

1189

1190

1191

1192

1193

1194

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.

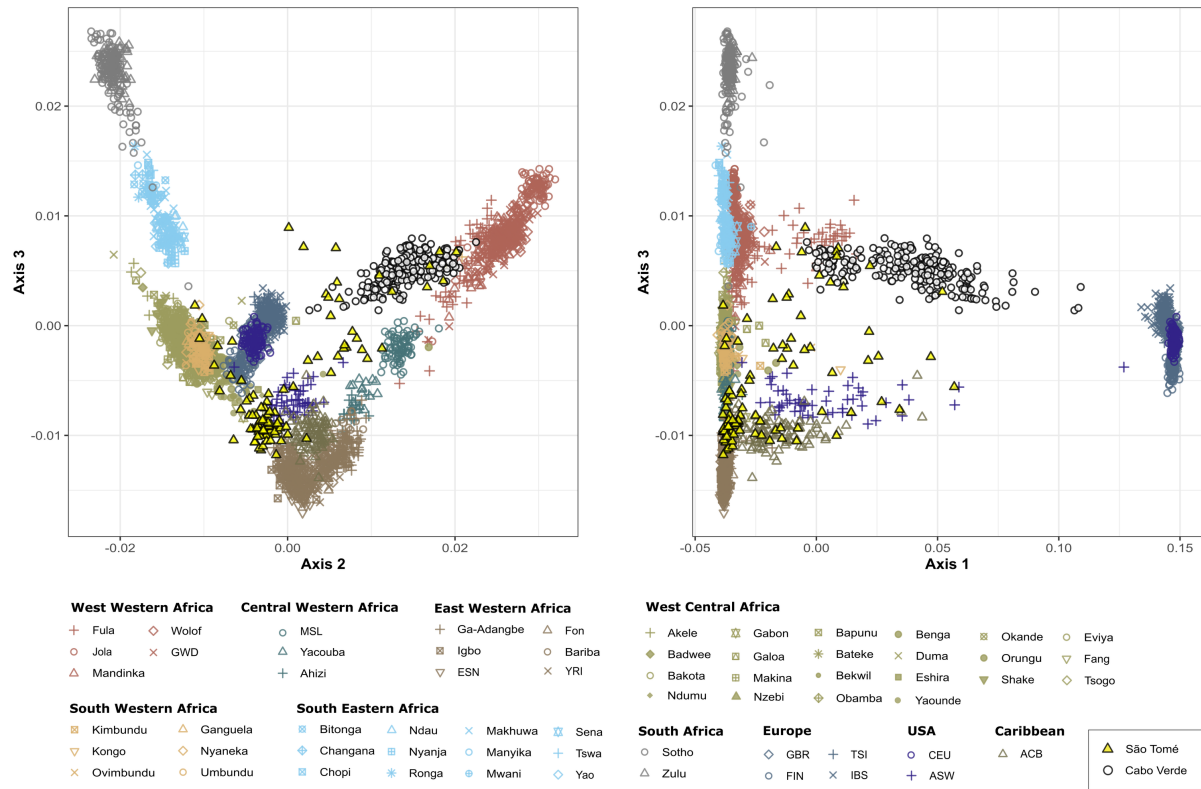
1200 **S2 Table. Overview of the Genotyping Quality Control (QC) Pipeline across four phases.**

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

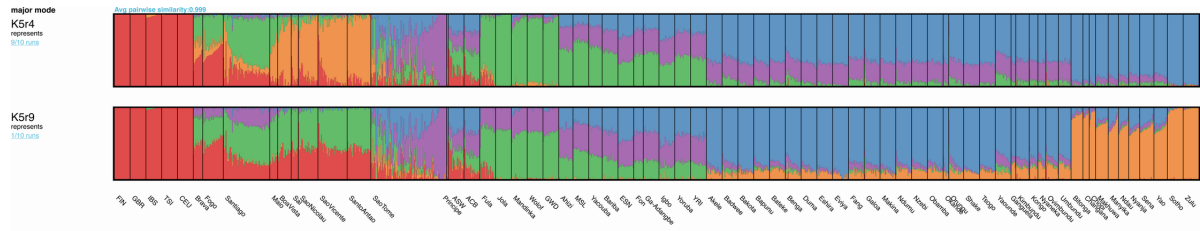


1211

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.

