

Article

Applications of Probe Capture Enrichment Next Generation Sequencing for Whole Mitochondrial Genome and 426 Nuclear SNPs for Forensically Challenging Samples

Shelly Y. Shih ¹ , Nikhil Bose ^{1,2}, Anna Beatriz R. Gonçalves ^{1,3}, Henry A. Erlich ¹ and Cassandra D. Calloway ^{1,2,*}

¹ Children's Hospital Oakland Research Institute, 5700 Martin Luther King Jr. Way, Oakland, CA 94609, USA; sshih@chori.org (S.Y.S.); nbose@ucdavis.edu (N.B.); herlich@chori.org (H.A.E.)

² Forensic Science Graduate Program, University of California, Davis, 1909 Galileo Ct. Ste. B, Davis, CA 95618, USA

³ Laboratório de Diagnósticos por DNA (LDD), Universidade do Estado do Rio de Janeiro (UERJ), Instituto de Biologia Roberto Alcântara Gomes (IBRAG), Rua São Francisco Xavier, number 524, Pavilhão Haroldo Lisboa da Cunha, 20550-900 Rio de Janeiro, RJ, Brazil; gonannab@gmail.com

* Correspondence: scalloway@chori.org

Received: 20 December 2017; Accepted: 17 January 2018; Published: 22 January 2018

Abstract: The application of next generation sequencing (NGS) for the analysis of mitochondrial (mt) DNA, short tandem repeats (STRs), and single nucleotide polymorphism (SNPs) has demonstrated great promise for challenging forensic specimens, such as degraded, limited, and mixed samples. Target enrichment using probe capture rather than PCR amplification offers advantages for analysis of degraded DNA since two intact PCR primer sites in the template DNA molecule are not required. Furthermore, NGS software programs can help remove PCR duplicates to determine initial template copy numbers of a shotgun library. Moreover, the same shotgun library prepared from a limited DNA source can be enriched for mtDNA as well as nuclear markers by hybrid capture with the relevant probe panels. Here, we demonstrate the use of this strategy in the analysis of limited and mock degraded samples using our custom probe capture panels for massively parallel sequencing of the whole mtgenome and 426 SNP markers. We also applied the mtgenome capture panel in a mixed sample and analyzed using both phylogenetic and variant frequency based bioinformatics tools to resolve the minor and major contributors. Finally, the results obtained on individual telogen hairs demonstrate the potential of probe capture NGS analysis for both mtDNA and nuclear SNPs for challenging forensic specimens.

Keywords: Next Generation Sequencing (NGS); Massively Parallel Sequencing (MPS); Probe Capture Target Enrichment; mitochondrial DNA (mtDNA); Single Nucleotide Polymorphism (SNP); Forensic Genetics; degraded DNA

1. Introduction

Next generation sequencing (NGS) has shown great promise in the analysis of DNA from degraded, limited, and mixed samples often encountered in forensic cases [1–3]. NGS platforms produce millions of reads for sequencing multiple samples in a single sequencing run, thereby making it possible for sequencing of the whole mitochondrial (mt) genome and proving valuable for analyzing nuclear single nucleotide polymorphism (SNPs) and short tandem repeats (STRs) [4–7]. Current NGS technologies are equipped with sequencing chemistry capable of generating 25 million–3 billion reads per sequencing run with an output of ~15–600 Gb file size, providing sufficient sequencing

depth for robust SNP and STR variant calling [6,8]. NGS analysis of STR polymorphism has revealed additional sequence polymorphisms within tandem repeats not identified by conventional capillary electrophoresis approaches [9,10]. Thus, NGS allows analysis of both length and sequence polymorphism on a single sequencing platform [11]. NGS applications have revolutionized not only STR and SNP typing, but also sequencing of small genomes such as the whole human mtgenome [12,13]. Analyzing the whole mtgenome, rather than just the hypervariable polymorphic regions HVI/HVII, would increase the discrimination power and alleviate the problem of common sequences found in the HVI/HVII regions shared by populations of the same origin [2]. In addition, NGS can increase resolution of mtgenome analysis by detecting low level-mixtures such as heteroplasmy (<5%), which would not be detectable using current standard analyses of mtDNA polymorphism by Sanger sequencing with its limit of detection at ~10–20% [13–18]. By utilizing this massively parallel sequencing aspect of NGS, the discrimination power can be increased with simultaneous analysis of multiple, various genetic markers including not just STRs, but also SNPs and mtDNA.

Sequencing SNPs and mtDNA can aid in genetic analysis of DNA from forensic samples when STR typing with capillary electrophoresis is not suitable, which is often the case for degraded DNA from forensic samples due to biological, environmental, or oxidative degradation. In these cases, the DNA is highly fragmented, and therefore both primer binding sites may not be present in most template DNA molecules, rendering the amplification of SNPs or STRs impractical [11,19]. The inability to amplify large amplicons would result in partial genetic profiles with dropout of alleles or loci [19,20]. Furthermore, mixtures, particularly of trace DNA, may represent a significant challenge for genetic analysis due to stochastic variation and potential allelic dropout [2,21]. Alternative approaches such as miniSTR analysis can be used for degraded DNA samples by designing primer binding sites that are closer to the target regions, thereby reducing the amplicon size to 80–150 bp [20,22,23]. Another useful approach is by typing SNPs with primers immediately adjacent to targeted single base variations, thereby reducing the amplicon size to as short as ~50 bp [11,24]. The short SNP amplicon sizes are particularly useful for analysis of DNA from degraded forensic specimens, such as those in mass disaster cases, and clinical samples, such as fetal cell free DNA (cfDNA) in maternal blood [11,25,26]. In addition to SNPs, the analysis of mitochondrial sequence polymorphism has proved useful when nuclear DNA is too degraded since mtDNA is more likely to be present due to its high copy number in forensic biological samples, such as telogen hairs roots, hair shafts, bones, teeth, and touch DNA [14,27]. Mitochondrial DNA, as a haploid lineage marker, is also valuable for the de-convolution of mixtures since each individual contributes only one sequence rather than two alleles. Presence of two nucleotides in a single mtgenome position would either indicate a mixture or presence of heteroplasmy [14,28]. The non-recombining uniparental inheritance of mtDNA is also useful for analyzing samples from mass disasters or missing person cases, in which a missing individual's mitochondrial haplotype can be directly compared to the haplotype of the individual's maternal relatives [13,14,27].

However, designing primers for mtgenome sequencing can be challenging for these densely polymorphic regions [29]. Furthermore, multiplexing SNPs using PCR based approaches still requires presence of intact primer binding sites. Therefore, enrichment of target DNA for forensically challenging samples by using an alternative method that does not depend on the presence of intact primer binding site would improve recovery of fragmented DNA prior to deep sequencing [4–6]. One alternative strategy is target enrichment using synthetic probes [30]. In general, the probe capture enrichment process consists of hybridizing biotinylated DNA or RNA probes to complementary DNA fragments in a shotgun library either in solution or on a solid surface [7]. The target DNA fragments are bound to the probes while the unbound DNA, primer dimer artifacts, and impurities are washed away [30–32]. These capture technologies using synthetic probes have been widely applied for exome capturing and sequencing in research and clinical laboratories [4,5,33,34].

Currently, several capture enrichment technologies are commercially available, including Agilent's capture platforms, Roche NimbleGen SeqCap EZ platforms, and Illumina TruSeq DNA and Nextera Exome platforms [34,35]. Several groups have conducted studies comparing some of these capture

enrichment platforms for their probe type, probe length, probe design, and target enrichment efficiency [34,36]. Agilent technology uses RNA probes with probe length ~120 bp while Illumina and NimbleGen technologies employ DNA probes that are, respectively, ~95 bp and ~55–105 bp in length [35,36]. The probe design for Agilent uses probes that are adjacent to each other to cover the region of interest. Illumina's design uses paired end reads that extend to cover the region of interest. NimbleGen's design uses a high density of probes that overlap target regions, resulting in a high redundancy of probes per base [34,35]. Clark et al. demonstrated that target enrichment efficiency was the highest for NimbleGen at 96.8% at $\geq 10\times$ read depth, followed by Illumina at 90.0% at $\geq 10\times$ read depth and Agilent at 89.6% at $\geq 10\times$ read depth [35]. Sulonen et al. also showed that NimbleGen had the highest target efficiency of the two capture platforms compared (NimbleGen and Agilent) [36]. Based on probe design and performance differences, one may choose one target enrichment platform over the other. Our evaluation of these results led to the decision of choosing Roche NimbleGen SeqCap EZ platform for our custom probe capture panels owing to its tiling approach for probe design and higher enrichment efficiency. With high redundancy of probes per base, this design can efficiently and uniformly target regions with high degree of polymorphism, such as the mtgenome regions and nuclear SNPs [29].

Hybrid probe capture NGS assays have proved useful for capturing and sequencing ancient DNA (aDNA) from human remains [37,38]. The DNA from ancient remains is often highly degraded due to chemical or environmental damages and can be highly contaminated with exogenous DNA from the microbes and organisms that colonize the remains in the burial environment. Exogenous DNA introduced in laboratories or by personnel during various stages of storage, cleaning, extraction, and library preparation can also be a source of contamination [39]. Thus, endogenous DNA from these ancient samples can comprise as low as ~1–5% of the total DNA, such as the cases with the Neanderthal DNA samples. Using the 454 pyro-sequencing platform, Green et al. successfully mapped the complete Neanderthal mtgenome from 70 Neanderthal bone samples [40–42]. However, to capture all target DNA fragments from these shotgun libraries without enrichment would require too many sequencing runs [39,43]. This limitation can be overcome by applying a probe capture enrichment method as demonstrated in the study by Reich et al. whereby using a target enrichment method resulted in recovery of ~70% endogenous DNA from an ancient Denisova phalanx sample dated ~50,000–~30,000 years ago [44].

