Title: CODIS STR template enrichment by affinity bead capture and its application in forensic DNA analysis

Award number: 2009-DN –BX-K181

Authors: Diane J Rowold Ph.D.*, Rachel E Balsam M.S. and Michael C Jablecki Ph.D. *

* These two authors contributed equally to manuscript.

Abstract

Short tandem repeat (STR) profiling is a workhorse in forensic analysis; however investigations involving sub-optimal evidentiary DNA samples are often hampered by incomplete and/or ambiguous Combined DNA Index System (CODIS) STR profiles arising from a number of factors. In such cases, polymerase chain reaction (PCR)-based analysis is challenging, impacted by analytical artifacts such as allele drop out (ADO), locus drop out (LDO), PCR stutter, heterozygote peak height imbalance or ambiguous multi-source peak patterns. These difficulties contribute to the growing backlog of under-processed forensic samples. Though several techniques have been and continue to be explored to address shortcomings in the current model, it is apparent that such methodologies need further improvement before they are sensitive or specific enough for routine application in forensic laboratories. The objective of this project was to adapt and explore a technical application as a potential front-end treatment of a forensic sample, particularly for DNA samples that are not currently amenable to conventional methods, to improve the probability that such samples could be evaluated with currently validated approaches.

A biotinylated oligonucleotide-streptavidin coated magnetic bead capture process was developed that allows for the multiplex capture and PCR amplification of CODIS specific STR loci. Recovered DNA materials from highly degraded DNA samples (DNAse I digestion) or significantly fragmented DNA samples (mechanically sheared) were assessed by a comparative analysis of the 13 established CODIS STR loci (TPOX**,** D3S1358**,** FGA**,** D5S818**,** CSF1PO, D7S820**,** D8S1179**,** TH01**,** vWA, D13S317**,** D16S539**,** D18S51 and D21S11) plus the amelogenin and D2S1338 STR loci. Head-to-head comparisons, at single to multiple loci, were used to assess whether the process impacted fragment length bias, two-contributor proportion analysis and eukaryotic versus prokaryotic specificity. Initial results with a minimally optimized system/process indicate that despite the loss of sizeable quantities (from 30-70%) of the specific sequence (as judged by peak height analysis from captured versus un-captured samples), there are no locus dropouts, minimal allelic dropouts (<1%) and statistically relevant CODIS STR

profiles are generated. Furthermore, it was demonstrated that PCR amplification/CE analysis of captured DNA samples did not introduce additional artifacts that might complicate CODIS STR analysis and was successful at quantitatively extracting eukaryotic STR specific alleles from a 5-fold excess prokaryotic background.

Executive Summary

Short tandem repeat (STR) profiling utilizing multiplex polymerase chain reaction (PCR) technology is the gold standard in forensic analysis. Such analyses have become robust, routine and cost effective when applied to pristine, single source, appropriately handled and processed samples. However forensic investigations involving sub-optimal evidentiary DNA samples are often hampered by incomplete and/or ambiguous Combined DNA Index System (CODIS) STR profiles arising from low template quantity, degradation (from age and exposure), mixed human source composition, and/or presence of PCR inhibitors. In these cases, PCR based analysis is challenging, impacted by analytical artifacts such allele drop out (ADO), locus

drop out (LDO), PCR stutter, heterozygote peak height imbalance or ambiguous multi-source peak patterns, and these difficulties contribute to the growing backlog of under-processed forensic samples.

A multiplex capture process capitalizing on the high affinity of the biotin/streptavidin interaction and magnetic bead separation strategy was adapted to assess the possibility of improving STR PCR amplification of CODIS specific loci for use in sub-optimal sample analysis. Published miniSTR primer sequences were used as biotinylated capture probes to minimize the distance between the targeted sequences and the microsatellite motifs and maximize the probability that the captured fragments would include the repeat region of interest. DNA sources (both human and bacterial) were subjected to DNAse I degradation and mechanical shearing to simulate non-optimal DNA samples and analysis of the process captured materials were compared head-to-head with un-captured controls. Identifiler Plus profiles at sixteen loci (13 CODIS STRs, amelogenin, D2S1338, and D19S433) were generated in replicate (ranging from 6 – 24) for each capture reaction to statistically evaluate the success of the target-specific capture experiments.

