

Boise State University ScholarWorks

Biology Faculty Publications and Presentations

Department of Biological Sciences

7-1-2013

Safeguarding Forensic DNA Reference Samples with Nullomer Barcodes

Jayita Goswami
Boise State University

Michael C. Davis
Boise State University

Tim Andersen
Boise State University

Abdelkrim Alileche
Boise State University

Greg Hampikian
Boise State University

NOTICE: This is the author's version of a work that was accepted for publication in *Journal of Forensic and Legal Medicine*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Journal of Forensic and Legal Medicine*, Volume 20, Issue 5, (July 2013), DOI: [10.1016/j.jflm.2013.02.003](https://doi.org/10.1016/j.jflm.2013.02.003).

Safeguarding Forensic DNA Reference Samples with Nullomer Barcodes

Jayita Goswami

Boise State University

Michael C. Davis

Boise State University

Tim Andersen

Boise State University

Abdelkrim Alileche

Boise State University

Greg Hampikian

Boise State University

Journal of Forensic and Legal Medicine [Volume 20, Issue 5, July 2013, Pages 513-519]

Abstract

Unintended transfer of biological material containing DNA is a concern to all laboratories conducting PCR analysis. While forensic laboratories have protocols in place to reduce the possibility of contaminating casework samples, there is no way to detect when a reference sample is mislabeled as evidence, or contaminates a forensic sample. Thus there is public concern regarding the safeguarding of DNA submitted to crime labs. We demonstrate a method of introducing an internal amplification control to reference samples, in the form of a nullomer barcode which is based upon sequences absent or rare from publically accessible DNA databases. The detection of this barcode would indicate that the source of analyzed DNA was from a reference sample provided by an individual, and not from an evidence sample. We demonstrate that the nullomers can be added directly to collection devices (FTA paper) to allow tagging during the process of sample collection. We show that such nullomer oligonucleotides can be added to existing forensic typing and quantification kits, without affecting genotyping or quantification results. Finally, we show that even when diluted a million-fold and spilled on a knife, the nullomer tags can be clearly detected. These tags support the National Research Council of the National Academy recommendation that “Quality control procedures should be designed to identify mistakes, fraud, and bias” in forensic science (National Academy of Sciences, 2009).

Keywords: Nullomers, Nullomer-tagged, Internal Amplification Controls (IAC), Internal Positive Controls (IPC), DNA Barcoding, Contamination, Forensic DNA, STR, Y-STR

1. Introduction

Unintended transfer of biological samples is an issue of great concern to all laboratories conducting sensitive analyses. This is particularly true for crime laboratories, where victims, suspects, and even investigators are asked to supply reference samples of their DNA for comparison to evidence profiles. Contamination, or unintended transfer of DNA, can happen at any time in the process of forensic DNA analysis. Instances of unintended transfer and mislabeling by personnel collecting and processing samples, while uncommon, have been documented¹⁻³, and concern about supplying reference samples has been raised by individuals and groups including those representing police officers in Connecticut and Missouri.^{4,5} This problem is likely to become more pronounced as forensic DNA techniques become ever more sensitive, and as databases grow with the increasing use of forensic DNA.⁶ While forensic DNA analysis is considered the “gold standard” of forensic science practice, there is the expectation that the

best possible practices of regulation and oversight be implemented. With regard to the forensic sciences in general, the 2009 report by the National Research Council of the National Academy recommends that “Quality control procedures should be designed to identify mistakes, fraud, and bias” as well as “confirm the continued validity and reliability of standard operating procedures and protocols.”¹ With this in mind, while recognizing that laboratory mistakes and mishandling are rare events, we present a means by which to provide an additional “safeguard” for reference samples used in forensic DNA analysis, employing sequences known to be rare or absent in nature^{7,8} as internal controls. These sequences (known as nullomers) could be used to both verify sample origin (as a reference sample, or any sample not collected from a crime scene) and as way to potentially track contamination if it occurs.

Studies have shown the propensity of DNA to be deposited by individuals via the mere touching of objects, such as DNA from fingerprints⁹ or the secondary transfer of touch-deposited DNA.⁹⁻¹¹ Ladd et al. (1999) found an average of 1-15 ng DNA recovered from touched objects (dependent on the object).¹⁰ Assuming that 1 ng of genomic DNA is the equivalent of genomes from ~170 cells¹², the swabs in the experiments of Ladd et al. picked up the DNA from ~170 to ~2500 cells. With modern techniques that can produce profiles from a few or even single cells¹³, it is expected that the detection of DNA from secondary and tertiary transfer will become more common. In the context of clinical genetic testing, a case of DNA contamination has been reported, and routine forensic STR typing was recommended to ensure that all DNA samples are truly from a single individual.¹⁴ Contamination of samples with amplified products has also been a concern, recognized since the early days of PCR.¹⁵ Laboratories which use enhanced techniques for low template DNA analysis have detected contamination in reagents, plastic ware and laboratory samples at levels that are below detection with less sensitive methods.¹⁶ While most crime labs have protocols in place to prevent PCR products from coming in contact with pre-amplified DNA samples, accidents can happen. Unfortunately, it is not presently possible to detect when a reference sample contaminates an evidentiary sample. A specific incident mentioned in the *Final Report of the Independent Investigator for the Houston PD Crime Lab* documents case in which a re-analysis indicated “that the victim’s reference sample was contaminated at some point in the handling of this sample.”³ The independent investigator concluded that the contamination probably occurred after extraction, and happened “most likely at the PCR amplification stage”.³

In order to safeguard against the accidental transfer or contamination of DNA samples collected from members of the public, we have developed a unique type of internal amplification control (IAC) that can serve as a universal tag and barcode and can be modified to further encode a variety of information, such as laboratory location, testing purpose, or date. These tags are based on the smallest sequences absent from all publicly available DNA databases (nullomers), first described by Hampikian and Andersen⁷, and nullomer technology is also being used to design small peptide drugs⁸. These nullomer IACs (also designated “nullomer barcodes”) can be embedded in collection devices such as swabs and FTA paper, thus allowing DNA samples to be tagged at the earliest stage in the DNA analysis process. The nullomer approach has generated recent interest in algorithms for counting and tracking biological sequences.^{17, 18} This paper describes, to the best of our knowledge, the first practical forensic application of nullomer sequences. Our results show that nullomer sequences can be used as an IAC, and as molecular tags and barcodes, successfully integrated into the multiplex PCR reactions of commercially available forensic profiling kits, and used along with PCR for sequencing. The use of IACs (sometimes designated internal positive controls, IPCs) is already commonplace for food and clinical microbiological testing,¹⁹ forensic quantification kits,²⁰ forensic human identification kits^{21,22,23}, and are particularly useful for detecting PCR inhibitors.²⁴ For PCR-based tests of food-borne pathogens, the European Standardization Committee has developed guidelines that require the presence of an IAC.²⁵ The nullomer approach described here is unique in two important ways: it can be used to distinguish between reference and evidentiary samples, and the tags are designed through an algorithm which identifies small sequences absent from the public databases of all sequenced organisms.⁸

2. Methods

2.1. Nullomer sequences and primer design

The algorithm of Hampikian and Andersen⁷ is able to process the entire set of biological sequence data found on NCBI’s web site in less than 8 hours, calculating the frequencies of all sequences up to length 17 (longer lengths can be calculated by our methods as well). Basically, all possible sequences up to a given length are generated, and each sequence is compared to those in the databases; any sequences that are not found in the databases are listed as nullomers. At length 17 there are (as of January 2011) approximately 700 million (695,038,288) absent sequences in the NCBI data bases.

From a list of absent 15-mers, we concatenated eight sequences to form a 120 bp nullomer “tag”. Several permutations of eight 15-mers were analyzed via the Oligoanalyzer tool available at the Integrated DNA Technologies (IDT) website (<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/>). Sequences were chosen to minimize the formation of secondary structure, and primer-binding regions were chosen so that the annealing temperature would match as closely as possible the 59 °C annealing temperature of the ABI kit protocols (58 °C for Powerplex™ Y kit). Most nullomer sequences have a high GC content; therefore the list of potential 15-mers was reduced to those having a GC content in the 40-60% range. However, since the primer binding region of the sequence is the determining factor in PCR specificity, the internal sequence is amenable to sequence modification, allowing nullomer tags to be used for DNA “cryptography” or barcodes. The sequences of the barcode primers used are shown in the supplementary material. The 120 bp construct was synthesized by IDT (Coralville, IA, USA) as two complimentary single stranded molecules. This complimentary pair was annealed to make a double stranded oligomer. After annealing, remaining single stranded molecules were removed with ExoSAP™ enzyme treatment. Concentration of nullomers (copies/uL) was calculated by measuring ds DNA concentration on a NanoDrop spectrophotometer, and calculating expected copy number from the molecular weight of each nullomer molecule. Primers were designed to yield amplicon sizes of 88, 90, and 114 bp. Several barcode concentrations were tested with different STR kits; we present the main results using 1,900 copies per PCR reaction (3,800 copies used in the Identifiler™ experiment shown in Fig. 1B), which gave barcode peaks at intensities comparable to the human STR alleles. In order to visualize the amplified nullomer peak on the 3130 Genetic Analyzer, the nullomer primers were ordered from IDT with a 6-FAM modification to the 5' end of one of the primer pair. In our experiments, we found that a rather high concentration of primers (2.5 uM) was useful for amplification of the nullomer tag such that the nullomer peaks approximated the size of normal amplicon peaks. Although higher than the standard range for PCR primers (0.1-0.5 uM), we saw no evidence for primer-dimer formation or non-specific amplification.

2.2 Co-amplification of forensic loci and Nullomer tags

We designed the tag (barcode) DNA amplicons to be of a size outside the range of human STR alleles, so that there is no confusion between the barcode and known STR allele peaks (See Fig. 1B and Fig. 3, and *Supplementary material*). Human DNA and barcode DNA was amplified according to manufacturer protocols (except for the addition of the nullomer barcode and barcode primers) with the following forensic DNA kits: Quantifiler Duo™ (Fig. 1A), Identifiler™ (Fig. 1B, Fig. S1-S5), Profiler Plus™ (Fig. S6 and S7), Yfiler™ (Fig. S8), and PowerPlex™ Y (Fig. S9 and S10)²⁶⁻³⁰. To test nullomer tag compatibility with mitochondrial DNA (mtDNA) sequencing, we amplified the HV-1 and HV-2 regions of human mtDNA in the presence of nullomers, with and without tag primers. Results from the Quantifiler Duo™ test, as well as results with Identifiler™ (performed in triplicate), were analyzed with unpaired *t*-tests and a one-way ANOVA, using GraphPad Prism version 5.00 for Windows, to test for significant differences, $p \leq 0.05$.

A common technique for the storage and processing of reference samples is to transfer a buccal swab sample to FTA paper and allow it to dry. DNA samples fixed onto FTA paper can be stored at room temperature, and then be extracted from FTA paper punches when needed. We treated an FTA card with a solution of nullomer DNA in nanopure H₂O, allowed it to dry, and then transferred DNA from a buccal swab to the FTA card.

