifically compared to  
559 other populations from the former Portuguese colonial empire [3,6].

560 This was somewhat unexpected compared to historical records, as admixture was generally  
561 tolerated at the very beginning of the settlement as part of Portugal's colonization efforts [49].  
562 However, historically, the African population in São Tomé, including both enslaved and free  
563 individuals, far outnumbered the European presence on the island [50], and to a larger extent  
564 than what has been reconstructed by historians in Cabo Verde [51]. Furthermore, a substantial

565 part of the Portuguese settlers left the island in the 18<sup>th</sup> century, a period of economic crisis  
566 between the two main Plantation Economies [52].

567 In our haplotype-sharing analyses, individuals with European descent in São Tomé may have  
568 clustered preferentially with individuals of Cabo Verdean descent, since Cabo Verdeans have  
569 relatively high levels of European haplotypic ancestry [6]. Therefore, we know that we  
570 overestimate European haplotypic ancestry in São Tomé when we do not consider Cabo  
571 Verdeans as sources of admixture (Fig 5A), while we may underestimate it if individuals with  
572 higher European ancestry are included in genetic clusters of Cabo Verdean descent (Fig 5C).  
573 Nevertheless, it is interesting to note that the majority of European haplotypic ancestry in São  
574 Tomé actually comes from Cabo Verdean admixed genomes.

575 The Plantation Economic system that was originally developed in São Tomé in the 16<sup>th</sup> century  
576 relied on strong social and marital segregation between free and enslaved-communities  
577 imposed by the socio-economically dominant colonists [49]. Together, these three  
578 phenomena, demography, migration, and socio-cultural constraints, likely explain the low  
579 absolute levels of shared ancestries with Europeans compared to most other populations  
580 whose history has been also intimately linked with the TAST, but where the Plantation  
581 Economy system and accompanying segregation was adopted substantially later than in São  
582 Tomé.

### 583 **Nested admixture between recently admixed populations**

584 Beyond the three groups previously discussed—the *Forros*, descendants of freed slaves, in  
585 the C3 cluster; the *Angolares*, descendants of runaway slaves, in the C1 cluster; and the Cabo  
586 Verdean *serviçais* and their descendants, in the C4 cluster—we also identified evidence of  
587 admixture among these groups within our sample, specifically in the C2 and C5 clusters.

## 588 **Admixture between *Angolares* and *Forros***

589 The nine individuals of the C2 cluster are genetically closer to the C3 cluster (Fig 3, 4A), and  
590 share numerous long IBD-tracts with both the *Angolares* of the C1 cluster and the *Forros* of  
591 the C3 cluster, as well as showing intermediate levels of medium and long ROH between  
592 these two groups (Fig 6C). C2 individuals may therefore result from admixture between the  
593 *Angolares* descendants of runaway slaves and the freed population of *Forros*, after the  
594 *Angolares* communities began to establish economic, social and cultural connections with the  
595 rest of the island population in the 19<sup>th</sup> century.

596 Previous studies on populations descended from maroon communities emphasize strong  
597 isolation and genetic differentiation [5,46], in line with their history of escaping slavery and  
598 seeking refuge in remote, isolated locations. However, they do not account for the potential  
599 admixture with surrounding populations over time. In contrast, our findings show that  
600 marooning isolation does not necessarily imply the absence of complex admixture processes,  
601 which in turn argues for further refinement of genetic expectations for descendants of  
602 “isolated” populations in the context of admixture in general and the TAST in particular.

## 603 **Admixture between Cabo Verdeans and São Toméans**

604 The immigrant indentured workers were largely confined to the *roças*, or plantation units,  
605 effectively isolating them from the local São Toméan population, who refused to engage in  
606 plantation work after the abolition of slavery . This period is thus marked by the emergence of  
607 new forms of discrimination and social segregation on the island [53]. We found evidence of  
608 varying degrees of recent common ancestry between the C4 and C5 clusters, of Cabo  
609 Verdean descent, and the C3 and C1 clusters, representing the *Forros* and *Angolares*,  
610 respectively (Fig 6A). The low amount of São Toméan genetic contribution to the haplotypic  
611 ancestry of the C4 cluster aligns with the history of discrimination (Fig 5C). However, the  
612 haplotypic ancestry of the C5 cluster suggest that substantial admixture occurred between

613 Cabo Verdeans and São Toméans (Fig 7), meaning that segregation may have been  
614 permeable or shaped by additional social factors.

## 615 **The genetic contributions from Mozambique and Angola**

616 In addition to Cabo Verdeans, Mozambican and Angolan contractual workers were also  
617 recruited to work on coffee and cacao plantations in São Tomé after the abolition of slavery  
618 [54,55]. While our São Toméan sample shows a substantial amount of haplotypic ancestry  
619 from Cabo Verde, we found relatively little genetic contribution from Mozambique, primarily in  
620 the C4 and C5 clusters (2% and 8%, respectively; Fig 5C). Assessing the impact of post-  
621 slavery migrations from Angola on the genetic diversity of São Tomé is more challenging, as  
622 South Western African ancestry in São Tomé may have multiple origins, including both TAST-  
623 related migrations in the 16<sup>th</sup> and 17<sup>th</sup> centuries, and more recent migrations of contractual  
624 workers from Angola in the late 19<sup>th</sup> century.

## 625 **Conclusions and perspectives**

626 The interplay of socio-cultural and economic factors on the island of São Tomé in the colonial  
627 context of the TAST likely influenced both reproductive isolation and gene flow between  
628 communities, with social constraints on admixture evolving over time. Our results suggest that  
629 three genetic clusters in São Tomé correspond to significant anthropological, historical, and  
630 linguistic categories - the *Angolares*, *Forros* and Cabo Verdeans - while two other clusters  
631 appear to arise from admixture among these groups. This observation reflects a gradual  
632 dismantling of a previously established genetic structure shaped by historical migrations and  
633 demographic events.

634 The intricate mosaic of genetic ancestries that we observed in São Tomé led us to define the  
635 concept of “nested admixture”, which refer to the gene flow between human groups that are  
636 recently admixed (Fig 8). In case the admixed source populations result from similar admixture  
637 histories, the genetic diversity of the resulting population may be composed of layers of

638 admixture patterns, making it difficult to identify the exact source populations of the ancestry  
639 tracts. Admixture inference methods generally perform best when the source populations are  
640 not recently admixed, are genetically distinct from each other, and differ significantly from the  
641 admixed population. Interestingly, recent advances have been made to address these  
642 challenges [56]. In the case of São Tomé, future studies could attempt to infer and date  
643 admixture between the *Forros* and the *Angolares*, using, for example, the C1 and C3 clusters  
644 as proxies of source populations for the C2 cluster. Similarly, it may be possible to disentangle  
645 the two main admixture processes - the founding of the Cabo Verdean population and  
646 subsequent admixture with São Toméans - for the C5 cluster. More generally, our work  
647 highlights the need to account for gene flow between recently admixed groups, following  
648 several previous investigations on populations descended from enslaved Africans on both  
649 sides of the Atlantic [2,6,57,58].

650 A genealogical perspective on admixture could further illuminate on the complex patterns of  
651 genetic ancestry in São Tomé. For instance, novel theoretical developments have investigated  
652 the relationship between the number of genetic and genealogical ancestors in the Afro-  
653 American population [9,59], which would be of major interest to further understand how  
654 admixture occurred over time in the different São Toméan genetic groups here identified.

655 Finally, the admixture process does not occur randomly but is influenced by social and cultural  
656 factors, especially in the colonial context of the TAST [8]. Among other consequences, we  
657 may have underestimated the number of generations since admixture by assuming random  
658 mating [33,60]. Previous genetic studies suggest that in some populations descended from  
659 enslaved Africans, female and male contributions from source populations were not equal,  
660 and mating occurred preferentially between individuals with a certain genetic ancestry [33,61–  
661 63]. Future studies will need to investigate the impact of sex-biased admixture and ancestry-  
662 related assortative mating on genetic diversity patterns and admixture inference in São Tomé.

## 663 **Materials and Methods**

### 664 **Genetic dataset**

#### 665 **Sampling strategies**

666 DNA sampling has been conducted by two different research groups on the two archipelagos,  
667 respectively. Sampling strategies are detailed in [6] for Cabo Verde and in [24] for São Tomé  
668 e Príncipe. Anthropological questionnaires, completed alongside DNA sampling, provide  
669 birthplace locations of each individual and their parents. DNA was collected with buccal swabs  
670 and extracted from saliva samples following standard protocols.

#### 671 **Genotyping and Quality Control**

672 We genotyped 361 DNA samples (261 samples from Cabo Verde and 100 samples from São  
673 Tomé e Príncipe) in 5 batches, separately, using different versions of the Illumina  
674 HumanOmni2.5Million-BeadChip genotyping array read with an iScan using BeadScan  
675 software at the OMICS platform of the Institut Pasteur (S1 Table). We conducted a  
676 comprehensive genotyping quality control in four phases (S2 Table).

677 In Phase 1, Genotype calling, we called genotypes using Illumina GenomeStudio Genotyping  
678 Module version 1.9.4 for each batch separately. We used GenCall version 7.0.0 with a low  
679 cutoff of 0.15. We removed markers with ambiguous positions on the human genome  
680 reference sequence, and we removed markers on sex chromosomes and mitochondrial DNA.  
681 We filtered autosomal SNPs that failed the 95% call rate and 0.2 cluster separation cutoffs,  
682 and samples that failed the 95% call rate cutoff. Transversions (A/T and G/C markers) and  
683 markers with duplicated positions were also removed.

684 In Phase 2, Batch merging, we only kept markers genotyped in all 5 batches using *ad hoc*  
685 Python scripts. To check the concordance rate of genotype calls, we genotyped 32 individuals  
686 separately on two, three, or four different batches. We removed SNPs that were called



687 differently on different batches for the same duplicated individual sample. We then removed  
688 the duplicated individual samples with the highest missing rate. Finally, we remapped markers  
689 positions based on rsIDs according to dbSNP Human Build 154 Release on GRCh38 genome  
690 assembly (Genome Reference Consortium Human Build 38). We removed markers whose  
691 rsIDs have multiple positions in the same build, or that do not have a position in this build. We  
692 finally retained a total of 2,104,297 autosomal markers mapped to the GRCh38 genome  
693 assembly (hg38) from 358 samples.