Similar to ancient DNA, DNA from forensic, mass disaster, or missing person case samples is often limited, highly degraded, and mixed. Recently, hybrid probe capture NGS technologies successfully recovered mtgenome DNA fragments from forensically challenging samples such as modern bones and hair shafts [1,2]. Templeton et al. showed that in-solution hybridization with DNA probes generated from human mtDNA yielded full coverage of the mtgenome from ~2500 years old bone samples with short mtDNA fragments (77 bp) and low mtDNA copies (350 copies/ μ L) [2]. Eduardoff et al. also demonstrated the efficiency, robustness, and sensitivity of the probe capture method on recovery of short mtDNA fragments (50 bp) from hair shafts and ancient solid tissues [1]. The primer extension capture (PEC) NGS method used by Eduardoff et al. was originally designed to target the mtDNA control region (CR). Their results demonstrated coverage of partial and full mtgenome, with noticeably lower coverage per base in regions outside of the CR, as expected [1]. In addition to biotinylated DNA probes, RNA probes can also be used for enriching limited and fragmented DNA from ancient and forensic specimens [33,39].

Here, we present the results obtained with our custom probe capture panels for target enrichment of the whole mtgenome and 426 nuclear SNPs in conjunction with deep sequencing on the Illumina MiSeq platform from a single shotgun library demonstrating the application to the analysis of limited, degraded, and mixed forensic samples [45,46]. Both custom probe capture panels were designed using the Roche NimbleGenSeqCap EZ custom probe capture assay. With its redundant, overlapping tiling design, the DNA probes can efficiently hybridize and capture highly fragmented DNA and densely polymorphic regions such as the HVI/HVII regions for clonal, massively parallel sequencing [34].

This redundancy of sample tiling using DNA probes greatly improves the capture of forensically challenging DNA samples such as DNA from telogen hair roots, telogen hair shafts, and touch-DNA. This strategy allows the capture of both mtDNA and nuclear SNP markers from a single shotgun DNA library without consuming additional DNA extracts of limited forensic biological samples.

2. Materials and Methods

2.1. Sample Preparation

Samples were prepared for the following studies as described below: mtDNA sensitivity study, nuclear DNA sensitivity and size selection, mock degradation, mtDNA mixture, and telogen hairs study.

2.1.1. Mitochondrial DNA Sensitivity Study

Dilutions targeting input DNA amounts of 1 ng, 100 pg, and 10 pg were prepared using the control human HL-60 DNA SRM2392-I (NIST, Gaithersburg, MD, USA). These samples were processed following our standard DNA fragmentation and library preparation as described below and depicted in Figure 1.

2.1.2. Nuclear DNA Sensitivity and Size Selection Study

The single source male control DNA NA24129 (Coriell Institute for Medical Research, Camden, NJ, USA) was mechanically fragmented using the Covaris® M220 Focused-ultrasonicator™ (Covaris, Woburn, MA, USA) to an average size of 175 bp with a range from 25 bp to 250 bp. To test short template sample analysis, size selection was carried out at ≤ 75 bp using the Pippin Prep® (Sage Science, Beverly, MA, USA). DNA libraries were then prepared for the ≤ 75 bp size selected fragments with 10 ng, 1 ng, and 0.5 ng DNA sample input amounts. The detailed procedure for the preparation of these size selected samples is described in the study by Bose et al. [46].

2.1.3. Mock Degradation Study

For this study, 1 ng of control DNA (K562 for mtDNA and 2800M for nuclear SNPs (Promega, Madison, WI, USA) was mechanically fragmented to 150 bp to simulate degraded samples following the manufacturer's protocol (Part No. 010166 Rev E, Covaris). These artificially degraded samples were then processed following our standard DNA fragmentation to 250 bp and library preparation as described below.

2.1.4. Mitochondrial DNA Mixture Study

For the mtDNA mixture, a two-person mixture was prepared in vitro using DNA from two previously extracted blood samples (Caucasian C163 and US Hispanic H104) of known mitochondrial haplotypes targeting a total input of ~300,000 mtDNA copies. The nuclear DNA and mtDNA amounts were quantified using a duplex real-time quantitative PCR (qPCR) assay [47,48]. The mixture samples were processed following our standard library preparation procedure as described below.

2.1.5. Telogen Hair Study

Two telogen hairs (shed hairs) were collected from female participants for the mock forensic sample study with the approval from the Institutional Review Board Administration (IRB # 852842) at the University of California, Davis, USA. The hairs were collected by combing and were screened for telogen roots using a transmitted light microscope. Two centimeters of the hairs were cut from the proximal end and labeled P1FTR2 and A1FTR1. The telogen hair roots were cleaned with 2% sodium dodecyl sulfate (SDS) followed by 10 min of sonication and a rinse of molecular grade water. The hair roots were then stained with Harris Hematoxylin (HE) (Richard-Allan Scientific 72704, Kalamazoo, MI, USA) and counted for nuclei under a transmitted light microscope [49]. For telogen hair P1FTR2,

>60 nuclei were observed and for hair A1FTR1, 20–60 nuclei were counted. DNA from these telogen hair roots was extracted using the QIAamp DNA Micro Kit (QIAGEN®, Valencia, CA, USA) following the manufacturer's protocol with elution in 50 µL of TE-4 (pH 8.0, 0.1 M EDTA). DNA extracts were quantified using a real-time Taqman™ Degradation qPCR assay to detect both human nuclear DNA and mtDNA by targeting two nuclear DNA makers, nuTH01 (~170–190 bp), nuCSF (67 bp), and a 283 bp mtDNA marker [47,48]. The mtDNA marker approximates the number of mtDNA copies in DNA extracts. DNA libraries were prepared using 1 ng DNA as estimated by the nuTH01 marker with ~270,000 and ~540,000 mtDNA copies for P1FTR2 and A1FTR1, respectively.

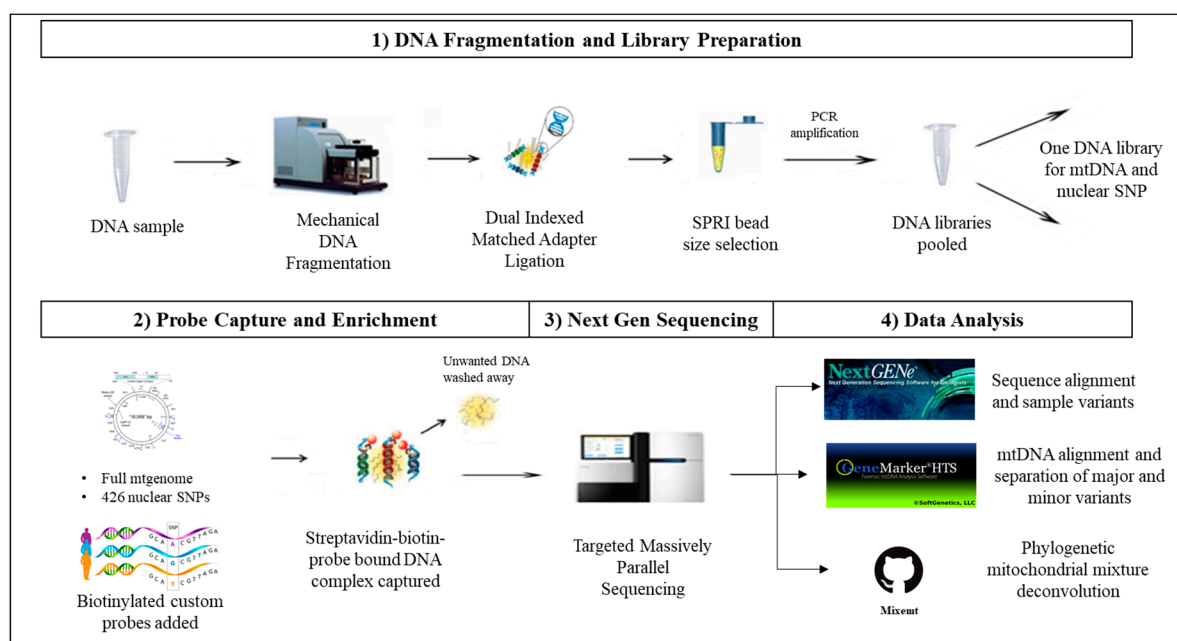


Figure 1. Novel probe capture next generation sequencing (NGS) assay for sequencing mitochondrial (mt) and nuclear DNA. To circumvent choosing between mitochondrial and nuclear analysis of degraded or limited forensic DNA samples, a capture enrichment based library preparation method can be implemented. In this method: (1) DNA libraries are prepared by fragmenting the DNA samples using Covaris® M220 Focused-ultrasonicator™ (Covaris). DNA fragments are ligated with dual-index adapters, and the amplified DNA libraries are size-selected using SPRIselect® beads (Beckman Coulter Life Sciences, Indianapolis, IN, USA). (2) Shotgun libraries are enriched using a DNA probe based method of targeted capture enrichment for the whole mtgenome and eight categories of nuclear single nucleotide polymorphism SNP markers (426 SNPs). (3) The enriched samples are then sequenced on a NGS platform. (4) The nuclear SNP data is analyzed using NextGENe (SoftGenetics LLC, State College, PA, USA), while the mitochondrial data is analyzed using NextGENe, GeneMarker® HTS (SoftGenetics LLC), and Mixemt [50].

