

Article

Complete Mitochondrial DNA Genome Variation in the Swedish Population

Kimberly Sturk-Andreaggi ^{1,2,3,*}, Martin Bodner ⁴, Joseph D. Ring ^{2,3}, Adam Ameer ¹, Ulf Gyllensten ¹, Walther Parson ^{4,5}, Charla Marshall ^{2,5} and Marie Allen ^{1,*}

¹ Department of Immunology Genetics and Pathology, Uppsala University, 751 08 Uppsala, Sweden; adam.ameur@igp.uu.se (A.A.); ulf.gyllensten@igp.uu.se (U.G.)

² Armed Forces Medical Examiner System's Armed Forces DNA Identification Laboratory (AFMES-AFDIL), Dover Air Force Base, DE 19902, USA; charla.k.marshall.civ@health.mil (C.M.)

³ SNA International, LLC, Alexandria, VA 22314, USA

⁴ Institute of Legal Medicine, Medical University of Innsbruck, 6020 Innsbruck, Austria; martin.bodner@i-med.ac.at (M.B.); walther.parson@i-med.ac.at (W.P.)

⁵ Forensic Science Program, The Pennsylvania State University, University Park, State College, PA 16801, USA

* Correspondence: kimberly.s.andreaggi.ctr@health.mil (K.S.-A.); marie.allen@igp.uu.se (M.A.)

Abstract: The development of complete mitochondrial genome (mitogenome) reference data for inclusion in publicly available population databases is currently underway, and the generation of more high-quality mitogenomes will only enhance the statistical power of this forensically useful locus. To characterize mitogenome variation in Sweden, the mitochondrial DNA (mtDNA) reads from the SweGen whole genome sequencing (WGS) dataset were analyzed. To overcome the interference from low-frequency nuclear mtDNA segments (NUMTs), a 10% variant frequency threshold was applied for the analysis. In total, 934 forensic-quality mitogenome haplotypes were characterized. Almost 45% of the SweGen haplotypes belonged to haplogroup H. Nearly all mitogenome haplotypes (99.1%) were assigned to European haplogroups, which was expected based on previous mtDNA studies of the Swedish population. There were signature northern Swedish and Finnish haplogroups observed in the dataset (e.g., U5b1, W1a), consistent with the nuclear DNA analyses of the SweGen data. The complete mitogenome analysis resulted in high haplotype diversity (0.9996) with a random match probability of 0.15%. Overall, the SweGen mitogenomes provide a large mtDNA reference dataset for the Swedish population and also contribute to the effort to estimate global mitogenome haplotype frequencies.

Keywords: mitochondrial DNA; next generation sequencing; whole genome sequencing; nuclear mitochondrial DNA segment (NUMT); Sweden; population data



Citation: Sturk-Andreaggi, K.; Bodner, M.; Ring, J.D.; Ameer, A.; Gyllensten, U.; Parson, W.; Marshall, C.; Allen, M. Complete Mitochondrial DNA Genome Variation in the Swedish Population. *Genes* **2023**, *14*, 1989. <https://doi.org/10.3390/genes14111989>

Academic Editor: Emiliano Giardina

Received: 3 September 2023

Revised: 4 October 2023

Accepted: 19 October 2023

Published: 25 October 2023



Copyright: © 2023 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

Mitochondrial DNA (mtDNA) sequencing plays an important role in forensic case-work, especially for the identification of human remains. Attributes of mtDNA such as a high copy number relative to nuclear DNA and matrilineal inheritance make the analysis a useful tool in decades-old cases that involve poor-quality DNA samples and no direct reference samples (e.g., [1–5]). Until recently, Sanger-type sequencing has been the “gold standard” for mtDNA analysis. However, this approach is expensive and labor-intensive, particularly for larger sample sets. For this reason, as well as for legal restrictions, forensic laboratories target the non-coding control region (CR) or the smaller hypervariable segments (HVSs) of the mtDNA rather than the entire mitochondrial genome (mitogenome). This results in limited discrimination power for mtDNA analysis due to common haplotypes observed in these smaller target regions. The most common European HVS (nps 16,024–16,365, 73–340) haplotype “263G 315.1C” is observed in approximately 1 in every 15 West Eurasian individuals in the European DNA Profiling

Group (EDNAP) mtDNA Population (EMPOP) database (v4/Release 13) [6]. The use of the entire CR (nps 16,024–16,569, 1–576) slightly improves the discrimination power for mtDNA analyses, but full mitogenome data are required to completely resolve common haplotypes [7–9].

Mitogenome data can be efficiently and cost-effectively produced in forensic laboratories using next generation sequencing (NGS) techniques. Commercial NGS kits offer enrichment approaches that use small amplicons to target the mitogenome, which are amenable to a range of sample qualities (e.g., [10,11]), including reference-type samples and degraded DNA samples. Benchtop platforms such as the Verogen MiSeq FGx Forensic Genomics System (San Diego, CA, USA) and the Ion Torrent S5 System (Thermo Fisher Scientific, Waltham, MA, USA) allow practicable access to NGS technology. Although mtDNA analysis using NGS has been implemented in forensic laboratories [11–14], the usage of mitogenome data is also limited due to insufficient information on haplotype frequencies. At the time of publication, there are fewer than 5000 forensic-quality mitogenomes searchable in Release 13 of the EMPOP v4 database [6]. Without appropriate haplotype frequency information, the requisite match statistics that provide the evidentiary weight of the evidence in a forensic case cannot be accurately estimated [15].

To date, only two Swedish mtDNA population datasets are available for forensic use [16,17]. However, each high-quality dataset consists of fewer than 300 samples and only analyzed the mtDNA CR. Both of the studies that generated these datasets used Sanger-type sequencing for data generation with Lembring et al. targeting the ~600 bps of the two HVS regions [16], while Tillmar et al. analyzed the full CR [17]. These Swedish mtDNA datasets showed lower HVS/CR haplotype diversity compared to those from other populations in Europe. With the analysis of the entire mitogenome, the number of unique haplotypes could increase dramatically, potentially to a 100% resolution of common HVS/CR haplotypes, as observed in other West Eurasian populations [7–9,18]. Studies have, moreover, concluded that there is no significant substructure within Sweden based on mtDNA HVS/CR and Y-chromosomal markers [16,17,19,20], which has also been confirmed with autosomal DNA studies [21,22]. An exception to the observed homogeneity of the Swedish population is seen in the indigenous Saami located in northern Sweden [16,17,23]. Specifically, mtDNA haplogroup U5b1 is observed at a much higher frequency in Saami than in the larger Swedish population [24]. Individuals from northern Sweden, including the Saami, have been shown to be genetically similar to Finnish people [25]. Swedes from the rest of the country are more genetically similar to other neighboring European populations to the south and west, such as Germans, Danes, and Norwegians [26–28].

In this study, a further and more detailed evaluation of mtDNA variation in Sweden at the highest resolution based on high-quality mitogenome haplotypes from whole genome sequencing (WGS) data was performed [29]. WGS datasets are a largely untapped resource for mitogenomes. The use of these data allows for a rapid expansion of mitogenome reference databases with no additional laboratory processing costs, though there would be some cost associated with bioinformatic analyses and review of the data. Because the sequencing is untargeted, the greatest challenge to the mitogenome analysis of WGS data is nuclear mtDNA segments (NUMTs) [30]. Since reads from both mtDNA and NUMTs co-align to the mtDNA reference genome due to their homology, it can be difficult to distinguish between authentic point heteroplasmies (PHPs) and low-level variants associated with NUMTs [31–33]. The feasibility of high-quality mitogenome haplotype generation from WGS data was previously assessed by Sturk-Andreaggi et al. [29]. In this previous study, the authors demonstrated that reliable mitogenome haplotypes could be generated from WGS data using a 10% minimum variant frequency (VF) threshold [29]. Applying this frequency threshold, NUMT interference was negligible as the proportion of heteroplasmic haplotypes, and the maximum number of PHPs observed per individual, were consistent with previous high-quality mitogenome datasets [7,18,34]. Based on these results, the SweGen mitogenomes produced from WGS data with a 10% frequency threshold are appropriate for population-level investigations and haplotype frequency estimations

for forensic purposes. The more than 900 Swedish haplotypes described in this study will substantially increase the number of high-quality mitogenomes available for forensic use, greatly improving the significance of mtDNA match statistics, which are dependent on the database size [15,35,36].