694 In Phase 3, Genetic relatedness, we removed genetically related individuals up to the 2<sup>nd</sup>  
695 degree using KING version 2.2.7 [64]. In Phase 4, SNP Quality control, we subjected the  
696 remaining autosomal SNPs to quality control filters using PLINK version 1.9 [65]. We removed  
697 markers with high missing data (10%), markers that deviate from Hardy-Weinberg equilibrium  
698 (HWE), and insertion/deletion markers. We finally retained 2,104,148 curated autosomal  
699 SNPs from 330 unrelated individuals, including 233 individuals sampled in Cabo Verde and  
700 97 individuals sampled in São Tomé e Príncipe.

## 701 **Merging with worldwide populations**

702 The resulting dataset of 330 family unrelated individuals from Cabo Verde and São Tomé e  
703 Príncipe was merged with 2504 samples from 26 populations worldwide included in the 1000  
704 Genomes project Phase 3 (International Genome Sample Resource (International Genome  
705 Sample Resource IGSR) [25], with 1307 samples from 14 African populations included in the  
706 African Genome Variation Project (EGA ID EGAS00001000959) [26], with 188 samples from  
707 15 populations in Mozambique and Angola (E-MTAB-8450) [28], and with 1366 samples from  
708 38 sub-Saharan African populations (EGA ID EGAS00001002078) [27] (S3 Table). For the  
709 last two datasets, we conducted the same steps of genotyping Quality Controls from raw  
710 genotyping data as described before. Moreover, we checked for genetic relatedness at the  
711 end of merging each of the four datasets. We retained polymorphic markers in the merged  
712 dataset which amounted 411,121 SNPs from 5423 family unrelated individuals.



## 713 **Population labels and geographical regions**

714 We assigned each individual sample from Cabo Verde and São Tomé e Príncipe to an island  
715 of birth based on individual birthplaces recorded in the family anthropology questionnaires.  
716 Out of the 233 individuals sampled in Cabo Verde, 225 were born on seven of the  
717 archipelago's islands, while seven were born outside of Cabo Verde, in particular two in São  
718 Tomé, three in Angola, one in Brazil, and one in Portugal. One individual has been excluded  
719 due to missing birthplace information. Out of 97 individuals sampled in São Tomé e Príncipe,  
720 almost all individuals were born in São Tomé, apart from one individual born in Gabon, and  
721 two individuals born in the island of Príncipe. Finally, we retained 96 individuals born in São  
722 Tomé, including the 94 individuals sampled in São Tomé and the two sampled in Cabo Verde.  
723 The dataset containing 323 individuals born in Cabo Verde and São Tomé e Príncipe will be  
724 referred to as CVSTP henceforth.

725 The 5423 genetically unrelated individuals from 107 populations in the final merged dataset  
726 were grouped into 16 distinct regions across the globe, including ten regions within Africa, with  
727 CVSTP identified separately, three regions in the Americas, and one region each in Europe,  
728 South Asia and East Asia.

## 729 **Population genetics descriptions**

### 730 **Allele Sharing Dissimilarity**

731 We investigated genetic diversity patterns between each pair of individuals based on  
732 successive subsets of populations in our dataset. We first computed a matrix of pairwise allele  
733 sharing dissimilarities including all the individuals and all SNPs in the merged dataset using  
734 the ASD software (<https://github.com/szpiech/asd>). Then we explored three axes of variation  
735 of Multi-Dimensional Scaling projections of various subsets of this ASD pairwise-matrix using  
736 the *cmdscale* function in R [66]. In particular, we removed East Asian, South Asian, South  
737 American, Puerto Rican PUR, Mexican American MXL, Central African, East African, and four

738 West-Central African hunter gatherers populations by subsampling the ASD matrix before  
739 conducting the MDS projection anew. Fig 2B report the first two dimensions of the ASD-MDS  
740 based on 411,121 autosomal SNPs of 3203 individuals from 77 populations, and the third  
741 dimension is presented in S1 Fig.

## 742 **Genetic clustering**

743 We used the software package ADMIXTURE version 1.3 [32] to explore further genetic  
744 resemblances among individuals. ADMIXTURE analysis is sensitive to sample size  
745 heterogeneities, therefore we randomly resampled without replacement 20 individuals for  
746 each population, and we removed 7 populations with less than 5 individuals, except for the  
747 island's populations of interest in Cabo Verde and São Tomé e Príncipe, where we retained  
748 all individuals for this analysis. This dataset comprising 1347 individuals from 70 populations  
749 is referred to as Working Dataset henceforth (S4 Table).

750 Following author's recommendations, we filtered the initial set of 411,121 autosomal SNPs for  
751 low Linkage Disequilibrium using the *--indep-pairwise* function in PLINK with a 50 SNP-window  
752 moving every 10 SNPs, and 0.1  $r^2$  cutoff. We thus run ADMIXTURE on 110,499 LD-pruned  
753 autosomal SNPs from the 1347 individuals in the Working Dataset. We performed 10  
754 independent runs of ADMIXTURE for values of  $K$  ranging from 2 to 7 (Fig 3). We used PONG  
755 [67] to define ADMIXTURE "modes" with a greedy approach for similarity threshold 0.95. The  
756 alternative ADMIXTURE mode for  $K=5$  is presented in S2 Fig. We conducted an evaluation of  
757 the cross-validation error across 10 distinct runs for 15 values of  $K$  and found that  $K=4$  yielded  
758 the lowest error (S3 Fig). ADMIXTURE results for  $K>7$  are reported in S4 Fig.

## 759 **Local-ancestry inferences**

### 760 **Phasing with Shapeit4**

761 The phasing of individual genotypes was performed using the Segmented Haplotype  
762 Estimation and Imputation tool SHAPEIT4 (version 4.2.2) [68]. We estimated haplotypes for

763 each autosomal chromosome separately using the HapMap Phase 3 Build GRCh38 genetic  
764 recombination map [69]. We used the default SHAPEIT4 parameters for phasing autosomal  
765 data: minimum phasing window length of 2.5 Mb, and a total of 15 MCMC iterations, including  
766 7 burn-in, 3 pruning, and 5 main iterations.

## 767 **Chromosome painting with ChromoPainter for fineSTRUCTURE**

768 We used the inferential algorithm implemented in ChromoPainter v2 [34] to paint each São  
769 Toméan genome as a combination of fragments received from other São Toméan individuals.  
770 We performed a first run of ChromoPainter to estimate nuisance parameters on four  
771 chromosomes using 10 EM iterations, obtaining  $N_e = 792.022979835471$ , and global mutation  
772 rate  $\mu = 0.0020978990636338$ . We thus run ChromoPainter using the estimated parameters  
773 on the whole dataset. We averaged the results for all individuals by chromosome.

## 774 **Clustering São Tomé individuals with fineSTRUCTURE**

775 We ran fineSTRUCTURE using the co-ancestry matrix based on the number of shared chunks  
776 between each pair of individuals previously computed with ChromoPainter. fineSTRUCTURE  
777 employs a Markov Chain Monte Carlo (MCMC) sampling approach to infer the population  
778 structure. During this process, the algorithm iteratively explores the space of possible  
779 population structures. First, 100,000 burn-in steps were performed to allow the algorithm to  
780 reach a stable state. Subsequently, 100,000 further iterations were performed, retaining  
781 samples every 10,000th iteration. Following the MCMC sampling, a tree representing the  
782 inferred relationships among individuals was constructed. The best state observed during the  
783 MCMC sampling, reflecting the optimal population structure, served as the initial state for tree  
784 inference. This step provided a hierarchical representation of the population structure, offering  
785 insights into the genetic relationships among individuals within the island of São Tomé. Finally,  
786 fineSTRUCTURE classified the 96 São Toméan individuals into 17 clusters. In order to  
787 increase the interpretability of subsequent analysis, and based on the haplotype sharing  
788 patterns of the co-ancestry matrices (Fig 4C-D), we reduced the number of identified groups

789 to five by cutting the dendrogram at height 4. For a more detailed representation of the  
790 fineSTRUCTURE dendrogram, see S5 Fig. The PCA in Fig 4B has been calculated on the co-  
791 ancestry matrix with the *pcares* function in R. The third and fourth Principal Components are  
792 reported in S6 Fig. The heatmaps in Fig 4C and Fig 4D are produced using the function  
793 provided by the authors in the *FinestructureDendrogram.R* script, and are capped at 450 and  
794 250 cM, respectively, for visualization purposes. For a visual representation of how  
795 consistently pairs of individuals are grouped together across different iterations of the MCMC  
796 process, refer to the pairwise coincidence matrix in S7 Fig.

## 797 **Chromosome painting with ChromoPainter for SOURCEFIND and** 798 **fastGLOBETROTTER**

799 We used the phased haplotype information to reconstruct the chromosomes of each individual  
800 from Africa, Europe and the Americas in the Working Dataset as a series of genomic segments  
801 inherited from a set of Donor individuals using ChromoPainter v2 [34]. As with the previous  
802 ChromoPainter run on the São Toméan sample, we first estimated nuisance parameters using  
803 10 E-M iterations, this time on a subset of the Working Dataset, following the author's  
804 recommendations. We randomly sampled 3 individuals per population thus subsampling  
805 1/10th of the entire dataset, and we selected 4 chromosomes: 1, 7, 14 and 21. We averaged  
806 the estimated values across chromosomes, weighted by chromosome size, over the 10  
807 replicate analyses. Finally, we used the a posteriori estimated nuisance parameters to run the  
808 ChromoPainter algorithm on the entire Working Dataset. We run ChromoPainter on 411,121  
809 autosomal SNPs from 1347 individuals of the Working Dataset to prepare input files for three  
810 distinct SOURCEFIND analyses.

811 For the first SOURCEFIND analysis, we run ChromoPainter setting each individual as both  
812 Donor and Recipient, except for the admixed individuals of interest Cabo Verde CV, São Tomé  
813 e Príncipe STP, African-Barbadians ACB, and African-Americans ASW, which are set as  
814 Recipient only. Importantly, each Donor population consists of a maximum of 20 individuals in

815 the Working Dataset. We obtained the averaged values of “recombination rate scaling  
816 constant”  $N_e = 228.3939$ , and “per site mutation rate”  $M = 0.00076$ . Finally, we combined  
817 painted chromosomes for each individual in each of CV, STP, ACB, and ASW populations,  
818 separately.