2.2. DNA Fragmentation and Library Preparation

All DNA samples, except the samples in the nuclear DNA sensitivity and size selection study, were mechanically sheared to 250 bp following the manufacturer's guideline for the Covaris® M220 Focused-ultrasonicator™ for the 50 µL screw capped tube. DNA libraries were then prepared using KAPA Hyper Prep Kit for Illumina® platforms (KAPA Biosystems, Roche®, San Francisco, CA, USA) following the manufacturer's protocol with optimized PCR cycle numbers (Table 1). Other modifications include two post-adaptor-ligation SPRI® cleanups at 0.8× bead-to-sample volume ratio and two post-amplification SPRI® cleanups at 1× bead-to-sample volume ratio with SPRIselect® reagent (Beckman Coulter Life Sciences, Indianapolis, IN, USA) [51]. For all samples, Matched Dual Index Adapters (Integrated DNA Technologies, Coralville, IA, USA) were used in DNA library

preparations. The size of the amplified DNA libraries was determined using the Agilent 2100 Bioanalyzer and Agilent 7500 DNA kit (Agilent Technologies) to ensure that all amplified products were within the optimal size range (300–700 bp) for sequencing.

Table 1. PCR cycle numbers for varying input DNA amounts.

Total DNA Amount	PCR Cycle Number
≤10 ng–1 ng	13
≤1 ng–500 pg	17
≤500 pg–50 pg	20
≤50 pg–10 pg	24

2.3. Probe Capture Enrichment and Next Generation Sequencing

Tecan[®] Microplate Reader Infinite[®] F200 instrument (Tecan, Männedorf, Switzerland) and Quant-iT PicoGreen[®] dsDNA Assay Kit (Thermo Fisher) were used to quantify the DNA sample library amount for subsequent pooling. For whole mtgenome capture, up to 24 sample libraries with unique adapter index sequences, were pooled at equal DNA amounts for a total of 1 µg of DNA. For nuclear SNP capture, a total of 1 µg was targeted by pooling equal DNA amounts of 9–11 samples depending on sequencing read depth requirements per sample.

The mtDNA custom probe capture panel was developed by Calloway et al. using proprietary software by NimbleGen targeting the mtDNA majority consensus sequence from a global mtDB-Human Mitochondrial Genome population Database [29]. The probe design uses unique probes with a tiling approach, resulting in high redundancy of probes overlapping target sequences. Custom parameters were used and the final design selected directly targets most of the mtgenome with only two small gaps at positions 2506–2513 (7 bp) and 2962–2972 (10 bp). The DNA probes are 50–100 bp in size with an average of 35 different probes per base. The design and hybridization conditions allow the probes to tolerate up to five mismatches within a DNA probe [29]. The nuclear SNP probe capture panel developed by Bose et al. consists of 426 SNPs with high target capture efficiency, including 41 ancestry, 135 identity, 73 microhaplotypes, 22 phenotypically informative, 24 tetra- and 28 tri-allelic, and 24 X and 79 Y SNPs (Table S1) [46]. The probes are 50–100 bp in size and span over SNP coordinates with ±50 bp. The SNP coordinates were determined for human genome version GRCh37/hg19 using the UCSC Genome Browser [46].

The probe capture experiments were carried out using the Roche NimbleGen SeqCap EZ Library Developer Library Kit following the manufacturer’s protocol for both mtDNA and nuclear SNP capture (NimbleGen SeqCap EZ Library SR User’s Guide V.5.1) (Roche). The hybridization samples were prepared by combining COT human DNA (Roche[®]), hybridization enhancing oligo pool, and the previously pooled multiplex DNA libraries. The COT human DNA fraction is largely highly repetitive DNA elements used to increase the specificity of hybrid probe capture. The hybridization samples were then concentrated using a Jouan RC 10.22 Vacuum Concentrator Centrifugal Evaporator following the manufacturer’s protocol (Jouan, Inc., Winchester, VA, USA). The concentrated multiplex DNA libraries were hybridized to the custom probes overnight for 16–20 h. The hybridized products were enriched and amplified using KAPA HiFi HotStart Ready Mix (KAPA Biosystems) and Dynabeads[™] M-270 Streptavidin (Thermo Fisher) following the manufacturer’s protocol (NimbleGen, Roche). The mtDNA hybridization products were amplified at 13 PCR cycles, and the nuclear SNP hybridization products were amplified at 15 PCR cycles.

The amplified enriched DNA library pool was quantified using both Agilent 2100 Bioanalyzer and Tecan[®] Microplate Reader Infinite[®] F200 instrument. The quantified enriched DNA library pool was then prepared using the “Preparing Libraries for sequencing on the MiSeq” protocol with a final concentration of 11 pM of DNA and 5–10% control phiX DNA. Sequencing on the Illumina MiSeq instrument was carried out following the manufacturer’s protocol (Illumina). The Illumina MiSeq Reagent v2 kit (500 cycles, 2 × 250 reads) was used for mtDNA sensitivity study, mock degradation

study, mixture study, and forensic type sample study. The Illumina MiSeq Reagent v2 Kit (300 cycles, 2×150 reads) was used for the size selection study.

2.4. Data Analysis

Samples in the mtDNA sensitivity study were analyzed with and without the PCR duplicates using both GeneMarker[®]HTS v.1.2.2 and NextGENe[®] v.2.4.2 (SoftGenetics LLC, State College, PA, USA). Nuclear SNPs were analyzed in the size selection and mock degradation studies using the NextGENe[®] software. Telogen hairs were analyzed using GeneMarker[®]HTS for the whole mtgenome and NextGENe[®] for SNP analysis.

Initial mtgenome sequencing analysis was carried out using the GeneMarker[®]HTS software by aligning the forward and reverse sequence reads (FASTQ) of each sample to the revised Cambridge Reference Sequence (rCRS) [52]. The software then outputs all forward and reverse reads (unmerged) into a single BAM file. Default motif was applied to help align data in problematic regions such as indels. Default identity setting was at 90% to prevent alignment of sequence reads that were <90% identical to the rCRS. Soft clipping at the 3' end was at $Q \leq 25$ to allow trimming of bases with quality scores below this threshold at the end of the reads. Additional custom filter settings were used: the variant percentage filter was set to $\geq 10\%$ to filter out variants that did not meet the 10% threshold. Variant allele coverage filter was set at ≥ 10 so that each variant detected exhibited at least tenfold coverage. Total coverage was set at ≥ 100 (with the duplicates) to ensure that there is at least $100\times$ coverage per base. In addition to the variant filters, allele score difference was set to ≤ 2.5 and the allele balance ratio was set to ≤ 5 for both SNP and Indel. Total reads, alignment percentage, average coverage, and the mtgenome coverage were obtained from the alignment statistics reports generated using GeneMarker[®]HTS. The PCR duplicate removal tool in GeneMarker[®]HTS removes PCR duplicates that have the same starting positions in the forward and reverse reads (reads not merged) [53].

Mitochondrial genome analysis using NextGENe[®] v.2.4.2 was carried out by first converting FASTQ files to FASTA files. Default alignment (default homology at 85%) was used for samples with PCR duplicates; custom alignment with homology at 90 was used for samples without PCR duplicates. Custom filter settings were applied to all samples: 10% mutation percentage, three SNP allele count, and $100\times$ total coverage. Depth of coverage and sequence read length distribution (average read length) were both obtained from reports generated using NextGENe[®] v.2.4.2.

In addition to GeneMarker[®]HTS, the sequenced reads for the mtDNA mixture samples were analyzed using Mixemt developed by Vohr et al. [54]. Prior to analysis using Mixemt, the raw sequence reads were trimmed for adapters, and the overlapping reads were merged using SeqPrep [49]. Both merged and unmerged sequence reads were then aligned separately to the Reconstructed Sapiens Reference Sequence (RSRS) using the Burrows–Wheeler Aligner (bwa) tool, which also converts the FASTQ files to SAM files [51,52]. SAMtools was then used to collapse the PCR duplicates and convert SAM files to BAM files [55]. The BAM files were used for Mixemt analysis, which assigns each sequence read to a haplogroup based on the probability of the read originating from a contributing haplogroup [50,54].

For nuclear SNPs analysis, raw sequence data of all samples were analyzed using the SoftGenetics[®] NextGENe[®] Software v2.4.1.1 (SoftGenetics LLC). The raw sequence data was first converted from FASTQ files to FASTA files. The FASTA files of each sample were then aligned to version GRCh37/hg19 of the human genome in NextGENe. Default alignment settings were used. Coverage and allele percent of targeted SNPs was generated in the NextGENe Mutation Report tool by specifying SNP genomic locations (as a BED file). Further sequencing statistics such as diploid read depth per SNP and percent SNP coverage were derived from the mutation reports [46]. SNP coverage percent was established based on alleles exhibiting $>10\times$ reads and SNP loci exhibiting $\geq 20\times$ reads, which is within the threshold necessary for confident SNP variation calling [4]. Locus and allele dropouts were determined

for size selected samples based on comparison of the sequencing data from experimental samples to the sequencing data from 25 ng references.

3. Results

3.1. Mitochondrial DNA Sensitivity Study

The sensitivity of the mtgenome probe capture assay was tested with HL60 samples at DNA amounts of 1 ng, 100 pg, and 10 pg. Sequencing data of the whole mtgenome was analyzed using GeneMarker[®] HTS with PCR duplicates included and removed. When PCR duplicates were included, all samples exhibited full coverage of the whole mtgenome at $>100\times$ read depth (Table 2). Even after removing PCR duplicates, both 1 ng and 100 pg samples exhibited full mtgenome coverage at $>100\times$ read depth (Figure 2). The 10 pg sample exhibited full coverage of the mtgenome at $>5\times$ read depth per base with an average coverage of $22\times$ without PCR duplicates (Figure 2, Table 2). These results also demonstrated the specificity of our custom probe capture NGS method; of the average total of $\sim 630,000$ reads per sample, $>88\%$ aligned to the rCRS reference (Table 2).

Table 2. Sequencing statistics of mitochondrial DNA sensitivity study samples.