2. Materials and Methods

2.1. Samples

The SweGen WGS data [25] for the 942 Swedish individuals from the TwinGene project [37] were analyzed in order to generate mitogenome haplotypes. These unrelated individuals were initially selected for the SweGen project as a dataset representative of the population density distribution across Sweden. However, no regional information within Sweden for these samples was available for this study.

2.2. Haplotype Generation

WGS data were previously generated from the SweGen individuals as described in [25]. The WGS reads that aligned to the revised Cambridge Reference Sequence (rCRS) [38,39] were previously analyzed through a robust analysis pipeline to overcome NUMT interference [29]. In short, analysis was performed in CLC Genomics Workbench v12.0.1 with AQME v2.1.1 tools [40], requiring a minimum read depth of 100X and a 10% minimum VF threshold for variant calling. Haplotypes were subjected to independent reviews by at least two analysts followed by stringent quality control (QC) procedures as described in Taylor et al. [18] and Sturk-Andreaggi et al. [29]. To assist in the QC assessment of the mitogenome haplotype, the AQME Mitochondrial Haplogrouper tool predicted the mtDNA haplogroup based on Phylotree Build 17 [41,42].

In the present study, the SweGen samples with five or more positions below 100X (incomplete) and possible mixtures excluded from the analysis in the Sturk-Andreaggi et al. study were re-evaluated [29]. A 20X minimum read depth was applied along with the 10% frequency threshold for these previously incomplete samples when requiring 100X coverage of the mitogenome. NUMT interference was suspected for at least five of the seven of the samples classified as possible mixtures in [29] due to their low average read depths (<700X) and reduced proportions of mtDNA in relation to nuclear DNA (<0.012%). Therefore, the data for these seven samples with full 100X mitogenome coverage were also reanalyzed with the 10% frequency threshold. If any previously excluded mitogenomes were considered high-quality after re-evaluation, they were reviewed by two analysts and ultimately added to the 917 samples that were included in the initial SweGen mtDNA dataset [29].

To ensure the SweGen data represented a “random” sampling of the Swedish population, shared haplotypes (ignoring indels and heteroplasmy) were identified, and haplotype groupings were evaluated for relatedness [43]. Kinship coefficients were calculated based on the nuclear data from Ameur et al. [25] to determine if any samples with shared haplotypes were related as parent-offspring, siblings, or second-degree relatives (e.g., avuncular, grandparent/grandchild). There were two first-degree relatives with the same mitogenome haplotype identified in Sturk-Andreaggi et al. [29], and one sample of this maternally related pair was previously removed from the SweGen data included in this study. This assessment was repeated after reanalysis of the incomplete and possibly mixed samples to ensure no additional related pairs were present. If any additional related groups were found, only one mitogenome haplotype from related individuals was included in the final SweGen dataset. This additional QC measure was implemented to avoid potential bias introduced by the inclusion of close (maternal) relatives in an mtDNA population sample [43].

The final dataset was then submitted to EMPOP for additional QC checks and confirmation of the AQME haplogroup predictions [6,44].

2.3. Data Analysis

Outputs from the CLC Genomics Workbench were exported to Excel (Microsoft, Redmond, WA, USA), and analysis metrics were calculated, including average VF and

average read depth. The metrics and other details were stored in Access (Microsoft), which was used to determine summary metrics. The distribution of PHPs, NUMT variants, and average read depth across the mitogenome was visualized using the circlize package v0.4.10 in R version 4.0.2 software [45,46]. Finalized haplotypes were uploaded to the Laboratory Information Systems Applications (LISA; Future Technologies Inc., Fairfax, VA, USA) database. Forensic and population genetic statistics, such as random match probability (RMP) and power of discrimination (haplotype diversity), were calculated based on pairwise comparisons performed in LISA. The comparisons, which ignored all indels, were performed using two approaches for matching: literal (e.g., a Y only matches a Y at a position) and pattern (e.g., a Y matches a C or T at a position). Additionally, summary statistics were calculated based on the two HVS regions (HVS1 nps 16,024–16,365 and HVS2 nps 73–340), CR (16,024–16,569, 1–576), and the entire mitogenome (1–16,569).

The SweGen dataset was compared to the two previous Swedish population datasets [16,17] at overlapping ranges (HVS for Lembring et al. [16] and CR for Tillmar et al. [17]). To ensure that the comparison of haplogroup composition was consistent (i.e., not impacted by the targeted region), haplogroups were assigned for HVS and CR range haplotypes using EMPOP v4 [44,47] for all datasets as applicable. Chi-squared tests with Yates's correction were used to assess whether the haplogroup distributions for the three datasets were statistically different. An alluvial diagram was generated using RAWGraphs [48] to visualize the impact of the target region on haplogroup assignments for the SweGen haplotypes.

3. Results

3.1. Overall Performance

The mtDNA reads extracted from WGS data of 942 Swedish individuals were analyzed to produce mitogenome haplotypes for population genetics and forensic purposes. A total of 917 haplotypes were previously identified as being of high-quality [29]. Of these, 858 had complete 100X coverage across the mitogenome, while the remaining 59 were nearly complete, with less than five positions below the 100X threshold. A lower read depth threshold (20X) was applied to 17 previously incomplete haplotypes (five or more positions with less than 100 reads). As a result of this reanalysis, the full mitogenomes of 16 samples were covered with at least 20 reads. One incomplete haplotype was still observed with eight positions below the 20X read depth threshold. This sample was previously discussed in Sturk-Andreaggi et al. [29], as the five J1c2 haplogroup variants preceding the HVS2 C-stretch (nps 185–295) appeared to impact the coverage in this region. Nevertheless, two other incomplete haplotypes with the same haplogroup produced full mitogenome coverage with the 20X threshold, thus ensuring representation of this haplogroup in the dataset. The seven samples flagged as possible mixtures in [29] were also re-evaluated using the 10% threshold frequency. As a result, one haplotype, which had complete 100X coverage, was reclassified as a single-source profile since no mixed positions exceeded the 10% frequency threshold, except for one high-frequency (~50%) PHP, and all other metrics (e.g., average VF) were consistent with those of other single-source samples. The classification for the other six possible mixtures did not change, as these haplotypes could not be confidently classified as single-source. Additionally, the sample identified in [29] as a first degree relative of another sample in the dataset with a shared haplotype remained excluded from the dataset. No other maternally related individuals were identified in the SweGen samples included in the final mtDNA dataset. After reassessment, 934 forensic-quality Swedish mitogenomes were characterized (Table 1).

Table 1. Summary of analysis metrics for the final SweGen mitochondrial DNA (mtDNA) dataset. Haplotypes are separated based on mitochondrial genome (mitogenome) coverage at 100X with complete, nearly complete (four or less positions below 100X), and incomplete (more than four positions below 100X) classifications. The average (avg) major nucleotide frequency is determined for all variant positions (All), as well as excluding both length and point heteroplasmy (No HP). The “Complete” 100X coverage group includes the one haplotype classified as a possible mixture in [29]. The data subset is a group of 100 SweGen samples that were selected as a smaller representation of the overall SweGen data. The subset was used to evaluate the distribution of coverage at each position in the mitogenome.

