

## Article

# The Genetic Structure of Cape Verdean Population Revealed by Y-Chromosome STRs

Rita Costa <sup>1</sup> , Jennifer Fadoni <sup>2,3</sup> , António Amorim <sup>2,3,4,5</sup>  and Laura Cainé <sup>1,3,4,\*</sup> <sup>1</sup> Faculty of Medicine, Porto University, 4200-319 Porto, Portugal; up201906725@fc.up.pt<sup>2</sup> National Institute of Legal Medicine and Forensic Sciences, I.P., North Branch, 4050-202 Porto, Portugal; jennifer.n.fadoni@inmlcf.mj.pt (J.F.); antonio.j.amorim@inmlcf.mj.pt (A.A.)<sup>3</sup> LAQV&REQUIMTE, Laboratory of Applied Chemistry, Department of Chemical Sciences, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal<sup>4</sup> National Institute of Legal Medicine and Forensic Sciences, I.P., Centre Branch, 3000-548 Coimbra, Portugal<sup>5</sup> Faculty of Sciences, Lisbon University, 1749-016 Lisboa, Portugal

\* Correspondence: laura.m.caine@inmlcf.mj.pt

## Abstract

**Background/Objectives:** Y-chromosomal short tandem repeats (Y-STR) are genetic markers widely used in forensic and population genetics. However, despite their importance, many populations remain under-represented in published studies and genetic databases. One such population is the Cape Verdean, which, despite its unique history of admixture between European and sub-Saharan African populations, continues to be under-represented in global Y-STR reference databases. This study aims to characterize the Y-STR haplotype diversity and paternal lineage composition of the Cape Verdean population using a high-resolution STR panel. **Methods:** A total of 143 unrelated Cape Verdean men were analyzed using a set of 26 Y-STR loci, including rapidly mutating markers. Allele and haplotype frequencies were calculated, along with standard forensic parameters such as gene and haplotype diversity. Paternal lineages were inferred, and genetic relationships with other populations were evaluated using distance-based and graphical methods. **Results:** A total of 135 haplotypes were detected, with 88.8% being unique, yielding a haplotype diversity of 0.999. The most common haplogroups reflected both West African and European ancestry. Genetic distance analysis positioned the Cape Verdean population between African and European groups, supporting its intermediate and admixed genetic background. **Conclusions:** This study provides the first high-resolution Y-STR dataset for Cape Verdeans, contributing valuable reference data for forensic casework and population genetic studies. The results highlight the utility of extended Y-STR panels in admixed populations and underscore the need to enhance the representation of admixed populations in international forensic reference databases.

**Keywords:** population genetics; haplotype; Y chromosome; Cape Verde; Argus Y28; forensic genetics



Academic Editor: Mitchell Mark Holland

Received: 3 August 2025

Revised: 21 August 2025

Accepted: 23 August 2025

Published: 25 August 2025

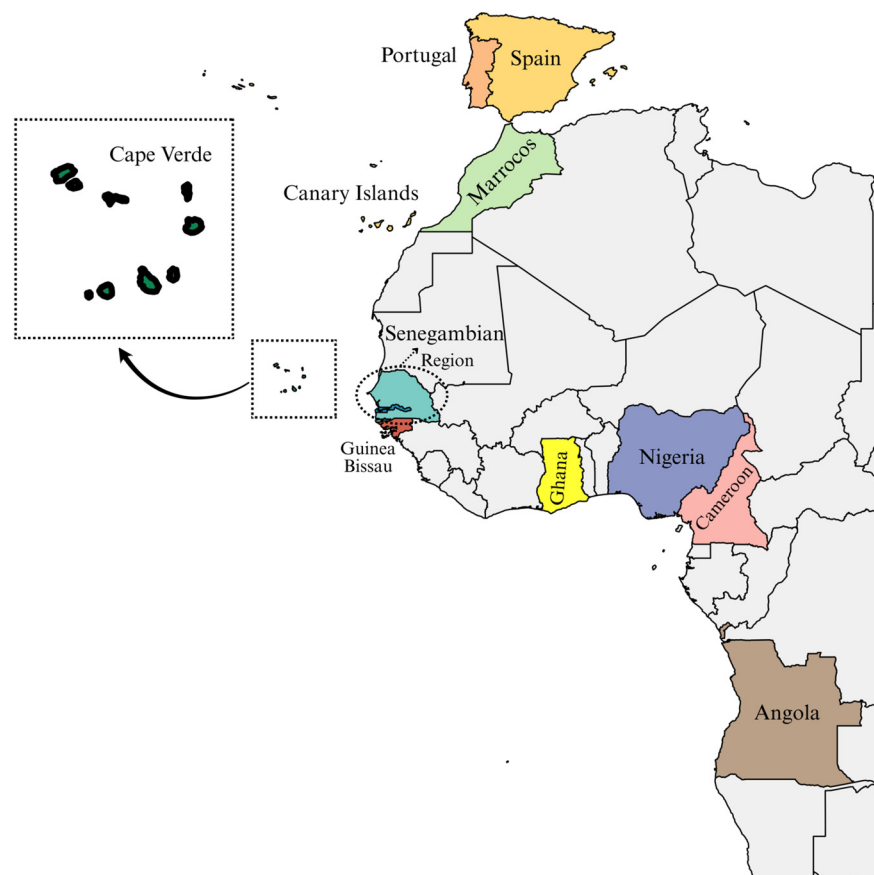
**Citation:** Costa, R.; Fadoni, J.; Amorim, A.; Cainé, L. The Genetic Structure of Cape Verdean Population Revealed by Y-Chromosome STRs. *Genes* **2025**, *16*, 999. <https://doi.org/10.3390/genes16090999>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Cape Verde is an archipelago located off the west coast of Africa in the Atlantic Ocean, discovered by Portuguese navigators in the 15th century. Permanent settlement began around 1460, driven by the arrival of a small number of European men, mainly Portuguese, and a much larger number of enslaved sub-Saharan Africans, primarily from the Senegambia region. Historical accounts also report the presence of Guanche individuals from the

Canary Islands among the first slaves. This founding population laid the groundwork for the unique demographic and genetic landscape observed in the archipelago today [1–5] (Figure 1).