Results indicated that all targeted loci are present in the captured samples (using manufacturer defined default peak height thresholds). Fewer labeled extraneous/spurious peaks are evolved in the captured versus the un-captured samples for both digested and sheared template sources. The decrease in these artifacts, which do not appear to be electrical spikes or dye blobs, is due, primarily, to the reduction in the overall signal which causes the residual signal from the spurious peaks to fall below the manufacturer's default thresholds (as determined by signal processing analysis).Thus, this apparent noise reduction cannot be attributed specifically as a beneficial outcome of the method. In the highly degraded sample (DNAse I treated with estimated fragment sizes distributed around a mean of ~150 bp), the mean allelic peak height is reduced by 30%-75% post-capture (varying as a function of amplicon length) for all but the amelogenin locus. In the case of the mechanically sheared DNA template (fragment size distribution around a mean of ~800 bp). The mean peak heights for five of the eight amplicons with mean size >225 bp are higher than at least one of those with shorter mean lengths. In addition, the allelic mean peak heights are higher compared to the post-capture DNAse I digested samples (fourteen versus six of the 15 selected loci exhibit mean peak heights $\geq$ 500 relative fluorescent units or RFUs for sheared and digested post-capture samples, respectively. Similarly, with the highly degraded DNAse I template there were two single ADO events (at different loci in two independent replicates) observed in the captured set

compared to none in the un-captured samples; however no ADOs were present with the sheared genomic DNA in either captured or un-captured controls.

The preservation of human mixtures proportions was evaluated by analysis of various mixture categories (Male: Female - 90:10, 70:30, 50:50, 30:70, 10:90). For each sample, the ratio of the sum of the peak heights for the major contributor alleles to the summed height of all peaks present was independently calculated at each locus with non-overlapping male and female alleles. High correlations are present in these ratios across the two treatment groups (captured r^2 = 0.990 and un-captured r^2 = 0.990) for all ten applicable loci and no significant variation was observed in the standard deviations of these ratios across all ten applicable loci. These results indicate the mixture proportions are highly preserved in the developed Affinity Capture technique.

The eukaryotic specificity of the multiplex Affinity Capture process was evaluated through a series of experiments introducing bacterial contamination to human genomic DNA samples. Only a single human amplicon (amelogenin) was assessed against the bacterial ribosomal 16 S locus. SYBR Green QPCR results demonstrated a mean 94% reduction of prokaryotic DNA in the captured samples at each of four mass ratios examined indicating that the decrease in bacterial component appears to be independent of the original human/bacterial genomic DNA quotient.

Overall, the data demonstrates that the prototypic Affinity Capture does not interfere with the conventional CODIS STR analysis by introducing artifacts. The post-captured profiles (at the 1500 copy number) are characterized by detectable allelic peak heights, the expected heterozygote balance and what appears to be a cleaner signal although as previously noted this may simply be as a result of overall signal reduction. It is possible that the reduction of background signal by Affinity Capture may allow for an increase in the number of PCR cycles in conventional CODIS STR analysis (for LCN or highly degraded evidentiary samples) however experiments designed to test this hypothesis have not be carried out. Additionally, the Affinity Capture process appears to preserve two contributor human mixture ratios and effectively reduce the prokaryotic DNA component of a sample, two critical concerns for both conventional downstream analysis as well as emerging DNA analysis technologies such as Next Generation Sequencing (NGS) mediated CODIS STR mixture de-convolution.

Though initial success has been achieved in the development of a 15 loci multiplex Affinity Capture process, it is apparent that such a technology needs further improvement before it is sensitive or specific enough for routine application in forensic laboratories. In the current iteration the process is not as effective at recovering longer versus shorter template molecules

so does not yet to confer much of an advantage over conventional treatment with respect to high quality, pristine genomic DNA subsequently fragmented by DNAse I or mechanical shearing. However, the case may be different with more problematic DNA frequently encountered at crime scenes. The probe binding efficiency of the longer DNA fragments needs to be increased across all targeted loci before this procedure can be used as a productive front end for conventional CODIS STR profiling as well as other technologies.

The authors envision that the optimized capture technique will be most useful for sub-optimal DNA samples not amenable to conventional preparation and/or analysis methods. Once optimized, this procedure will be able to be packaged in a kit format and can be performed in the pre-amplification area of the local crime laboratory.