HL 60 Samples	Input mtDNA Copies	Total Reads	Alignment (%)	PCR Duplicates Included		PCR Duplicates Removed	
				Avg. Read Length (bp)	Avg. Coverage Per Base	Avg. Read Length (bp)	Avg. Coverage Per Base
1 ng	200,000	604,052	89.66%	190	6474	181	1226
100 pg	20,000	631,716	89.17%	200	7043	178	183
10 pg	2000	656,556	88.54%	195	7197	157	22

Avg., Average.

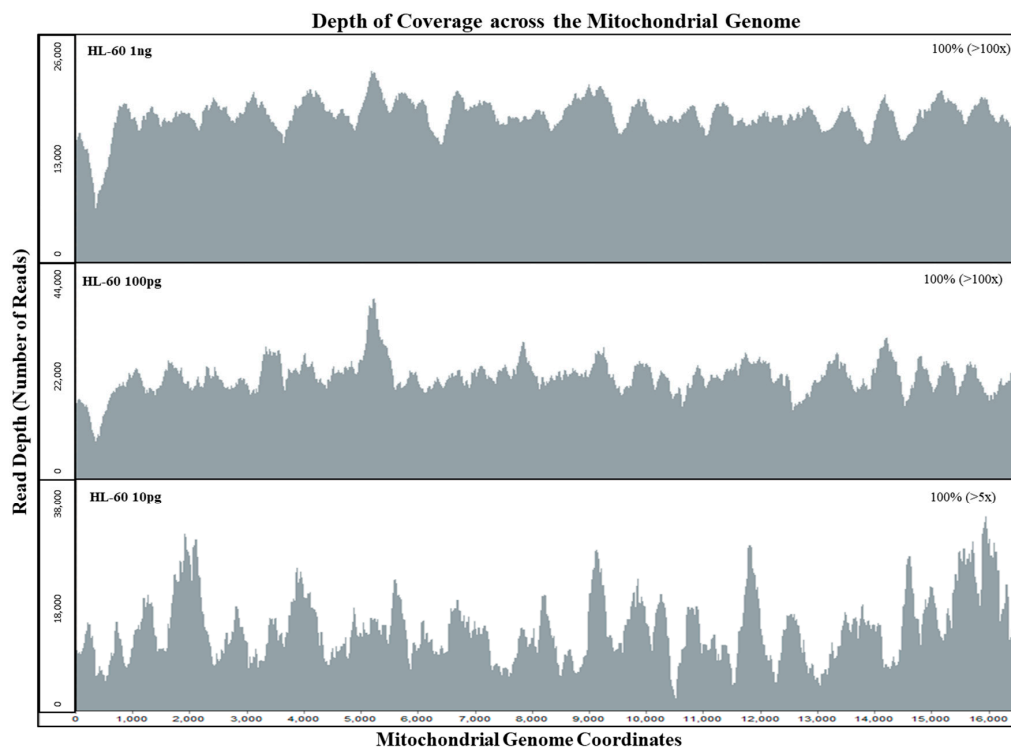


Figure 2. Depth of coverage across the whole mitochondrial genome for mtDNA sensitivity study samples. Limiting dilution of HL60 control DNA samples were tested with input DNA amounts of 1 ng, 100 pg, and 10 pg for the mtDNA sensitivity study. All HL60 samples exhibited full coverage of the mtgenome at $>5\times$ read depth (PCR duplicates removed using NextGENE[®] v.2.4.2).

3.2. Nuclear DNA Sensitivity and Size Selection Study

The sensitivity of our custom SNP probe capture NGS assay to recover nuclear SNPs from short DNA fragments was tested with ≤ 75 bp size selected samples at input DNA amounts of 10 ng, 1 ng, and 0.5 ng. All size selected samples exhibited an average of ~ 4.2 million reads per sample. On average, the diploid read depth per SNP was ~ 1830 per SNP and the average read length per sample was ~ 79 bases (Table 3). Over 99% of SNPs were captured and sequenced with diploid read depth per SNP allele $\geq 10\times$ (Figure 3). The depth of coverage was uniformly distributed across all SNPs captured (Figure 3). Comparison with the reference genotype demonstrated no allele drop-ins, and the highest allele drop-out was observed for the ≤ 75 bp 0.5 ng sample with six SNP alleles that were undetected. At 1 ng input, the number of SNPs without dropouts was nearly equal between the size selected samples and the non-size selected controls. Even after removing the duplicates, the diploid read depths for the 1 ng and 0.5 ng samples were similar at an average of $\sim 200\times$ read depth per SNP. Our results show that over 99% of SNPs were covered for 10 ng and 1 ng and 98.8% of the SNPs were covered for the 0.5 ng sample. Removing PCR duplicates resulted only in 1–2 additional SNP locus dropout (Table 3). These results demonstrate that allelic dropout is more dependent on input DNA copy numbers rather than the size of the DNA targets.

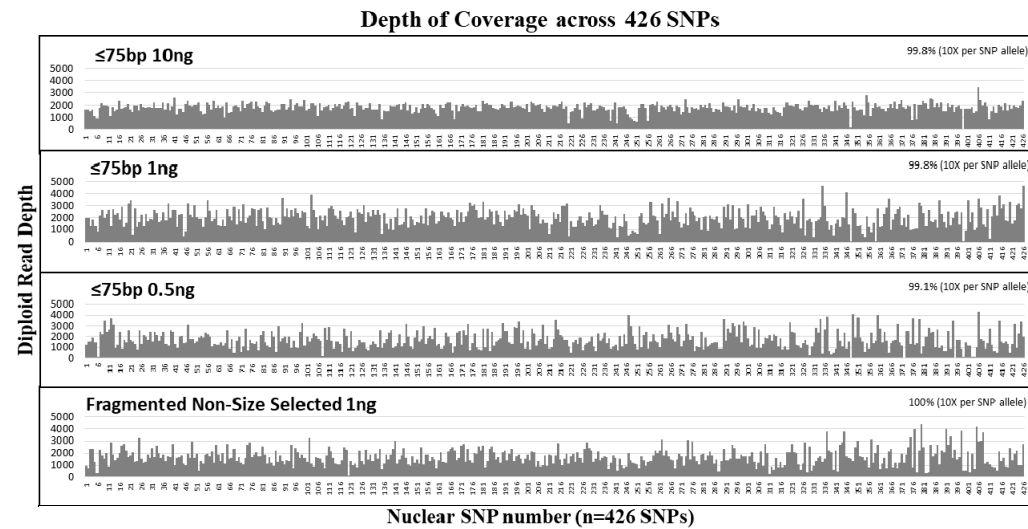


Figure 3. Average diploid read depth per SNP from the size selection study. Mechanically fragmented NA24129 control DNA sample was size selected at ≤ 75 bp and tested at 10 ng, 1 ng, and 0.5 ng. The diploid reads of each of the 426 SNPs (*y*-axis) was plotted in order of their SNP number (see Table S1 for details of SNPs corresponding to each number (*x*-axis)). All samples exhibited >99% coverage of 426 nuclear SNPs at $>10\times$ allelic read depth per SNP.

Table 3. Sequencing statistics for nuclear SNPs size selection study.

Sample	Input Amount (ng)	Total Reads	PCR Duplicates Included			Locus/Allele dropout	PCR Duplicates Removed				
			Avg. Read Length (bp)	Avg. Diploid Read Depth per SNP	SNP Coverage (<i>n</i> = 426 SNPs)		Total Reads	Avg. Read Length (bp)	Avg. Diploid Read Depth per SNP	SNP Coverage (<i>n</i> = 426 SNPs)	Locus/Allele dropout
≤ 75 bp	10	4,606,916	78	1811 \pm 377	425 (99.8%)	1/0	2,476,408	78	517 \pm 93	424 (99.5%)	2/0
	1	4,431,052	79	1978 \pm 746	425 (99.8%)	1/2	2,638,898	78	217 \pm 97	423 (99.3%)	3/1
	0.5	3,679,914	78	1721 \pm 820	422 (99.1%)	4/6	2,199,585	78	190 \pm 115	421 (98.8%)	5/10
Control	1	3,648,297	130*	1690 \pm 697	426 (100%)	0/3	2,259,909	127	336 \pm 154	425 (99.8%)	1/3

* As a MiSeq Reagent v2 kit (300-cycles, 2×150 reads) was used, the maximum sequencing read length was 150 bases.

3.3. Mock Degradation Study

The recovery of fragmented DNA for both mtDNA and nuclear SNPs using our custom probe capture systems was assessed in the mock degradation study. DNA was mechanically fragmented to an average of 150 bp to mimic degraded DNA and samples exhibited expected size distribution of sequenced reads with average read lengths of 146–147 bp for mtDNA and nuclear SNPs (Table 4, Figure 4). Full coverage of the mtgenome was achieved for both the control and mock degraded mtDNA samples at $>100\times$ read depth per base (duplicates not removed), demonstrating capture and sequencing efficiency of DNA fragments independent of fragment size (Table 4).

The total number of reads recovered from the mock degraded and control samples for nuclear SNPs was 2–2.5 million. The average diploid read depth per SNP for the mock degraded sample was 968 and 498 for the control. The two-fold lower read depth for the 1 ng nuclear DNA control resulted in a lower than expected SNP coverage, $\sim 84.5\%$ compared to 96% of the SNPs covered with $>10\times$ reads per allele for the mock degraded sample (Table 4). However, even at low diploid read depth, the 96% SNP recovery for the mock degraded sample demonstrates the utility of the system for capturing and sequencing highly fragmented DNA samples.

Table 4. Sequencing statistics for mock degraded and control samples at 1 ng.