100X Coverage	Sample Count	Avg Mapped mtDNA Reads	Avg Read Depth	Avg Major Nucleotide Frequency	
				All	No HP
Complete	859	297,369.6	2363.8	98.0	99.6
Nearly Complete	59	82,393.4	652.5	97.5	99.0
Incomplete	16	69,264.9	538.4	95.7	98.4
<i>All</i>	934	279,882.2	2224.4	98.0	99.5
<i>Subset</i>	100	288,141.7	2271.6	97.9	99.5

Mitogenome coverage at 100X was influenced by the number of mapped mtDNA reads, which had a linear relationship ($R^2 = 0.998$) to the average read depth (Figure S1a). Overall, almost 280,000 reads on average were used to generate a mtDNA haplotype, ranging from approximately 25,000 to over 1.2 million reads. Complete 100X coverage of the mitogenome was obtained from samples with average read depths as low as 502X; however, the majority of the complete haplotypes had average read depths greater than 1275X (Figure S1b). With the exception of seven outliers, all nearly complete and incomplete haplotypes had average read depths of less than 1200X. In the study by Sturk-Andreaggi et al. [29], the average read depth was correlated to the proportion of mtDNA reads in relation to nuclear DNA reads in the WGS data rather than total WGS reads. Therefore, the nearly complete and incomplete samples had a lower mtDNA proportion in the WGS data (i.e., relative mtDNA copy number [49–51]) than the samples with complete 100X haplotypes.

The distribution of coverage was evaluated for 100 samples, which were representative of the overall SweGen dataset (Table 1). The average read depth at each position was calculated based on the read depths observed in the data subset. This overall average read depth was 2328X and ranged from 411X to 2805X (Figure 1). Read depths were relatively consistent across the mitogenome, with approximately 10% variation in the overall average read depth and an interquartile range of 240X (Figure S2). However, there were 855 positions (5.2% of the mitogenome) with substantially lower read depths (less than 1883X). Most (760; 88.8%) of these low-coverage positions were localized to four regions: nps 217–598 (373), 3490–3634 (65), 10,891–11,043 (120), and 13,674–14,071 (202). The large proportion (43.6%) of the positions below 100X in the nearly complete and incomplete haplotypes occurred across nps 217–598 due to the presence of the three C-stretches (nps 303–309, 456–469, 568–576). In fact, several C-stretches are present in all four low-coverage regions, and HiSeq chemistry has previously been shown to exhibit poor sequencing in homopolymer regions [52]. As a result, reads in these regions were of generally poor quality, and thus read depths were reduced. Sequencing errors specific to polycytosine residues combined with the high-stringency mapping parameters employed to eliminate NUMTs are likely to have contributed to the low coverage observed in these regions.

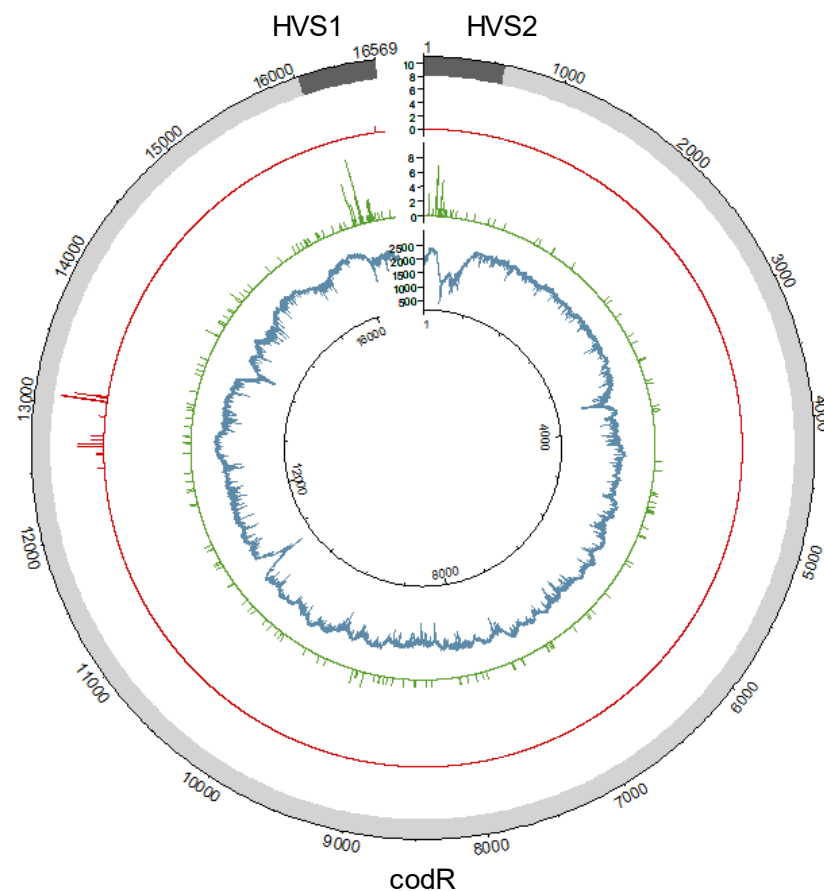


Figure 1. Distribution of the average read depth observed in a subset of 100 haplotypes (inner plot; blue), point heteroplasmies observed with the 10% frequency threshold (middle plot; green), and variants associated with nuclear mitochondrial DNA segments detected at 10% or higher frequencies (outer plot; red).

The average frequency of the major nucleotide at all variant positions (average VF) was used to assess the quality of the haplotypes. This value is expected to exceed 98% in single-source, high-quality mitogenomes when PHP and LHP positions are ignored based on the level of observed background noise in the SweGen data. The average VF (excluding PHP and LHP) averaged 99.5% across the 934 SweGen mitogenomes (Table 1). Moreover, less than 1% of the SweGen haplotypes had average VFs below 98% (two complete, three nearly complete, and four incomplete). As shown in Figure S3, slightly lower average VFs were observed in the nearly complete (99.0%) and incomplete (98.5%) haplotypes compared to complete haplotypes (99.6%). This likely relates to the reduced average read depths observed in the non-complete (nearly complete and incomplete) haplotypes and increased observation of NUMT interference. In fact, the 31 NUMT-associated variants detected above the 10% frequency threshold were exclusively observed in haplotypes without full 100X coverage of the mitogenome. There were 13 NUMT variants in eight nearly complete haplotypes and an additional 27 NUMT variants detected in five incomplete haplotypes, all with average read depths of less than 550X. NUMT variants were detected above the 10% frequency threshold at multiple positions between nps 12,501 and 13,105 as well as at np 16,496 (Figure 1), which is consistent with the NUMT hotspots identified in [29]. Although NUMT-associated variants were not observed above the 10% frequency threshold in most haplotypes, NUMT interference was detectable above background noise (2% frequency) in nearly 40% (373) of the 934 SweGen mitogenomes. The presence of NUMT reads in the mtDNA alignments reduces the frequency of the major nucleotide at variant positions, thereby reducing the average VF. Since NUMT interference was correlated with lower

average read depths, it is not unexpected that slightly lower average VFs were observed in the low-coverage data of the nearly complete and incomplete haplotypes (Table 1).

3.2. Variants and Heteroplasmy

Overall, 23,857 variants including substitutions and indels, both homo- and heteroplasmic, were observed in the 934 SweGen mitogenome haplotypes, averaging 25.5 variants per haplotype. There was one sample with no differences from rCRS, belonging to haplogroup H2a2a1, and a maximum of 55 variants (i.e., differences from the rCRS) was detected in a T2f1a1 haplotype, which included nine deletions at nps 8281–8289. A total of 21,442 substitutions were observed at 1499 positions across the mitogenome. There were 355 deletions and 2060 insertions reported at 25 positions.

There were 236 PHPs observed in 205 (21.9%) of the 934 SweGen mitogenome haplotypes. Most (177; 86.3%) of these haplotypes contained a single PHP. Two PHPs were observed in 25 haplotypes, and 3 haplotypes had three PHPs, which was the maximum number of PHPs observed in a single haplotype. The 236 PHPs were detected at 189 nucleotide positions (Figure 1). The majority (146; 61.9%) of the PHPs were detected in the coding region (codR), and these PHPs were located at 142 different positions. Of the 146 codR PHPs, 94.5% ($n = 138$) were observed once, and 4 were seen twice. In contrast, there were 90 CR PHPs detected at just 47 different positions. There were 13 CR positions at which PHPs were observed in more than one haplotype, ranging from two to nine occurrences. The most frequently observed PHP was 16192Y ($n = 9$), followed by 152Y ($n = 7$) and 16093Y ($n = 6$). Other heteroplasmic hotspots were observed at nps 146, 204, and 16,189 (all occurring five times). Overall, 181 (76.7%) of the PHPs involved transitions with 80 G/A (R) and 101 C/T (Y). There were eight heteroplasmic transversions: 593K, 955M, 4385W, 5625W, 13718S, 14020K, 16294S, and 16524W. There was one PHP (13105R with the G detected at 15.3%) reported in one haplotype, even though 13105G was also identified as a NUMT variant and was consequently removed from two incomplete haplotypes (with the G observed at 19.0% and 12.8%). As discussed above, these incomplete haplotypes had substantially lower average read depths (201X and 305X), resulting in increased NUMT interference, even above the 10% frequency threshold. The NUMT-associated 13105A was observed in-phase with other NUMT variants in this hotspot region [29]. Conversely, the haplotype that included the 13105R PHP had a high average read depth (>3000X) and no indication of NUMT interference, even below the 10% frequency threshold.