Main Manuscript

Introduction

Statement of Problem

Short tandem repeat (STR) profiling utilizing multiplex polymerase chain reaction (PCR) technology is the gold standard in forensic analysis (1). Such analyses have, become robust, routine and cost effective when applied to pristine, single source, appropriately handled and processed samples. However forensic investigations involving sub-optimal evidentiary DNA samples are often hampered by incomplete and/or ambiguous CODIS STR profiles arising from low template quantity, degradation (from age and exposure), mixed human source composition, and/or presence of PCR inhibitors. In these cases, PCR based analysis is challenging, impacted by analytical artifacts such allele drop out (ADO), locus drop out (LDO), PCR stutter, heterozygote peak height imbalance or ambiguous multi-source peak patterns, and these difficulties contribute to the growing backlog of under-processed forensic samples.

Literature Review

Degraded DNA

Cells in postmortem tissue undergo death by apoptotic or necrotic processes that damage and degrade the DNA via catabolic reactions (2). DNA damage and fragmentation also results from exposure to environmental radiation (abasic lesions, thymine dimers) (2, 3, 4) chemical insults (oxidation, hydrolysis, base modification), and bacterial activity, so it is reasonable to assume that degenerative pathways will impact biological deposits as well (2).

The success of a CODIS STR profile depends, not only upon total quantity of human genomic DNA added to the multiplex PCR reaction, but on the relative proportions of intact versus degraded DNA template as well (5). DNA degradation, if sufficiently severe, could be considered a form of "low-template" DNA since the amount of intact template in a fixed mass of sample nucleic acid available for PCR amplification is likely substantially less than that of a pristine sample.

One validated approach to CODIS STR analysis for interrogating degraded samples involves the use of primers that lie closer to the microsatellite repeat region (referred to as miniSTR primers), thereby shortening the amplicons by 25 to 198 bases (6) and increasing the probability of amplifying an intact STR locus. The allelic bin sizes of these loci remain unchanged which enables direct comparison of allelic length variants regardless of amplicon size. Additionally, the existing CODIS STR databases established with conventional primer sequences may still be referenced. However, the commercial miniSTR kit most widely used in the United States, AmpFlSTR® MiniFiler™[i] yields an incomplete CODIS profile including only eight of the established 13 CODIS STR loci (CSF1PO, D13S317, D16S539, D18S51, D21S11, D2S1338, D7S820 and FGA) thus requiring at least two different multiplex reactions to generate a full CODIS STR complement. Furthermore, the discriminatory power of the additional loci included in the commercial 15 STR multiplex kits remains untapped (i.e. D8S1179 and D3S13358 in AmpFlSTR Identifiler®[i], or Penta D and Penta E in PowerPlex®[ii] 16). Another emerging miniSTR PCR approach to analyzing degraded DNA utilizes a 26 loci set, including non-CODIS STRs, in which the amplicons are in the range of 100-300 bp (7).

Mixed Sample/Multiple-Genomic DNA Mixtures

Some progress has been made in the de-convolution of CODIS STR mixtures, primarily through the development of expert systems (8, 9, 10, 11). Previous efforts have focused on the quantitative interpretation of resulting electropherograms, statistical ranking of possible genotypes, and implementation of automated database searches to generate a list of potential matches. Many of the successful mixture de-convolultion cases using these strategies are limited to two person combinations in the absence of confounding artifacts (8,11). Lucas and co-workers (11) reported partial resolution of three person mixtures, but only after the manual pre-

[i] TaqMan ,GeneMapper, LIZ, GeneAmp , ABI PRISM, AmpliTaq Gold, Identifiler, Identifiler Plus, AmpFlSTR and are registered trademarks and POP-6 and MiniFiler are trademarks of Applied Biosystems by Life Technologies, Foster City, CA.
[ii] PowerPlex is a registered trademark of Promega Corporation, Madison, WI.

selection of peaks and subtraction of the reference sample. Even so, many technical challenges remain before a statistically robust de-convolution of human DNA mixtures becomes routine (8).

Effective resolution of CODIS STR mixtures is often hindered by artifacts generated by (or during) the PCR amplification step as well as the subsequent amplicon separation and detection during capillary electrophoresis (CE) (12,13). Some PCR artifacts, including non-proportional amplification of contributors across loci, preferential amplifications of the shorter alleles (13, 14) and stutter (12, 15), may foil the effective use of relative peak heights and areas to assign alleles to a particular contributor per locus as well as the subsequent linkage of genotypes across multiple loci to generate a CODIS STR profile. In addition, CE based fragment analysis cannot distinguish between co-migrating alleles that vary in sequence (16, 17). One approach to mixture de-convolution is to avoid the limitations of non-clonal multiplex PCR with the use of a technology that employs a clonal PCR amplification of the mixed sample such as Next Generation Sequencing (NGS). Although, such technologies are not yet ready for routine application to the large backlog of samples awaiting forensic analysis (18), they may be useful in the future as a method of resolving individual CODIS STR profiles from mixed human samples. In order to employ these technologies effectively for CODIS STR mixture de-convolution, the loci of interest must be preferentially selected over the remainder of the genome to minimize the cost, labor and electronic data storage space and possible signal swamping from an abundant prokaryotic DNA component (19).