Mock Degradation Study for mtDNA and nuclear SNPs (Input Amount 1 ng)				
Sample Type	Average Read Length (bp)	Total Reads	Average Depth of Coverage	Coverage (%)
mtDNA Mock Degraded	146	990,269	7148	100 (100 \times)
mtDNA Control	211	328,256	4130	100 (100 \times)
Nuclear DNA Mock Degraded	147	2,614,853	968 *	409 (96.0%) **
Nuclear DNA Control	194	1,988,568	498 *	360 (84.5%) **

* Average diploid read depth per SNP, ** percent (%) SNP coverage.

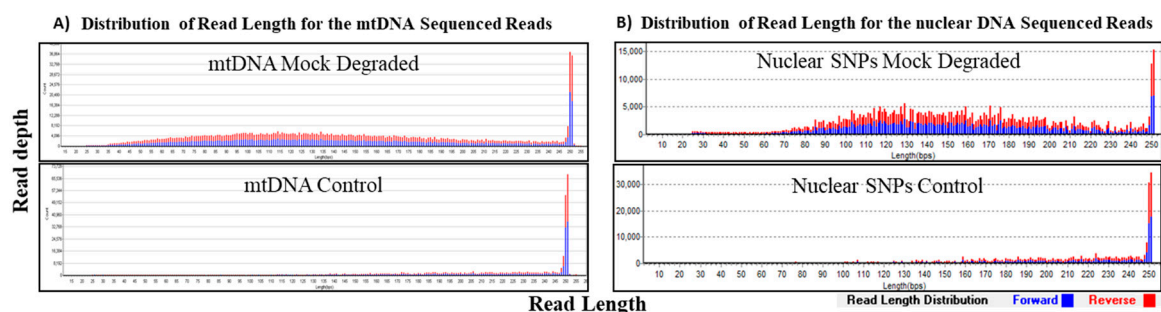


Figure 4. Sequencing results of mock degraded and control samples at 1 ng. Distribution of read length (x-axis) at corresponding coverage (y-axis) was determined for the mock degraded mtDNA sample and the control sample. Forward reads are in blue and the reverse reads are in red. (A) The average read length for the control mtDNA sample was ~ 200 bp and that of mock degraded mtDNA sample was ~ 150 bp. (B) For the nuclear SNPs control sample, the average read length was ~ 200 bp and ~ 150 bp for the mock degraded nuclear SNPs sample.

3.4. Mitochondrial DNA and Nuclear SNPs Recovery from Telogen Hairs

The performance of the mtgenome and nuclear SNP probe capture panels on forensic type samples was assessed using single shotgun DNA libraries constructed from telogen hairs captured in parallel. Analysis of the mtDNA sequencing results for the telogen hair P1FTR2 yielded 100% coverage of the mtgenome at $>100\times$ read depth per base (Figure 5B). High specificity of on-target alignment for hair P1FTR2 was achieved; $>92\%$ of $\sim 3,900,000$ reads were aligned to the rCRS reference (Table 5). Full coverage of the mtgenome was also achieved for the telogen hair A1FTR1 at $>100\times$ read depth per base with $>96\%$ alignment of ~ 14 million reads (Figure 5B, Table 5).

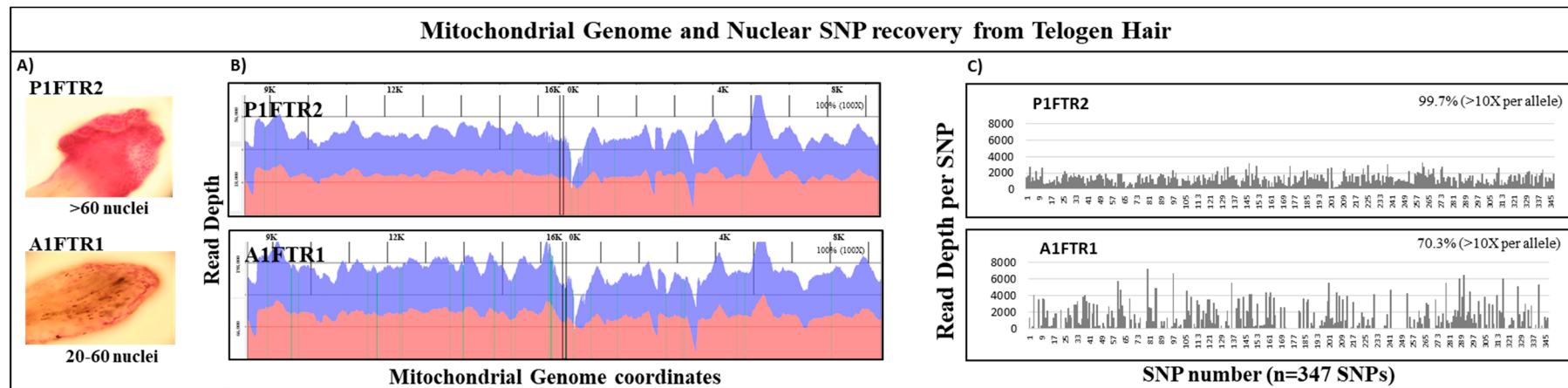


Figure 5. Recovery of mtDNA and nuclear SNPs from telogen hairs. (A) Microscopic images of stained telogen hair roots revealed that P1FTR2 exhibited >60 nuclei count while A1FTR1 exhibited 20–60 nuclei count. (B) mtgenome read depth plotted against mtgenome coverage coordinates was shown for P1FTR2 and A1FTR1. Blue coverage represents forward reads; red coverage represents reverse reads. Both P1FTR2 and A1FTR1 exhibited 100% coverage of the mtgenome at >100 × read depth per base. (C) SNP coverage for P1FTR2 and A1FTR1 showed that P1FTR2 exhibited near 100% SNP coverage while A1FTR1 exhibited ~70% SNP coverage. Refer to Table S1 for details of the SNP numbers.

Table 5. Sequencing statistics of mtDNA and nuclear SNPs recovery from telogen hairs.

Sample	Nuclei Count		mtDNA Recovery					Nuclear SNPs Recovery				
	Number Nuclei	mtDNA Copies	Avg. Read Length (bp)	Total Reads	Alignment (%)	Avg. Coverage	Coverage ** (%)	nuDNA amt. (ng)	Total Reads	Avg. Read Length (bp)	Avg. Coverage	SNP coverage ***
P1FTR2	>60	270,000	185	3,926,604	92.92%	46,624	100	1	3,741,800	204	1035	346 (99.7%) *
A1FTR1	20–60	540,000	193	14,015,748	96.14%	184,491	100	1	4,083,881	217	1068	244 (70.3%) *

* As samples were of female origin, the number of SNPs for analysis comprised of 347 autosomal and X SNPs. ** >100 × coverage. *** >10 × coverage per allele.

Nuclear SNP analysis of the same two hair samples revealed no reads for the Y chromosome SNPs as expected since both hairs were collected from female participants. By excluding the number of Y SNPs (79 SNPs) from the 426 total SNPs, 347 autosomal and X SNPs were analyzed for the two hairs. Sample P1FTR2 with >60 nuclei yielded ~3 million reads with an average diploid read depth of ~1000 reads per SNP while sample A1FTR1 with ~20–60 nuclei yielded ~4 million reads with an average diploid read depth ~1000 reads per SNP (Figure 5A, Table 5). Out of the 347 autosomal and X SNPs, near 100% (346 SNPs) were recovered for P1FTR2 with >60 nuclei and ~70% (244 SNPs) were recovered for A1FTR1 with ~20–60 nuclei at >10× allelic read depth (Figure 5C, Table 5). Using our two custom probe capture panels, one can provide NGS data for both mtDNA and nuclear DNA with high sensitivity and specificity from a single shotgun library for limited forensically challenging samples, such as these two telogen hairs.

3.5. Mitochondrial DNA Mixture Study

In the mtDNA mixture study, the performance of the capture assay on recovery of minor contributor mtDNA sequences was assessed with a two-person in vitro mixture. DNA from two contributors of different haplogroups were used to prepare the 80:20 mixture (Caucasian C163 and US Hispanic H104). Analysis of the two-person mixture yielded full coverage of the mtgenome at >100× read depth per base. Sequencing results showed ~86% of the ~1 million reads were aligned to the rCRS reference with an average coverage of ~13,215 coverage per base (data not shown).

Similar contributor proportion estimation of the 80:20 mixture was observed using both the variant frequency based GeneMarker[®]HTS software and the phylogenetic based Mixemt software. The major and minor mutation reports from the GeneMarker[®]HTS analysis were used for a haplogroup search in EMPOP (EDNAP mtDNA Population Database) [56]. The major contributor with haplogroup K2a6 was estimated to be 79.58% and the minor contributor haplogroup C1b11 at 18.68% using the GeneMarker[®]HTS software (Table 6). All variants pertaining to the minor contributor, including both phylogenetic and private mutations, were accurately assigned by GeneMarker[®]HTS (Figure 6A). Mixemt directly yielded major contributor haplogroup K2a6 at 82.2% and minor contributor haplogroup C1b11 at 17.8%. While Mixemt successfully assigned mtDNA sequence fragments to the contributing major and minor haplogroups based on the phylogenetic variants (marked by triangles in Figure 6B), the software does not consider private mutations (marked by circles in Figure 6B). Overall, both GeneMarker[®]HTS and Mixemt were successful in separating and estimating the major and minor contributors in the observed 80:20 mixture. These results demonstrated the robustness of our probe capture panel to capture minor contributor sequences with ~20% of ~300,000 mtDNA copies. In general, future mixture de-convolution can be most reliably achieved by using both frequency based, when possible, and phylogenetic based software to better characterize and understand private and phylogenetic variants in the human mtgenome.

Table 6. Haplogroup determination and contributor proportions estimation for a contrived 80:20 mixture using GeneMarker[®]HTS and Mixemt Software.