Length heteroplasmy (LHP) was observed in 682 (73%) of the 934 SweGen mitogenome haplotypes. Although most of the haplotypes with LHP ($n = 523$) showed length variation in only one region, a portion of haplotypes had two ($n = 150$) or three ($n = 9$) regions exhibiting LHP. The region in which LHP was most often detected was the HVS2 C-stretch ($n = 599$), followed by the HVS1 C-stretch ($n = 172$), HVS3 C-stretch ($n = 46$), and AC-stretch at nps 513–524 ($n = 17$). Thirteen haplotypes displayed LHP in the C-stretches preceding nps 460 ($n = 4$), 960 ($n = 5$), 5899 ($n = 2$), and 8276 ($n = 2$). These LHP regions are well documented and typically ignored for forensic comparisons [6,15,44,53]. The LHP in the remaining three haplotypes was in other homopolymeric regions at which LHP is less frequently observed (i.e., the C-stretches preceding nps 356, 498, and 7471). However, the length variation observed in these regions was only evident in two haplotypes (i.e., the inclusion of a 356.1C and 498.1C) due to the reporting of the major length molecule. It is important to note that the reporting of PHPs in and around homopolymer regions was difficult due to post-homopolymer errors observed in HiSeq X sequencing [52]. As a result, low-level variants were consistently observed in the polyadenine stretches preceding the C-stretches of the HVS1 and HVS2 regions. In fact, 302M was observed in nearly all haplotypes, as was 16183M in samples with the 16189C variant, regardless of the complexity of the LHP. As a result, no haplotypes included a 302M or 16183M (or 16182M, etc.), and the major nucleotide at the position was reported. However, other PHPs in other flanking homopolymer regions were included in the haplotype when no or minimal length variation was observed. SweGen mitogenome haplotypes included 17 PHPs in or flanking the HVS1

C-stretch (at nps 16,188, 16,189, 16,192, 16,193, and 16,195). Additionally, PHPs at np 316 at the end of the HVS2 C-stretch and at np 955 at the beginning of another C-stretch at nps 956–960 were reported.

3.3. Population Composition

A total of 821 mitogenome haplotypes were observed in the final SweGen mtDNA dataset (934 samples) when PHPs were treated literally and indels were ignored (Table 2). The most common mitogenome haplotype was observed seven times in the dataset (0.7% of the population), regardless of the match approach (i.e., literal or pattern) for comparisons (Table S1). This haplotype was assigned to haplogroup H2a1n (146C 263G 309.xC 315.1C 750G 951A 4659A 8860G 15326G 16354T). The next most common haplotype, which was seen six times (0.6%) in the dataset, was assigned to haplogroup U8a1a1a (73G 263G 282C 309.1C 315.1C 750G 1438G 1811G 2706G 3738T 4129G 4769G 5240G 6392C 6455T 7028T 7055G 8860G 9365T 9698C 10733T 11150A 11467G 11719A 12135A 12308G 12372A 13145A 14766T 15326G 16209C 16342C). A haplotype observed five times (0.5%) was assigned to another U haplogroup (U5b1b1a) and comprised the following variants: 73G 150T 263G 309.xC 315.1C 750G 1438G 2706G 3197C 4769G 5656G 7028T 7385G 7768G 8860G 9477A 10927C 11467G 11719A 12308G 12372A 12618A 13617C 14182C 14766T 15326G 16144C 16189C 16270T. There were five haplotypes seen four times in the SweGen dataset (0.43% of the population), including three haplotypes assigned to H haplogroups (H1c1a, H2a5, H6c) plus two haplotypes belonging to haplogroups V and W1a (Table S1). The remaining 67 other shared haplotypes (literal match approach and ignoring indels) were seen only two or three times in the SweGen population (Table S1). The observed haplotype frequencies correspond to a haplotype diversity of 0.9996 with the literal approach and 0.9993 when the pattern match approach is applied (Table 2).

Table 2. Haplotype diversity estimates for the SweGen population ($n = 934$). The number of haplotypes was determined based on pairwise comparison performed with both literal and pattern matching for point heteroplasmies. The summary statistics including random match probabilities (RMP) were calculated for three different ranges: hypervariable segments (HVS; nps 16,024–16,365, 73–340), control region (CR; nps 16,024–16,569, 1–576), and the entire mitochondrial genome (mtG; nps 1–16,569). Indels were ignored in all comparisons.

Range	Match Type	Total Haplotypes	Unique Haplotypes (Proportion Unique)	Observed RMP (%)	Empirical RMP (%)	Haplotype Diversity
HVS	Literal	531	393 (74.0%)	0.71	0.60	0.9940
	Pattern	473	324 (68.5%)	1.14	1.03	0.9897
CR	Literal	583	447 (76.7%)	0.47	0.37	0.9963
	Pattern	524	372 (71.0%)	0.67	0.57	0.9943
Mitogenome	Literal	821	746 (90.9%)	0.15	0.04	0.9996
	Pattern	750	629 (83.7%)	0.17	0.07	0.9993

The mitogenome haplogroup composition of the SweGen dataset is similar to that of other Western European populations (Figure 2). Most haplotypes were assigned to haplogroup H (44.6%), which included subhaplogroups H1 (1.9%), H2a1n (1.1%), and H1b (1.0%). Haplogroup U was also observed in a large proportion of the SweGen samples (18.0%), with 12 (1.3%) haplotypes assigned to both U5b1b1a and U8a1a1a subhaplogroups (Table S1). In the SweGen dataset, U5b1 subhaplogroups were assigned to 3.9% ($n = 36$) of the haplotypes. Haplogroups T (10.6%) and J (10.4%) were also observed at high frequencies in the population, specifically subhaplogroups T1a1 (1.4%), T2b (1.2%), J1b1a1 (1.2%), and J2a1a1 (1.1%). Other common European haplogroups K (5.9%), V (3.0%), I (2.7%), HV (1.6%), X (1.4%), and W (1.0%) were represented (Table S1). There were a few

haplotypes assigned to rarer European haplogroups N1a1a1a2 (0.2%) and Z1a1a (0.3%). In addition, several non-European haplogroups were observed once in the dataset (0.1%), including D4i, G3a3, and M1a1b1. These specific D and G haplogroups are common in Eastern Asia [54], while M1a1 haplogroups are predominantly observed in North and East Africa [55].

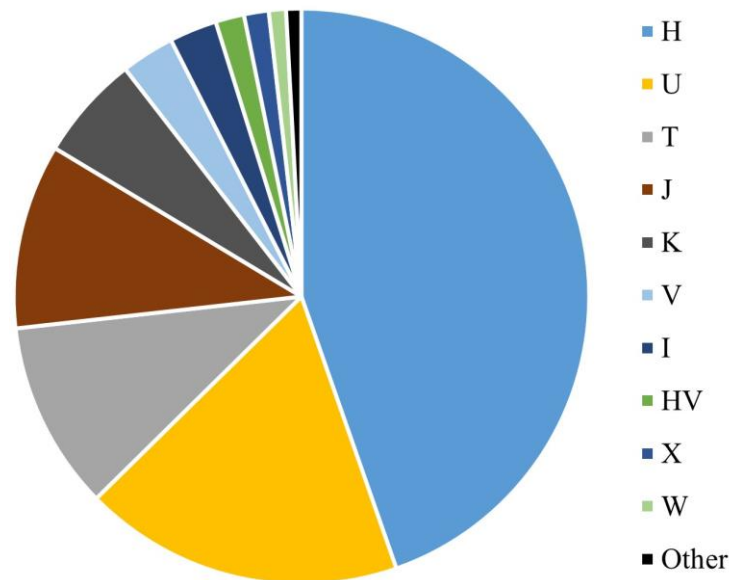


Figure 2. Macrohaplogroup breakdown for the SweGen mitochondrial genome data. The “Other” category (black) includes haplogroups D, G, M, N, and Z.