819 For the second and third SOURCEFIND analysis, and for the fastGLOBETROTTER analysis,  
820 we run ChromoPainter setting all individuals as both Donor and Recipient, except for the  
821 admixed individuals of interest STP, ACB, and ASW, which are set as Recipient only.  
822 Importantly, the nine CV populations are set as both Donor and Recipient. We obtained the  
823 averaged values of “recombination rate scaling constant”  $N_e = 193.2902$ , and “per site  
824 mutation rate”  $M = 0.00040$ . Finally, we combined painted chromosomes for each individual  
825 in each of STP, ACB, and ASW populations, separately.

826 Finally, we run ChromoPainter setting all individuals as both Donor and Recipient to have a  
827 symmetric matrix for PCA computation (S9 Fig). PCA was computed using the *eigen* function  
828 in R on the normalized matrix of chunks counts considering all individuals in the Working  
829 Dataset.

## 830 **Estimating possible sources for the admixed populations using** 831 **SOURCEFIND**

832 We applied the model-based method SOURCEFIND [35] to estimate the shared haplotypic  
833 ancestry among populations based on chromosome painting. We modeled the copying vector  
834 of each “Target” admixed individual, obtained with the previous ChromoPainter analysis, as a  
835 weighted mixture of copying vectors from a set of “Surrogate” individuals. The Target and  
836 Surrogate individuals copy from the same set of Donors in the ChromoPainter analysis, in  
837 particular all the populations included in the Working Dataset except for the Target populations  
838 of interest. Since the set of Donors does not contain any Target population, this is a “regional”  
839 analysis according to the definition given in [43].

840 We run three separate analyses to model different hypotheses of admixture. In the first  
841 analysis (Fig 5A), we used CV, STP, ACB, and ASW populations as Targets, and all other  
842 populations in the dataset as Surrogates. In this case, the sets of Donor and Surrogate  
843 populations coincides. We aggregated results obtained for all individuals in the CV, STP, ACB,  
844 and ASW Target populations, separately.

845 In the second analysis (Fig 5B), we used the STP, ACB, and ASW population as Target, and  
846 all other populations in the dataset as Surrogates, including CV island populations. Indeed,  
847 we aim to estimate also this time the relative contribution of CV populations to the haplotypic  
848 ancestry of Target populations. To reduce the influence of differences in sample size among  
849 CV birth-island populations, each CV Surrogate birth-island population is composed of a  
850 random sample of 20 individuals per island, and all individuals if the island population is  
851 composed of less than 20 individuals. We aggregated results obtained for all individuals in the  
852 STP, ACB, and ASW Target populations, separately.

853 In the third analysis (Fig 5C), we used the five São Toméan genetic clusters as separate  
854 Targets, and we used the same set of Surrogate populations as the second analysis. Indeed,  
855 we aim to estimate the contribution of these Surrogate populations to the haplotypic ancestry  
856 of each sub-population sample in São Tomé. We aggregated results obtained for all  
857 individuals in the C1, C2, C3, C4, and C5 Target populations, separately.

858 We conducted 20 independent SOURCEFIND runs for each analysis. We did not allow for  
859 “selfcopying” (self.copy.ind), meaning that each Target individual is modelled as a mixture of  
860 Surrogate individuals only. We allowed for a maximum of 8 Surrogates (num.surrogates), with  
861 4 expected number of Surrogates (exp.num.surrogates), for each MCMC iteration, and we  
862 divided each Target individuals genome in 100 slots (num.slots) with possibly different  
863 ancestry. In other words, each of these 100 slots can be claimed by any one of the maximum  
864 8 Surrogates selected in a given MCMC iteration. We considered 400,000 MCMC iterations  
865 (num.iterations), we discarded the first 100,000 as “burn-in” (num.burnin), and we sampled an  
866 MCMC iteration every 10,000 (num.thin). Therefore, the final results consists of 30 MCMC

867 samples following the formula  $M = (\text{num.iterations} - \text{num.burnin}) / \text{num.thin}$ . For each Target  
868 population, we retained as the final result the inferred contributions to the mixture model of the  
869 MCMC iteration with the highest posterior probability over 20 runs.

870 We created categories of São Toméans based solely on genetic measures. Previous studies  
871 have also used genetic criteria to cluster individuals of the reference panel into genetically-  
872 resembling groups for use as proxies of source populations in admixture models [3,35].  
873 Instead, in our reference panel, we maintained the categories defined by the study that first  
874 described each population sample, as we believe that this approach facilitates the  
875 anthropological interpretation of genetic results. Nevertheless, we faced challenges in  
876 distinguishing between genetic contributions from populations that share much haplotypic  
877 ancestry (S8 Fig).

## 878 **Dating admixture events with fastGLOBETROTTER**

879 We used fastGLOBETROTTER to infer admixture using as Surrogates the populations  
880 imputed by SOURCEFIND for each Target São Toméan cluster, and dated admixture events  
881 based on the LD decay patterns among haplotypes matching any pair of Surrogates. The  
882 fastGLOBETROTTER algorithm also checks if the data fits a single rate of decay, which  
883 suggests a single date of admixture, or multiple rates of decay, which suggests multiple dates  
884 of admixture.

885 The same co-ancestry matrix used for the third SOURCEFIND analysis was employed, in  
886 which each individual Donor and Target genome is allowed to copy from each other. This co-  
887 ancestry matrix is utilized by fastGLOBETROTTER to infer sources, in a manner analogous  
888 to SOURCEFIND. Additionally, a separate file was prepared in which each Target genome is  
889 copying from each individual genome in the Donor population samples. This copying vector  
890 file is employed by fastGLOBETROTTER to date admixture events. I used as Donors all the  
891 populations of the reference panel, including the Cabo Verdean populations. I used as  
892 Surrogates the populations imputed by SOURCEFIND for each São Toméan cluster.



893 Following author's recommendations, I first run fastGLOBETROTTER separately for each  
894 group setting "null.ind: 1" in the parameter file, inferring admixture proportions, dates, and  
895 sources over 5 iterations, and performing 100 bootstrap re-samples to generate confidence  
896 intervals around the estimated dates. In all bootstrap resamples for all Target São Toméan  
897 cluster, no date estimate was  $\leq 1$ , nor  $\geq 400$ . Therefore, the p-value for  
898 evidence of "any detectable admixture" is 0 for each Target São Toméan cluster. I thus  
899 presented in the Results the output of a second fastGLOBETROTTER run, setting "null.ind:  
900 0". The results of the first and second run were consistent in terms of best guess of admixture  
901 dating.

902 The "null individual analysis" mitigates signals in the LD decay curve due to strong genetic  
903 drift. In the "null individual" step described by [43], GLOBETROTTER constructs a "null"  
904 coancestry curve, representing the likelihood that two DNA segments, separated by a certain  
905 genetic distance and originating from different Target individuals, match a specific pair of  
906 Surrogate populations. This process aims to eliminate linkage disequilibrium (LD) decay  
907 signals in the coancestry curves that are not due to admixture, thus improving the accuracy of  
908 date estimates for admixture events. GLOBETROTTER then adjusts the coancestry curve  
909 using the "null" coancestry curve before estimating admixture dates and proportions. This step  
910 is implemented in both GLOBETROTTER and fastGLOBETROTTER. However,  
911 fastGLOBETROTTER is better able to deal with atypically long chunks inferred by  
912 ChromoPainter in drifted groups. Specifically, fastGLOBETROTTER detects if the left end of  
913 the coancestry curve is affected by long chunks, and thus removes this part of the curve prior  
914 to model fitting. Despite these corrections, the authors advise caution, as strong genetic drift  
915 can still influence the results.

916 According to thresholds set by the authors, all São Toméan clusters better fit one pulse of  
917 admixture, apart from the C5 cluster which fit multiple pulses. In particular, for the C5 cluster,  
918 the additional goodness-of-fit ( $R^2$ ) explained by adding a second date versus assuming only  
919 a single date of admixture, taking the maximum value across all inferred coancestry curves,



920 is 0.44, which is slightly higher than the threshold of 0.35 suggested by the authors. However,  
921 the authors further suggest to visually inspect the coancestry curves, where the x-axis gives  
922 genetic distance in cM, and the y-axis gives the probability of copying from a pair of Surrogate  
923 populations at a pair of DNA segments separated by a certain cM distance. For all pairs of  
924 Surrogate populations of the C5 cluster, there is an increased probability to copy DNA  
925 segments separated by small cM distances from any pair of Surrogates. However, the 1-date  
926 curve fit the data as well as the 2-date curve starting from 2 cM distances. We thus decided  
927 to retain the 1-date results for C5, since the 2-date results account for a pulse of admixture  
928 between Cabo Verdean populations 197 generations ago, which is highly improbable since  
929 Cabo Verdean populations have been founded in the late 15<sup>th</sup> century as per historical records  
930 [51].

931 To estimate confidence intervals for the date of admixture, we performed 100 bootstrap  
932 resampling, and we calculated the quantiles of this distribution (2.5 and 97.5 percentiles) to  
933 determine the lower and upper bounds of the confidence interval, respectively (Fig 7C).

## 934 **Recent shared ancestry tracts**

### 935 **Long identical by descent (IBD) tracts**

936 We used hapIBD to generate shared identity-by-descent (IBD) segments from the phased  
937 data of 96 São Toméan and 225 Cabo Verdean unrelated individuals. We used the same  
938 genetic distance map used for phasing, and we employed the seed-and-extend algorithm  
939 implemented in hap-IBD with default parameters. Initially, seed segments are identified as  
940 identity-by-state (IBS) segments longer than 2 cM. While IBS segments indicate that two  
941 individuals share a genetic segment, IBD segments specifically represent shared ancestry,  
942 where the shared segment is inherited from a common ancestor. These segments are  
943 extended iteratively, considering another long IBS segment for the same haplotype pair,  
944 separated by a short non-IBS gap. Default parameters included a maximum non-IBS gap of

945 1,000 base pairs and a minimum extension length of 1 cM. By allowing short non-IBS gaps,  
946 the method accounts for discordant alleles that may result from genotyping errors.