2.3 Mock contamination experiments with Nullomer-tagged reference and post-amplification samples

We setup a mock unintended transfer of Identifiler™-amplified DNA (with nullomer barcode and nullomer primers co-amplified) to a mock evidentiary weapon (knife). The amplified DNA (~ 0.5 uL) was allowed to dry, and the knife swabbed later with a wet sterile Omni-swab (Whatman). DNA was extracted from the swab and analyzed with the Identifiler™ kit, with nullomer primers added to the PCR reaction.

Another mock contamination experiment was performed, with barcode-tagged reference DNA mixed with a forensic DNA extraction from the surface of a plastic coffee cup lid. A nullomer tag was added to a sample of “reference DNA” extracted from the buccal swab of a volunteer (reference DNA concentration was 36 ng/uL, as measured with a Nanodrop Spectrophotometer). The reference DNA sample was augmented with ~80,000 copies of barcode

(~2,000 copies of nullomer barcode per uL of reference DNA), and then 1.0 microliter of tagged reference sample was added to the forensic swab, prior to DNA extraction. The tagged DNA was extracted along with the forensic sample, to simulate a contamination event involving reference DNA. The nullomer primers used for this experiment generated a 113-bp amplicon.

3. Results

Adding the nullomer tag to a human buccal swab did not affect the quantification of extracted DNA (Fig. 1A). The kit used in our experiment (Quantifiler® Duo) is commonly employed by forensic laboratories to determine the DNA concentration for both the total human and male fraction (Y-chromosome) of a sample. We demonstrate that the nullomer tag does not negatively impact quantification of either total human or male fraction DNA (unpaired t-test, $p \leq 0.05$)

When amplified with the STR alleles of the forensic kits we tested, the 90 base pair nullomer barcode appears in the electropherogram as an additional peak outside the first “bin set” (regions where peaks from alleles from the various loci are known to occur). There were no differences in the DNA profile of individuals when barcode DNA was added. The nullomer DNA does not interfere with STR genotyping of individuals (Fig. 1B).

When we extracted and amplified DNA from nullomer-treated FTA paper, we obtained a profile which includes the barcode tag, identifying the DNA as coming from a reference source (Fig. S11), and not from evidence.

Barcode PCR products were co-amplified with mtDNA targets (when barcode primers were added to the reactions), and detected as distinct bands of ~90 bp (Fig. 1C, Fig. S12). Sanger sequencing of the amplified mtDNA was not affected by the presence of the nullomer barcode, whether the barcode was added during initial PCR using HV1 and HV2 primer sets, or if added to the sequencing reactions using HV1 or HV2 amplicons as templates (Fig. S13).

Amplification of barcode DNA along with human DNA in an Identifiler™ reaction did not adversely affect the amplification of the forensic loci, as shown by a comparison of mean peak heights (\pm SEM) for each allele, with and without the nullomer tag added and amplified (Fig. 2).

The analysis of the DNA collected from the mock forensic contamination event showed a clear signal of nullomer DNA, at two different dilutions. A 10^5 -fold dilution of the PCR product showed amplification of nullomer barcode DNA along with a partial profile of the transferred human amplicons (Fig. 2, top electropherogram). After a 10^6 -fold dilution, the nullomer barcode could be detected; even though the human profile was lost at our signal threshold of 100 RFUs (Fig. 2, lower electropherogram, Fig. S14).

Nullomer-tagged reference DNA can be detected when it contaminates a forensic DNA sample (Fig. 3a). Allelic drop-out was observed in this experiment, as is commonly seen in amplifications of low levels of DNA, but there was no correlation with barcode treatment. The amplification of the contaminated DNA mixture was carried out 5 times with and without nullomer barcode primers. A typical example with the barcode primers is shown (Fig. 3a), and two amplifications of the same extract without barcode primers (Fig. 3b, Fig. 3c). Some allelic drop out occurred whether or not the nullomer barcode was amplified (Table 1), however more extensive validation will have to be performed to optimize and validate particular nullomer tags.

We have shown that an artificial DNA barcode can be used in conjunction with forensic genetic analysis kits, without affecting DNA quantification, STR amplification, profile determination, or mitochondrial sequence, using standard protocols. DNA profiles were obtained by amplification of 0.5 – 1.0 nanograms of genomic DNA in the presence of 500-8000 copies of nullomer barcodes. We have stored the barcodes at 4° C for one year in TE buffer without affecting amplification and detection, and used FTA paper with dried barcodes for six months without any noticeable effect on amplification.

4. Discussion

Extrinsic DNA can enter the laboratory through contaminated reagents, disposables, centrifuges, and water baths.^{15, 16, 31} Reagent controls and routine monitoring can detect many of these examples, however, contamination of evidence with DNA from reference samples taken from suspects, or the switching of reference samples, is more difficult to detect, and may in fact implicate an innocent person in a crime.² The nullomer tags described in this paper were developed to assure the public that their reference samples can be marked so as to prevent false incrimination. Further development of the tag technology can be used to code individual samples to further safeguard the public.

The initial tags described here could be added to collection kits so that reference samples are “safeguarded” from the point of collection. Physical and chemical modifications of the tags could be used to stabilize them further;³¹ though we demonstrate here that even unmodified synthetic DNA is sufficient.

Synthetic DNA barcoding has been used to make positive amplification controls for applications in a wide variety of fields, such as clinical microbiological testing³³ and food pathogen testing,¹⁹ and has been proposed for a variety of tagging and tracing protocols.^{34,35} Internal positive controls are already an important part of forensic genotyping and qPCR quantification kits.^{20, 23-30} In those kits, the IPC allows the forensic analyst to assess the presence of contaminating PCR inhibitors, such as hematin or humic acids.²⁵ Unlike our nullomer tagging protocol, current tags are present in the PCR reaction components, and so cannot distinguish reference from evidentiary samples. IPCs currently used in forensic kits have been designed such that the IPC sequence was checked against GenBank to ensure uniqueness—but the methods of sequence design are not always reported.³⁶ Non-human sequences have been used for this purpose, such as a portion of Sea Pansy (*Renilla reniformis*) luciferase gene³⁷, a hydra-specific (*Hydra vulgaris*) actin gene segment^{22,23}, or rat DNA.³⁸ In some commercial kits (e.g. the Quantifiler® Duo kit), the IPC is a trade secret and is simply designated a “synthetic polynucleotide” in the published literature.²⁰ Our approach of building from these small absent sequences is novel. While these sequences may eventually appear in a natural database, it is highly unlikely that concatamers based on them will—and for such a sequence to arise and work with our primers, and produce the proper sized amplicon, would be quite remarkable. An easy verification for checking that the peak does correspond to the correct nullomer tag would be sequencing the PCR products with barcode primers, which would reveal the exact nullomer sequence in a tagged sample. While we don’t envision sequencing every reference sample, the nullomer tags could be used to trace the source of contamination. This should be a rare event, but it would be valuable to trace the origin of suspected contamination. The nullomer barcode that we have designed is 120 bases long, and since only ~1,900 copies are needed per PCR reaction, the additional cost to sampling kits would be minimal. These barcodes could easily be added to FTA paper, liquid buffer, cotton swabs, or other components of human DNA sampling kits. While DNA has the power to free the wrongfully convicted,³⁹ it can also be the route of forensic error as illustrated by a recent, highly publicized wrongful conviction and incarceration, due to mislabeling of DNA samples.² As DNA sampling and archiving becomes routine, the public needs to be assured that their DNA is being properly collected, stored and interpreted.^{1,5, 40, 41} We propose that nullomer markers can support the National Research Council’s recommendations for strengthening and improving forensic science in the United States.¹

Acknowledgments

That authors wish to thank Josu Zubizarreta for his help in processing some of the samples. This work was funded by a grant from the US Department of Defense, Defense Threat Reduction Agency under contract number W81XWH-07-1-000, DNA Safeguard.

Disclosure

One of the authors (Greg Hampikian) has also applied for a patent covering the nullomer applications described in this article.

References

1. National Academy of Sciences, *Strengthening Forensic Science in the United States: A Path Forward*, The National Academies Press, Washington, DC, 2009, pp. 1-254.
2. J. Valley, Metro reviewing DNA cases after error led to wrongful conviction, *Las Vegas Sun*, July 7, 2011. Available from <http://www.lasvegassun.com/news/2011/jul/07/dna-lab-switch-led-wrongful-conviction-man-who-ser/> (accessed May 3, 2012).
3. M. R. Bromwich, (2006), Final Report of the Independent Investigator for the Houston Police Department Crime laboratory and Property Room, Houston, TX, <http://www.hpdlabinvestigation.org/>
4. C. Byers, St. Louis police riled over collection of their DNA, *STLtoday.com*, November 14, 2011. Available from http://www.stltoday.com/news/local/crime-and-courts/city-police-riled-over-collection-of-own-dna/article_8adfcf0a-2469-52d9-adad-d20f3805a883.html (accessed May 3, 2012).
5. D. Collins, Police cite privacy concerns over their own DNA, *CNSNews.com*, October 19, 2011. Available from <http://cnsnews.com/news/article/police-cite-privacy-concerns-over-their-own-dna> (accessed May 3, 2012).
6. M. E. Smith, Lets make the DNA identification database as inclusive as possible., *The Journal of Law, Medicine, and Ethics* 34 (2006), 385-389.
7. G. Hampikian, T. Andersen, Absent sequences: nullomers and primes, *Pac Symp Biocomputing* 12 (2007), 355-366. Available from <http://psb.stanford.edu/psb-online/proceedings/psb07/hampikian.pdf> (accessed May 3, 2012).
8. A. Alileche, J. Goswami, B. Bourland, M. Davis and G. Hampikian, Nullomer derived anticancer peptides (NulloPs): differential lethal effects on normal and cancer cells in vitro, *Peptides* (2012), in press, <http://dx.doi.org/10.1016/j.peptides.2012.09.015>
9. R. A. van Oorschot and M. K. Jones, DNA fingerprints from fingerprints, *Nature* 387 (1997), 767.
10. C. Ladd, M. S. Adamowicz, M. T. Bourke, C. A. Scherczinger and H. C. Lee, A systematic analysis of secondary DNA transfer, *J Forensic Sci* 44 (1999), 1270-1272.
11. A. Lowe, C. Murray, J. Whitaker, G. Tully and P. Gill, The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces, *Forensic Science International* 129 (2002), 25-34.
12. J. Butler, *Forensic DNA Typing: Biology, Technology, and Genetics of STR markers*, 2 ed, Academic Press, 2005.
13. S. Bruck, H. Evers, F. Heidorn, U. Muller, R. Kilper and M. A. Verhoff, Single cells for forensic DNA analysis--from evidence material to test tube, *J Forensic Sci* 56 (2011), 176-180.
14. G. L. Axler-DiPerte, E. Wurmbach, Z. M. Budimlija, B. Jian, F. Fogt and M. Prinz, Forensic STR analysis reveals DNA contamination previously undetected during clinical analysis of chronically inflamed tissues, *Forensic Science International: Genetics Supplement Series* 2 (2009), 14-16.
15. C. Urban, F. Gruber, M. Kundi, F. G. Falkner, F. Dorner and T. Hammerle, A systematic and quantitative analysis of PCR template contamination, *J Forensic Sci* 45 (2000), 1307-1311.
16. C. A. Scherczinger, C. Ladd, M. T. Bourke, M. S. Adamowicz, P. M. Johannes, R. Scherczinger, T. Beesley and H. C. Lee, A systematic analysis of PCR contamination, *J Forensic Sci* 44 (1999), 1042-1045.
17. A. J. Pinho, P. J. Ferreira, S. P. Garcia and J. M. Rodrigues, On finding minimal absent words, *BMC Bioinformatics* 10 (2009), 137.
18. Z.-D. Wu, T. Jiang and W.-J. Su, Efficient computation of shortest absent words in a genomic sequence, *Information Processing Letters* 110 (2010), 596-601.
19. D. Rodriguez-Lazaro, M. Pla, M. Scotti, H. J. Monzo and J. A. Vazquez-Boland, A novel real-time PCR for *Listeria monocytogenes* that monitors analytical performance via an internal amplification control, *Appl Environ Microbiol* 71 (2005), 9008-9012.
20. M. Barbisin, R. Fang, C. E. O'Shea, L. M. Calandro, M. R. Furtado and J. G. Shewale, Developmental Validation of the Quantifiler® Duo DNA Quantification Kit for Simultaneous Quantification of Total Human and Human Male DNA and Detection of PCR Inhibitors in Biological Samples, *Journal of Forensic Sciences* 54 (2009), 305-319.
21. Wurmb-Schwarz, N.v., et al., A simple Duplex-PCR to evaluate the DNA quality of anthropological and forensic samples prior short tandem repeat typing. *Legal Medicine* 2004. 6(2): p. 80-88.
22. Wurmb-Schwarz, N.v., et al., A new multiplex-PCR comprising autosomal and y-specific STRs and mitochondrial DNA to analyse highly degraded material. *Forensic Science International Genetics* 2009. 3(2), 96-103.