The haplogroups observed in the SweGen dataset were consistent with previously published Swedish mtDNA data [16,17]. However, differences in the sequencing range of each Swedish dataset impacted the level of haplogroup refinement. For example, there was a high proportion of R0 haplogroups (18.4%) observed in the HVS data of Lembring et al. [16], whereas a greater proportion of H haplotypes (44.6%) were observed in the SweGen mitogenome dataset (Table S1). This is explained by the increased haplogroup resolution possible with sequencing larger portions of the mitogenome, resulting in the refinement of the R0 haplotypes from HVS data to H haplogroups with entire mitogenomes (Figure S4). A comparison of updated haplogroup predictions based on HVS regions confirmed this since the updated haplogroup predictions utilized the same prediction tool and target region for the three Swedish datasets (Figure 3). This analysis showed no significant difference in HVS haplogroup frequencies between the three Swedish population datasets ($p > 0.13$), although the SweGen still had a slightly higher proportion of H haplogroups than R0, in contrast to the other two Swedish datasets. When comparing the CR haplogroup frequencies between the Tillmar et al. [17] and SweGen datasets, a range that was not sequenced for the Lembring et al. dataset [16], there was no significant difference ($p = 0.88$; Figure S5). Minor differences in haplogroup proportions are not unexpected and likely the result of sampling variation as well as the size of the populations, as the SweGen dataset is over three times larger than the other two Swedish datasets.

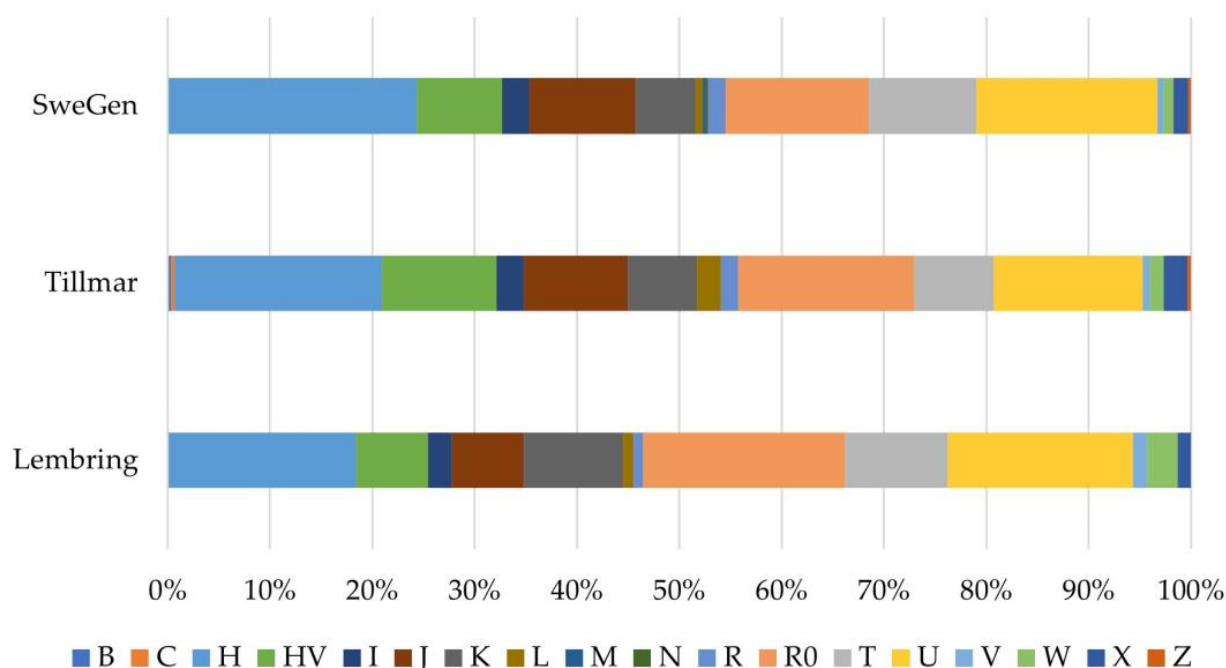


Figure 3. Macrohaplogroup breakdown of the SweGen, Tillmar et al. [17], and Lembring et al. [16] Swedish datasets based on the hypervariable segment regions (nps 16,024–16,365, 73–340), the largest overlapping range for all three datasets, using EMPOP v4 for haplogroup assignment.

4. Discussion

The observed heteroplasmy in this dataset was consistent with the findings in previous studies that produced forensic-quality mitogenome data [7,18], providing the basis for applying the 10% minimum VF threshold, as discussed in Sturk-Andreaggi et al. [29]. Furthermore, the SweGen mitogenomes presented here are notably different than those from questionable datasets that contain high numbers of PHPs per haplotype, codR heteroplasmic hotspots, and high frequencies of heteroplasmic transversions [56]. There were six heteroplasmic hotspots (observed five times or more), all of which were observed in the CR and were previously reported by Irwin et al. [57]. Of note, 16192Y was observed more frequently in this dataset compared to other studies. Due to its presence in a homopolymer region, heteroplasmy detected at np 16,192 was scrutinized to ensure that reported PHPs were not caused by misalignment. It is therefore likely that the higher PHP rate at np 16,192 seen in the SweGen dataset is the result of a high frequency of U5 haplogroups (10.3%) in which 16192T is a diagnostic mutation [42]. In terms of LHP, it is important to understand the impact of the sample preparation method and sequencing platform on the homopolymeric regions when analyzing NGS data [58]. These factors may impact the interpretation and reporting of these regions, such as the inclusion of PHPs or reporting of the major length molecule [53]. Procedures then need to be adjusted based on the methods used for the data generation to ensure the production of high-quality haplotypes.

The mitogenome haplotype diversity observed in the SweGen dataset (0.9996) is similar to that of other Western European populations [59,60]. Compared with the U.S. populations of European ancestry [7,18], the Swedes exhibit a higher proportion of shared haplotypes. As expected, the number of unique haplotypes decreases when considering smaller ranges (from 629 to 372 for the CR and 324 for the HVS), resulting in lower RMPs and haplotype diversities than for the entire mitogenome. The CR haplotype diversity for the SweGen population (0.9963 literal, 0.9943 pattern) was consistent with that of the Tillmar et al. Swedish CR dataset (0.996) [17]. In addition, the HVS population statistics for the SweGen data using the pattern approach (RMP of 1.14% and haplotype diversity of 0.9897) were similar to the findings of Lembring et al. (RMP of 1.39% and haplotype diversity of 0.9895) [16]. Regardless of range, when the pattern match approach was used

for comparisons, the RMPs increased, while the haplotype diversities decreased slightly due to the increase in non-unique haplotypes.

Ameur et al. noted a small representation of Finnish ancestry in the nuclear DNA of the SweGen individuals [25], which was further supported by the mtDNA analyses. Haplogroups U5b1 and W1a, which are common within Finland (both ~6%) [61], were observed in 3.9% and 1.7% in the SweGen dataset, respectively. Additionally, haplogroup U5b1 was observed in higher proportions in northern Swedes [17], specifically in the Saami population [24]. Interestingly, U5b1 has also been documented in Viking Age Scandinavians [62]. We expected to observe this haplogroup in the SweGen dataset as it was sampled to represent the genetic diversity within the Swedish population, though no regional information was known for these samples. Ameur et al. also detected an East Asian bias observed in the SweGen nuclear data [25], which was also seen in the mitogenome data with two East Asian mitogenome haplotypes observed (i.e., D4i and G3a3 haplotypes). The presence of non-European maternal lineages in the SweGen dataset is not unexpected given the findings from previous studies [17,23]. In particular, Lappalainen et al. showed slightly elevated frequencies of East Asian and Native American haplogroups as well as African and Near East influence [23]. The observation of non-European haplogroups in the Swedish population is likely the result of increased immigration from more distant countries during the end of last century [63].