Targeted DNA template enrichment may be accomplished by one of several front-end selection methods. Affinity bead capture is one of the most effective (see discussion below). Others include less desirable options such as targeted PCR (may increase amplification bias for the reasons discussed above) and cell-based cloning, a messy, time consuming process that involves the culturing of bacterial strains.

Bead-based Affinity Capture

Bead based DNA capture is a well-established technology that has been utilized in several arenas to address limited specific sequence copy number analysis in threat detection and diagnostic systems (20) as well as a front-end for emerging DNA technologies such as NGS (21). Affinity capture employing bead technology can be either non-specific or targeted. Non-specific technologies such as Charge Switch®[iii] Nucleic Acid Purification isolate DNA (and/or RNA) regardless of sequence (22). Targeted or sequence-specific capture techniques

[iii] Charge Swtich and Dynabeads are registered trademarks of Invitrogen by Life technologies, Carlsbad CA

usually involve the use of a reverse-complementary strand of DNA functionalized to attach to a bead. One example and the approach employed in this study is the hybridization of the biotinylated capture oligomer to the target DNA followed by capture of the entire complex on streptavidin-coated beads (23). An alternate method first coats the beads with oligomers and then captures the target DNA onto the bead (20).

Affinity bead capture can be performed on a solid substrate or in solution (24). Previous research (25) demonstrates that the later strategy, displays a higher specific binding capability and thus, requires far less DNA template (a 40 fold reduction) (25). The more effective template enrichment of the solution versus solid phase affinity capture is believed to be due, in part, to this system's ability to benefit from higher concentrations of probe which are free to search the entire reaction volume for targets as opposed to being tethered on a chip (25). Another scientific group (26) reported that solution phase capture is much more efficient than mixed phase (solid and solution) and determined that the lower limit of recovery with the former is 100 copies of target DNA molecules at zeptamolar concentrations.

Project Objective

The primary objective of this study was to adapt and develop a targeted Affinity Capture protocol specific for the STR regions (plus the amelogenin locus) of human DNA samples utilized in CODIS analysis and to evaluate this approach as a tool to facilitate CODIS STR analysis of two types of sub-optimal evidentiary samples, degraded DNA and mixed source/multi-genomic DNA. The aim of Phase I of this project was to explore the feasibility of a specific affinity capture process (Affinity Capture) and demonstrate that, as a front-end treatment of a forensic sample, the process does not introduce additional artifacts that would further complicate downstream analysis such as conventional multiplex PCR with CE or NGS.

Materials and Methods

Source DNA

Three sources of DNA were used for the studies reported here: two human DNA samples, one male and one female, and a non-pathogenic bacterial genome (*Staphylococcus epidermidis* or SE), a common inhabitant of the human oral cavity. The two human genomic DNA samples were purchased from the Coriell Cell Repository, Camden, NJ, and their respective STR profiles are depicted in Figure 1. The prokaryotic DNA sample was purchased from the American Type Culture Collection (Manassas, VA). Simulation of degraded DNA was accomplished through two methods: a limited DNAse I digestion of human specific DNA

(estimated mean length of 150 bp) (5) (Figure 2) and mechanical shearing (Covaris Inc. of Woburn, MA) of the human and prokaryotic DNA (3 distributions of mean length, 300, 800 and/or 500 bp (Figure 3)).

Oligonucleotides

All sequence specific primer/probe oligonucleotides used in this study (with the exception of commercial multiplex kits) were custom ordered from Integrated DNA Technologies (Coralville, IA). The sequence and structures are shown in Table1.

Capture Probes

MiniSTR primer sequences were used to minimize the distance between the targeted sequences and the microsatellite motifs and maximize the probability that the captured fragments would include the repeat region of interest. The specific oligonucleotide sequences used to capture the relevant fragments from the source material were based on 28 miniSTR primer sequences (6) encompassing 14 STR loci (TPOX, D2S1338, D3S1358, FGA, D5S818, CSF1PO, D7S820, D8S1179, TH01, vWA, D13S317, D16S539, D18S51 and D21S11). The original primer pair for the amelogenin locus was also included (27). Each capture probe was derivatized at the five prime (5') terminus with a biotin moiety for streptavidin bead capture and incorporated a 15 base poly thymidine spacer between the biotin-streptavidin complex and the three prime (3') end of the bound DNA template fragment.