23. Zahra, N., et al., Development of internal amplification controls (IACs) for DNA profiling with the AmpFLSTR® SGM Plus® amplification kit. *Electrophoresis*, 2011. 32(11): p. 1371-1378.
24. J. Hoorfar, B. Malorny, A. Abdulmawjood, N. Cook, M. Wagner and P. Fach, Practical Considerations in Design of Internal Amplification Controls for Diagnostic PCR Assays, *Journal of Clinical Microbiology* 42 (2004), 1863-1868.
25. J. Hoorfar, N. Cook, B. Malorny, M. Wagner, D. De Medici, A. Abdulmawjood, P. Fach, Making Internal Amplification Control Mandatory for Diagnostic PCR, *J. Clin. Microbiol.* 41 (2003), 5835.
26. Applied Biosystems by Life Technologies™. Quantifiler® Duo Human DNA Quantification Kit User's Manual, 2005. Pub. No. 4391294.
27. Applied Biosystems by Life Technologies™. AmpFLSTR® Identifiler® PCR Amplification Kit User's Manual, 2011. Pub. No. 4323291.
28. Applied Biosystems by Life Technologies™. Profiler Plus® and Profiler Plus® ID PCR Amplification Kit User Guide, 2012. Pub. No. 4476688 Revision B 08.14.2012.
29. Applied Biosystems by Life Technologies™. AmpFLSTR® Yfiler® PCR Amplification Kit, User Guide, 2012. Pub. No. 4358101 Rev. H.
30. Promega Corporation. PowerPlex™ Y system technical manual, 2012. Part #TMD018.
31. R. Daniel and R. A. H. van Oorschot, An investigation of the presence of DNA on unused laboratory gloves, *Forensic Science International: Genetics Supplement Series* 3 (2011), e45-e46.
32. S. Meng, S. Zhan and J. Li, Nuclease-resistant double-stranded DNA controls or standards for hepatitis B virus nucleic acid amplification assays, *Virology* 6 (2009), 226.
33. A. Das, E. Spackman, D. Senne, J. Pedersen, D.L. Suarez, Development of an Internal Positive Control for Rapid Diagnosis of Avian Influenza Virus Infections by Real-Time Reverse Transcription-PCR with Lyophilized Reagents, *J. Clin. Microbiol.* 44 (2006), 3065-3073.
34. D. Heider and A. Barnekow, DNA-based watermarks using the DNA-Crypt algorithm, *BMC Bioinformatics* 8 (2007), 176.
35. I.H. Sabir, S. Haldorsen, J. Torgersen, P. Aleström, S. Gaut, H. Colleuille, T.S. Pedersen, N.-O. Kitterød, Synthetic DNA tracers: examples of their application in water related studies, in: A. Dassargues (Ed.), *Tracers and Modelling in Hydrogeology*, International Association of Hydrological Sciences, Wallingford, UK (2000), pp. 159-166.
36. K. L. Swango, M. D. Timken, M. D. Chong and M. R. Buoncristiani, A quantitative PCR assay for the assessment of DNA degradation in forensic samples, *Forensic Sci Int* 158 (2006), 14-26.
37. G. A. Benson, Improved Quantitation of Human DNA using Quantitative Template Amplification Technology, M.S. Thesis, Oklahoma State University, 2007.
38. M. T. Costello and J. W. Schumm, A single assay for human-specific quantification of less than 1 pg DNA and detection of the presence of PCR inhibitors in forensic samples, *International Congress Series* 1288 (2006), 753-755.
39. G. Hampikian, E. West and O. Akselrod, The genetics of innocence: analysis of 194 U.S. DNA exonerations, *Annu Rev Genomics Hum Genet* 12 (2011), 97-120.
40. I. E. Dror and G. Hampikian, Subjectivity and bias in forensic DNA mixture interpretation, *Science & Justice* 51 (2011), 204-208.
41. R. E. Gaensslen, Should biological evidence or DNA be retained by forensic science laboratories after profiling? No, except under narrow legislatively-stipulated conditions., *The Journal of Law, Medicine, and Ethics* 34 (2006), 375-379.

Figure Legends

Fig. 1a. Nullomer Tag does not interfere with Quantifiler DUO™ DNA quantification.

A standard curve was generated using control DNA supplied with the Quantifiler DUO™ kit from Applied Biosystems. The DNA standard was diluted, and the real-time PCR performed, according to the manufacturer protocol. Three sets of standards were made without the nullomer tag, and three sets were made with nullomer tag added to the reaction mix. Number of cycles to reach the quantification threshold (C_q) is shown, for DNA with (○) and without (□) 1.9×10^3 copies of nullomer barcode. Results show mean \pm SEM.

Fig. 1b. STR profile of female genomic DNA in the presence and absence of the nullomer barcode, amplified with the Identifiler™ kit.

Top electropherogram, STR amplification without the nullomer barcode. Bottom electropherogram, STR amplification in the presence of nullomer barcode (3.8×10^3 copies). Size of each amplified product is given in base pairs; the locus is indicated by labels above the peaks. The y-axis is in RFUs, and is scaled according to maximum peak height. The nullomer peak is the smallest fragment (90 bp amplicon).

Fig. 1c. HV2 region of mitochondrial DNA from male (left) and female (right) amplified in the presence and absence of the nullomer barcode.

Mitochondrial PCR product was visualized on a 3% agarose, ethidium bromide stained gel. HV2 product amplified properly with nullomer barcode (with and without nullomer primers added to the HV2 PCR reaction).

Fig 2. Average peak heights (per locus) of human DNA amplified with the Identifiler™ kit, with and without nullomer barcode and FAM-labeled primers added.

Experiment performed in triplicate. Results shown \pm SEM. White bars: control. Black bars: with 3,800 copies of nullomer barcode-1 added. Inset: average peak heights of all alleles. No significant differences between control and treatment peak heights, per locus, and over all loci (unpaired t-test, $p \leq 0.05$).

Fig. 3. Contamination of evidence with amplified DNA is detected with nullomer barcode, even when diluted 1,000,000 fold.

Top panel and inset: DNA from amplified STR profile (amplified with nullomer barcode) was diluted 100,000 fold in water, and then 1ul of the dilution was applied to a newly purchased knife. This knife was swabbed, and amplified according to the STR kit manufacturer's instructions (with the addition of nullomer primers). The STR profile from the contaminated knife shows both the contaminating profile, and the nullomer barcode peak. Bottom Panel Electropherogram: original PCR product was diluted 1:1,000,000 in water, and 1 uL of the diluted product was added to a newly purchased knife. The knife was swabbed and processed as above. The nullomer barcode amplicon is prominent, although no alleles from the amplified human profile are detected at this dilution.

Fig. 4. Contamination of forensic DNA sample with reference DNA is detected with nullomer barcode amplification.

Fig. 4a. Identifiler™ profile (blue channel) of DNA from coffee cup (forensic sample) that was contaminated with reference DNA.

The presence of nullomer tag peak indicates that a contamination event took place. Allele numbers highlighted in yellow indicate the contaminating reference DNA.

Fig. 4b and 4c. Identifiler™ profile of same mixture, except that nullomer primers were not added to the PCR reaction.

Note that stochastic effects from low amount of template DNA result in allele drop-out, with or without nullomer amplification (see text for more details).