**Figure 1.** Geographic location of Cape Verde and the populations that contributed to the genetic landscape of the archipelago. Only countries with available Y-STR haplotype data are represented. The map was created using MapChart (<https://www.mapchart.net>, accessed on 2 July 2025).

Over the centuries, Cape Verde has experienced intense migratory flows, mainly involving sub-Saharan Africans and Europeans. This admixture has led to an asymmetric genetic pattern, with a predominance of African maternal lineages and European paternal lineages, as documented by previous genetic studies [6,7].

Beyond its unique demographic history, the Cape Verdean population maintains a strong historical and cultural connection with Portugal, reflected in significant and sustained migration [8,9]. According to the 2023 Migration and Asylum Report, Cape Verdeans are the third-largest foreign community legally residing in Portugal, with 48,885 residents, accounting for 4.7% of the total foreign population. However, this figure does not fully capture the true demographic weight of Cape Verdeans due to the increasing acquisition of Portuguese nationality and ongoing migration to other countries [10,11].

These migration patterns and historical admixture have important implications in forensic genetics. Genetic studies indicate that the Cape Verdean population exhibits substructure and diverges significantly from other sub-Saharan African groups. Notably, short tandem repeat (STR) marker analysis has revealed differences between the northern (Barlavento) and southern (Sotavento) islands, likely due to genetic drift and limited gene flow. Such differentiation underscores the importance of establishing population-specific reference data, particularly for forensic applications [2,6].

Y-chromosomal STRs (Y-STRs) are commonly used in forensic casework involving male individuals, such as sexual assault investigations and paternity testing [12]. Among them, rapidly mutating Y-STRs (RM Y-STRs), which have elevated mutation rates (greater than  $10^{-2}$ ), are especially valuable for distinguishing closely related men, as they provide higher resolution in pedigree analysis and in complex forensic cases [13,14].

The interpretation of Y-STR profiles relies on the availability of comprehensive population data. For this purpose, international databases like the Y Chromosome Haplotype Reference Database (YHRD) are essential tools that enable global comparison of haplotypes and support statistical evaluation of matches in forensic casework [15,16]. However, the Cape Verdean population remains under-represented in the YHRD, especially with respect to data generated from expanded multiplex kits such as the Argus Y-28, which includes both conventional and rapidly mutating loci, with only 117 haplotypes currently registered for 17 loci and none for 23 or more loci (<https://yhrd.org> (accessed on 20 August 2025)) [15,17–19].

Although previous studies have described the genetic characteristics of the Cape Verdean population, none have provided high-resolution data on allele and haplotype frequencies using expanded Y-STR panels, limiting their forensic applicability [2,6].

The aim of this study is to determine allele frequencies, haplotype diversity, and forensic parameters for 26 Y-STR loci in a sample of men residing in continental Portugal with Cape Verdean paternal ancestry, using an Argus Y-28 QS kit (QIAGEN, Hilden, Germany), and to contribute these data to the YHRD. This will enhance the representation of the Cape Verdean population in international forensic reference databases and support reliable Y-STR-based identification in medico-legal contexts.

## 2. Materials and Methods

### 2.1. Sample Collection

A total of 143 samples (135 dried blood spots and 8 buccal swabs) from unrelated men of Cape Verdean paternal ancestry were selected and analyzed. Paternal ancestry was assessed based on the individual's identification documents, which confirmed their place of birth. Information regarding paternal lineage was further obtained from the individual's self-reported account of their father's origin. The samples, collected between 2012 and 2022 from paternity cases available at INMLCF, I.P., were fully anonymized prior to analysis, with no personal information of the individuals being accessible, except for their paternal ancestry. They remained stored until their use in this study (2024–2025). Buccal swabs were stored at room temperature in paper envelopes in a dry and dark environment until DNA extraction.

This study received ethical approval from the Ethics Committee of the National Institute of Legal Medicine and Forensic Sciences. The approval was granted on 25 July 2024, under the reference code CE-25/2024. The use of these samples complied with Portuguese legislation (Law No. 12/2005, of January 26) and with INMLCF regulations, which permit the use of anonymized samples stored for more than two years after casework completion.

### 2.2. DNA Extraction and Y-STR Fragment Analysis

The saliva samples collected using buccal swabs were extracted with a PrepFiler Express™ Forensic DNA Extraction Kit (Thermo Fisher Scientific, Waltham, MA, USA), following the manufacturer's protocol. The blood samples on FTA® cards were processed by direct amplification, without any prior DNA extraction procedure. Amplification of 26 Y-STR loci was performed using an Investigator® Argus Y-28 QS PCR kit (QIAGEN, Hilden, Germany) on Applied Biosystems GeneAmp® PCR System 9700 thermal cyclers (Thermo Fisher Scientific, Waltham, MA, USA). Reactions followed the manufacturer's

protocol, with two modifications to optimize reagent use and prevent overamplification: the reaction volume was halved to 12.5  $\mu\text{L}$ , and the number of PCR cycles was reduced to 28.