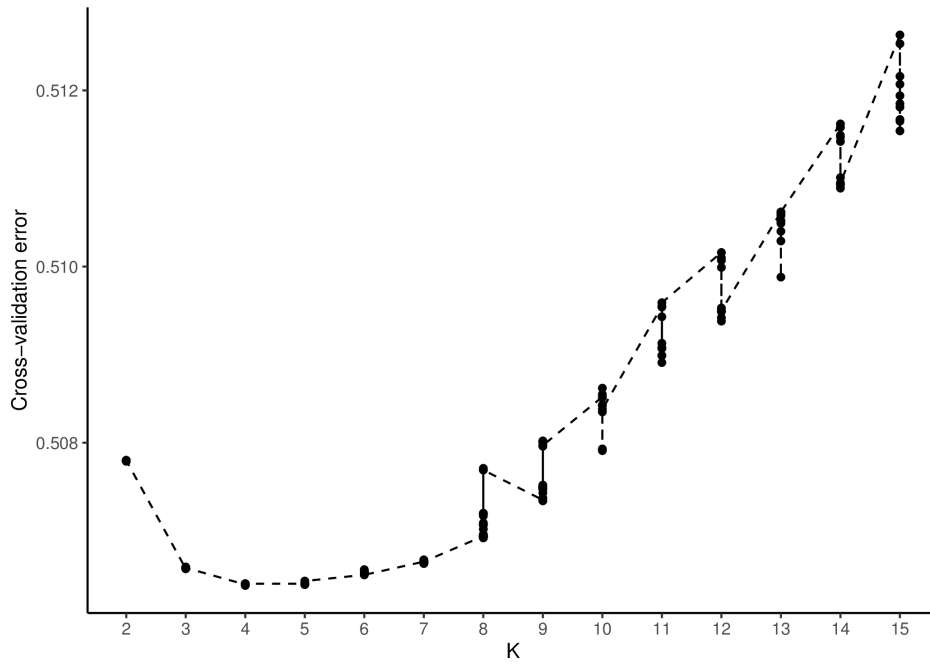
1217

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



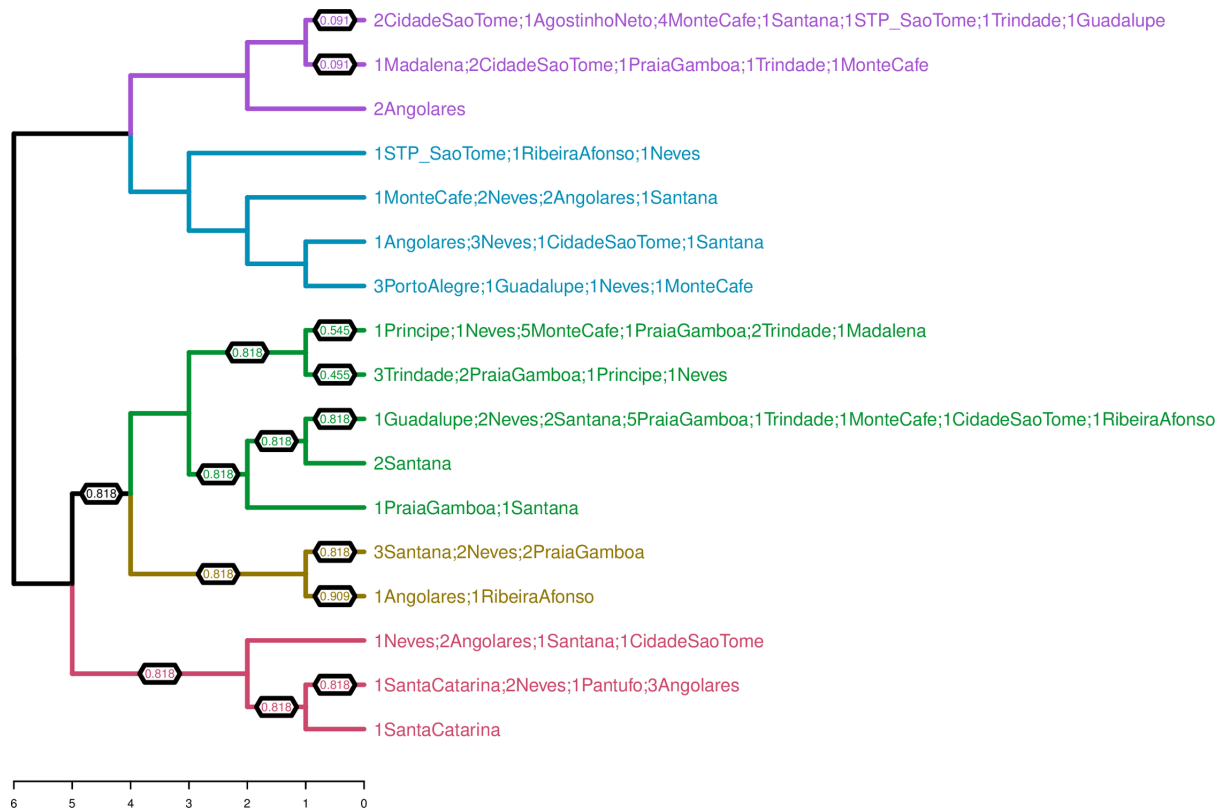