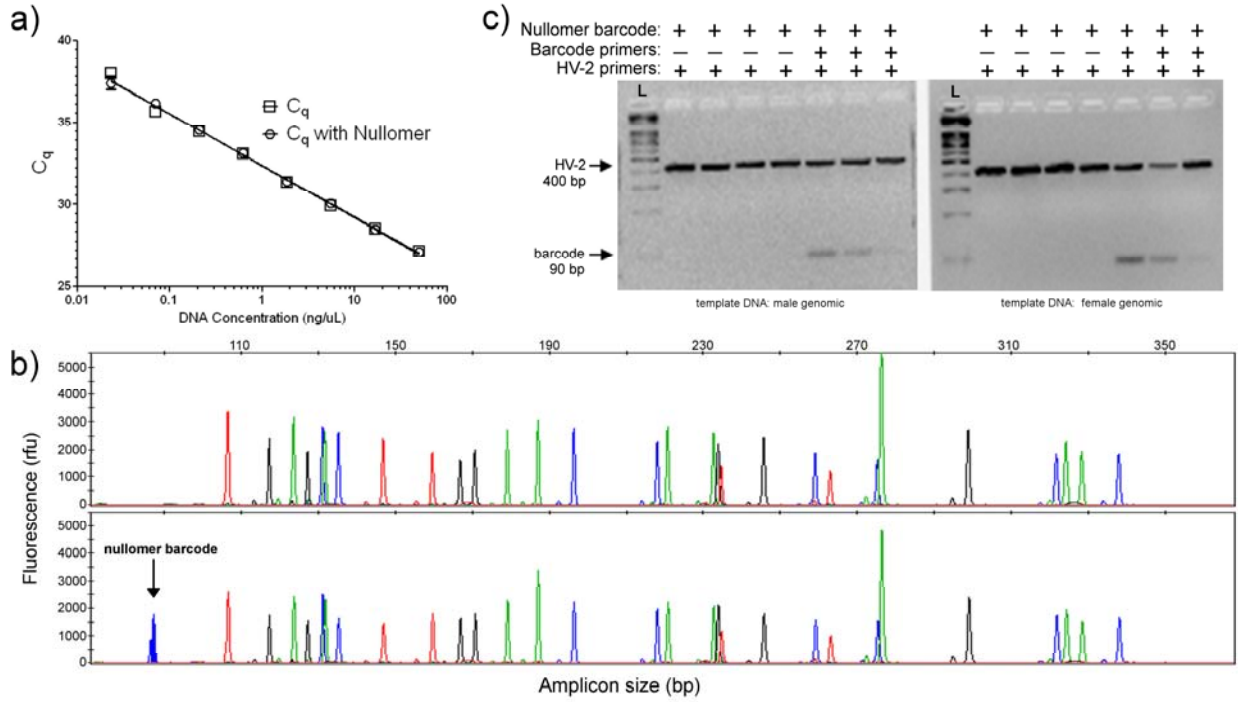


Fig. 1

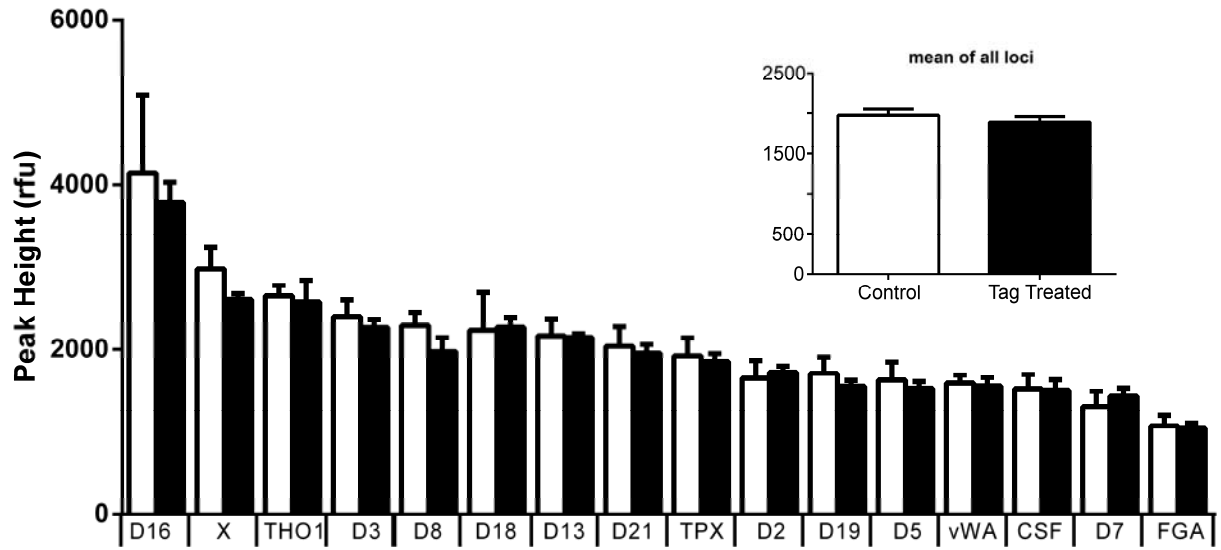


Fig. 2

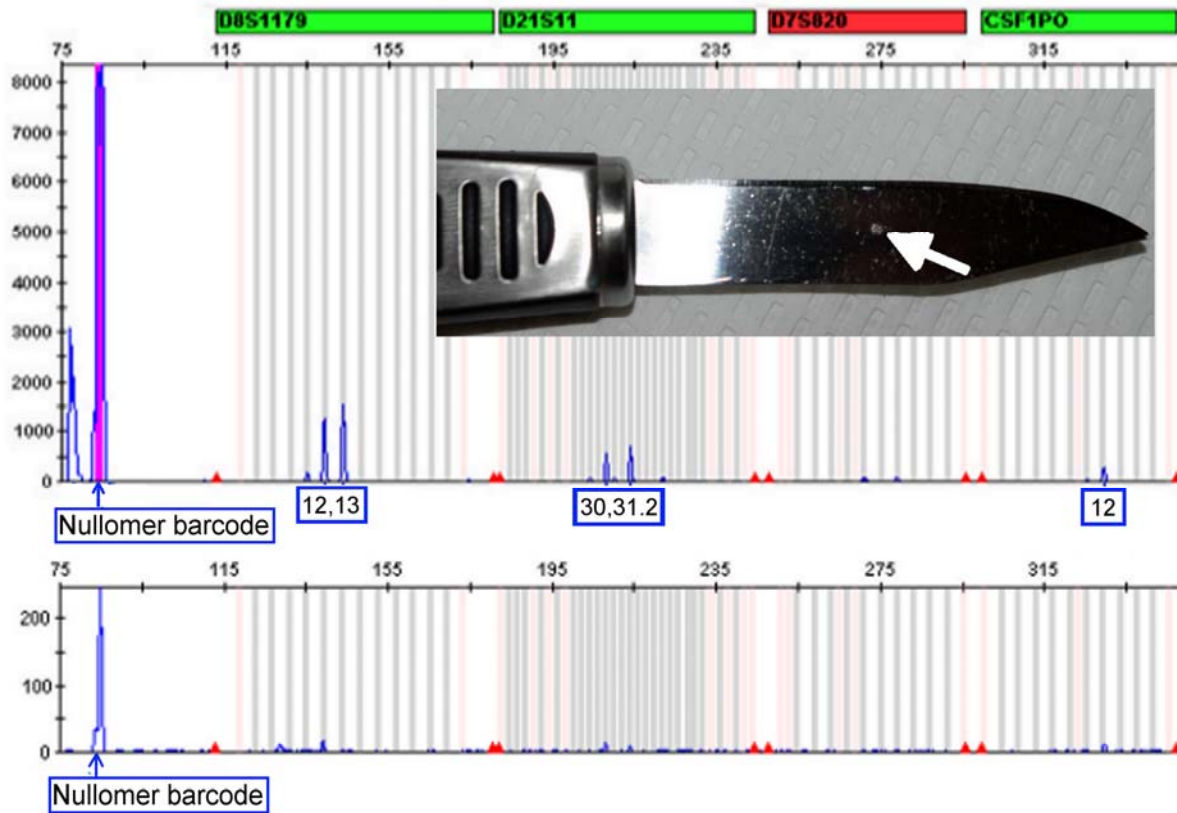


Fig. 3

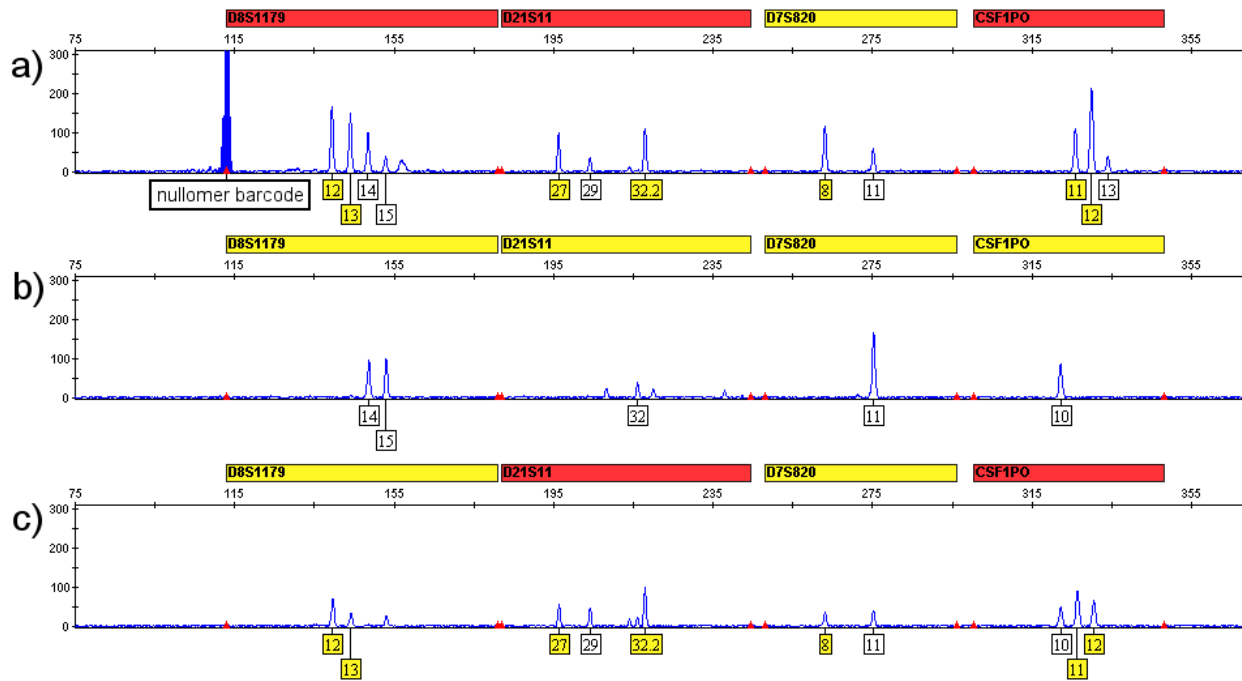


Fig. 4

Table 1. No significant difference in detectable alleles with or without Nullomer barcode. Stochastic allelic drop out, in 5 Identifiler runs with and without nullomer barcode primers for a low template DNA sample is shown in the table. Results are shown for the cup mixture profiles (Fig. 4) using Identifiler, with and without barcode primers. The detection threshold of 35 rfu cut-off was used for this low template DNA analysis. The results show stochastic drop out, with and without barcode primers. No significant difference between treatments was found, using χ^2 test; $p \leq 0.05$, $df = 13$. D7 and Amel were not used in χ^2 analysis (zero values).

# alleles dropped out	D8	D21	D7	CSF	D3	THO1	D13	D16	D2	D19	vWA	TPOX	D18	Amel	D5	FGA
Without nullomer primers	4	6	0	5	4	2	3	5	8	4	5	3	6	0	5	4
with nullomer primes	0	4	0	5	2	4	3	3	7	4	1	1	6	0	4	5

Supporting Online Material

This PDF file includes:

Supplementary Methods
Nullomer and Primer sequences
Supplementary Figures 1–13

Supplementary Methods

Nullomer Barcode Identification

Candidate nullomer sequences for barcodes were identified by examining all of the DNA sequence data available from the National Center for Biotechnology Information (NCBI) website using a robust software algorithm we developed for this purpose (Hampikian and Andersen 2007). The algorithm, which is capable of processing the entire data set in about 8 hours, simply counts the number of possible sequences up to a maximum specified length, and outputs the sequences that have a zero count. Currently, there are approximately 700 million DNA 17-mers that have a zero count, i.e. are not found in any of the DNA data available from NCBI (and 2,358,580 absent 16-mers as well as 34 absent 15-mers), which makes it possible to construct an enormous number of unique DNA based tags using relatively short sequences.