For more details, the PCR reaction was performed using the following program: pre-denaturation at 96 °C for 12 min; denaturation at 96 °C for 10 s, annealing at 61.5 °C for 85 s, and extension at 72 °C for 5 s, for 28 cycles; and final extension at 68 °C for 5 min and 60 °C for 5 min, with maintenance at 10 °C. Control DNA 9948 (QIAGEN, Hilden, Germany) was used as the positive control and ddH<sub>2</sub>O as a negative control (without DNA) for each batch of Y-STR fragment analysis. The Investigator<sup>®</sup> Argus Y-28 QS PCR kit included 20 standard Y-STR loci (DYS389I, DYS391, DYS389II, DYS533, DYS390, DYS458, DYS393, DYS19, DYS437, DYS460, YGATAH4, DYS448, DYS439, DYS549, DYS438, DYS456, DYS643, DYS635, DYS385, DYS392), 6 rapidly mutating markers (DYS449, DYS481, DYS570, DYS576, DYS518, DYS627), and quality sensors QS1 and QS2. PCR products were prepared for capillary electrophoresis (CE) by adding 1  $\mu\text{L}$  of each PCR product to a mixture of 12  $\mu\text{L}$  Hi-Di<sup>™</sup> Formamide (Thermo Fisher Scientific, Waltham, MA, USA) and 0.5  $\mu\text{L}$  of DNA Size Standard 24plex (BTO) (QIAGEN, Hilden, Germany). The reaction plate was then heat denatured at 95 °C for 3 min and subsequently cooled using a thermal cycler set to 4 °C. Capillary electrophoresis was carried out on an ABI 3500 Genetic Analyzer (Thermo Fisher Scientific, Waltham, MA, USA). DNA typing was carried out following the manufacturer's protocol, using the provided locus panel, allele bins, and allele designations based on the supplied allelic ladder. The results were processed and analyzed using GeneMapper ID-X version 1.6 software (Thermo Fisher Scientific, Waltham, MA, USA), following the reference allelic ladder for genotype assignment. Reagent blanks were included alongside each batch of samples to verify the accuracy and reliability of the analyses. Additionally, samples that showed apparent null alleles, off-ladder (OL) peaks, or mutations in specific markers were reamplified to confirm the observed results.

### 2.3. Statistical Analyses

The number of distinct haplotypes, the frequency of unique haplotypes, discrimination capacity (DC), haplotype match probability (HMP), and haplotype diversity (HD) were calculated using a direct counting method in Microsoft Office Excel (version 16.94). The haplotype diversity (HD) was calculated applying the formula  $HD = n(1 - \sum p_i^2)/(n - 1)$ , where  $n$  is the sample size, and  $p_i$  is the frequency of the  $i$ -th haplotype [20]. The discrimination capacity (DC) was calculated as the ratio between the number of different haplotypes and the total number of individuals in the sample. The frequency of unique haplotypes was determined as the ratio between the number of unique haplotypes and the total number of individuals analyzed in the sample. The haplotype match probability (HMP) was calculated by summing the squared relative frequencies of all observed haplotypes ( $HMP = \sum p_i^2$ ), where  $p_i$  represents the frequency of the  $i$ -th haplotype. STRAF software (STR Analysis for Forensics) (version 2.2.2) was used to calculate forensic parameters, including allele frequencies at each locus, gene diversity (GD), polymorphism information content (PIC), and power of discrimination (PD) [21]. Pairwise genetic distances ( $R_{ST}$ ), estimated in Arlequin v3.5.2.2 software, were used to measure genetic distances between the samples from this study and the other compiled populations. For inter-population comparisons, the DYS389II allele length was calculated by subtracting the repeat count at DYS389I from that at DYS389II. For this analysis, only 17 Y-STR markers common to all population datasets were used. Using R statistical software version 4.0, a multidimensional scaling (MDS) plot was generated to visualize the relationships between populations based on the calculated matrix of pairwise genetic distances. Haplogroup prediction was performed using NevGen Y-DNA Haplogroup Predictor (<https://www.nevgen.org>, accessed on 16 June

2025) [22], assigning the haplogroup with the highest membership probability based on Y-STR haplotype.

### 3. Results

#### 3.1. Analysis of Haplotype and Allele Diversity (and Forensic Parameters)

A total of 135 different haplotypes were obtained from 143 Cape Verdean samples, of which 127 (88.8%) were unique. The complete haplotype data for all 26 Y-STR loci are detailed in Supplementary Table S1. The allelic diversity observed across the loci ranged from 4 distinct alleles at DYS391 and DYS460 to 13 distinct alleles at DYS449, as shown in Supplementary Figure S1. Allele frequency data for all loci are available in Supplementary Table S2. The haplotype match probability (HMP) was calculated to be 0.0078, and the haplotype diversity (HD) was found to be 0.999. The corresponding allelic frequencies varied from 0.0070 to 0.6643. The lowest GD value was 0.497 in locus DYS391 and the highest was 0.859 in locus DYS627. Among the male Cape Verdean samples, biallelic patterns were observed. These included the DYS389II locus (alleles 29/30 and 28/29), DYS437 (alleles 14/15), DYS448 (alleles 19/20, 20/21, and 19/22), and DYS439 (alleles 10/11), suggesting possible duplication events or structural variations on the Y chromosome. Additionally, microvariant alleles such as 17.2 and 16.2 were detected at the DYS458 locus, further highlighting the genetic diversity and mutational dynamics within the studied population. No meaningful differences in profile completeness were observed between saliva and blood samples. Nonetheless, blood stains occasionally exhibited slightly higher signal intensities and a slightly elevated level of background peaks. These variations did not affect the overall quality or the interpretation of the profiles.