1224

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

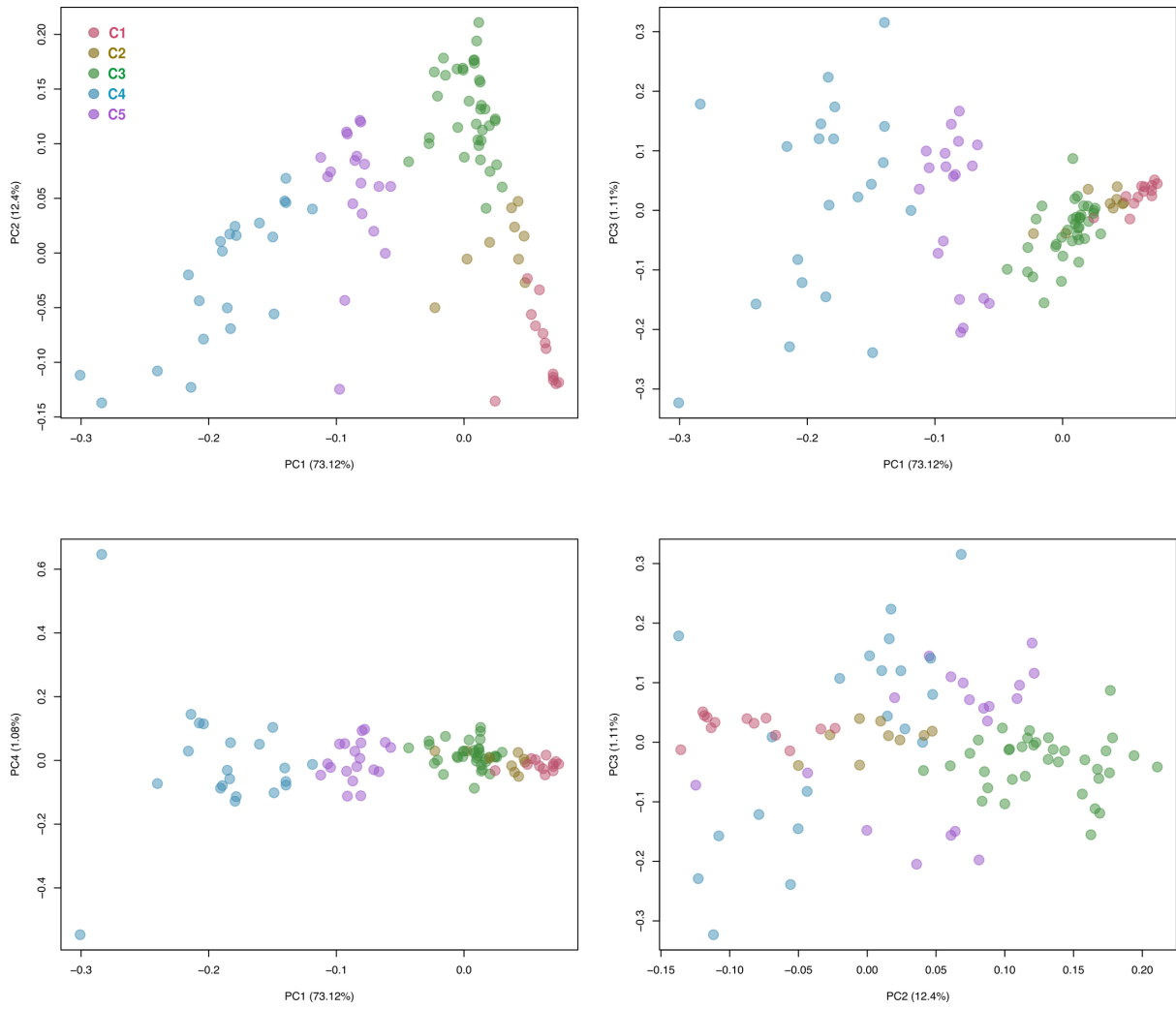
1227 from $K=7$.