Nullomer Barcode Construction

To construct the barcodes for our experiments, eight 15-mer nullomer sequences were arranged in tandem. Sequences were chosen to minimize the formation of secondary structure, by running sequences in the OligoAnalyzer tool on the Integrated DNA Technologies (IDT) website (<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/>). Terminal sequences (primer binding regions) were chosen so that the annealing temperature would match as closely as possible the 59 °C annealing temperature of the ABI kit protocols (58 °C for Powerplex-Y kit). The 120 bp construct was synthesized by IDT (Coralville, IA, USA) as two complimentary single stranded molecules. This complimentary pair was annealed to make it double stranded, using a thermal gradient from 95°C to 75°C. After annealing the nullomer strands, ExoSAP™ enzyme treatment was used to cleave all the remaining single stranded molecules. Concentration of nullomers (copies/uL) was calculated by measuring ds DNA concentration on a NanoDrop spectrophotometer, and calculating expected copy number from the molecular weight of each nullomer molecule. Primers were designed to yield amplicon sizes of 88, 90, and 114 bp. Several barcode concentrations were tested with different STR kits; we present the results using 1,900 copies per PCR reaction, which gave barcode peaks at intensities comparable to the human STR alleles.

Real-Time PCR Quantification

For DNA quantification a DNA dilution series was made, according to the manufacturer's protocol, with the male genomic DNA standard supplied with the Duo kit. The nullomer barcode (without barcode primers) was added to each of the qPCR reactions. PCR was performed on an Eppendorf Mastercycler realplex instrument and data analyzed with Realplex software and GraphPad Prism 5 software.

STR and Y-STR Amplification

The compatibility of nullomer barcodes was tested with STR kits from ABI (Identifiler™, Yfiler™, Profiler Plus™) and Promega (PowerPlex™ Y). The Profiler Plus™ kit amplifies 10 loci, and the Identifiler amplifies 16 loci (including amelogenin). The Yfiler™ kit amplifies 17 Y-STR loci, and PowerPlex™ Y amplifies 12 Y-STR loci. For STR reactions, ABI and Promega's standard protocols were strictly followed, e.g. thermal cycling, reaction buffer, and primer mix concentrations. Approximately 1,900 nullomer barcode amplicons were added to each 1 ng human DNA samples extracted from buccal swabs. Barcode primers were at 2.5 μM final concentration per PCR reaction. The amplicons were analyzed in an ABI 3130 genetic analyzer, with Pop-4™ polymer, using Gene Mapper ID-X software.

Human Mitochondrial DNA Sequencing

Three different concentrations of nullomer barcode molecules were added to sequencing reactions: 962, 1900, and 3800 copies. Amplicons were visualized on a 3% agarose gel to verify both bands present (HV1 and HV2), along with the nullomer barcode bands. For sequencing reactions, nullomer barcodes were added to the reactions but not the barcode primers. Pop-4™ polymer and Big Dye 3.1® terminator chemistry were used for sequencing. PCR products were purified by ExoSap (USB) kit and also with the Bigdye® XTerminator™ kit (ABI) before sequencing. 3-10 ng DNA were used for each cycle sequencing reaction. Both the HV1 and HV2 control regions were analyzed to verify that the nullomer barcode did not interfere with mitochondrial sequencing reactions.

DNA extraction from swab and FTA paper

Approximately 1.5×10^8 copies of the barcodes were dissolved in 100 µl of double distilled water applied to 572.6 mm² FTA paper (Fitzco Inc, Minnesota, USA), and allowed to dry overnight. Then cells from a male and female donor's buccal swabs were applied to the FTA paper by gently rubbing each swab against the card. Five punches (2 mm diameter) were taken from each FTA card. Assuming an even distribution of nullomer barcode solution, we estimate that each punch had approximately 1.5×10^7 copies of the nullomer barcode. DNA extraction was performed using Qiagen's DNA Mini kit (according to manufacturer's protocol). The amount of total extracted DNA was quantified on a Nanodrop ND-1000. Quantification was visually checked by agarose gel electrophoresis with Lonza's quant ladder, and visualized with a Gel Doc XR imager (BIO-RAD) using Quantity One 4.6.5 software (BIO-RAD). One nanogram of extracted DNA was used to generate a profile using the Identifiler kit supplemented with forward and reverse barcode primers.

Mock "unintended transfer" of amplified DNA to a knife in the laboratory

DNA was extracted from a buccal swab (Whatman sterile OmniSwab), transferred to FTA paper containing the nullomer barcode (as above) and amplified with ABI's Identifiler kit. All reactions were processed according to manufacturer's instructions with the addition of the nullomer primers to the PCR reaction. After STR amplification, 1-10 µl of post amplified DNA was applied to several knives to see the abundance and stability of nullomer barcodes. Post PCR reactions were also diluted 100,000, and 1,000,000 times before applying to the knife and swabbing. Each knife was swabbed with a sterile omniswab, and DNA was extracted with Qiagen's DNA mini kit using a 15 minute, 56° C incubation in the extraction buffer. The extracted DNA was amplified with the Identifiler™ kit.

Mock contamination of nullomer-tagged reference DNA to a forensic DNA PCR reaction

We also set up a mock unintended transfer of barcode-tagged reference DNA to a forensic DNA extraction from the surface of a plastic coffee cup lid. Several swabs were taken from the area around the opening of the lid. To simulate a contamination event, a microliter of tagged "reference DNA" was added to the extraction buffer (barcode 2 was used in this experiment). As a negative control, reference DNA was added without nullomer tag. The reference DNA was at a concentration of 36 ng/uL, and it was tagged with ~80,000 copies/uL of nullomer 2 (~2,000 copies per uL). The extraction was performed on a Qiagen EZ1 Biorobot using the EZ1 DNA Tissue kit and the EZ1 DNA Forensic Card protocols. The extracted DNA was amplified with the Identifiler™ kit, both with and without nullomer primers added.

Sequences for Nullomer barcodes and primers used in this study.

Nullomer Barcode 1

TAC TAG GCG ACT CGA CGG TCT TAC GCG TTA CGT CCG ACT ATA GAG CTT AGA TTA GCG ACG
TTA GAC CTA TCG CGC CTT AGA TTA GCG ACG CTA GCG TAC GCT ACG GTC CTA ACG CGC TAT

Nullomer Barcode 2

ATA CTA GAC CGC TCG ATC CGA CCT AGC GTA CCT AGT ACG TTA CGA CGA CTA AGC ATA CCG
CTA ACT AAG TTC GCG ACG ATA GTC TAA CGC CGG TCT TAC GCG TTA CAT CGG ACT AGT ACC