#### 3.2. Haplogroup Distribution

Out of the 143 Y-STR profiles analyzed, haplogroups were predicted for 120 individuals with a probability greater than 50% using the NevGen Y-DNA Haplogroup Predictor (<https://www.nevgen.org>, accessed on 16 June 2025) [22]. Within this subset, a total of 13 distinct haplogroups were identified: E1a, E1b1a, E1b1b, E3a, G2a2b1, I2a2a, I2a2b, J1a2a1a2, J2b2a, L1a, L1b, R1b, and T. The most frequently observed haplogroups were R1b and E1b1a (Table 1). Interestingly, haplogroups R and E showed similar overall frequencies in the sample, with R1b representing 44% and the combined E sub-haplogroups accounting for 45%. This suggests a balanced genetic contribution from both European (R) and African (E) paternal lineages. The remaining haplogroups appeared in lower frequencies, reflecting additional genetic diversity from Middle Eastern and Asian origins.

**Table 1.** Y-chromosome haplogroup and ancestry frequencies in Cape Verde individuals, with predicted haplogroups ( $n = 120$ ).

Haplogroup (H)	Sub-Haplogroup (SH)	Frequency SH (%)	Frequency H (%)	Ancestry
E	E1a	3%	45%	African
	E1b1a	28%		
	E1b1b	14%		
	E3a	1%		
L	L1a	1%	2%	Asian
	L1b	1%		

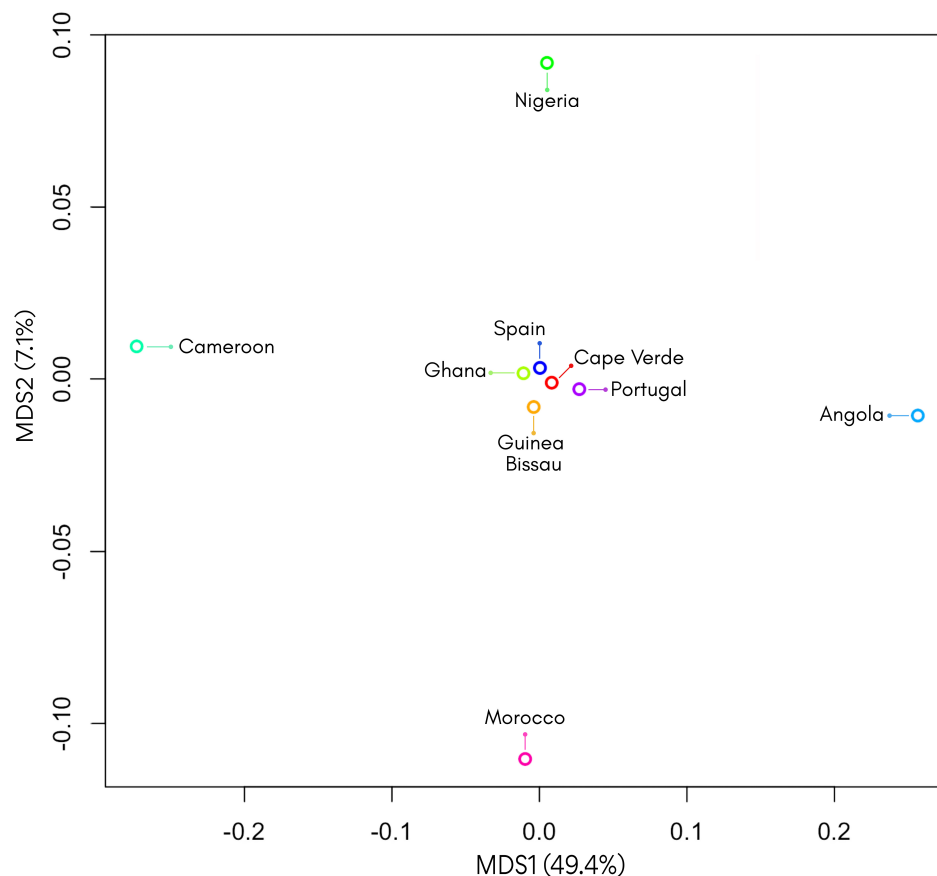
Table 1. Cont.

Haplogroup (H)	Sub-Haplogroup (SH)	Frequency SH (%)	Frequency H (%)	Ancestry
I	I2a2a	1%	4%	European
	I2a2b	3%		
R	R1b	44%	44%	
G	G2a2b1	1%	1%	
J	J1a2a1a2	2%	4%	Middle Eastern
	J2b2a	3%		
T	-	1%	1%	