947 We found a total of 259,624 IBD segments shared between the 323 unrelated individuals from  
948 São Tomé and Príncipe. We filtered for IBD segments larger than 18 cM, taking into account  
949 the last 8 generations of recombination according to previous calculations [2]. By summing  
950 the length of all shared IBD segments greater than 18 cM, we calculated the cumulative IBD  
951 sharing between each pair of individuals. We thus summed the cumulative length of IBD  
952 segments shared between individuals within and between each population, specifically the 5  
953 genetic clusters in São Tomé and the 7 populations for islands of birth in Cabo Verde. We  
954 plotted the resulting matrix as a heat map, capped at 10,000 cM (S10 Fig). In Fig 6A, we built  
955 a network based on this matrix using the “networkx” package in python.

## 956 **Runs of Hemizyosity (ROH)**

957 We called runs of homozygosity (ROH) with GARLIC [37], considering the five genetic clusters  
958 in São Tomé as separate populations, as well as 59 individuals from the island of Santiago, in  
959 Cabo Verde, 20 Gambian GWD individuals, 20 Iberian IBS individuals, 11 Kongo individuals,  
960 and 20 Yoruba YRI individuals. For each population separately, we ran GARLIC using the  
961 weighted logarithm of the odds (wLOD) [70], with a genotyping-error rate of 0.001, and using  
962 the same genetic distance map used for phasing, window sizes ranging from 30 to 90 SNPs  
963 in increments of 10 SNPs, 100 resampling to estimate allele frequencies, and all other GARLIC  
964 parameters set to default values.

965 GARLIC calculates the ROH using window sizes for ROH detection and length boundaries for  
966 ROH class detection separately for each population, depending on the specific allele  
967 frequencies and distributions of ROH lengths, respectively. We calculated the sum of ROH for  
968 each class length, for each individual (Fig 6B). We also calculated the sum of the length of all  
969 medium and long ROH, and the number of medium and long ROH, for each individual (Fig  
970 6C).

## 971 **Acknowledgements**

972 The authors would like to thank all the São Toméan participants who contributed to this study.  
973 We also thank the "Paléogénomique et génétique moléculaire" (P2GM) platform at the  
974 Muséum National d'Histoire Naturelle, Musée de l'Homme, for their assistance in handling  
975 biological samples and generating genetic data, and the BIOMICS platform at the Institut  
976 Pasteur for carrying out the genotyping analyses.

## 977 **Funding**

978 This project was partially funded by the French "Agence Nationale pour la Recherche" (ANR)  
979 under grant ANR METHIS 15-CE32-0009-1. MC was supported by the Fundação para a  
980 Ciência e a Tecnologia (FCT), Portugal. ZAS was supported by the National Institute of  
981 General Medical Sciences under award number R35GM146926.

## 982 **Ethic statement**

983 Research sampling protocols followed the Declaration of Helsinki guidelines and informed  
984 consent was obtained from all subjects involved in the sampling in São Tomé and Cabo Verde.  
985 The study of the São Toméan sample was undertaken with the support and permission of the  
986 Provincial Government of the Ministry of Health of the Democratic Republic of São Tomé and  
987 Príncipe, and the Provincial Government of Príncipe.  
988 For the Cabo Verdean sample, research and ethics authorizations were provided by the  
989 Ministério da Saúde de Cabo Verde (228/DGS/11), and the French ethics committees and  
990 CNIL (Declaration n°1972648).

## 991 **Data availability**

992 The novel genetic data presented here can be accessed via the European Genome-phenome  
993 Archive (EGA) database accession number #XX (pending) upon request to the corresponding  
994 Data Access Committee. The dataset can be shared provided that future envisioned studies  
995 comply with the informed consents provided by the participants, and in agreement with  
996 institutional ethics committee's recommendations applying to this data.

## 997 **References**

- 998 [1] Eltis D, Richardson D. Atlas of the Transatlantic Slave Trade. Yale University Press; 2010.
- 999 [2] Baharian S, Barakatt M, Gignoux CR, Shringarpure S, Errington J, Blot WJ, et al. The  
1000 Great Migration and African-American genomic diversity. PLOS Genetics  
1001 2016;12:e1006059. <https://doi.org/10.1371/journal.pgen.1006059>.
- 1002 [3] Ongaro L, Scliar MO, Flores R, Raveane A, Marnetto D, Sarno S, et al. The Genomic  
1003 Impact of European Colonization of the Americas. Current Biology 2019;29:3974-  
1004 3986.e4. <https://doi.org/10.1016/j.cub.2019.09.076>.
- 1005 [4] Micheletti SJ, Bryc K, Ancona Esselmann SG, Freyman WA, Moreno ME, Poznik GD, et  
1006 al. Genetic Consequences of the Transatlantic Slave Trade in the Americas. The  
1007 American Journal of Human Genetics 2020;107:265–77.  
1008 <https://doi.org/10.1016/j.ajhg.2020.06.012>.
- 1009 [5] Almeida J, Fehn A-M, Ferreira M, Machado T, Hagemeyer T, Rocha J, et al. The Genes  
1010 of Freedom: Genome-Wide Insights into Marronage, Admixture and Ethnogenesis in the  
1011 Gulf of Guinea. Genes 2021;12:833. <https://doi.org/10.3390/genes12060833>.
- 1012 [6] Laurent R, Szpiech ZA, da Costa SS, Thouzeau V, Fortes-Lima CA, Dessarps-Freichay  
1013 F, et al. A genetic and linguistic analysis of the admixture histories of the islands of Cabo  
1014 Verde. eLife 2023;12:e79827. <https://doi.org/10.7554/eLife.79827>.

- 1015 [7] Gravel S. Population Genetics Models of Local Ancestry. *Genetics* 2012;191:607–19.  
1016 <https://doi.org/10.1534/genetics.112.139808>.
- 1017 [8] Fortes-Lima C, Verdu P. Anthropological genetics perspectives on the transatlantic slave  
1018 trade. *Human Molecular Genetics* 2020;30. <https://doi.org/10.1093/hmg/ddaa271>.
- 1019 [9] Mooney JA, Agranat-Tamir L, Pritchard JK, Rosenberg NA. On the number of genealogical  
1020 ancestors tracing to the source groups of an admixed population. *Genetics*  
1021 2023;224:iyad079. <https://doi.org/10.1093/genetics/iyad079>.
- 1022 [10] Fortes-Lima C, Laurent R, Thouzeau V, Toupance B, Verdu P. Complex genetic admixture  
1023 histories reconstructed with Approximate Bayesian Computation. *Molecular Ecology*  
1024 *Resources* 2021;21:1098–117. <https://doi.org/10.1111/1755-0998.13325>.
- 1025 [11] Curtin PD. *The Rise and Fall of the Plantation Complex: Essays in Atlantic History*.  
1026 Cambridge: Cambridge University Press; 1990.
- 1027 [12] Caldeira AM. Learning the Ropes in the Tropics: Slavery and the Plantation System on  
1028 the Island of São Tomé. *African Economic History* 2011;39:35–71.
- 1029 [13] Caldeira AM. The Island Trade Route of São Tomé in the 16th Century: Ships, Products,  
1030 Capitals. *Rivista Istituto Storia Europa Mediterranea RiMe* 2021:55–76.  
1031 <https://doi.org/10.7410/1507>.
- 1032 [14] Caldeira AM. *Escravos e traficantes no império português: o comércio negreiro português*  
1033 *no Atlântico durante os séculos XV a XIX*. Lisboa: A Esfera dos Livros; 2013.
- 1034 [15] Carreira A. *Migrações nas ilhas de Cabo Verde*. Lisboa: Universidade Nova; 1976.
- 1035 [16] Carreira A. *Cabo Verde: formação e extinção de uma sociedade escravocrata (1460-*  
1036 *1878)*. Bissau: Centro de Estudos da Guiné Portuguesa; 1972.
- 1037 [17] Seibert G. Crioulização em Cabo Verde e São Tomé e Príncipe: divergências históricas e  
1038 identitárias. *Afro-Ásia* 2014;49:41–70.

- 1039 [18]Seibert G. A questão da origem dos Angolares de São Tomé. Instituto Superior de  
1040 Economia e Gestão 1998;5.
- 1041 [19]Lorenzino G. The Angolar Creole Portuguese of São Tomé: its grammar and  
1042 sociolinguistic history. Munchen: LINCOM Europa; 1998.
- 1043 [20]Hagemeijer T. The Gulf of Guinea Creoles: Genetic and typological relations. Journal of  
1044 Pidgin and Creole Languages 2011;26:111–54.
- 1045 [21]Seibert G. Colonialismo em São Tomé e Príncipe: hierarquização, classificação e  
1046 segregação da vida social. Anuário Antropológico 2015;40:99–120.  
1047 <https://doi.org/10.4000/aa.1411>.
- 1048 [22]Dados Distritais e Nacional Recenseamento 2012. Instituto Nacional de Estatística 2012.  
1049 [https://www.ine.st/index.php/publicacao/documentos/category/71-dados-distritais-e-](https://www.ine.st/index.php/publicacao/documentos/category/71-dados-distritais-e-nacional-recenseamento-2012)  
1050 [nacional-recenseamento-2012](https://www.ine.st/index.php/publicacao/documentos/category/71-dados-distritais-e-nacional-recenseamento-2012) (accessed May 20, 2024).
- 1051 [23]Tomas G, Seco L, Seixas S, Faustino P, Lavinha J, Jorge Rocha, et al. The peopling of  
1052 São Tomé (Gulf of Guinea): Origins of slave settlers and admixture with the Portuguese.  
1053 Human Biology 2002;74:397–411. <https://doi.org/10.1353/hub.2002.0036>.
- 1054 [24]Coelho M, Coia CAV, Luiselli D, Useli A, Hagemeijer T, Amorim A, et al. Human  
1055 Microevolution and the Atlantic Slave Trade: A Case Study from São Tomé. Current  
1056 Anthropology 2008;49:134–43. <https://doi.org/10.1086/524762>.
- 1057 [25]Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global  
1058 reference for human genetic variation. Nature 2015;526:68–74.  
1059 <https://doi.org/10.1038/nature15393>.
- 1060 [26]Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K,  
1061 et al. The African Genome Variation Project shapes medical genetics in Africa. Nature  
1062 2015;517:327–32. <https://doi.org/10.1038/nature13997>.