Primer1_114F

TAG GCG ACT CGA CGG TCT TAC GCG TTA CGT

Primer1_114R

GCG CGT TAG GAC CGT AGC GTA CGC TAG CGT

Primer1_88F

TAC TAG GCG ACT CGA CGG TC

Primer1_88R

TCG CTA ATC TAA GGC GCG ATA GGT C

Primer1_90F

TAC TAG GCG ACT CGA CGG TC

Primer1_90R

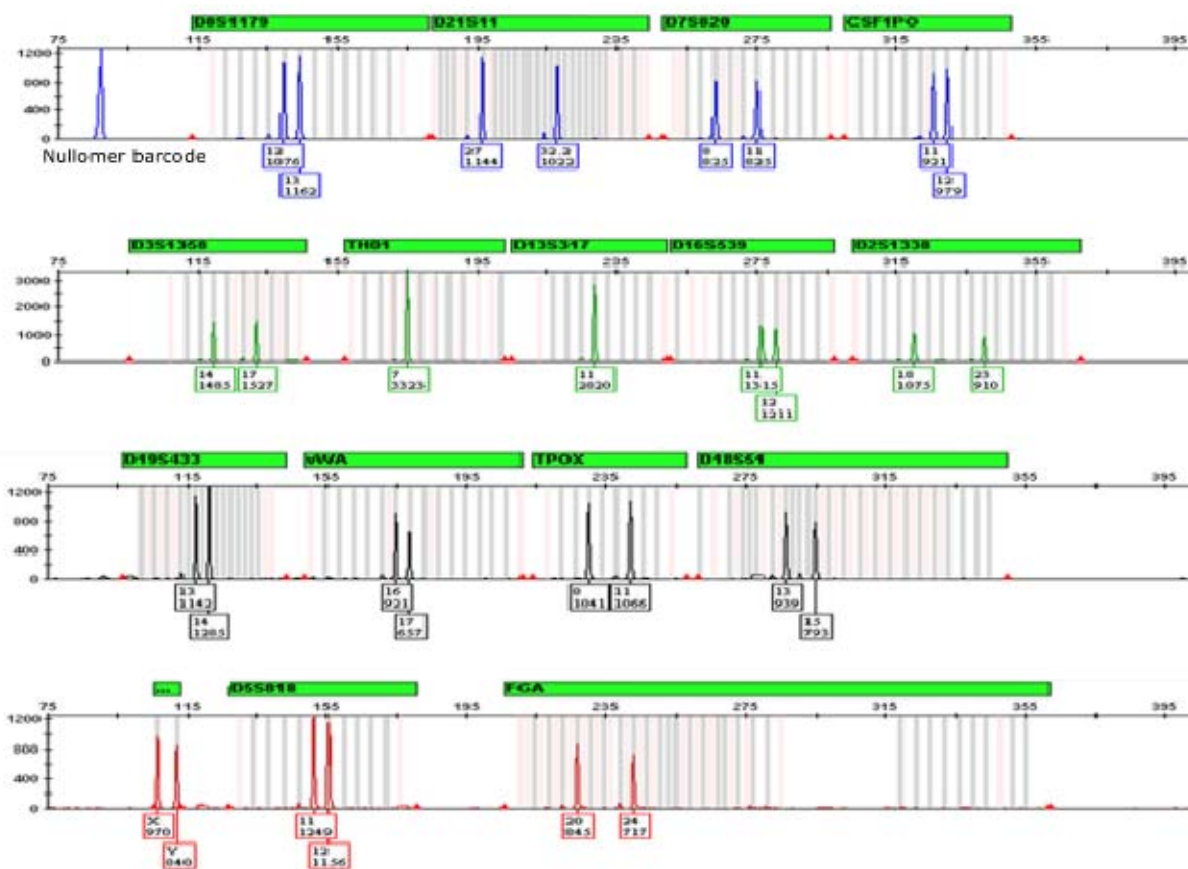
CGT CGC TAA TCT AAG GCG CGA

Primer2_115F

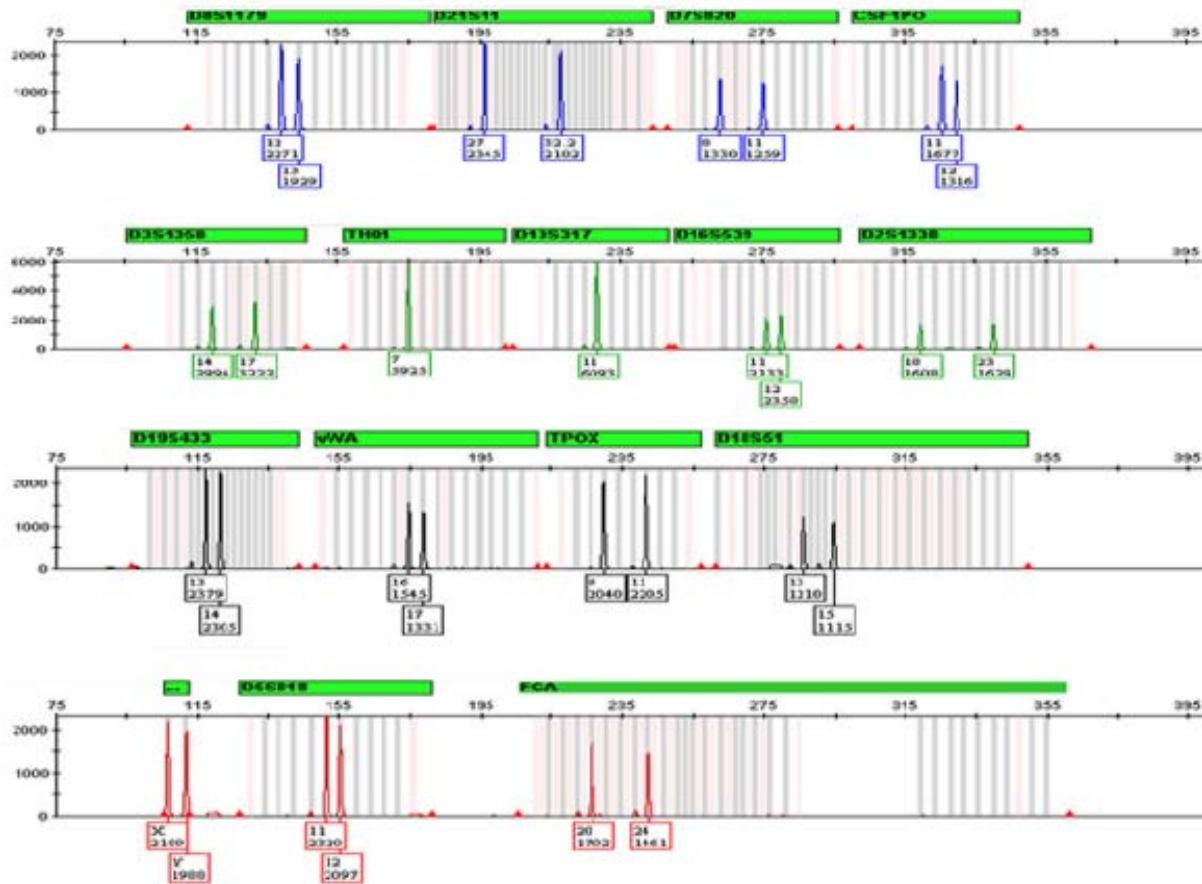
ACT AGA CCG CTC GAT CCG ACC T

Primer2_110R

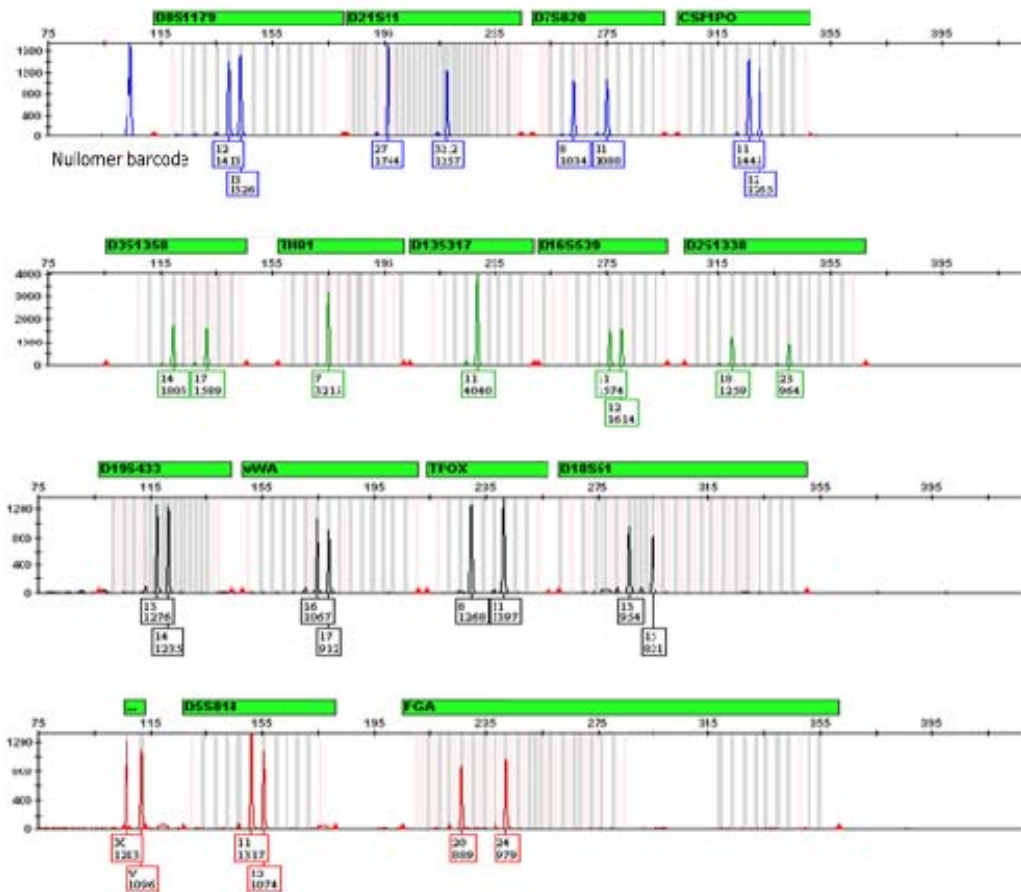
ACT AGT CCG ATG TAA CGC GTA AGA CCG GC



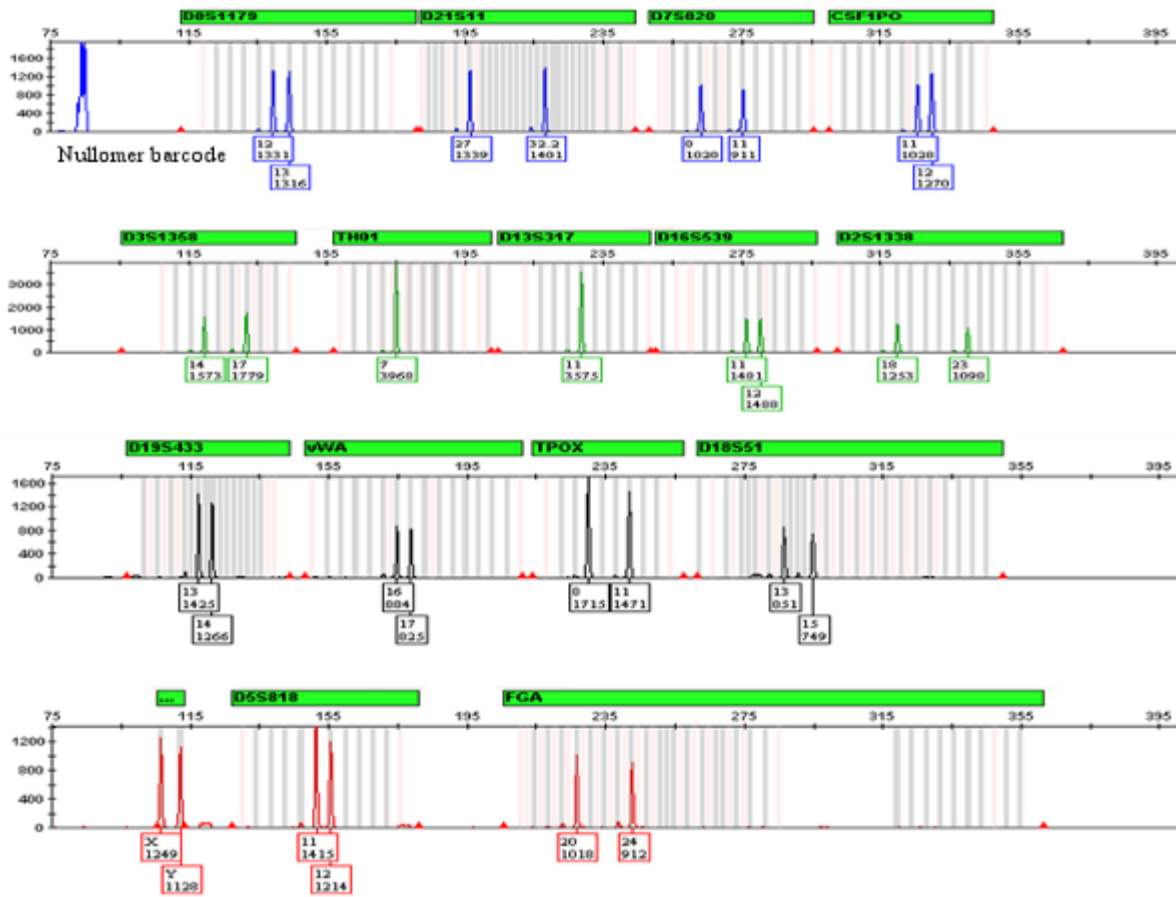
Supplementary Figure 1. STR profile of male genomic DNA amplified with Identifiler™ kit in presence of the 90 bp nullomer barcode. Size of each amplified product is given in bp, the locus is indicated by labels above the peaks. Labels below peaks indicate the number of repeats (allele) and the relative fluorescence units (RFU) value. The y-axis is in RFUs, and is scaled according to maximum peak height.



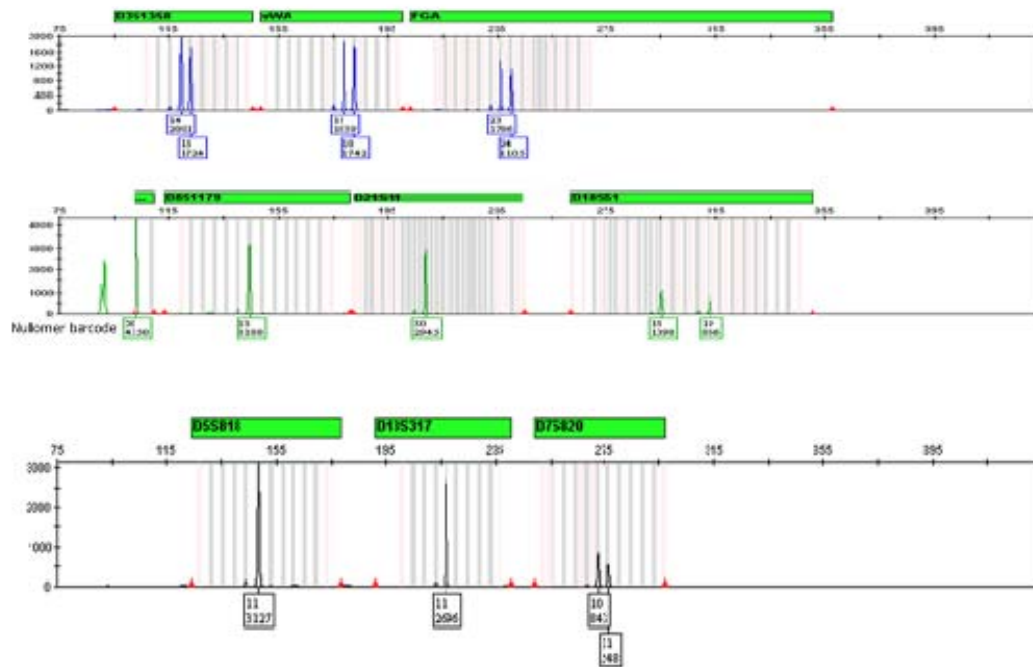
Supplementary Figure 2. STR Profile of male genomic DNA amplified with the Identifiler™ kit in absence of barcode. Size of each amplified product is given in bp, the locus is indicated by labels above the peaks. Labels below peaks indicate the number of repeats (allele) and the relative fluorescence units (RFU) value. The y-axis is in RFUs, and is scaled according to maximum peak height.