5. Conclusions

The SweGen WGS dataset was used for the determination of mtDNA haplotype frequency estimates. The high haplotype diversity and haplogroup distribution observed in the SweGen population were consistent with existing Swedish datasets [16,17]. Moreover, the observation of haplogroups characteristic of northern Swedish and Finnish populations in the SweGen data confirmed this finding from the analysis of the nuclear DNA [25]. The nuclear DNA from the SweGen WGS data was also beneficial in this study as it was readily available for kinship analyses to identify close maternal relatives with shared mitogenome haplotypes [43]. Although it is possible to identify distant relatives from WGS data [64], it may be necessary to determine the degrees of relatedness that should be included in mtDNA reference datasets to reflect the haplotype distribution in the population properly. In the end, over 900 high-quality mitogenome haplotypes were generated from WGS data with no additional costs, except for those associated with the analysis. This large dataset of Swedish mitogenomes is now available for population genetic studies as well as forensic applications.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes14111989/s1>, Table S1: The haplogroup breakdown using the literal match approach for the 934 SweGen haplotypes included in the final mitochondrial genome dataset; Figure S1: Graphical description of the average read depths observed in the SweGen dataset; Figure S2: The distribution of read depth for the 16,569 positions of the mitochondrial genome based on the average observed in a subset of 100 representative SweGen haplotypes; Figure S3: The box-and-whisker plot presents the distribution of average variant frequency for each coverage classification group; Figure S4: Alluvial diagram representing the change in haplogroup assignments for the SweGen dataset for hypervariable segment (HVS) regions to control region (CR) and the CR to entire mitochondrial genome (mitogenome). Colored bands denote haplotypes that were assigned to different haplogroups between one region and the next region (i.e., HVS to CR, CR to mitogenome), whereas gray bands indicate no change in the haplogroup assignment.; Figure S5: Macrohaplogroup breakdown the SweGen and Tillmar et al. [17] Swedish datasets based on the control region (nps 16,024–16,365, 73–340) using EMPOP v4 for haplogroup assignment.

Author Contributions: Conceptualization, K.S.-A. and M.A.; methodology, K.S.-A.; formal analysis, J.D.R., W.P., M.B. and K.S.-A.; investigation, K.S.-A.; resources, A.A. and U.G.; data curation, K.S.-A.; writing—original draft preparation, K.S.-A.; writing—review and editing, K.S.-A., M.B. and M.A.; visualization, K.S.-A.; supervision, C.M. and M.A.; project administration, K.S.-A. and M.A.; funding

acquisition, K.S.-A., W.P. and C.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Institute of Justice, grant number NIJ-17-RO-0219. WGS was funded by Science for Life Laboratory (SciLifeLab) as a National Project, supported by the Knut and Alice Wallenberg Foundation (2014.0272), and the Swedish Research Council (PI:UG). Illumina sequencing was performed by the National Genomics Infrastructure (NGI), which is hosted by SciLifeLab in Stockholm and Uppsala. Additionally, this work received support from the European Union grant agreement number 779485-STEFA-ISFP-2016-AG-IBA-ENFSI.

Institutional Review Board Statement: The TwinGene study was conducted according to the guidelines of the Declaration of Helsinki, and the study was approved by the regional ethics committee (Regionala Etikprövningsnämnden, Stockholm, dnr 2007-644-31, dnr 2014/521-32). The use of these data for mitogenome haplotype generation was also reviewed by the Institutional Review Board Office under the U.S. Army Medical Research and Materiel Command's Office of Research Protections (IRBO-M-10185).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The SweGen mitogenome haplotypes are presented in Table S1 and are available upon request.

Acknowledgments: The authors would like to thank Cassandra Taylor, Jennifer Higginbotham, Courtney Cavagnino (SNA International, AFMES-AFDIL), and Andreas Tillmar (National Board of Forensic Medicine, Sweden) for analysis support; Jon Norris (Future Technologies Inc.), Timmathy Cambridge, and James Ross (AFMES) for technical assistance; April Higgins (SNA International), and Timothy McMahon and Col. Alice Briones (ret.) (AFMES) for administrative and logistical support. The assertions herein are those of the authors and do not necessarily represent the official position of the United States Department of Defense, the Defense Health Agency, or its entities, including the Armed Forces Medical Examiner System. Any mention of commercial products occurred for scientific transparency and should not be viewed as an endorsement of the product or manufacturer.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of the data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Nelson, K.; Melton, T. Forensic Mitochondrial DNA Analysis of 116 Casework Skeletal Samples. *J. Forensic Sci.* **2007**, *52*, 557–561. [[CrossRef](#)] [[PubMed](#)]
2. Melton, T.; Dimick, G.; Higgins, B.; Lindstrom, L.; Nelson, K. Forensic Mitochondrial DNA Analysis of 691 Casework Hairs. *J. Forensic Sci.* **2005**, *50*, 73–80. [[CrossRef](#)] [[PubMed](#)]
3. Edson, S.M.; Ross, J.P.; Coble, M.D.; Parsons, T.J.; Barritt, S.M. Naming the Dead—Confronting the Realities of Rapid Identification of Degraded Skeletal Remains. *Forensic Sci. Rev.* **2004**, *16*, 63–90. [[PubMed](#)]
4. Holland, M.M.; Parsons, T.J. Mitochondrial DNA Sequence Analysis—Validation and Use for Forensic Casework. *Forensic Sci. Rev.* **1999**, *11*, 21–50.
5. Berger, C.; Parson, W. Mini-midi-mito: Adapting the amplification and sequencing strategy of mtDNA to the degradation state of crime scene samples. *Forensic Sci. Int. Genet.* **2009**, *3*, 149–153. [[CrossRef](#)]
6. Parson, W.; Dür, A. EMPOP—A forensic mtDNA database. *Forensic Sci. Int. Genet.* **2007**, *1*, 88–92. [[CrossRef](#)]
7. Just, R.S.; Scheible, M.K.; Fast, S.A.; Sturk-Andreaggi, K.; Röck, A.W.; Bush, J.M.; Higginbotham, J.L.; Peck, M.A.; Ring, J.D.; Huber, G.E.; et al. Full mtGenome reference data: Development and characterization of 588 forensic-quality haplotypes representing three U.S. populations. *Forensic Sci. Int. Genet.* **2015**, *14*, 141–155. [[CrossRef](#)]
8. King, J.L.; LaRue, B.L.; Novroski, N.M.; Stoljarova, M.; Seo, S.B.; Zeng, X.; Warshauer, D.H.; Davis, C.P.; Parson, W.; Sajantila, A.; et al. High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. *Forensic Sci. Int. Genet.* **2014**, *12*, 128–135. [[CrossRef](#)]
9. Bodner, M.; Amory, C.; Olivieri, A.; Gandini, F.; Cardinali, I.; Lancioni, H.; Huber, G.; Xavier, C.; Pala, M.; Fichera, A.; et al. Helena's Many Daughters: More Mitogenome Diversity behind the Most Common West Eurasian mtDNA Control Region Haplotype in an Extended Italian Population Sample. *Int. J. Mol. Sci.* **2022**, *23*, 6725. [[CrossRef](#)]
10. Holt, C.L.; Stephens, K.M.; Walichiewicz, P.; Fleming, K.D.; Forouzmand, E.; Wu, S. Human Mitochondrial Control Region and mtGenome: Design and Forensic Validation of NGS Multiplexes, Sequencing and Analytical Software. *Genes* **2021**, *12*, 599. [[CrossRef](#)]