### 3.3. Genetic Distance Analysis Based on $R_{ST}$ Distances

To explore the genetic relationships between the Cape Verdean population and other African and European groups, genetic distances were estimated based on the sum of squared size differences ( $R_{ST}$ ) between the haplotype distributions observed for 17 Y-STR loci. Comparative analysis could not be performed for the entire set of loci due to the absence of corresponding population data. However, for 17 loci in common (DYS389I, DYS391, DYS389II, DYS390, DYS458, DYS393, DYS19, DYS437, YGATAH4, DYS448, DYS439, DYS438, DYS456, DYS635, DYS392, DYS385a, DYS385b), comparisons were possible using published data from Guinea Bissau [23], Ghana [24], Nigeria [25], Cameroon [26], Angola [27], Spain [25], Portugal [28], and Morocco [29]. The pairwise  $R_{ST}$  values and corresponding  $p$ -values between the Cape Verdean population and the eight reference populations are displayed in Supplementary Table S3. The resulting distances were visualized through a multidimensional scaling (MDS) plot (Figure 2). The lowest genetic distances were observed with Cameroon ( $R_{ST} = 0.01230$ ), Morocco ( $R_{ST} = 0.01747$ ), Portugal ( $R_{ST} = 0.02098$ ), and Angola ( $R_{ST} = 0.02100$ ), indicating greater genetic affinity of Cape Verde with these African and Iberian populations. Although Spain showed a slightly negative  $R_{ST}$  value ( $-0.00347$ ), this difference was not statistically significant ( $p = 0.88288$ ). The highest genetic distances were recorded with Ghana ( $R_{ST} = 0.04314$ ), Guinea-Bissau ( $R_{ST} = 0.03635$ ), and Nigeria ( $R_{ST} = 0.03307$ ). Among the pairwise comparisons, only some  $R_{ST}$  distances involving the Cape Verdean population were statistically significant ( $p < 0.05$ ), as presented in Supplementary Table S3. The statistical analysis of  $R_{ST}$  distances revealed significant genetic differentiation between the Cape Verdean population and six of the eight populations compared.  $p$ -values below the conventional threshold of 0.05 were observed in the comparisons with Ghana ( $p = 0.00000$ ), Nigeria ( $p = 0.02703$ ), Cameroon ( $p = 0.01802$ ), Angola ( $p = 0.01802$ ), Portugal ( $p = 0.00000$ ), and Morocco ( $p = 0.03604$ ), indicating genetic divergence from these groups. In contrast, the comparisons with Guinea-Bissau ( $p = 0.0901$ ) and Spain ( $p = 0.8829$ ) did not reach statistical significance, suggesting no strong evidence of genetic differentiation from these populations. This may be consistent with a relatively closer genetic relationship, particularly with Iberian and some West African populations. The two-dimensional MDS plot provided an initial visualization of genetic relationships. In this representation, a cluster can be observed in which Cape Verde is positioned near Spain, Portugal, Ghana, and Guinea-Bissau, suggesting relative genetic proximity among these populations. However, the proportion of variance explained (PVE) was relatively low, with the first dimension accounting for 49.4% and the second for only 7.1%. These low cumulative values indicate that the 2D configuration does not accurately preserve the original genetic distances. Therefore, a three-dimensional MDS plot was generated (Supplementary Figure S2) to enhance the spatial resolution and improve the interpretation

of population structure. Together, these results highlight the importance of combining genetic distances with statistical support and appropriate dimensional representations when interpreting population affinities. This integrative approach contributes to a more accurate understanding of the paternal genetic structure and the intermediate position of the Cape Verdean population between African and European lineages.



**Figure 2.** Multidimensional scaling (MDS) plot based on  $R_{ST}$  genetic distances among Cape Verde and reference populations. The analysis was performed using 17 shared Y-STR loci. The plot illustrates the genetic affinities between Cape Verde and other European and African populations, showing Cape Verde clustering closely with Guinea-Bissau, Spain, Ghana, and Portugal, while more distant populations include Angola, Cameroon, Nigeria, and Morocco. The percentages on the axes correspond to the proportion of variance explained (PVE) by each dimension.

#### 4. Discussion

The genetic analysis of 143 unrelated Cape Verdean men using 26 Y-STR loci from an Argus Y-28 QS kit revealed a high level of haplotype diversity ( $HD = 0.999$ ), with 135 different haplotypes and 88.8% being unique. These values reflect a high discrimination capacity, confirming the forensic utility of this multiplex system in admixed populations such as Cape Verde. The presence of only a few ( $n = 8$ ) repeated haplotypes among unrelated individuals supports the system's ability to differentiate between male individuals, including those with potential close kinship, particularly due to the inclusion of rapidly mutating Y-STRs that enhance resolution. The practical value of rapidly mutating Y-STRs lies in their elevated mutation rates, which generate novel allelic differences even between close paternal relatives, such as brothers or father–son pairs. This mutational dynamic reduces the probability of observing identical haplotypes among relatives and therefore increases the power of Y-STR analysis to exclude or distinguish individuals in forensic contexts. Such resolution is particularly important in cases involving sexual assault with multiple male

contributors or in kinship testing scenarios where conventional Y-STR panels may yield indistinguishable profiles [14].

The detection of microvariant alleles and biallelic patterns in this study can be attributed to well-known mutational mechanisms affecting Y-STRs. Microvariants are most often generated by slippage events during DNA replication, where partial repeat units are inserted or deleted, producing intermediate allele sizes [13]. In contrast, duplicated alleles at certain loci may reflect structural rearrangements of the Y chromosome, such as segmental duplications or gene conversion events, which have been described in other populations [16,30]. Although relatively rare, these phenomena are relevant in forensic practice as they may influence allele designation and the interpretation of Y-STR profiles in casework.