SYBR Green QPCR Assay Primers

The same oligonucleotide sequences, minus the 5' biotin label and poly-thymidine linker, were employed in the SYBR Green QPCR assay. The primer pair used for analysis of bacterial contamination amplifies a section of a bacterial 16S ribosomal gene (28) and exhibits 100% sequence complimentary to the corresponding regions in the *Staphylococcus epidermidis* (SE) DNA.

Affinity Bead Capture

The Affinity Capture process followed the general protocol outlined in Figure 4 based on the DynaBeads® M-280[iii] manufacturer's recommended protocol with several modifications to accommodate a primer extension (PE) process delineated in previous research (MCJ,

unpublished results) and ABI AmpliTaq Gold®[i] guidelines. Briefly, the poly T biotinylated specific probe(s) were added to the fragmented genomic DNA sample, denatured for 5 minutes (Figure 4, Step 3), hybridized and primer extended to improve the duplex stability and improve the efficiency of the capture process (Figure 4, Step 4). The biotin containing duplexes were then captured via streptavidin coated magnetic Dynabeads (Figure 4, Step 5) and washed to remove the background non-specific DNA (Figure 4, Step 6). The probe-captured fragments were released by heat denaturation at 90^{o}C in a small volume of TE buffer (Figure 4, Steps 7 and 8).

QPCR Quantification

The SYBR Green QPCR assay was executed according to vendor specifications on the DNA Engine Opticon® 2[iv] (annealing temperature = 55^{0}C) to estimate copy number of targeted loci. The prokaryotic-specific QPCR amplification was performed under the same conditions with the exception of the final primer concentration (250 µM).

CODIS STR Profiling

Identifiler Plus® profiles at 16 loci (13 CODIS STRs, amelogenin, D2S1338, and D19S433) were generated in replicate (6 to 12 replicates) for each capture reaction to statistically evaluate the success of the target-specific capture experiments. These were amplified on a 96-Well GeneAmp®[i] PCR System 9700 as per vendor's instructions for Identifiler Plus. CE of the amplified samples was performed at the San Diego State University DNA Core Facility on an ABI PRISM®[i] 3100 Genetic Analyzer using the POP-6™[i] polymer, 50 centimeter capillary and LIZ®[i] 600 internal size standard. Fragment length determination and allelic assignment was accomplished with GeneMapper®[i] version 4.0.

Statistical Analysis

Statistical analysis of SYBR Green and Identifiler Plus results involved calculation of the mean and some measure of variance (standard deviation, standard error and/or coefficient of variation (CV)) of the replicate sample sets with respect to various QPCR and Identifiler Plus profile metrics. In all cases, captured and un-captured samples were compared to one another to characterize artifacts specifically introduced by the capture process.

[iv] Opticon is a registered trademark of Bio-Rad, Hercules CA

Results and Analysis

Affinity Capture Assay

The Affinity Capture process was developed as a 15 loci multiplex. Initial experiments designed to optimize experimental parameters in the Affinity Capture protocol involved the titration of both degraded human DNA template and primer amount in the capture reaction and post-captured template quantity required by the Identifiler Plus PCR. Preliminary results indicated that with the level of degradation exhibited by the DNAse I treated sample (mean size 150 bp as seen in Figure1) (Table 2), at the 150 template copy Identifiler Plus input level (equivalent to 1 ng genomic DNA) both the captured and un-captured samples contained a high number of LDOs events (81 and 31, respectively for each of 12 replicate samples). LDO artifacts increased two fold and six fold in captured and un-captured samples, respectively when Identifiler Plus DNA input is reduced from 150 to 15 template copies (179 and 172 LDO events for, captured and un-captured sets, respectively). The outcome of probe titration experiments suggests that differences in percent target recovery among 100, 75 and 25 nanomolar (nm) probe for the four loci (amelogenin, CSF1PO, FGA and D7S820) are comparable (Table 3) but that with the 12.5 nm probe is reduced (capture yields for the four loci range from 42% to 66% of the means generated from the percent target recoveries at the three higher probe levels). Thus, the 25 nm probe concentration was selected for input into the Affinity Capture process. Additionally, it can be seen by the Identifiler Plus profiles and associated table in Figure 5 that primer extension improves the overall capture process across loci (1.1 to 2.2 fold increase in locus – specific mean peak heights) and therefore primer extension was utilized as an obligatory step in all subsequent experimentation.