- 1063 [27]Patin E, Lopez M, Grollemund R, Verdu P, Harmant C, Quach H, et al. Dispersals and  
1064 genetic adaptation of Bantu-speaking populations in Africa and North America. *Science*  
1065 2017;356:543–6. <https://doi.org/10.1126/science.aal1988>.
- 1066 [28]Semo A, Gayà-Vidal M, Fortes-Lima C, Alard B, Oliveira S, Almeida J, et al. Along the  
1067 Indian Ocean Coast: Genomic Variation in Mozambique Provides New Insights into the  
1068 Bantu Expansion. *Molecular Biology and Evolution* 2020;37:406–16.  
1069 <https://doi.org/10.1093/molbev/msz224>.
- 1070 [29]Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL. High  
1071 resolution of human evolutionary trees with polymorphic microsatellites. *Nature*  
1072 1994;368:455–7. <https://doi.org/10.1038/368455a0>.
- 1073 [30]Beleza S, Campos J, Lopes JR, Araújo II, Almada AH, Correia e Silva A, et al. The  
1074 Admixture Structure and Genetic Variation of the Archipelago of Cape Verde and Its  
1075 Implications for Admixture Mapping Studies. *PLOS ONE* 2012;7.  
1076 <https://doi.org/10.1371/journal.pone.0051103>.
- 1077 [31]Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human  
1078 Demographic History Impacts Genetic Risk Prediction across Diverse Populations.  
1079 *American Journal of Human Genetics* 2017;100:635–49.  
1080 <https://doi.org/10.1016/j.ajhg.2017.03.004>.
- 1081 [32]Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in  
1082 unrelated individuals. *Genome Research* 2009;19:1655–64.  
1083 <https://doi.org/10.1101/gr.094052.109>.
- 1084 [33]Korunes KL, Soares-Souza GB, Bobrek K, Tang H, Araújo II, Goldberg A, et al. Sex-biased  
1085 admixture and assortative mating shape genetic variation and influence demographic  
1086 inference in admixed Cabo Verdeans. *G3 Genes|Genomes|Genetics* 2022;12:jkac183.  
1087 <https://doi.org/10.1093/g3journal/jkac183>.

- 1088 [34]Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of Population Structure using  
1089 Dense Haplotype Data. *PLOS Genetics* 2012;8:e1002453.  
1090 <https://doi.org/10.1371/journal.pgen.1002453>.
- 1091 [35]Chacón-Duque J-C, Adhikari K, Fuentes-Guajardo M, Mendoza-Revilla J, Acuña-Alonzo  
1092 V, Barquera R, et al. Latin Americans show wide-spread Converso ancestry and imprint  
1093 of local Native ancestry on physical appearance. *Nature Communications* 2018;9:5388.  
1094 <https://doi.org/10.1038/s41467-018-07748-z>.
- 1095 [36]Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic  
1096 patterns of homozygosity in worldwide human populations. *American Journal of Human  
1097 Genetics* 2012;91:275–92. <https://doi.org/10.1016/j.ajhg.2012.06.014>.
- 1098 [37]Szpiech ZA, Blant A, Pemberton TJ. GARLIC: Genomic Autozygosity Regions Likelihood-  
1099 based Inference and Classification. *Bioinformatics* 2017;33:2059–62.  
1100 <https://doi.org/10.1093/bioinformatics/btx102>.
- 1101 [38]Szpiech ZA, Mak ACY, White MJ, Hu D, Eng C, Burchard EG, et al. Ancestry-Dependent  
1102 Enrichment of Deleterious Homozygotes in Runs of Homozygosity. *American Journal of  
1103 Human Genetics* 2019;105:747–62. <https://doi.org/10.1016/j.ajhg.2019.08.011>.
- 1104 [39]Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: windows  
1105 into population history and trait architecture. *Nature Reviews Genetics* 2018;19:220–34.  
1106 <https://doi.org/10.1038/nrg.2017.109>.
- 1107 [40]Wangkumhang P, Greenfield M, Hellenthal G. An efficient method to identify, date, and  
1108 describe admixture events using haplotype information. *Genome Research*  
1109 2022;32:1553–64. <https://doi.org/10.1101/gr.275994.121>.
- 1110 [41]Hagemeijer T, Rocha J. Creole languages and genes: the case of São Tomé and Príncipe.  
1111 *Faits de Langues* 2019;49:167–82. <https://doi.org/10.1163/19589514-04901011>.
- 1112 [42]Gonçalves R, Hagemeijer T. O português num contexto multilingue: o caso de São Tomé  
1113 e Príncipe. *Revista Científica da UEM: Série Letras e Ciências Sociais* 2015;1.



- 1114 [43]Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, et al. A genetic atlas  
1115 of human admixture history. *Science* 2014;343:747–51.  
1116 <https://doi.org/10.1126/science.1243518>.
- 1117 [44]Seibert G. Tenreiro, Amador e os Angolares ou a reinvenção da história de São Tomé. In:  
1118 Silva MC da, Saraiva C, editors. *As Lições de Jill Dias : Antropologia, História, África e*  
1119 *Academia*, Lisboa: Etnográfica Press; 2013, p. 171–85.  
1120 <https://doi.org/10.4000/books.etnograficapress.529>.
- 1121 [45]Mooney JA, Huber CD, Service S, Sul JH, Marsden CD, Zhang Z, et al. Understanding  
1122 the Hidden Complexity of Latin American Population Isolates. *The American Journal of*  
1123 *Human Genetics* 2018;103:707–26. <https://doi.org/10.1016/j.ajhg.2018.09.013>.
- 1124 [46]César Fortes-Lima, Fortes-Lima C, Antoine Gessain, Gessain A, Andrés Ruiz-Linares,  
1125 Ruiz-Linares A, et al. Genome-wide Ancestry and Demographic History of African-  
1126 Descendant Maroon Communities from French Guiana and Suriname. *American Journal*  
1127 *of Human Genetics* 2017;101:725–36. <https://doi.org/10.1016/j.ajhg.2017.09.021>.
- 1128 [47]Pool JE, Nielsen R. Inference of Historical Changes in Migration Rate From the Lengths  
1129 of Migrant Tracts. *Genetics* 2009;181:711–9.  
1130 <https://doi.org/10.1534/genetics.108.098095>.
- 1131 [48]Nascimento A. *Poderes e Quotidiano nas Roças de S. Tomé e Príncipe. De finais de*  
1132 *oitocentos a meados de novecentos*. Lousa: Tipografia Lousanense; 2002.
- 1133 [49]Caldeira AM. *Mestiçagem, estratégias de casamento e propriedade feminina no*  
1134 *arquipélago de São Tomé e Príncipe nos séculos XVI, XVII e XVIII*. *ARQUIPÉLAGO*  
1135 *História* 2007;11–12:49–71.
- 1136 [50]Henriques I de C. *São Tomé e Príncipe: a invenção de uma sociedade*. Lisboa: Vega;  
1137 2000.
- 1138 [51]Albuquerque L de, Santos M. *História geral de Cabo Verde*. 2nd ed. Lisboa: Instituto  
1139 Nacional de Investigação Cultural; 2001.

- 1140 [52]Lucas PG. The demography of São Tomé and Príncipe (1758–1822): preliminary  
1141 approaches to an insular slave society. *Anais De Historia De Alem-Mar* 2015;XVI:51–78.
- 1142 [53]Bouchard M-E. Scaling proximity to whiteness: Racial boundary-making on São Tomé  
1143 Island. *Ethnography* 2023;24:197–126. <https://doi.org/10.1177/1466138120967373>.
- 1144 [54]Nascimento A. O sul da diáspora: cabo-verdianos em plantações de S. Tomé Príncipe e  
1145 Mocambique. Praia: Edição da Presidência da República de Cabo Verde; 2003.
- 1146 [55]Nascimento A. Exile and contract: journeys of the mozambicans to S. Tomé and Príncipe  
1147 (1940 to 1960). Centro de História da Universidade de Lisboa; 2023.  
1148 <https://doi.org/10.51427/10451/58340>.
- 1149 [56]Browning SR, Waples RK, Browning BL. Fast, accurate local ancestry inference with  
1150 FLARE. *The American Journal of Human Genetics* 2023;110:326–35.  
1151 <https://doi.org/10.1016/j.ajhg.2022.12.010>.
- 1152 [57]Moreno-Estrada A, Gravel S, Zakharia F, McCauley JL, Byrnes JK, Gignoux CR, et al.  
1153 Reconstructing the Population Genetic History of the Caribbean. *PLOS Genetics*  
1154 2013;9:e1003925. <https://doi.org/10.1371/journal.pgen.1003925>.
- 1155 [58]Fortes-Lima C, Bybjerg-Grauholm J, Marin-Padrón LC, Gomez-Cabezas EJ, Bækvad-  
1156 Hansen M, Hansen CS, et al. Exploring Cuba's population structure and demographic  
1157 history using genome-wide data. *Scientific Reports* 2018;8:11422.  
1158 <https://doi.org/10.1038/s41598-018-29851-3>.
- 1159 [59]Agranat-Tamir L, Mooney JA, Rosenberg NA. Counting the genetic ancestors from source  
1160 populations in members of an admixed population. *Genetics* 2024;226:iyae011.  
1161 <https://doi.org/10.1093/genetics/iyae011>.
- 1162 [60]Zaitlen N, Huntsman S, Hu D, Spear M, Eng C, Oh SS, et al. The Effects of Migration and  
1163 Assortative Mating on Admixture Linkage Disequilibrium. *Genetics* 2017;205:375–83.  
1164 <https://doi.org/10.1534/genetics.116.192138>.