1228



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



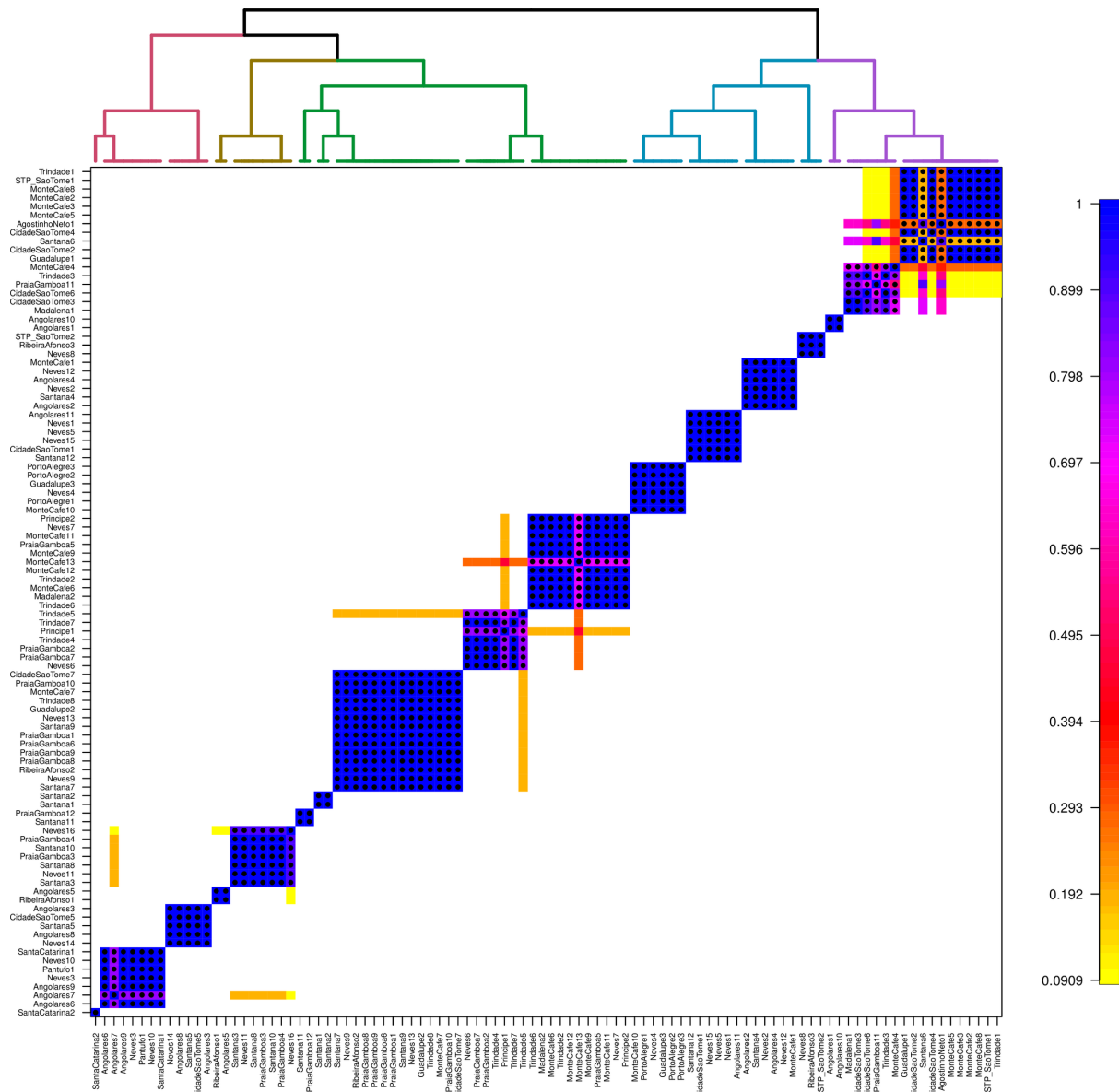
1238

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