11. Cihlar, J.C.; Amory, C.; Lagace, R.; Roth, C.; Parson, W.; Budowle, B. Developmental Validation of a MPS Workflow with a PCR-Based Short Amplicon Whole Mitochondrial Genome Panel. *Genes* **2020**, *11*, 1345. [CrossRef] [PubMed]
12. Marshall, C.; Sturk-Andreaggi, K.; Daniels-Higginbotham, J.; Oliver, R.S.; Barritt-Ross, S.; McMahon, T.P. Performance evaluation of a mitogenome capture and Illumina sequencing protocol using non-probative, case-type skeletal samples: Implications for the use of a positive control in a next-generation sequencing procedure. *Forensic Sci. Int. Genet.* **2017**, *31*, 198–206. [CrossRef] [PubMed]
13. Peck, M.A.; Sturk-Andreaggi, K.; Thomas, J.T.; Oliver, R.S.; Barritt-Ross, S.; Marshall, C. Developmental validation of a Nextera XT mitogenome Illumina MiSeq sequencing method for high-quality samples. *Forensic Sci. Int. Genet.* **2018**, *34*, 25–36. [CrossRef]
14. Cuenca, D.; Battaglia, J.; Halsing, M.; Sheehan, S. Mitochondrial Sequencing of Missing Persons DNA Casework by Implementing Thermo Fisher's Precision ID mtDNA Whole Genome Assay. *Genes* **2020**, *11*, 1303. [CrossRef] [PubMed]
15. Scientific Working Group on DNA Analysis Methods Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories (23 April 2019). Available online: https://www.swgdam.org/_files/ugd/4344b0_f61de6abf3b94c52b28139bff600ae98.pdf (accessed on 1 September 2023).
16. Lembring, M.; Oven, M.; Montelius, M.; Allen, M. Mitochondrial DNA analysis of Swedish population samples. *Int. J. Leg. Med.* **2013**, *127*, 1097–1099. [CrossRef] [PubMed]
17. Tillmar, A.O.; Coble, M.D.; Wallerström, T.; Holmlund, G. Homogeneity in mitochondrial DNA control region sequences in Swedish subpopulations. *Int. J. Leg. Med.* **2010**, *124*, 91–98. [CrossRef]
18. Taylor, C.R.; Kiesler, K.M.; Sturk-Andreaggi, K.; Ring, J.D.; Parson, W.; Schanfield, M.; Vallone, P.M.; Marshall, C. Platinum-Quality Mitogenome Haplotypes from United States Populations. *Genes* **2020**, *11*, 1290. [CrossRef]
19. Karlsson, A.O.; Wallerström, T.; Götherström, A.; Holmlund, G. Y-chromosome diversity in Sweden—A long-time perspective. *Eur. J. Hum. Genet. EJHG* **2006**, *14*, 963–970. [CrossRef]
20. Zerjal, T.; Beckman, L.; Beckman, G.; Mikelsaar, A.V.; Krumina, A.; Kucinskis, V.; Hurles, M.E.; Tyler-Smith, C. Geographical, Linguistic, and Cultural Influences on Genetic Diversity: Y-Chromosomal Distribution in Northern European Populations. *Mol. Biol. Evol.* **2001**, *18*, 1077–1087. [CrossRef]
21. Salmela, E.; Lappalainen, T.; Liu, J.; Sistonen, P.; Andersen, P.M.; Schreiber, S.; Savontaus, M.; Czene, K.; Lahermo, P.; Hall, P.; et al. Swedish Population Substructure Revealed by Genome-Wide Single Nucleotide Polymorphism Data. *PLoS ONE* **2011**, *6*, e16747. [CrossRef]
22. Humphreys, K.; Grankvist, A.; Leu, M.; Hall, P.; Liu, J.; Ripatti, S.; Rehnström, K.; Groop, L.; Klareskog, L.; Ding, B.; et al. The Genetic Structure of the Swedish Population. *PLoS ONE* **2011**, *6*, e22547. [CrossRef]
23. Lappalainen, T.; Hannelius, U.; Salmela, E.; Von Döbeln, U.; Lindgren, C.M.; Huoponen, K.; Savontaus, M.; Kere, J.; Lahermo, P. Population Structure in Contemporary Sweden—A Y-Chromosomal and Mitochondrial DNA Analysis. *Ann. Hum. Genet.* **2009**, *73*, 61–73. [CrossRef] [PubMed]
24. Tambets, K.; Rootsi, S.; Kivisild, T.; Help, H.; Serk, P.; Loogväli, E.; Tolk, H.; Reidla, M.; Metspalu, E.; Pliss, L.; et al. The Western and Eastern Roots of the Saami—The Story of Genetic “Outliers” Told by Mitochondrial DNA and Y Chromosomes. *Am. J. Hum. Genet.* **2004**, *74*, 661–682. [CrossRef] [PubMed]
25. Ameer, A.; Dahlberg, J.; Olason, P.; Vezzi, F.; Karlsson, R.; Martin, M.; Viklund, J.; Kähäri, A.K.; Lundin, P.; Che, H.; et al. SweGen: A whole-genome data resource of genetic variability in a cross-section of the Swedish population. *Eur. J. Hum. Genet. EJHG* **2017**, *25*, 1253–1260. [CrossRef] [PubMed]
26. Lao, O.; Lu, T.T.; Nothnagel, M.; Junge, O.; Freitag-Wolf, S.; Caliebe, A.; Balascakova, M.; Bertranpetit, J.; Bindoff, L.A.; Comas, D.; et al. Correlation between Genetic and Geographic Structure in Europe. *Curr. Biol.* **2008**, *18*, 1241–1248. [CrossRef] [PubMed]
27. Novembre, J.; Johnson, T.; Bryc, K.; Kutalik, Z.; Boyko, A.R.; Auton, A.; Indap, A.; King, K.S.; Bergmann, S.; Nelson, M.R.; et al. Genes mirror geography within Europe. *Nature* **2008**, *456*, 274. [CrossRef]
28. McEvoy, B.P.; Montgomery, G.W.; McRae, A.F.; Ripatti, S.; Perola, M.; Spector, T.D.; Cherkas, L.; Ahmadi, K.R.; Boomsma, D.; Willemsen, G.; et al. Geographical structure and differential natural selection among North European populations. *Genome Res.* **2009**, *19*, 804–814. [CrossRef]
29. Sturk-Andreaggi, K.; Ring, J.D.; Ameer, A.; Gyllenstein, U.; Bodner, M.; Parson, W.; Marshall, C.; Allen, M. The Value of Whole-Genome Sequencing for Mitochondrial DNA Population Studies: Strategies and Criteria for Extracting High-Quality Mitogenome Haplotypes. *Int. J. Mol. Sci.* **2022**, *23*, 2244. [CrossRef]
30. Marshall, C.; Parson, W. Interpreting NUMTs in forensic genetics: Seeing the forest for the trees. *Forensic Sci. Int. Genet.* **2021**, *53*, 102497. [CrossRef]
31. Woerner, A.E.; Cihlar, J.C.; Smart, U.; Budowle, B. Numt identification and removal with RtN. *Bioinformatics* **2020**, *36*, 5115–5116. [CrossRef]
32. Cihlar, J.C.; Strobl, C.; Lagacé, R.; Muenzler, M.; Parson, W.; Budowle, B. Distinguishing mitochondrial DNA and NUMT sequences amplified with the precision ID mtDNA whole genome panel. *Mitochondrion* **2020**, *55*, 122–133. [CrossRef] [PubMed]
33. Ring, J.D.; Sturk-Andreaggi, K.; Alyse Peck, M.; Marshall, C. Bioinformatic removal of NUMT-associated variants in mitotiling next-generation sequencing data from whole blood samples. *Electrophoresis* **2018**, *39*, 2785–2797. [CrossRef] [PubMed]
34. Li, M.; Schröder, R.; Ni, S.; Madea, B.; Stoneking, M. Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 2491–2496. [CrossRef] [PubMed]