- 1165 [61]Ongaro L, Molinaro L, Flores R, Marnetto D, Capodiferro MR, Alarcón-Riquelme ME, et  
1166 al. Evaluating the Impact of Sex-Biased Genetic Admixture in the Americas through the  
1167 Analysis of Haplotype Data. *Genes* 2021;12:1580.  
1168 <https://doi.org/10.3390/genes12101580>.
- 1169 [62]Mas-Sandoval A, Mathieson S, Fumagalli M. The genomic footprint of social stratification  
1170 in admixing American populations. *eLife* 2023;12:e84429.  
1171 <https://doi.org/10.7554/eLife.84429>.
- 1172 [63]Kim J, Edge MD, Goldberg A, Rosenberg NA. Skin deep: The decoupling of genetic  
1173 admixture levels from phenotypes that differed between source populations. *American*  
1174 *Journal of Physical Anthropology* 2021;175:406–21. <https://doi.org/10.1002/ajpa.24261>.
- 1175 [64]Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship  
1176 inference in genome-wide association studies. *Bioinformatics* 2010;26:2867–73.  
1177 <https://doi.org/10.1093/bioinformatics/btq559>.
- 1178 [65]Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation  
1179 PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 2015;4:s13742-  
1180 015-0047–8. <https://doi.org/10.1186/s13742-015-0047-8>.
- 1181 [66]R: The R Project for Statistical Computing. The R Foundation n.d. [https://www.r-](https://www.r-project.org/)  
1182 [project.org/](https://www.r-project.org/).
- 1183 [67]Behr AA, Liu KZ, Liu-Fang G, Nakka P, Ramachandran S. pong: fast analysis and  
1184 visualization of latent clusters in population genetic data. *Bioinformatics* 2016;32:2817–  
1185 23. <https://doi.org/10.1093/bioinformatics/btw327>.
- 1186 [68]Delaneau O, Zagury J-F, Robinson MR, Marchini JL, Dermitzakis ET. Accurate, scalable  
1187 and integrative haplotype estimation. *Nature Communications* 2019;10:5436.  
1188 <https://doi.org/10.1038/s41467-019-13225-y>.

1189 [69]Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating  
1190 common and rare genetic variation in diverse human populations. *Nature* 2010;467:52–  
1191 8. <https://doi.org/10.1038/nature09298>.

1192 [70]Blant A, Kwong M, Szpiech ZA, Pemberton TJ. Weighted likelihood inference of genomic  
1193 autozygosity patterns in dense genotype data. *BMC Genomics* 2017;18:928.  
1194 <https://doi.org/10.1186/s12864-017-4312-3>.

1195

## 1196 **Supporting Information**

1197 **S1 Table. Genotyping details for the São Toméan and Cabo Verdean sample batches.**

1198 Batch codes, sampling locations, genotyping years, and the Illumina manifest used for each  
1199 batch.

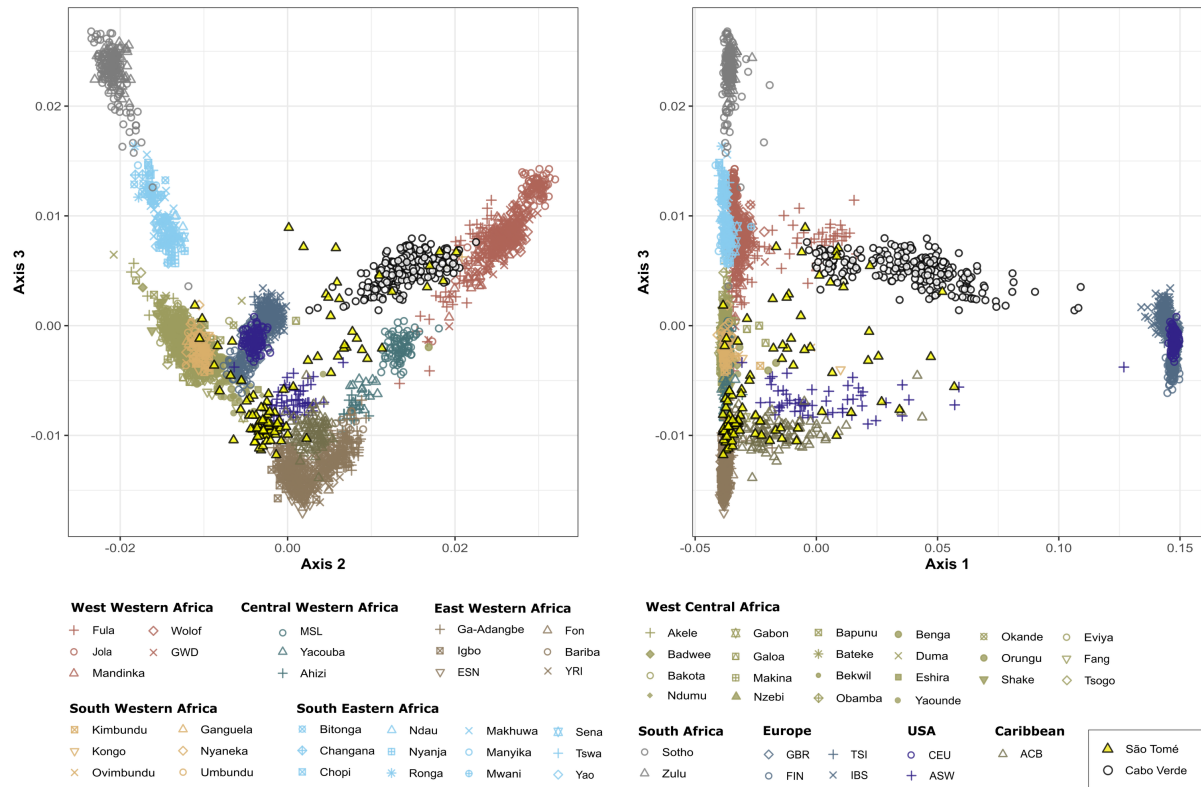
1200 **S2 Table. Overview of the Genotyping Quality Control (QC) Pipeline across four phases.**

1201 Sequential steps involved in each phase of the genotyping QC, including genotype calling  
1202 (Phase 1), batch merging (Phase 2), genetic relatedness filtering (Phase 3), and population  
1203 genetics QC (Phase 4). Specific actions, such as removing ambiguous markers, markers on  
1204 sex chromosomes, duplicates, and markers with low call rates, are noted alongside the  
1205 number of markers and samples retained after each step.

1206 **S3 Table. Population datasets included in this study.** The original publication source for  
1207 each population dataset.

1208 **S4 Table. Population datasets included in the Working Dataset.** Populations and number  
1209 of individual samples included in analyses presented in this study.

1210

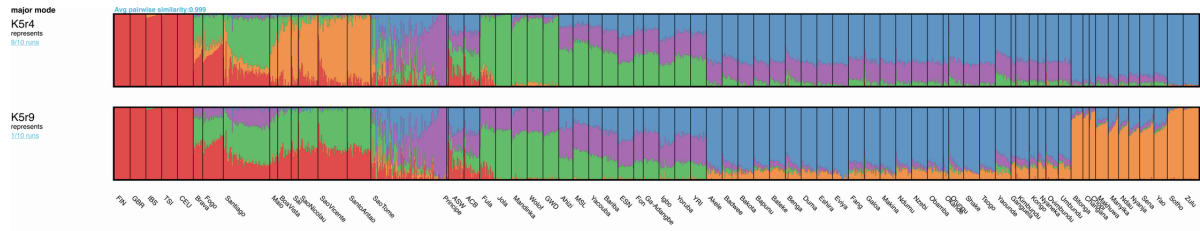


1211

1212 **S1 Fig. Multi-Dimensional Scaling (MDS) analysis.** This figure extends Fig 2B by including  
 1213 the third axis of variation in the MDS projection of pairwise allele sharing dissimilarities (ASD,  
 1214 Bowcock et al. 1994). The MDS includes São Toméans, Cabo Verdeans, and various African,  
 1215 American, and European populations, with the projection based on 3203 individuals and  
 1216 411,121 autosomal SNPs.

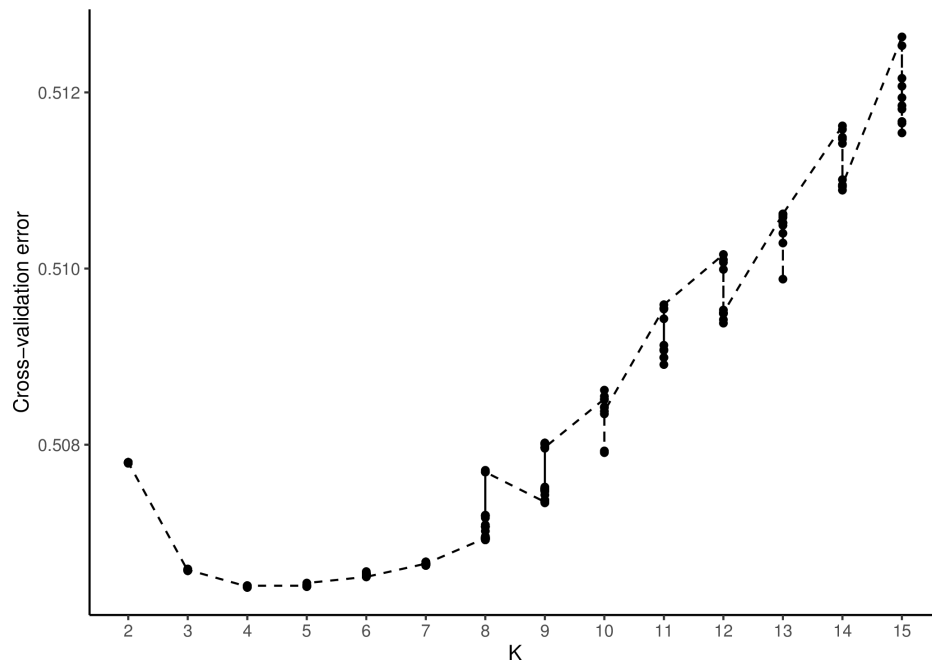
1217

1218



1219 **S2 Fig. Alternative ADMIXTURE mode for  $K=5$ .** Alternative admixture mode for  $K=5$ ,  
1220 differing from the major mode reported in Fig 3 that represents 9 out of 10 independent  
1221 ADMIXTURE runs. The average similarity between this alternative mode and the major mode  
1222 is 0.814581, as calculated with PONG.