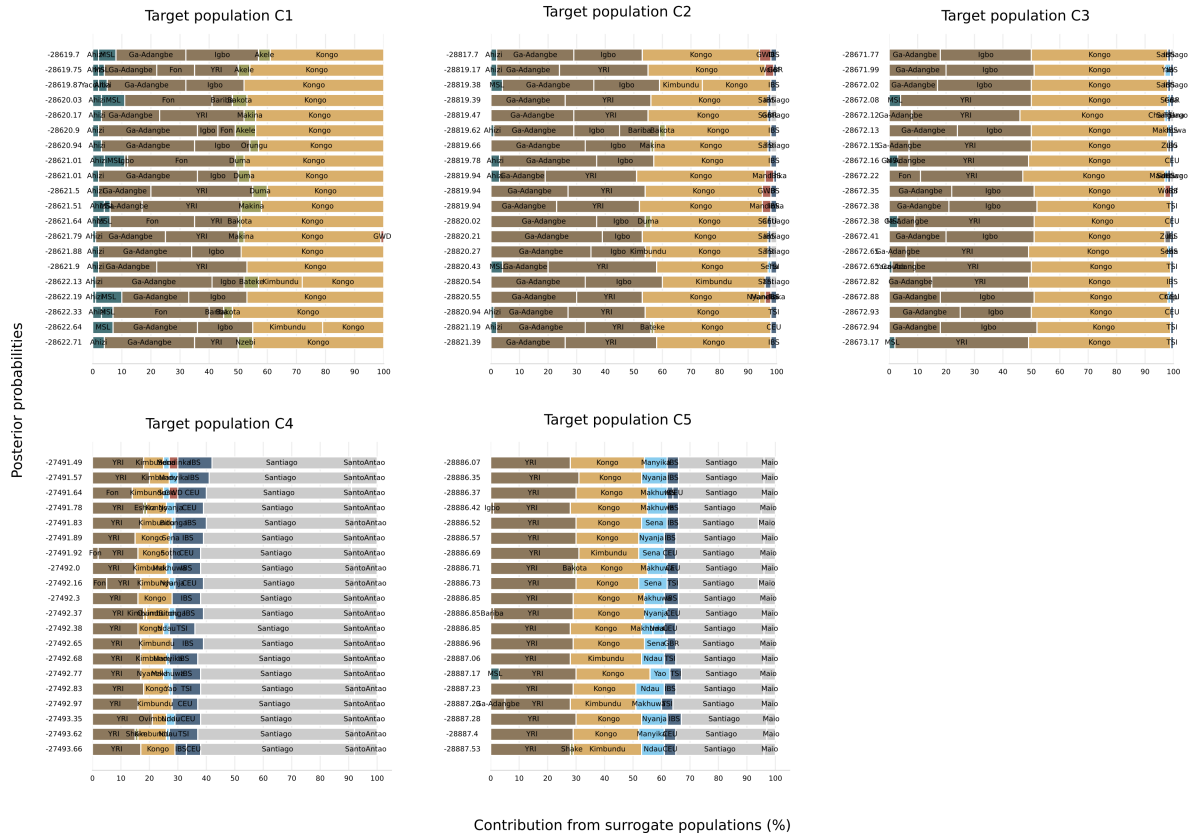
1249 corresponding pair of individuals is almost always grouped together in the same cluster,

1250 indicating strong genetic similarity.