From a forensic perspective, the availability of population-specific Y-STR reference data is essential for accurate statistical interpretation in casework. This need is particularly critical for admixed populations such as Cape Verdeans, whose paternal genetic profiles result from asymmetric historical admixture between European (predominantly R1b) and West African (mainly E1b1a and E1b1b) lineages [6,7]. In addition to the admixture events associated with the transatlantic slave trade, the observed genetic affinities with Iberian populations may also reflect more recent migratory flows during the 20th century, when thousands of Cape Verdeans settled in Portugal and other parts of Europe for economic and political reasons [10,11]. While the primary source of enslaved individuals brought to the archipelago during the transatlantic slave trade is historically associated with the Senegambian region [5], haplotypes from that area are under-represented or even absent in current Y-STR databases, such as YHRD [15,16]. This limits both comparative analyses and forensic resolution, especially when evaluating the statistical weight of evidence in investigations involving biogeographic ancestry inference. This absence of available reference data prevented the direct inclusion of populations from Senegal and Gambia in the RST-based comparisons, despite their well-documented historical relevance for the genetic makeup of Cape Verdeans. Additionally, many existing studies on African and admixed populations rely on panels with a reduced number of Y-STR loci or use datasets with limited haplotypic resolution, which restricts the robustness of comparative population analyses [31]. This limitation is particularly problematic for admixed populations or those with recent common paternal ancestors, where finer resolution is needed to distinguish close male relatives. In contrast, the use of 26 Y-STR loci in this study, including rapidly mutating markers, enabled the detection of rare allelic patterns such as microvariants and biallelic loci [13,14]. These markers are particularly informative in complex forensic scenarios, including sexual assault or kinship testing, where male individuals of the same paternal lineage must be distinguished. The genetic distinctiveness of Cape Verdeans also has practical implications in forensic casework. Y-STR profiles from individuals with Cape Verdean ancestry may be misclassified if interpreted against reference populations that do not adequately reflect their genetic background. This risk has been highlighted in studies emphasizing the forensic implications of misinterpretation in cases involving under-represented populations, particularly when Y-STR evidence is central to the investigation [32]. It should be noted that haplogroup prediction based on Y-STR profiles using NevGen is inherently probabilistic and particularly prone to misclassification in admixed populations such as Cape Verdeans. While SNP-based genotyping would provide more reliable assignments, the present results should be regarded as probabilistic inferences rather than definitive classifications. The RST-based comparisons performed in this study provide further insight into the genetic structure of the Cape Verdean population within a broader African–European context. The multidimensional scaling (MDS) analysis provided a visual approximation of genetic relationships among the studied populations. However,

due to the limited amount of genetic variance represented in the two-dimensional configuration, the spatial distribution observed should be interpreted with caution. To mitigate this limitation, a three-dimensional MDS plot was generated, offering a more accurate depiction of the underlying genetic structure. In this higher-resolution model, the Cape Verdean population appears to occupy an intermediate position between Iberian and West/Central African groups. While this distribution does not reflect strict geographic proximity, it is consistent with the historical and demographic processes that shaped the paternal genetic landscape of Cape Verde, including founder effects, asymmetric admixture, and successive migratory events. Although this study focuses on Cape Verdean individuals residing in Portugal, the findings may not be directly generalizable to the population living in the Cape Verde islands due to possible migration effects, integration processes, and genetic drift within the diaspora. Moreover, the sample size ( $n = 143$ ) reflects the limited availability of Cape Verdean paternal lineage samples accessible within the studied diaspora context, which inevitably constrains the statistical power of some analyses. These findings contribute valuable high-resolution data to the global Y-STR reference framework, particularly for under-represented admixed populations, and underscore the need to broaden the geographic and demographic scope of future studies [25,30,33]. Future studies should aim to include a broader sampling of individuals from specific islands, as well as from Cape Verdean communities in other regions of the diaspora, to build a more comprehensive and representative genetic dataset for forensic and population genetic purposes [8,9].

## 5. Conclusions

This study presents the first comprehensive analysis of 26 Y-STR loci in a sample of Cape Verdean men residing in Portugal, revealing high haplotype diversity and excellent discriminatory power. These findings underscore the effectiveness of the Argus Y-28 QS kit in forensic applications involving admixed populations. The detection of unique haplotypes, including rare allelic variants such as microvariants and biallelic patterns, highlights the added value of using extended Y-STR panels, especially those including rapidly mutating loci, for resolving cases involving close paternal relatives. Given the genetic distinctiveness and under-representation of Cape Verdean populations in international forensic databases, this dataset strengthens the basis for more reliable statistical evaluations in forensic casework and ancestry inference involving individuals of Cape Verdean descent. To ensure fair and accurate interpretation of Y-STR evidence, it is essential to expand population-specific reference databases, particularly for admixed and minority populations. Future research should prioritize the inclusion of samples from different Cape Verdean islands and other diaspora communities to refine our understanding of their genetic structure and improve global forensic practices. In addition, our findings highlight the importance of expanding forensic reference databases to include samples from diaspora communities. When island-based sampling is limited, diaspora populations can provide valuable insights into the genetic structure of under-represented groups, thereby strengthening the reliability of Y-STR data for both forensic and population genetic applications. Incorporating such data will ultimately contribute to more equitable forensic investigations and support judicial systems with scientifically robust genetic evidence.

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/genes16090999/s1>. Table S1. Haplotype database of Cape Verde population. Table S2. Allele frequency in Cape Verde population. Table S3.  $R_{ST}$  genetic distances and corresponding  $p$ -values between the Cape Verdean population and eight reference populations based on 17 shared Y-STR loci. Figure S1. Allele frequency distributions of 26 Y-STR loci in the Cape Verde population. The  $x$ -axis represents allele values per locus; the  $y$ -axis indicates the observed allele

frequencies. Figure S2. Interactive 3D MDS plot based on RST genetic distances between the Cape Verdean population and eight reference populations using 17 shared Y-STR loci.

**Author Contributions:** Conceptualization: L.C. and A.A.; methodology, R.C. and J.F.; writing—original draft preparation, R.C. and J.F.; writing—review and editing, L.C. and A.A.; supervision, J.F., L.C. and A.A.; funding acquisition, L.C. and A.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Instituto Nacional de Medicina Legal e Ciências Forenses, Portugal.

**Institutional Review Board Statement:** This study received ethical approval from the Ethics Committee of the National Institute of Legal Medicine and Forensic Sciences. The approval was granted on 25 July 2024, under the reference code CE-25/2024.

**Informed Consent Statement:** Informed consent was waived due to the retrospective nature of the study and use of anonymized data.