1223



1224

1225 **S3 Fig. Cross-validation error of 10 independent ADMIXTURE runs for K from 2 to 15.**

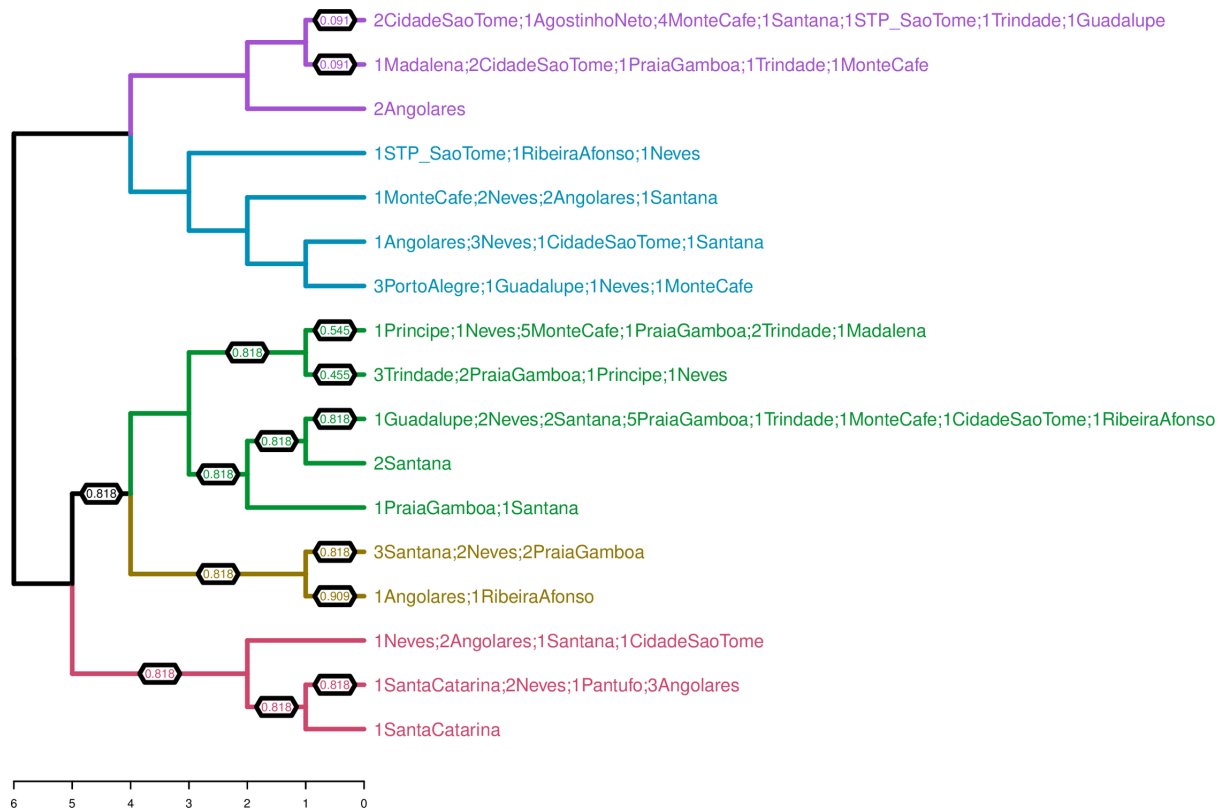
1226 The cross-validation error for 10 independent ADMIXTURE runs begins to increase starting

1227 from  $K=7$ .

1228



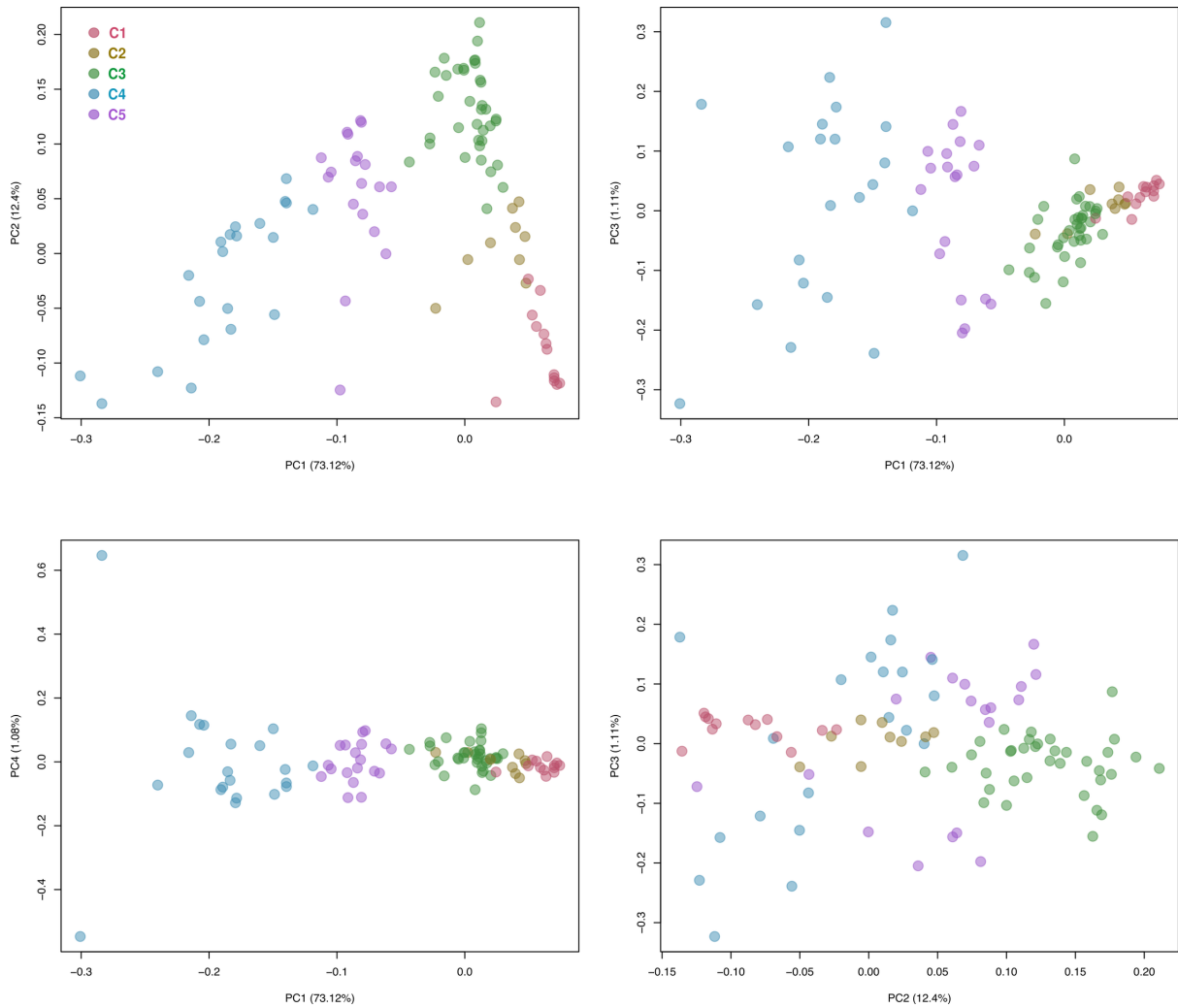




1233

1234 **S5 Fig. fineSTRUCTURE dendrogram of the São Toméan sample.** The numbers on the  
1235 edges of the dendrogram give the proportion of MCMC iterations for which each population  
1236 split is observed (only displayed when the proportion is below 1).

1237



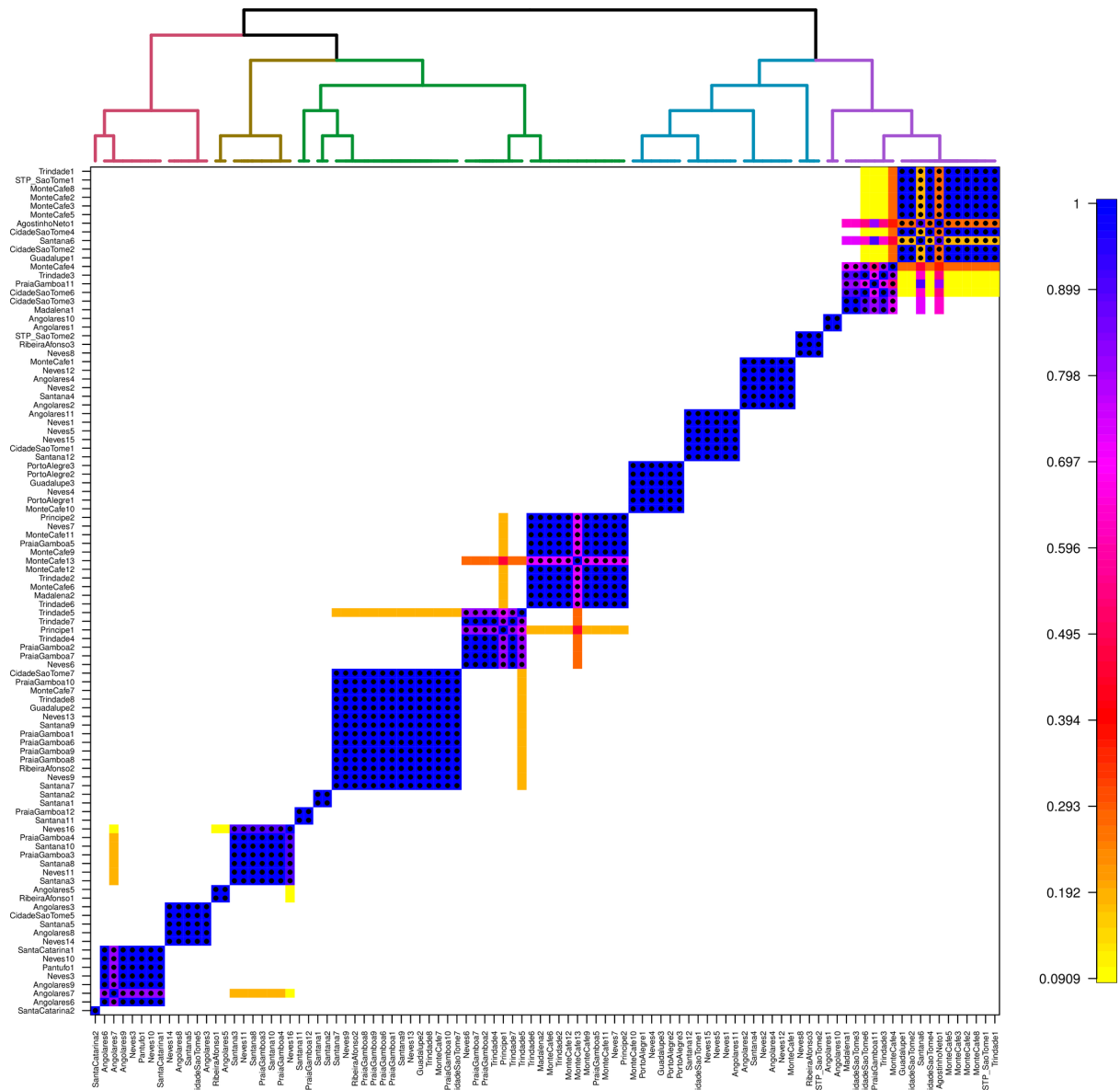
1238

1239 **S6 Fig. Principal Component Analysis (PCA) based on the co-ancestry matrix of the**

1240 **São Toméan sample.** This figure expands upon Fig 4B by including additional principal

1241 components, specifically PC3 and PC4.