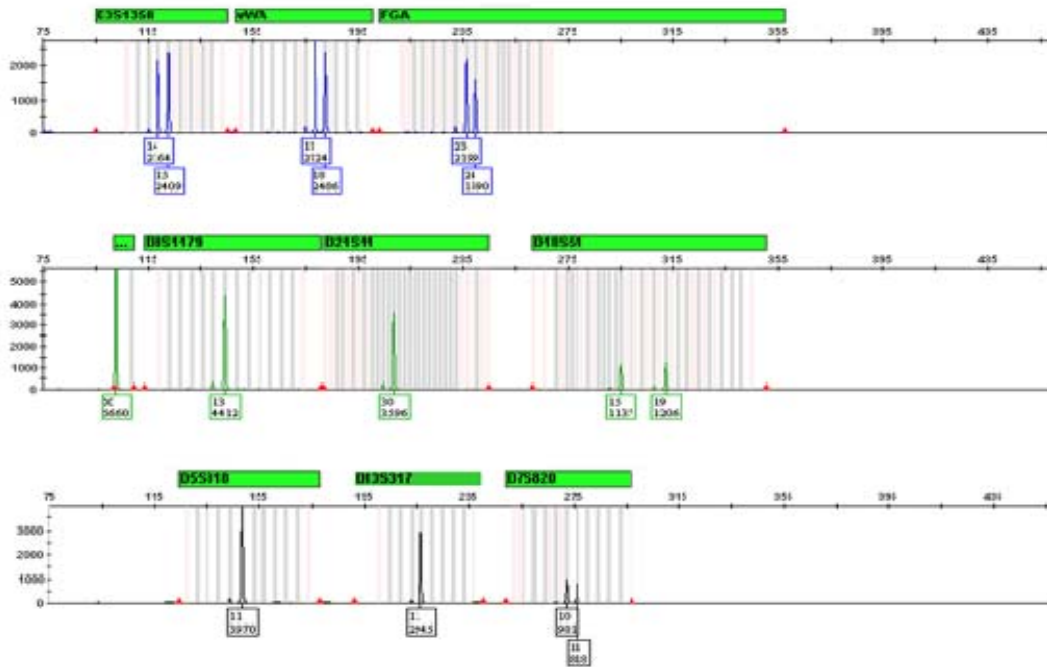
Supplementary Figure 3. STR profile of male genomic DNA amplified with the Identifiler™ kit in presence of the 114 bp nullomer barcode. Size of each amplified product is given in bp, the locus is indicated by labels above the peaks. Labels below peaks indicate the number of repeats (allele) and the relative fluorescence units (RFU) value. The y-axis is in RFUs, and is scaled according to maximum peak height.



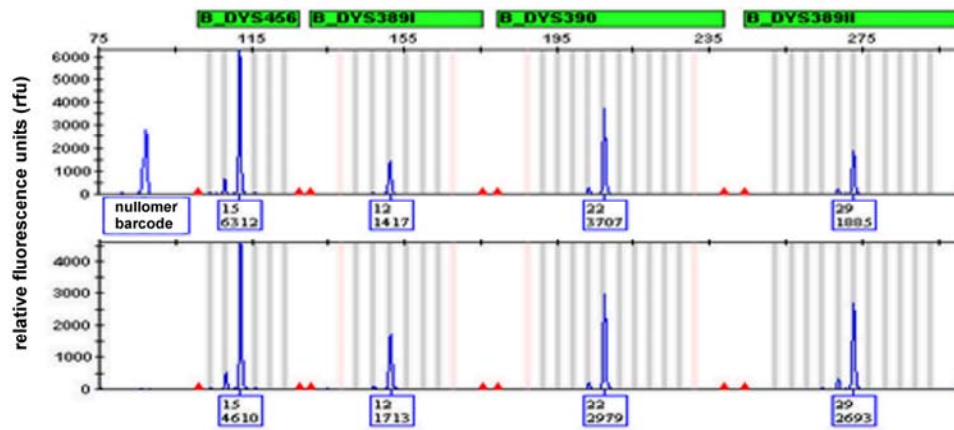
Supplementary Figure 4. STR profile of male genomic DNA amplified with Identifier™ kit along with an 88 bp nullomer barcode. Size of each amplified product is given in bp, the locus is indicated by labels above the peaks. Labels below peaks indicate the number of repeats (allele) and the relative fluorescence units (RFU) value. The y-axis is in RFUs, and is scaled according to maximum peak height.



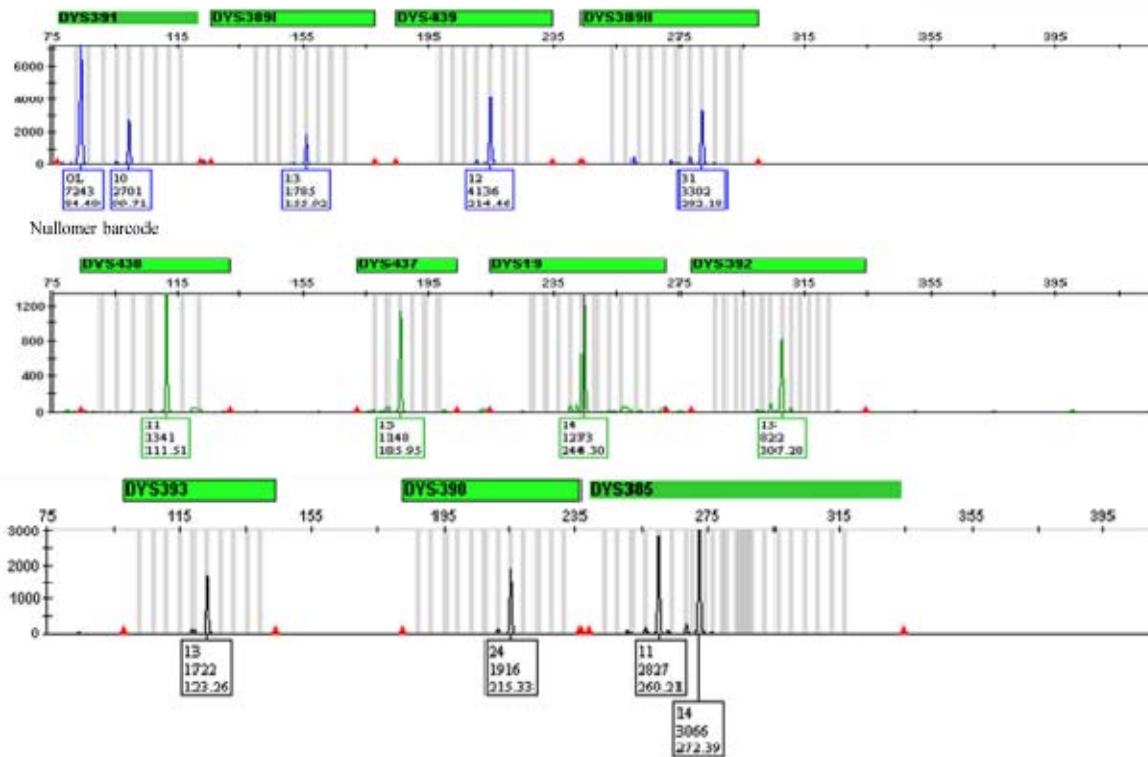
Supplementary Figure 5. STR Profile of female genomic DNA amplified with the Profiler Plus™ kit in the presence of 90 bp nullomer barcode. Size of each amplified product is given in bp, the locus is indicated by labels above the peaks. Labels below peaks indicate the number of repeats (allele) and the relative fluorescence units (RFU) value. The y-axis is in RFUs, and is scaled according to maximum peak height.



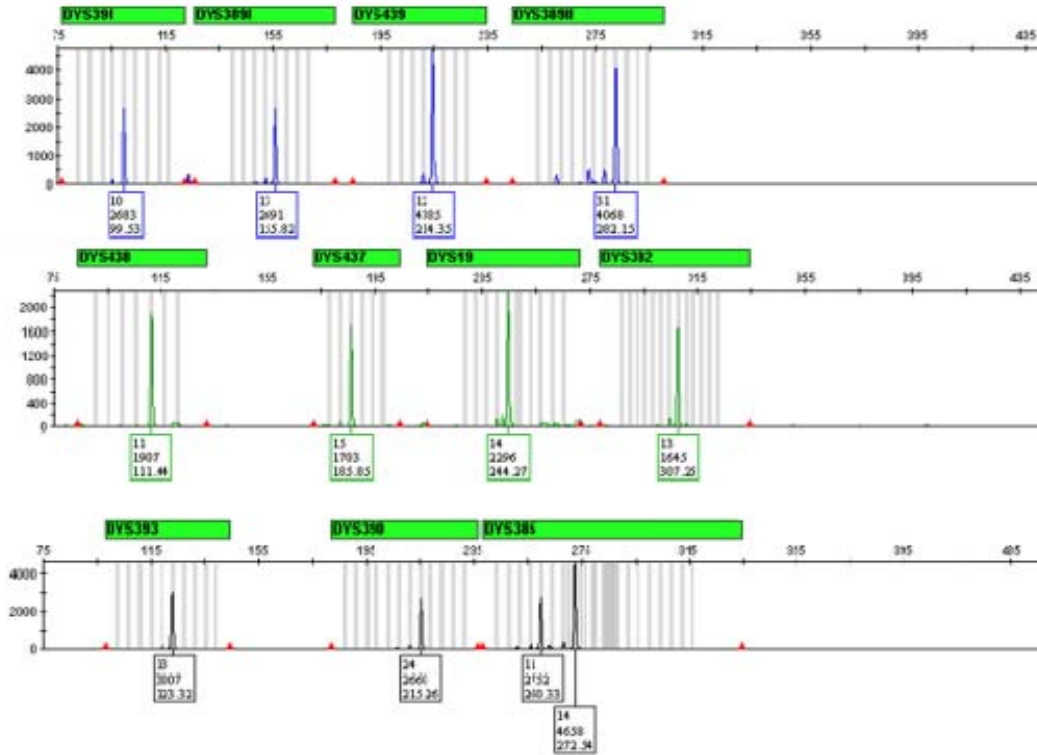
Supplementary Figure 6. STR Profile of female genomic DNA amplified with the Profiler Plus™ kit in absence of nullomer barcode. Size of each amplified product is given in bp, the locus is indicated by labels above the peaks. Labels below peaks indicate the number of repeats (allele) and the relative fluorescence units (RFU) value. The y-axis is in RFUs, and is scaled according to maximum peak height.