**Data Availability Statement:** The original contributions presented in this study are included in this article/Supplementary Material. Further inquiries can be directed to the corresponding author(s).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

CE	Capillary Electrophoresis
DC	Discrimination Capacity
GD	Gene Diversity
HD	Haplotype Diversity
HMP	Haplotype Match Probability
INMLCF, I.P.	Instituto Nacional de Medicina Legal e Ciências Forenses, Instituto Público
MDS	Multidimensional Scaling
NGS	Next-Generation Sequencing
OL	Off-Ladder
PCR	Polymerase Chain Reaction
PIC	Polymorphism Information Content
PVE	Proportion of Variance Explained
QS	Quality Sensor
RM Y-STRs	Rapidly Mutating Y-Chromosomal Short Tandem Repeats
STR	Short Tandem Repeat
YHRD	Y Chromosome Haplotype Reference Database
Y-STR	Y-Chromosomal Short Tandem Repeat

## References

1. Gonçalves, R.; Fernandes, A.T.; Brehm, A. Cabo Verde islands: Different maternal and paternal heritage testifies the nature of its first settlers. *Int. Congr. Ser.* **2004**, *1261*, 372–373. [[CrossRef](#)]
2. Fernandes, A.T.; Velosa, R.; Jesus, J.; Carracedo, A.; Brehm, A. Genetic differentiation of the Cabo Verde archipelago population analysed by STR polymorphisms. *Ann. Hum. Genet.* **2003**, *67*, 340–347. [[CrossRef](#)] [[PubMed](#)]
3. Laurent, R.; Szpiech, Z.A.; da Costa, S.S.; Thouzeau, V.; Fortes-Lima, C.A.; Dessarps-Freichy, F.; Lémée, L.; Utgé, J.; Rosenberg, N.A.; Baptista, M.; et al. A genetic and linguistic analysis of the admixture histories of the islands of Cabo Verde. *eLife* **2023**, *12*, e79827. [[CrossRef](#)]
4. Parra, E.J.; Ribeiro, J.C.; Caeiro, J.L.; Riveiro, A. Genetic structure of the population of Cabo Verde (west Africa): Evidence of substantial European admixture. *Am. J. Phys. Anthr.* **1995**, *97*, 381–389. [[CrossRef](#)]
5. Brehm, A.; Pereira, L.; Bandelt, H.J.; Prata, M.J.; Amorim, A. Mitochondrial portrait of the Cabo Verde archipelago: The Senegambian outpost of Atlantic slave trade. *Ann. Hum. Genet.* **2002**, *66*, 49–60. [[CrossRef](#)]
6. Beleza, S.; Campos, J.; Lopes, J.; Araújo, I.I.; Hoppfer Almada, A.; Correia e Silva, A.; Parra, E.J.; Rocha, J. The admixture structure and genetic variation of the archipelago of Cape Verde and its implications for admixture mapping studies. *PLoS ONE* **2012**, *7*, e51103. [[CrossRef](#)]