1242



1243

1244 **S7 Fig. Pairwise coincidence matrix of São Toméan individual samples.** The coincidence

1245 matrix is used to summarize the results of fineSTRUCTURE's Markov Chain Monte Carlo

1246 (MCMC) clustering process. It captures how consistently pairs of individuals are grouped

1247 together across different iterations of the MCMC process. The coloring represents the average

1248 pairwise coincidence across MCMC samples. If the value is close to 1, it means that the

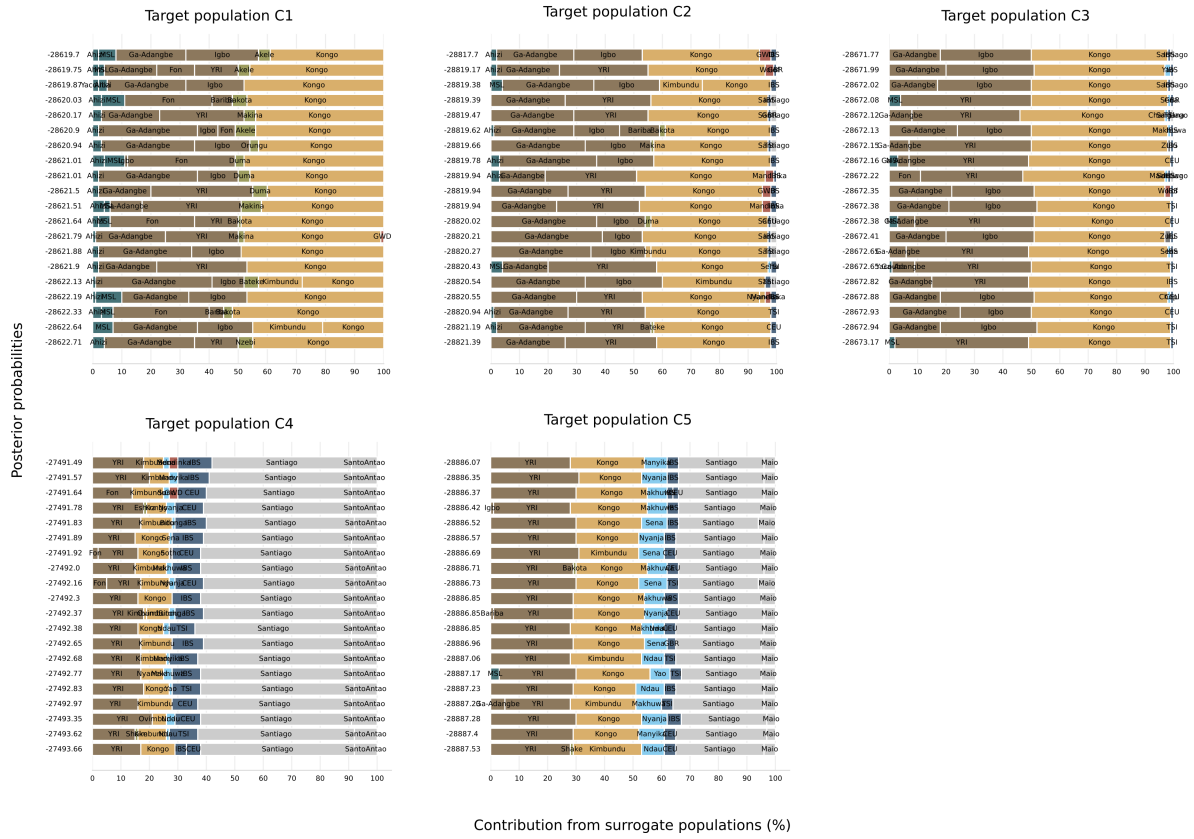
1249 corresponding pair of individuals is almost always grouped together in the same cluster,

1250 indicating strong genetic similarity.

1251

1252

MCMC iterations with the highest likelihood over 20 independent SOURCEFIND runs



1253

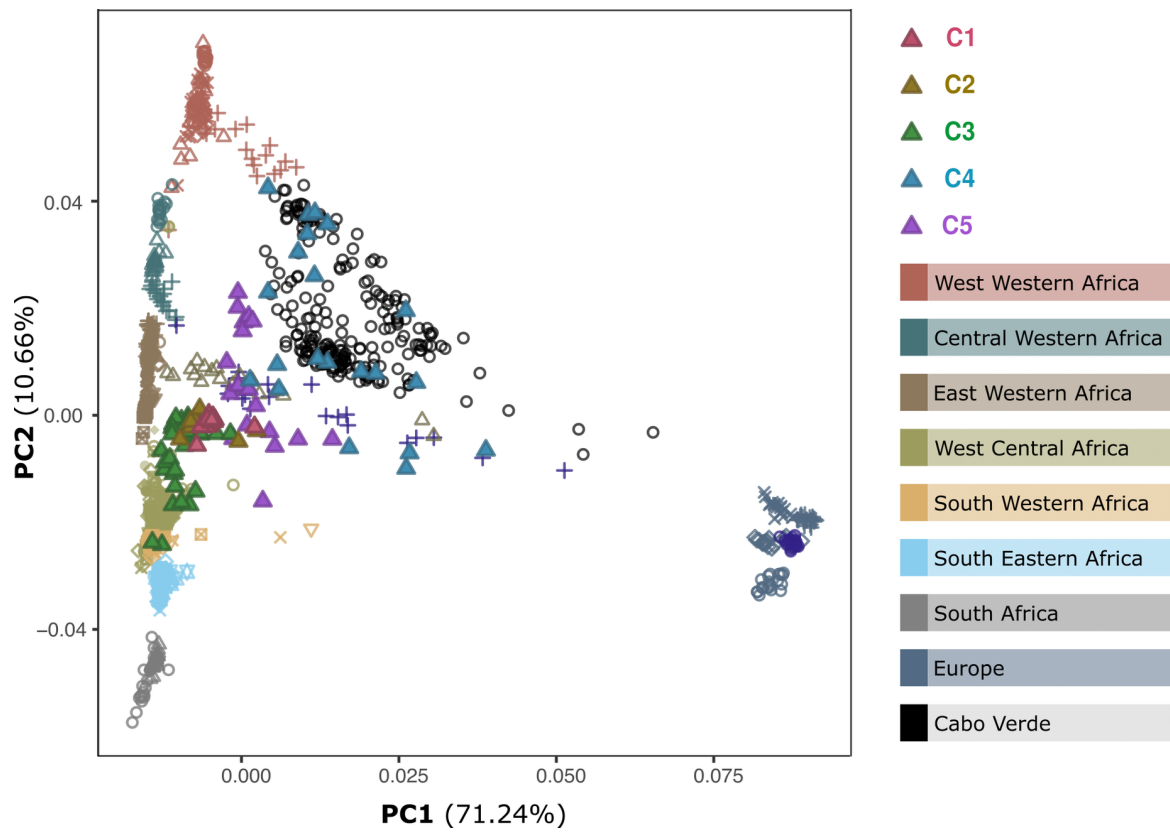
1254

1255

1256

1257

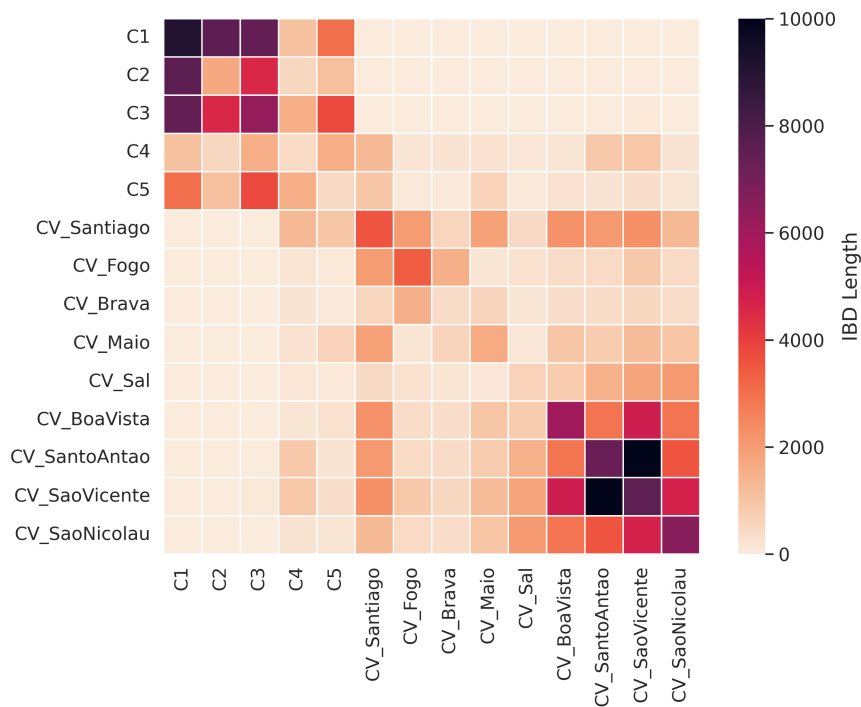
**S8 Fig. MCMC iterations with the highest likelihood over 20 independent SOURCEFIND runs.** The results are ordered by posterior probability and illustrate both the consistency and variability in the source estimates across different runs.



1258

1259 **S9 Fig. Haplotype-based PCA.** Principal Component Analysis (PCA) based on chromosome  
1260 painting with Chromopainter2 using all the 1347 individuals of the Working Dataset as both  
1261 Donors and Recipients.

1262



1263

1264 **S10 Fig. Heatmap of the cumulative length of long Identical by Descent (IBD) tracts.**

1265 Alternative representation of the results shown in Fig 6A, displaying the cumulative length of  
1266 IBD tracts longer than 18 cM, that are shared within and between samples from São Tomé  
1267 (C1-C5) and Cabo Verde.