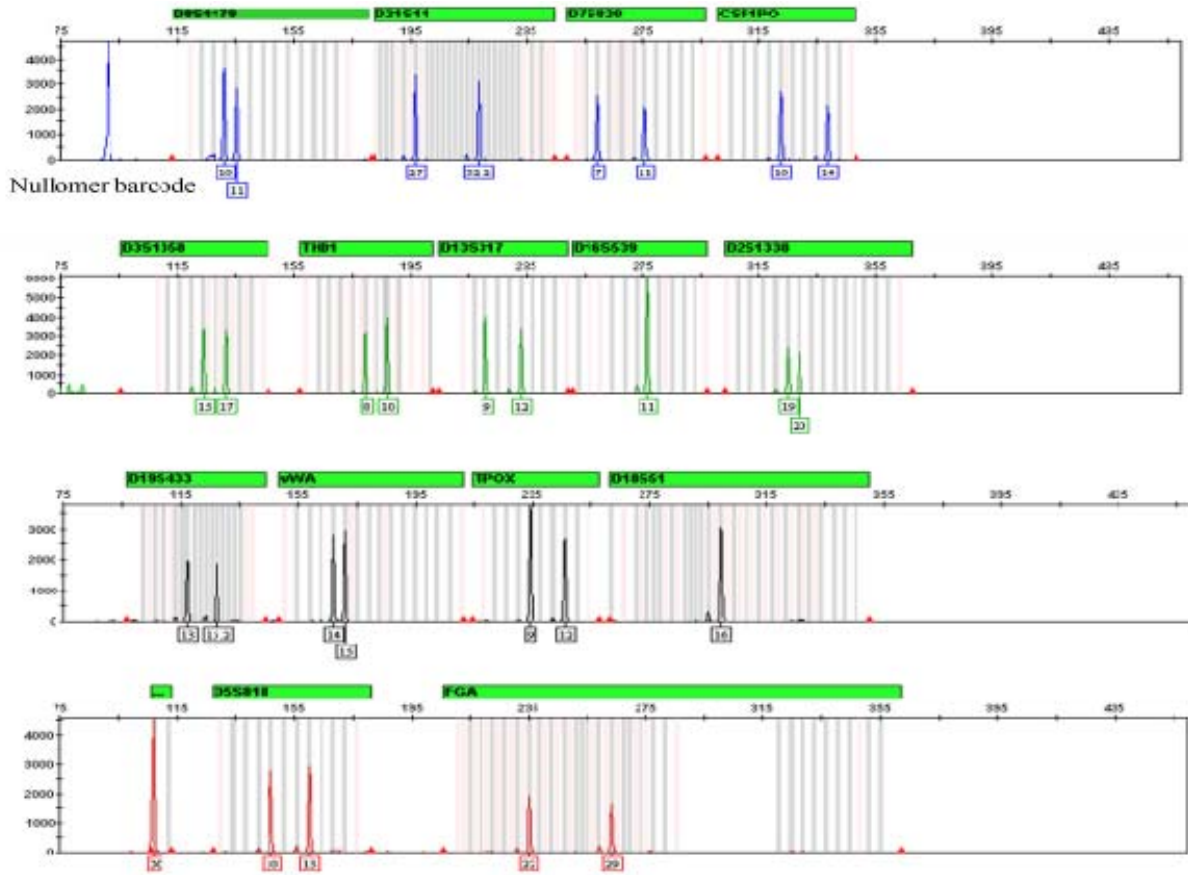
Supplementary Figure 7. Y-STR profile (blue channel) of male genomic DNA amplified with Yfiler™ kit in presence (above) and absence (below) of nullomer barcode (1.9×10^3 copies). Size of each amplified product is given in bp, the locus is indicated by labels above the peaks. Labels below peaks indicate the number of repeats (allele) and the relative fluorescence units (RFU) value. The y-axis is in RFUs, and is scaled according to maximum peak height.



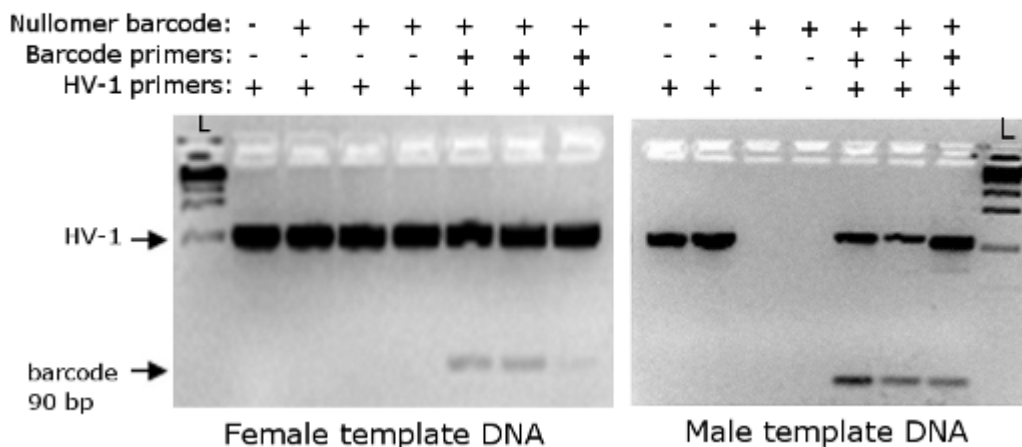
Supplementary Figure 8. Y-STR profile of male genomic DNA amplified with PowerPlex™ Y kit in presence of nullomer barcode. Barcode is recognized as off-ladder (OL) peak by the forensic analysis software (GeneMapper ID). Size of each amplified product is given in bp, the locus is indicated by labels above the peaks. Labels below peaks indicate the number of repeats (allele) and the relative fluorescence units (RFU) value. The y-axis is in RFUs, and is scaled according to maximum peak height.



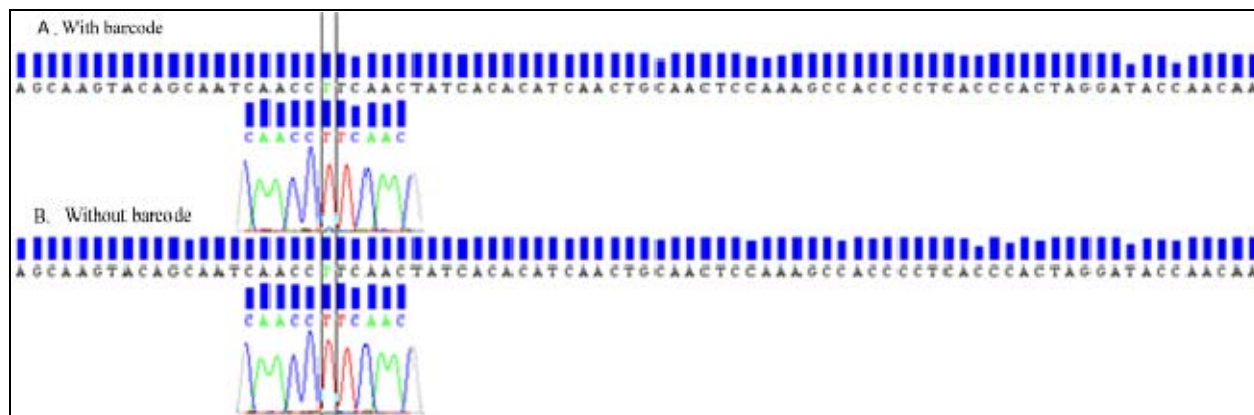
Supplementary Figure 9. Male DNA amplified with PowerPlex™ Y kit in presence of nullomer barcode. No nullomer barcode primers present in this reaction. The y-axis scale is in relative fluorescence units (RFU). The label for each peak indicates the number of repeats (allele) and the RFU value.



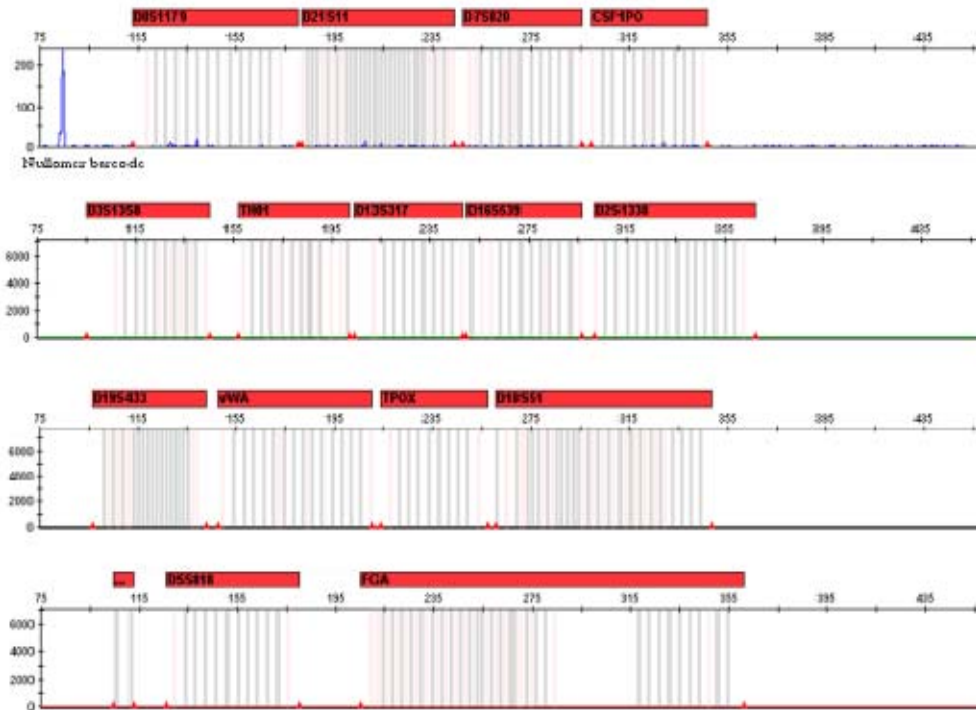
Supplementary Figure 10. DNA extracted from nullomer barcode-impregnated FTA paper, amplified with the Identifiler™ kit. Five punches of FTA paper were used for extraction. The y-axis indicates the corresponding RFU value for each peak.



Supplementary Figure 11. Human mitochondrial DNA amplified with nullomer barcode. Mitochondrial PCR product was visualized on 3% agarose gel. Presence or absence of target (HV-1) and barcode amplification products are indicated for the different primer combinations used in each reaction (bands at about 500 and 100 bp, respectively).



Supplementary Figure 12. Sequence of HV-1 region from female volunteer, from DNA sample tagged with nullomer barcode. 1,900 copies Nullomer barcode added to initial PCR reaction, and another 1,900 copies added to sequencing reactions (Big Dye 3.1 kit, ABI). No nullomer primers were added. The sequences obtained are identical; with nearly identical QV scores for each read (blue bars). Electropherogram detail is shown around the position 16223, where a T is common in the individual's mitochondrial haplogroup (in the Cambridge reference sequence, 16223 is a C).



Supplementary Figure 13. Mock STR/PCR spill on knife. Contamination of evidence with amplified DNA is detected with nullomer barcode, even when diluted 10^6 -fold. Original PCR product was diluted 1: 10^6 in water, and 1 μ L of the diluted product was added to a newly purchased knife. The knife was swabbed and processed as above. The nullomer barcode amplicon is prominent, although no alleles from the amplified human profile are detected.