Mixture	Haplogroups		Contributor Proportions (%)	
	GeneMarker [®] HTS	Mixemt	GeneMarker [®] HTS	Mixemt
C163 (Major)	K2a6	K2a6	79.58	82.2
H104 (Minor)	C1b11	C1b11	18.68	17.8

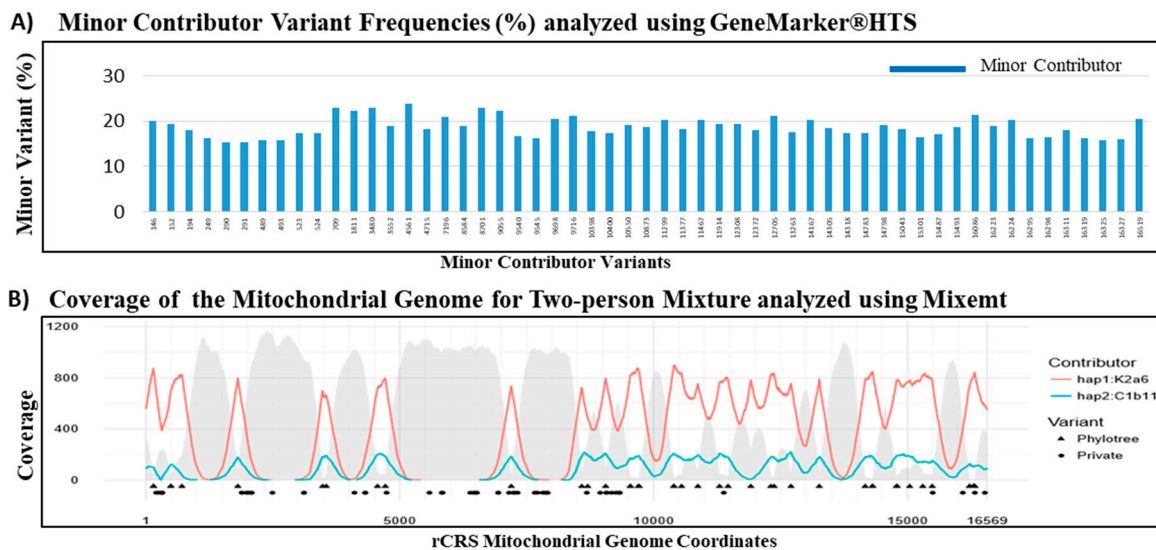


Figure 6. De-convolution of mtDNA mixtures using GeneMarker®HTS and Mixemt Software. **(A)** The variant frequency based software GeneMarker®HTS successfully separated all minor contributor variants from the 80:20 two-person mixture. **(B)** The phylogenetic based software Mixemt also successfully assigned the sequence reads to major and minor contributors based on the phylogenetic variants (triangles). Red line represents the depth of coverage for reads containing phylogenetic informative variants of haplogroup K2a6, teal line represents the depth of coverage for reads containing phylogenetic informative variants of haplogroup C1b11, and gray areas indicate depth of coverage for the reads that were unassigned. Private mutations (circles) did not affect Mixemt in assigning sequence reads to their respective contributors.

4. Discussion

The clonal and massively parallel features of NGS offer many advantages for the analysis of forensic samples that contain degraded, limited, and/or mixed DNA. One advantage of NGS is that it provides a single platform capable of analyzing mtDNA sequence polymorphism, STRs, and SNPs. However, most strategies for target enrichment rely on the construction of NGS libraries using PCR amplification. Although a relatively simple and efficient method, target enrichment by PCR is not as effective with samples of short DNA fragments as PCR requires the presence of two intact primer binding sites in the template molecule [57].

The hybrid capture method of target enrichment discussed here is capable of analyzing short DNA fragments and therefore, represents a valuable alternative strategy to PCR amplification. Previous studies using PCR based NGS systems have demonstrated limited success in recovering SNPs for samples with template sizes <100 bases [11]. Our experiments using size selected DNA fragments (≤ 75 bp) and samples sheared to simulate degraded DNA demonstrated that our hybrid probe capture NGS system is capable of recovering and sequencing very short fragments for both mtDNA and a panel of 426 nuclear SNPs (Tables 3 and 4). In DNA (1 ng input) fragmented to 150 bp (mock degraded), the average of the mtDNA sequence read size distribution was 146 bp, the shortest reads were around 35 bp and 100% coverage (based on $>100\times$ read counts per base) of the mtgenome was achieved. Similarly, for nuclear SNPs, the average size of the sequence read distribution was 147 bp, and the shortest reads were around 35 bp, and with 1 ng input, 96% coverage (based on read counts $>10\times$ per base) of 426 SNPs was achieved (Table 4). Included in the SNPs with no locus or allele dropout were ≥ 129 identity informative SNPs (with high heterozygosity); this number of SNPs is much higher than the 50–60 high heterozygosity SNPs that are required for a profile comparable in discriminatory power to the 13 STR [24,58–60]. Target enrichment based on a PCR amplicon of, for example, 100 bp would have failed to amplify most of the DNA fragments in the sample. In general, probe capture

methods for library preparation combined with the clonal sequencing aspect of NGS are particularly well suited to analysis for mixed samples with short DNA fragments such as the maternal plasma used in non-invasive prenatal testing [26,61].

Although the specificity of hybrid capture is less than that of PCR amplification, our system achieved 100% coverage ($>100\times$ read depth per base) for the mtgenome at 10 pg input DNA and an on-target rate of $>88\%$ (aligned to the rCRS reference). Off-target reads are then removed using bioinformatics software tools during the alignment process and specificity can be further increased if necessary by adjusting filters and increasing homology cut-offs [62]. In addition to the ability to analyze short DNA fragments, the hybrid capture method of enriching targets from a shotgun library can allow the removal of PCR duplicates, based on the unique start positions of the forward and reverse reads [53]. This feature is critical in evaluating the potential for stochastic variation in sequence read counts and, consequently, in the establishment of thresholds for interpreting sequencing data [33,63]. In Figure 2, the average sequence read coverage per base for the mtDNA genome for an input of 100 pg and of 10 pg was virtually identical (7043 and 7191). However, with PCR duplicates removed, the average number of sequence reads per base was 183 and 22 for the 100 pg and 10 pg samples, respectively. By bioinformatically removing PCR duplicates, we can determine if the sequence data of the initial template DNA molecules were within the stochastic range. Setting thresholds based on the initial copy number can reduce the stochastic effects posed by possible allele dropouts with limited DNA samples during the probe capture enrichment process [15,33].

The NGS analysis of mtDNA sequences has been reported as an effective strategy for the de-convolution of the DNA mixtures characteristic of many forensic specimen [64,65]. The overlapping sequence reads derived from the randomly fragmented DNA in a shotgun library allow for efficient sequence assembly in a single source sample. However, assembling full genome sequences from the short reads obtained by NGS remains challenging for mixtures. In a 90:10 two-person mixture, the high frequency sequence reads can be presumed to be derived from the major contributor and the low frequency reads from the minor contributor. However, if the contributions from the two individuals were more balanced, or if there were more than two contributors to the mixture, then the frequency based assignment of sequence reads to individual contributor becomes problematic. A phylogenetic based software, Mixemt, co-estimates the number of contributors and their contributing proportions, and then assigns sequence reads to each haplogroup based on phylogenetically informative sites. In Figure 6, GeneMarker[®]HTS and Mixemt were both successful in resolving the major and minor contributor haplogroups in the two-person mixture. Furthermore, Vohr et al. demonstrated the ability to resolve a three-person in silico mixture using Mixemt [54]. However, since de-convolution of mixtures using Mixemt is based on phylogenetic variant information, private mutations are not always assigned during this process unless they are found on the same read as phylogenetically informative SNP. While variant frequency based software, such as Genemarker[®]HTS, can resolve mixtures at higher resolution, it is currently impossible to resolve mixtures involving more than two contributors [54]. Thus, the de-convolution of forensic mixtures can be most reliably achieved by using both frequency based and phylogenetic based software to better characterize and understand private and phylogenetic variants.

One of the advantages of the probe capture NGS strategy for analysis of forensically challenging samples is that a single shotgun library construction is needed for both the whole mtgenome and the nuclear SNP panel. This feature is illustrated in the analysis of two telogen hairs (Figure 5). Analysis of the mtDNA sequencing results for both telogen hairs yielded 100% coverage of the mtgenome at $>100\times$ read depth with $>92\%$ aligned to the rCRS reference (Figure 5). Nuclear SNP analysis (347 autosomal and X SNPs) of the same two hairs yielded an average diploid read depth of ~ 1000 reads per SNP. Overall, $\sim 99\%$ of the nuclear SNP markers were recovered for sample P1FTR2 and $>70\%$ of SNP markers were recovered for sample A1FTR1 at $>10\times$ allelic read depth (Table 5). Although SNP loci dropout occurred in both telogen hairs, the partial SNP profile combined with whole mtgenome sequence would increase the discrimination power and likely approach or exceed the discriminatory

power of a 13 STR loci profile based on the number of the SNPs recovered (>50–60 high heterozygosity SNPs) [24,58–60].

Thus, although the DNA probe capture strategy of target enrichment is less specific than PCR amplification, it provides many advantages for analysis of forensic samples, especially for samples with fragmented DNA. Probe capture enrichment is more effective at recovering and generating sequence reads from short DNA fragments and allows for removal of PCR duplicates to better quantitate the starting copies of the template DNA molecules. As demonstrated here, our custom probe capture panels can provide sensitive analyses of both mtDNA and nuclear DNA from one single shotgun library constructed from a limited forensically challenging sample.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4425/9/1/49/s1>. Table S1: Information Corresponding to SNP Number from Figures 3 and 5.