35. Andersen, M.M.; Balding, D.J. Assessing the Forensic Value of DNA Evidence from Y Chromosomes and Mitogenomes. *Genes* **2021**, *12*, 1209. [[CrossRef](#)]
36. Clopper, C.J.; Pearson, E.S. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika* **1934**, *26*, 404–413. [[CrossRef](#)]
37. Magnusson, P.K.E.; Almqvist, C.; Rahman, I.; Ganna, A.; Viktorin, A.; Walum, H.; Halldner, L.; Lundstrom, S.; Ullen, F.; Langstrom, N.; et al. The Swedish Twin Registry: Establishment of a biobank and other recent developments. *Twin. Res. Hum. Genet.* **2013**, *16*, 317–329. [[CrossRef](#)]
38. Andrews, R.M.; Kubacka, I.; Chinnery, P.F.; Lightowlers, R.N.; Turnbull, D.M.; Howell, N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **1999**, *23*, 147. [[CrossRef](#)]
39. Anderson, S.; Bankier, A.T.; Barrell, B.G.; de Bruijn, M.H.L.; Coulson, A.R.; Drouin, J.; Eperon, I.C.; Nierlich, D.P.; Roe, B.A.; Sanger, F.; et al. Sequence and organization of the human mitochondrial genome. *Nature* **1981**, *290*, 457–465. [[CrossRef](#)]
40. Sturk-Andreaggi, K.; Peck, M.A.; Boysen, C.; Dekker, P.; McMahon, T.P.; Marshall, C.K. AQME: A forensic mitochondrial DNA analysis tool for next-generation sequencing data. *Forensic Sci. Int. Genet.* **2017**, *31*, 189–197. [[CrossRef](#)]
41. van Oven, M.; Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **2009**, *30*, E386–E394. [[CrossRef](#)]
42. van Oven, M. PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Sci. Int. Genet. Suppl. Ser.* **2015**, *5*, e392–e394. [[CrossRef](#)]
43. Bodner, M.; Irwin, J.A.; Coble, M.D.; Parson, W. Inspecting close maternal relatedness: Towards better mtDNA population samples in forensic databases. *Forensic Sci. Int. Genet.* **2011**, *5*, 138–141. [[CrossRef](#)]
44. Huber, N.; Parson, W.; Dür, A. Next generation database search algorithm for forensic mitogenome analyses. *Forensic Sci. Int. Genet.* **2018**, *37*, 204–214. [[CrossRef](#)] [[PubMed](#)]
45. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
46. Gu, Z.; Gu, L.; Eils, R.; Schlesner, M.; Brors, B. circlize implements and enhances circular visualization in R. *Bioinformatics* **2014**, *30*, 2811–2812. [[CrossRef](#)] [[PubMed](#)]
47. Dür, A.; Huber, N.; Parson, W. Fine-Tuning Phylogenetic Alignment and Haplogrouping of mtDNA Sequences. *Int. J. Mol. Sci.* **2021**, *22*, 5747. [[CrossRef](#)] [[PubMed](#)]
48. Mauri, M.; Elli, T.; Caviglia, G.; Uboldi, G.; Azzi, M. RAWGraphs: A Visualisation Platform to Create Open Outputs. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter*; ACM: New York, NY, USA, 2017; pp. 28:1–28:5. [[CrossRef](#)]
49. Filograna, R.; Mennuni, M.; Alsina, D.; Larsson, N. Mitochondrial DNA copy number in human disease: The more the better? *FEBS Lett.* **2021**, *595*, 976–1002. [[CrossRef](#)]
50. Chu, H.; Hsiao, W.W.L.; Tsao, T.T.H.; Chang, C.; Liu, Y.; Fan, C.; Lin, H.; Chang, H.; Yeh, T.; Chen, J.; et al. Quantitative assessment of mitochondrial DNA copies from whole genome sequencing. *BMC Genom.* **2012**, *13* (Suppl. S7), S5. [[CrossRef](#)] [[PubMed](#)]
51. Longchamps, R.J.; Castellani, C.A.; Yang, S.Y.; Newcomb, C.E.; Sumpter, J.A.; Lane, J.; Grove, M.L.; Guallar, E.; Pankratz, N.; Taylor, K.D.; et al. Evaluation of mitochondrial DNA copy number estimation techniques. *PLoS ONE* **2020**, *15*, e0228166. [[CrossRef](#)]
52. Stoler, N.; Nekrutenko, A. Sequencing error profiles of Illumina sequencing instruments. *NAR Genom. Bioinform.* **2021**, *3*, lqab019. [[CrossRef](#)]
53. Parson, W.; Gusmão, L.; Hares, D.R.; Irwin, J.A.; Mayr, W.R.; Morling, N.; Pokorak, E.; Prinz, M.; Salas, A.; Schneider, P.M.; et al. DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing. *Forensic Sci. Int. Genet.* **2014**, *13*, 134–142. [[CrossRef](#)]
54. Tanaka, M.; Cabrera, V.M.; González, A.M.; Larruga, J.M.; Takeyasu, T.; Fuku, N.; Guo, L.; Hirose, R.; Fujita, Y.; Kurata, M.; et al. Mitochondrial Genome Variation in Eastern Asia and the Peopling of Japan. *Genome Res.* **2004**, *14*, 1832–1850. [[CrossRef](#)]
55. González, A.M.; Larruga, J.M.; Abu-Amero, K.K.; Shi, Y.; Pestano, J.; Cabrera, V.M. Mitochondrial lineage M1 traces an early human backflow to Africa. *BMC Genom.* **2007**, *8*, 223. [[CrossRef](#)] [[PubMed](#)]
56. Just, R.S.; Irwin, J.A.; Parson, W. Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. *Forensic Sci. Int. Genet.* **2015**, *18*, 131–139. [[CrossRef](#)]
57. Irwin, J.A.; Saunier, J.L.; Niederstätter, H.; Strouss, K.M.; Sturk, K.A.; Diegoli, T.M.; Brandstätter, A.; Parson, W.; Parsons, T.J. Investigation of Heteroplasmy in the Human Mitochondrial DNA Control Region: A Synthesis of Observations from More Than 5000 Global Population Samples. *J. Mol. Evol.* **2009**, *68*, 516–527. [[CrossRef](#)] [[PubMed](#)]
58. Sturk-Andreaggi, K.; Parson, W.; Allen, M.; Marshall, C. Impact of the sequencing method on the detection and interpretation of mitochondrial DNA length heteroplasmy. *Forensic Sci. Int. Genet.* **2020**, *44*, 102205. [[CrossRef](#)] [[PubMed](#)]
59. Stoljarova, M.; King, J.L.; Takahashi, M.; Aaspõllu, A.; Budowle, B. Whole mitochondrial genome genetic diversity in an Estonian population sample. *Int. J. Leg. Med.* **2016**, *130*, 67–71. [[CrossRef](#)]
60. Malyarchuk, B.; Litvinov, A.; Derenko, M.; Skonieczna, K.; Grzybowski, T.; Grosheva, A.; Shneider, Y.; Rychkov, S.; Zhukova, O. Mitogenomic diversity in Russians and Poles. *Forensic Sci. Int. Genet.* **2017**, *30*, 51–56. [[CrossRef](#)]
61. Översti, S.; Onkamo, P.; Stoljarova, M.; Budowle, B.; Sajantila, A.; Palo, J.U. Identification and analysis of mtDNA genomes attributed to Finns reveal long-stagnant demographic trends obscured in the total diversity. *Sci. Rep.* **2017**, *7*, 6193. [[CrossRef](#)]

62. Margaryan, A.; Lawson, D.J.; Sikora, M.; Racimo, F.; Rasmussen, S.; Moltke, I.; Cassidy, L.M.; Jorsboe, E.; Ingason, A.; Pedersen, M.W.; et al. Population genomics of the Viking world. *Nature* **2020**, *585*, 390–396. [[CrossRef](#)]
63. *Foreign-Born by Country of Birth, Sex and Year of Immigration*; Statistics Sweden: Solna, Sweden, 2022.
64. Kling, D.; Tillmar, A. Forensic genealogy-A comparison of methods to infer distant relationships based on dense SNP data. *Forensic Sci. Int. Genet.* **2019**, *42*, 113–124. [[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.