1251

1252

MCMC iterations with the highest likelihood over 20 independent SOURCEFIND runs



1253

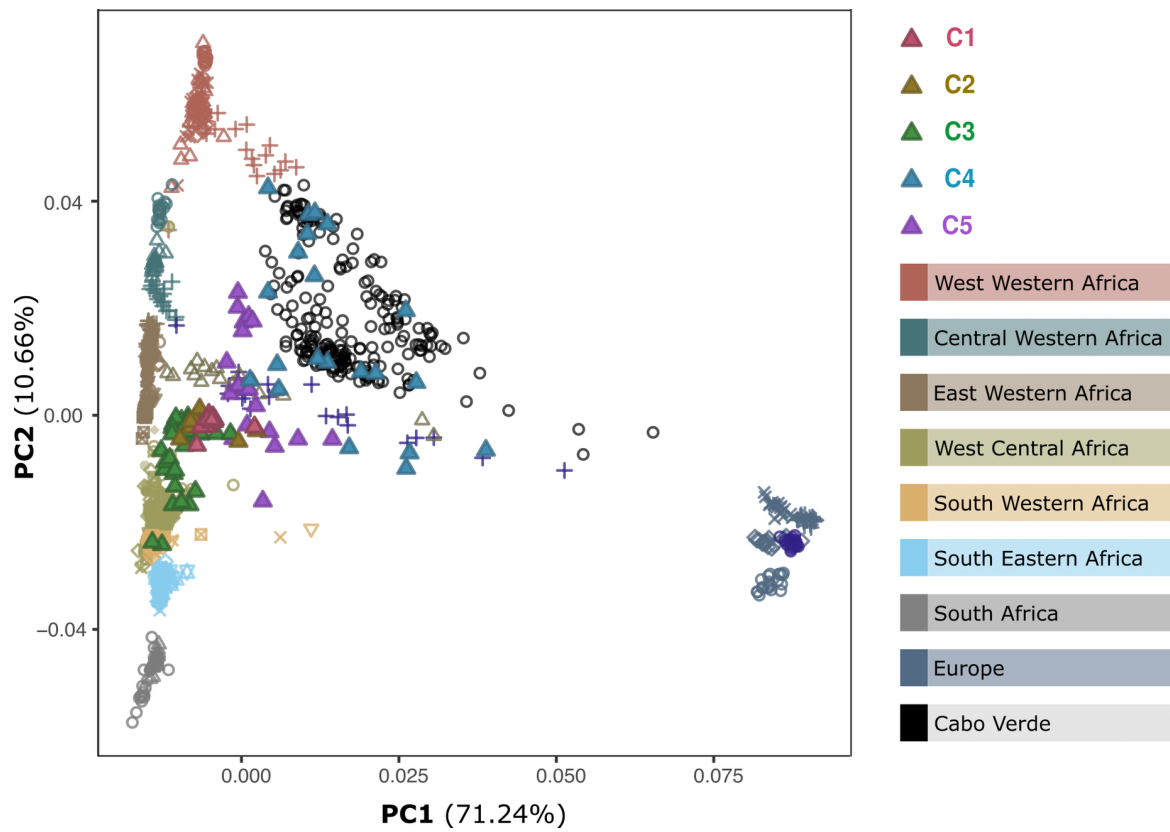
1254

1255

1256

1257

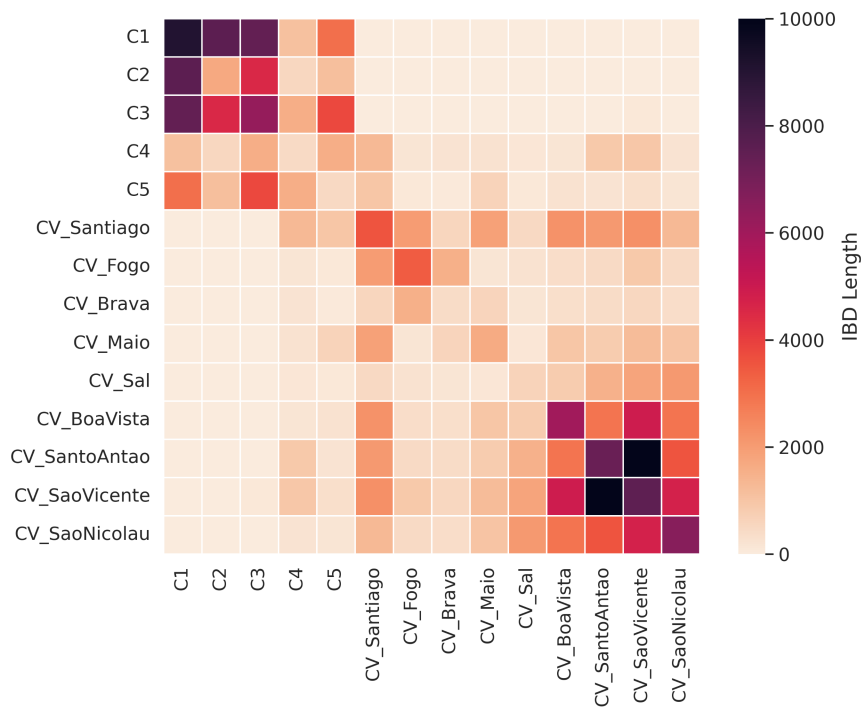
S8 Fig. MCMC iterations with the highest likelihood over 20 independent SOURCEFIND runs. The results are ordered by posterior probability and illustrate both the consistency and variability in the source estimates across different runs.



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.