Acknowledgments: This research was supported by award No. 2013-MU-MU-K044, awarded by the National Institute of Justice, Office of Justice Programs, US. Department of Justice to Cassandra Calloway. The findings, recommendations, and conclusions expressed belong to the authors and do not necessarily reflect the official positions or policies of Department of Justice. Anna Beatriz R. Gonçalves received fellowship funding from the Capes Foundation within the Ministry of Education, Brazil. The authors would like to thank Rachel Gordon for assisting in in vitro preparation of the mtDNA sensitivity samples; Guillermina Almada for assisting in preparation and staining of the telogen hair samples; and Sam Vohr for generating the mtgenome haplogroup coverage graph for the contrived mixture sample. Finally, the authors express gratitude to George Sensabaugh for his insightful advice throughout this research.

Author Contributions: C.D.C. conceived and designed the mtDNA experiments. C.D.C. and N.B. conceived and designed the nuclear DNA experiments. N.B. and A.B.R.G. performed the experiments. S.Y.S. and N.B. performed bioinformatics processing. S.Y.S., N.B., and C.D.C. analyzed the data. S.Y.S., N.B., H.A.E. and C.D.C. wrote the paper. All authors reviewed the final draft.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Eduardoff, M.; Xavier, C.; Strobl, C.; Casas-Vargas, A.; Parson, W. Optimized mtDNA control region primer extension capture analysis for forensically relevant samples and highly compromised mtDNA of different age and origin. *Genes* **2017**, *8*, 237. [[CrossRef](#)] [[PubMed](#)]
2. Templeton, J.E.; Brotherton, P.M.; Llamas, B.; Soubrier, J.; Haak, W.; Cooper, A.; Austin, J.J. DNA capture and next-generation sequencing can recover whole mitochondrial genomes from highly degraded samples for human identification. *Investig. Genet.* **2013**, *4*, 26. [[CrossRef](#)] [[PubMed](#)]
3. Molto, J.E.; Loreille, O.; Mallott, E.K.; Malhi, R.S.; Fast, S.; Daniels-Higginbotham, J.; Marshall, C.; Parr, R. Complete mitochondrial genome sequencing of a burial from a romano-christian cemetery in the Dakhleh Oasis, Egypt: Preliminary indications. *Genes* **2017**, *8*, 262. [[CrossRef](#)] [[PubMed](#)]
4. Mertes, F.; Elsharawy, A.; Sauer, S.; van Helvoort, J.M.; van der Zaag, P.J.; Franke, A.; Nilsson, M.; Lehrach, H.; Brookes, A.J. Targeted enrichment of genomic DNA regions for next-generation sequencing. *Brief. Funct. Genom.* **2011**, *10*, 374–386. [[CrossRef](#)] [[PubMed](#)]
5. Avila-Arcos, M.C.; Cappellini, E.; Romero-Navarro, J.A.; Wales, N.; Moreno-Mayar, J.V.; Rasmussen, M.; Fordyce, S.L.; Montiel, R.; Vielle-Calzada, J.P.; Willerslev, E.; et al. Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Sci. Rep.* **2011**, *1*, 74. [[CrossRef](#)] [[PubMed](#)]
6. Cummings, N.; King, R.; Rickers, A.; Kaspi, A.; Lunke, S.; Haviv, I.; Jowett, J.B. Combining target enrichment with barcode multiplexing for high throughput SNP discovery. *BMC Genom.* **2010**, *11*, 641. [[CrossRef](#)] [[PubMed](#)]
7. Rohland, N.; Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **2012**, *22*, 939–946. [[CrossRef](#)] [[PubMed](#)]
8. Buermans, H.P.; den Dunnen, J.T. Next generation sequencing technology: Advances and applications. *Biochim. Biophys. Acta* **2014**, *1842*, 1932–1941. [[CrossRef](#)] [[PubMed](#)]
9. Chen, C.M.; Sio, C.P.; Lu, Y.L.; Chang, H.T.; Hu, C.H.; Pai, T.W. Identification of conserved and polymorphic STRs for personal genomes. *BMC Genom.* **2014**, *15* (Suppl. 10), S3. [[CrossRef](#)] [[PubMed](#)]

10. Koboldt, D.C.; Steinberg, K.M.; Larson, D.E.; Wilson, R.K.; Mardis, E.R. The next-generation sequencing revolution and its impact on genomics. *Cell* **2013**, *155*, 27–38. [[CrossRef](#)] [[PubMed](#)]
11. Gettings, K.B.; Kiesler, K.M.; Vallone, P.M. Performance of a next generation sequencing SNP assay on degraded DNA. *Forensic Sci. Int. Genet.* **2015**, *19*, 1–9. [[CrossRef](#)] [[PubMed](#)]
12. Chaitanya, L.; Ralf, A.; van Oven, M.; Kupiec, T.; Chang, J.; Lagace, R.; Kayser, M. Simultaneous whole mitochondrial genome sequencing with short overlapping amplicons suitable for degraded DNA using the ion torrent personal genome machine. *Hum. Mutat.* **2015**, *36*, 1236–1247. [[CrossRef](#)] [[PubMed](#)]
13. 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](#)] [[PubMed](#)]
14. Parson, W.; Gusmao, 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](#)] [[PubMed](#)]
15. Mamanova, L.; Coffey, A.J.; Scott, C.E.; Kozarewa, I.; Turner, E.H.; Kumar, A.; Howard, E.; Shendure, J.; Turner, D.J. Target-enrichment strategies for next-generation sequencing. *Nat. Methods* **2010**, *7*, 111–118. [[CrossRef](#)] [[PubMed](#)]
16. Bar, W.; Brinkmann, B.; Budowle, B.; Carracedo, A.; Gill, P.; Holland, M.; Lincoln, P.J.; Mayr, W.; Morling, N.; Olaisen, B.; et al. Guidelines for mitochondrial DNA typing. DNA Commission of the International Society for Forensic Genetics. *Vox Sang.* **2000**, *79*, 121–125. [[CrossRef](#)] [[PubMed](#)]
17. Calloway, C.D.; Reynolds, R.L.; Herrin, G.L., Jr.; Anderson, W.W. The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. *Am. J. Hum. Genet.* **2000**, *66*, 1384–1397. [[CrossRef](#)] [[PubMed](#)]
18. Tsiatis, A.C.; Norris-Kirby, A.; Rich, R.G.; Hafez, M.J.; Gocke, C.D.; Eshleman, J.R.; Murphy, K.M. Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: Diagnostic and clinical implications. *J. Mol. Diagn.* **2010**, *12*, 425–432. [[CrossRef](#)] [[PubMed](#)]
19. Alaeddini, R.; Walsh, S.J.; Abbas, A. Forensic implications of genetic analyses from degraded DNA—A review. *Forensic Sci. Int. Genet.* **2010**, *4*, 148–157. [[CrossRef](#)] [[PubMed](#)]
20. Opel, K.L.; Chung, D.T.; Drabek, J.; Butler, J.M.; McCord, B.R. Developmental validation of reduced-size STR Miniplex primer sets. *J. Forensic Sci.* **2007**, *52*, 1263–1271. [[CrossRef](#)] [[PubMed](#)]
21. Evett, I.W.; Buffery, C.; Willott, G.; Stoney, D. A guide to interpreting single locus profiles of DNA mixtures in forensic cases. *J. Forensic Sci. Soc.* **1991**, *31*, 41–47. [[CrossRef](#)]
22. Butler, J.M.; Shen, Y.; McCord, B.R. The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J. Forensic Sci.* **2003**, *48*, 1054–1064. [[CrossRef](#)] [[PubMed](#)]
23. Dixon, L.A.; Dobbins, A.E.; Pulker, H.K.; Butler, J.M.; Vallone, P.M.; Coble, M.D.; Parson, W.; Berger, B.; Grubwieser, P.; Mogensen, H.S.; et al. Analysis of artificially degraded DNA using STRs and SNPs—results of a collaborative European (EDNAP) exercise. *Forensic Sci. Int.* **2006**, *164*, 33–44. [[CrossRef](#)] [[PubMed](#)]
24. Budowle, B. SNP typing strategies. *Forensic Sci. Int.* **2004**, *146*, S139–S142. [[PubMed](#)]
25. Pakstis, A.J.; Speed, W.C.; Fang, R.; Hyland, F.C.; Furtado, M.R.; Kidd, J.R.; Kidd, K.K. SNPs for a universal individual identification panel. *Hum. Genet.* **2010**, *127*, 315–324. [[CrossRef](#)] [[PubMed](#)]
26. Schwarzenbach, H.; Hoon, D.S.; Pantel, K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat. Rev. Cancer* **2011**, *11*, 426–437. [[CrossRef](#)] [[PubMed](#)]
27. Budowle, B.; Wilson, M.R.; DiZinno, J.A.; Stauffer, C.; Fasano, M.A.; Holland, M.M.; Monson, K.L. Mitochondrial DNA regions HVI and HVII population data. *Forensic Sci. Int.* **1999**, *103*, 23–35. [[CrossRef](#)]
28. 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](#)]
29. Calloway, C.D.; Kim, H.; Erlich, H.A. Resolution of DNA mixtures and analysis of degraded DNA using the 454 DNA sequencing technology. Available online: <https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=271240> (accessed on 7 December 2017).
30. Winters, M.; Monroe, C.; Barta, J.L.; Kemp, B.M. Are we fishing or catching? Evaluating the efficiency of bait capture of CODIS fragments. *Forensic Sci. Int. Genet.* **2017**, *29*, 61–70. [[CrossRef](#)] [[PubMed](#)]
31. Maricic, T.; Whitten, M.; Paabo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* **2010**, *5*, e14004. [[CrossRef](#)] [[PubMed](#)]