7. Gonçalves, R.; Rosa, A.; Freitas, A.; Fernandes, A.; Kivisild, T.; Villems, R.; Brehm, A. Y-chromosome lineages in Cabo Verde Islands witness the diverse geographic origin of its first male settlers. *Hum. Genet.* **2003**, *113*, 467–472. [[CrossRef](#)]
8. Resende, A.; Amorim, A.; da Silva, C.V.; Ribeiro, T.; Porto, M.J.; Costa Santos, J.; Afonso Costa, H. Study of genetic markers of CODIS and ESS systems in a population of individuals from Cabo Verde living in Lisboa. *Int. J. Leg. Med.* **2017**, *131*, 119–121. [[CrossRef](#)] [[PubMed](#)]
9. Amorim, A.; Marques-Santos, R.; Vieira-Silva, C.; Afonso-Costa, H.; Espinheira, R.; Ferreira-Gomes, P.; Costa-Santos, J. Genetic portrait of a native population of Cabo Verde living in Lisboa. *Forensic Sci. Int. Genet.* **2012**, *6*, e166–e169. [[CrossRef](#)]
10. Gabinete de Estratégia e Estudos. População estrangeira com estatuto legal de residente em Portugal—Cabo Verde. 2023. Available online: <https://www.gee.gov.pt/pt/lista-publicacoes/estatisticas-de-imigrantes-em-portugal-por-nacionalidade/paises/Cabo%20Verde/3947-populacao-estrangeira-com-estatuto-legal-de-residente-em-portugal-cabo-verde/file> (accessed on 5 May 2025).
11. AIMA I.P.—DPEE—Direção de Planeamento, Estudos e Estatística. *Relatório de Migrações e Asilo 2023*; AIMA: London, UK, 2024.
12. Kayser, M. Forensic use of Y-chromosome DNA: A general overview. *Hum. Genet.* **2017**, *136*, 621–635. [[CrossRef](#)] [[PubMed](#)]
13. Ballantyne, K.N.; Goedbloed, M.; Fang, R.; Schaap, O.; Lao, O.; Wollstein, A.; Choi, Y.; van Duijn, K.; Vermeulen, M.; Brauer, S.; et al. Mutability of Y-chromosomal microsatellites: Rates, characteristics, molecular bases, and forensic implications. *Am. J. Hum. Genet.* **2010**, *87*, 341–353. [[CrossRef](#)] [[PubMed](#)]
14. Ballantyne, K.N.; Keerl, V.; Wollstein, A.; Choi, Y.; Zuniga, S.B.; Ralf, A.; Vermeulen, M.; de Knijff, P.; Kayser, M. A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. *Forensic Sci. Int. Genet.* **2012**, *6*, 208–218. [[CrossRef](#)]
15. Y-Chromosome Haplotype Reference Database (YHRD). Available online: <https://yhrd.org> (accessed on 5 May 2025).
16. Roewer, L. The Y-Short Tandem Repeat Haplotype Reference Database (YHRD) and Male Population Stratification in Europe—Impact on Forensic Genetics. *Forensic Sci. Rev.* **2003**, *15*, 165–172. [[PubMed](#)]
17. QIAGEN. *Investigator® Argus Y-28 QS Handbook*; QIAGEN: Hilden, Germany, 2022. Available online: <https://www.qiagen.com/in/resources/download.aspx?id=cf7ae42a-72bc-4d3d-812d-684d0625a179&lang=en> (accessed on 5 May 2025).
18. Steffen, C.R.; Huszar, T.I.; Borsuk, L.A.; Vallone, P.M.; Gettings, K.B. A multi-dimensional evaluation of the ‘NIST 1032’ sample set across four forensic Y-STR multiplexes. *Forensic Sci. Int. Genet.* **2022**, *57*, 102655. [[CrossRef](#)]
19. Haarkötter, C.; Isabel Medina-Lozano, M.; Vinueza-Espinosa, D.C.; Saiz, M.; Gálvez, X.; Carlos Álvarez, J.; Antonio Lorente, J. Evaluating the efficacy of three Y-STRs commercial kits in degraded skeletal remains. *Sci. Justice* **2024**, *64*, 543–548. [[CrossRef](#)]
20. Nei, M.; Tajima, F. Genetic drift and estimation of effective population size. *Genetics* **1981**, *98*, 625–640. [[CrossRef](#)]
21. Gouy, A.; Zieger, M. STRAF—A convenient online tool for STR data evaluation in forensic genetics. *Forensic Sci. Int. Genet.* **2017**, *30*, 148–151. [[CrossRef](#)]
22. Cetkovic Gentula, M.; Nevski, A. NEVGEN Y-DNA Haplogroup Predictor. Available online: <https://www.nevgen.org> (accessed on 16 June 2025).
23. Carvalho, M.; Brito, P.; Bento, A.M.; Gomes, V.; Antunes, H.; Costa, H.A.; Lopes, V.; Serra, A.; Balsa, F.; Andrade, L.; et al. Paternal and maternal lineages in Guinea-Bissau population. *Forensic Sci. Int. Genet.* **2011**, *5*, 114–116. [[CrossRef](#)] [[PubMed](#)]
24. Kofi, A.E.; Hakim, H.M.; Khan, H.O.; Ismail, S.A.; Ghansah, A.; David, A.A.; Mat, N.F.C.; Chambers, G.K.; Edinur, H.A. Population data of 23 Y chromosome STR loci for the five major human subpopulations of Ghana. *Int. J. Leg. Med.* **2020**, *134*, 1313–1315. [[CrossRef](#)]
25. Purps, J.; Siegert, S.; Willuweit, S.; Nagy, M.; Alves, C.; Salazar, R.; Angustia, S.M.; Santos, L.H.; Anslinger, K.; Bayer, B.; et al. A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Sci. Int. Genet.* **2014**, *12*, 12–23. [[CrossRef](#)]
26. Della Rocca, C.; Cannone, F.; D’Atanasio, E.; Bonito, M.; Anagnostou, P.; Russo, G.; Barni, F.; Alladio, E.; Destro-Bisol, G.; Trombetta, B.; et al. Ethnic fragmentation and degree of urbanization strongly affect the discrimination power of Y-STR haplotypes in central Sahel. *Forensic Sci. Int. Genet.* **2020**, *49*, 102374. [[CrossRef](#)]
27. Palha, T.; Gusmão, L.; Ribeiro-Rodrigues, E.; Guerreiro, J.F.; Ribeiro-Dos-Santos, A.; Santos, S. Disclosing the genetic structure of Brazil through analysis of male lineages with highly discriminating haplotypes. *PLoS ONE* **2012**, *7*, e40007. [[CrossRef](#)]
28. Pontes, M.L.; Cainé, L.; Abrantes, D.; Lima, G.; Pinheiro, M.F. Allele frequencies and population data for 17 Y-STR loci (AmpF $\ell$ STR $^{\text{®}}$  Y-filer $^{\text{™}}$ ) in a Northern Portuguese population sample. *Forensic Sci. Int.* **2007**, *170*, 62–67. [[CrossRef](#)]
29. Aboukhalid, R.; Bouabdellah, M.; Abbassi, M.; Bentayebi, K.; Elmzibri, M.; Squalli, D.; Amzazi, S. Haplotype frequencies for 17 Y-STR loci (AmpFISTRY-filer) in a Moroccan population sample. *Forensic Sci. Int. Genet.* **2010**, *4*, e73–e74. [[CrossRef](#)]
30. Gusmão, L.; Butler, J.M.; Carracedo, A.; Gill, P.; Kayser, M.; Mayr, W.R.; Morling, N.; Prinz, M.; Roewer, L.; Tyler-Smith, C.; et al. DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. *Int. J. Leg. Med.* **2006**, *120*, 191–200. [[CrossRef](#)]
31. Rosa, A.; Ornelas, C.; Brehm, A.; Villems, R. Population data on 11 Y-chromosome STRs from Guiné-Bissau. *Forensic Sci. Int.* **2006**, *157*, 210–217. [[CrossRef](#)]

32. Costa, R.; Fadoni, J.; Amorim, A.; Cainé, L. Y-STR Databases-Application in Sexual Crimes. *Genes* **2025**, *16*, 484. [[CrossRef](#)]
33. Excoffier, L.; Laval, G.; Schneider, S. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **2007**, *1*, 47–50. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.