32. Mokry, M.; Feitsma, H.; Nijman, I.J.; de Bruijn, E.; van der Zaag, P.J.; Guryev, V.; Cuppen, E. Accurate SNP and mutation detection by targeted custom microarray-based genomic enrichment of short-fragment sequencing libraries. *Nucleic Acids Res.* **2010**, *38*, e116. [[CrossRef](#)] [[PubMed](#)]
33. Gnirke, A.; Melnikov, A.; Maguire, J.; Rogov, P.; LeProust, E.M.; Brockman, W.; Fennell, T.; Giannoukos, G.; Fisher, S.; Russ, C.; et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **2009**, *27*, 182–189. [[CrossRef](#)] [[PubMed](#)]
34. Chilamakuri, C.S.; Lorenz, S.; Madoui, M.A.; Vodak, D.; Sun, J.; Hovig, E.; Myklebost, O.; Meza-Zepeda, L.A. Performance comparison of four exome capture systems for deep sequencing. *BMC Genom.* **2014**, *15*, 449. [[CrossRef](#)] [[PubMed](#)]
35. Clark, M.J.; Chen, R.; Lam, H.Y.; Karczewski, K.J.; Chen, R.; Euskirchen, G.; Butte, A.J.; Snyder, M. Performance comparison of exome DNA sequencing technologies. *Nat. Biotechnol.* **2011**, *29*, 908–914. [[CrossRef](#)] [[PubMed](#)]
36. Sulonen, A.M.; Ellonen, P.; Almusa, H.; Lepisto, M.; Eldfors, S.; Hannula, S.; Miettinen, T.; Tyynismaa, H.; Salo, P.; Heckman, C.; et al. Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol.* **2011**, *12*, R94. [[CrossRef](#)] [[PubMed](#)]
37. Enk, J.M.; Devault, A.M.; Kuch, M.; Murgua, Y.E.; Rouillard, J.M.; Poinar, H.N. Ancient whole genome enrichment using baits built from modern DNA. *Mol. Biol. Evol.* **2014**, *31*, 1292–1294. [[CrossRef](#)] [[PubMed](#)]
38. Cruz-Davalos, D.I.; Llamas, B.; Gauntz, C.; Fages, A.; Gamba, C.; Soubrier, J.; Librado, P.; Seguin-Orlando, A.; Pruvost, M.; Alfarhan, A.H.; et al. Experimental conditions improving in-solution target enrichment for ancient DNA. *Mol. Ecol. Resour.* **2017**, *17*, 508–522. [[CrossRef](#)] [[PubMed](#)]
39. Carpenter, M.L.; Buenrostro, J.D.; Valdiosera, C.; Schroeder, H.; Allentoft, M.E.; Sikora, M.; Rasmussen, M.; Gravel, S.; Guillen, S.; Nekhrizov, G.; et al. Pulling out the 1%: Whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *Am. J. Hum. Genet.* **2013**, *93*, 852–864. [[CrossRef](#)] [[PubMed](#)]
40. Green, R.E.; Briggs, A.W.; Krause, J.; Prufer, K.; Burbano, H.A.; Siebauer, M.; Lachmann, M.; Paabo, S. The Neandertal genome and ancient DNA authenticity. *EMBO J.* **2009**, *28*, 2494–2502. [[CrossRef](#)] [[PubMed](#)]
41. Green, R.E.; Krause, J.; Ptak, S.E.; Briggs, A.W.; Ronan, M.T.; Simons, J.F.; Du, L.; Egholm, M.; Rothberg, J.M.; Paunovic, M.; et al. Analysis of one million base pairs of Neanderthal DNA. *Nature* **2006**, *444*, 330–336. [[CrossRef](#)] [[PubMed](#)]
42. Green, R.E.; Malaspina, A.S.; Krause, J.; Briggs, A.W.; Johnson, P.L.; Uhler, C.; Meyer, M.; Good, J.M.; Maricic, T.; Stenzel, U.; et al. A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **2008**, *134*, 416–426. [[CrossRef](#)] [[PubMed](#)]
43. Briggs, A.W.; Good, J.M.; Green, R.E.; Krause, J.; Maricic, T.; Stenzel, U.; Lalueza-Fox, C.; Rudan, P.; Brajkovic, D.; Kucan, Z.; et al. Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **2009**, *325*, 318–321. [[CrossRef](#)] [[PubMed](#)]
44. Reich, D.; Green, R.E.; Kircher, M.; Krause, J.; Patterson, N.; Durand, E.Y.; Viola, B.; Briggs, A.W.; Stenzel, U.; Johnson, P.L.; et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **2010**, *468*, 1053–1060. [[CrossRef](#)] [[PubMed](#)]
45. Cuenca, D. Optimization and Validation of a Probe Capture NGS Assay for Sequencing the Whole Mitochondrial Genome on Forensically Relevant Samples. Master's Thesis, Forensic Science, University of California, Davis, CA, USA, 2013.
46. Bose, N. Development of a Nuclear SNP Probe Capture Assay for Massively Parallel Sequencing of Degraded and Mixed DNA Samples. Master's Thesis, Forensic Science, University of California, Davis, CA, USA, 2016.
47. Timken, M.D.; Swango, K.L.; Orrego, C.; Buoncristiani, M.R. A duplex real-time qPCR assay for the quantification of human nuclear and mitochondrial DNA in forensic samples: Implications for quantifying DNA in degraded samples. *J. Forensic Sci.* **2005**, *50*, 1044–1060. [[CrossRef](#)] [[PubMed](#)]
48. Aceves, M. Development of a Real-Time qPCR Assay for the Evaluation of Human Nuclear and Mitochondrial DNA for Use in Missing Persons Casework. Master's Thesis, Forensic Science, University of California, Davis, CA, USA, 2012.
49. Almada, G. Improving Telogen Hair Analysis by Predicting Nuclear and Mitochondrial DNA Success for Massively Parallel Sequencing Using Microscopic and qPCR Methods. Master's Thesis, Forensic Science, University of California, Davis, CA, USA, 2017.
50. Mixemt. Available online: <https://github.com/svoehr/mixemt> (accessed on 13 November 2017).

51. Shih, S. Characterization of Germline Heteroplasmy in Mother-Offspring Pairs Using Next Generation Sequencing. Master's Thesis, Forensic Science, University of California, Davis, CA, USA, 2017.
52. 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](#)]
53. GeneMakerHTS Software Quick Start Guide. Available online: <http://www.softgenetics.com/GeneMarkerHTS.php> (accessed on 19 January 2017).
54. Vohr, S.H.; Gordon, R.; Eizenga, J.M.; Erlich, H.A.; Calloway, C.D.; Green, R.E. A phylogenetic approach for haplotype analysis of sequence data from complex mitochondrial mixtures. *Forensic Sci. Int. Genet.* **2017**, *30*, 93–105. [[CrossRef](#)] [[PubMed](#)]
55. Minaschek, G.; Bereiter-Hahn, J.; Bertholdt, G. Quantitation of the volume of liquid injected into cells by means of pressure. *Exp. Cell Res.* **1989**, *183*, 434–442. [[CrossRef](#)]
56. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595. [[CrossRef](#)] [[PubMed](#)]
57. Chou, L.S.; Liu, C.S.; Boese, B.; Zhang, X.; Mao, R. DNA sequence capture and enrichment by microarray followed by next-generation sequencing for targeted resequencing: Neurofibromatosis type 1 gene as a model. *Clin. Chem.* **2010**, *56*, 62–72. [[CrossRef](#)] [[PubMed](#)]
58. Budowle, B.; van Daal, A. Forensically relevant SNP classes. *Biotechniques* **2008**, *44*, 603–608, 610. [[CrossRef](#)] [[PubMed](#)]
59. Chakraborty, R.; Stivers, D.N.; Su, B.; Zhong, Y.; Budowle, B. The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing systems. *Electrophoresis* **1999**, *20*, 1682–1696. [[CrossRef](#)]
60. Kidd, K.K.; Pakstis, A.J.; Speed, W.C.; Grigorenko, E.L.; Kajuna, S.L.; Karoma, N.J.; Kungulilo, S.; Kim, J.J.; Lu, R.B.; Odunsi, A.; et al. Developing a SNP panel for forensic identification of individuals. *Forensic Sci. Int.* **2006**, *164*, 20–32. [[CrossRef](#)] [[PubMed](#)]
61. Chan, K.C.; Zhang, J.; Hui, A.B.; Wong, N.; Lau, T.K.; Leung, T.N.; Lo, K.W.; Huang, D.W.; Lo, Y.M. Size distributions of maternal and fetal DNA in maternal plasma. *Clin. Chem.* **2004**, *50*, 88–92. [[CrossRef](#)] [[PubMed](#)]
62. Holland, M.M.; Pack, E.D.; McElhoe, J.A. Evaluation of GeneMarker((R)) HTS for improved alignment of mtDNA MPS data, haplotype determination, and heteroplasmy assessment. *Forensic Sci. Int. Genet.* **2017**, *28*, 90–98. [[CrossRef](#)] [[PubMed](#)]
63. Porreca, G.J.; Zhang, K.; Li, J.B.; Xie, B.; Austin, D.; Vassallo, S.L.; LeProust, E.M.; Peck, B.J.; Emig, C.J.; Dahl, F.; et al. Multiplex amplification of large sets of human exons. *Nat. Methods* **2007**, *4*, 931–936. [[CrossRef](#)] [[PubMed](#)]
64. Kim, H.; Erlich, H.A.; Calloway, C.D. Analysis of mixtures using next generation sequencing of mitochondrial DNA hypervariable regions. *Croat. Med. J.* **2015**, *56*, 208–217. [[CrossRef](#)] [[PubMed](#)]
65. Holland, M.M.; McQuillan, M.R.; O'Hanlon, K.A. Second generation sequencing allows for mtDNA mixture deconvolution and high resolution detection of heteroplasmy. *Croat. Med. J.* **2011**, *52*, 299–313. [[CrossRef](#)] [[PubMed](#)]