The effectiveness of the capture/amplification process was evaluated via a series of side-by-side analyses comparing the Identifiler Plus performance of human genomic DNA samples subjected to capture versus their un-captured counterparts. Representative Identifiler Plus electropherograms generated from captured and un-captured digested human DNA based on approximately 1500 template copies (equivalent to 10 ng of genomic DNA) per Identifiler Plus reaction are depicted in Figure 6. Results indicated that all 15 targeted loci are present in the captured samples (using the default Applied Biosystems peak height threshold of 50 relative fluorescent units or RFUs). The electropherograms in Figure 6 and the data in Table 7 indicate that, in addition to all 15 target loci being present, fewer labeled extraneous peaks are reported in the captured samples versus the un-captured samples for both digested and sheared template using automated software on existing analysis systems. This difference is most likely simply due to the overall reduction in signal intensity in the processed samples (see following

Peak Height analysis), causing the signal from the spurious peaks to fall below the manufacturer's default thresholds. The extraneous peak data listed in Table 7 are designated by Identifiler Plus loci and labeled by corresponding allelic sizes if the identified length matches a locus compatible peak on the allelic ladder, or OL if the peak length does not exactly match a known ladder marker. The observation that many of the extraneous peaks migrate at or very close to known allelic lengths strongly suggests that the source of the "erroneous" or "spurious peaks" is likely not related to the DNA source or the Affinity Capture process. These artifacts also do not appear to represent electrical spikes or dye blobs. Preliminary signal processing analysis suggests other sources of origin not covered in the scope of this project. Further examination of the captured versus un-captured DNAse I digested samples reveals that the mean allelic peak height is reduced post-capture across all loci with the exception of the amelogenin locus (Figure 7, Panel A Table 4). This observation is depicted graphically by plotting the allelic mean peak heights by locus per treatment (Figure 7, Panel Al) and sorting the loci by amplicon length on the horizontal axis. The mean peak height is reduced by 30%-75% (varying as a function of amplicon length) post-capture for all but the amelogenin locus. These results also indicate preferential capture of some loci over others (see results of Fragment Length Bias). Although an overall decay in signal strength from the shortest to the longest amplicons is a typical phenomenon associated with the multiplex PCR of degraded DNA templates and seen in the un-captured set as well, it is much more evident in the captured samples (Figure 6, Figure 7 and Table 4). However, the coefficient of variation (CV) of allelic peak heights is comparable across treatments (0.18 to 0.41 and 0.20 to 0.33 for captured and un-captured, respectively) (Table 4) which indicates Affinity Capture results maintain reproducibility of peak height across experiments as seen in the un-captured samples.

In the case of the sheared DNA template, the capture-specific length effects are less pronounced. The mean peak heights for five of the eight amplicons with mean lengths >225 bp are higher than at least one of those with shorter mean lengths. In addition, the allelic mean peak heights are higher compared to the post-capture DNAse I digested samples (fourteen versus six of the 15 selected loci exhibit mean peak heights $\geq$ 500 RFUs for sheared and digested post-capture samples, respectively) (Figure 7 Panel B versus Panel A). However, they are, in general, still lower than those of the un-captured set (Figure 7, Table 5). Nonetheless, the allelic peak heights CV of the captured sheared set are considerably less than those determined for the captured digested samples and compare favorably to those of un-captured sheared counterparts (0.04 to 0.28 and 0.07 to 0.20, for captured and un-captured samples, respectively) (Table 5). This also indicates comparable reproducibility levels.

Two ADO events were observed in the Affinity Captured set with the DNAse I digested template (2 different loci in 2 separate replicates) (Table 2 and Table 4). None were seen in the un-captured samples and no ADOs resulted from Affinity Capture of the sheared genomic DNA. Heterozygote ratios calculated from mean peak height data of the digested samples were analyzed, omitting ADOs, as a metric to determine whether or not the Affinity Capture process increases allelic imbalance and dropout. The mean heterozygote ratios of both treatment groups are above 70% for all selected loci (Figure 8, panel A), which is typically considered the threshold below which indicates the presence of more than one DNA contributor (29). The mean locus-specific heterozygote ratios for the captured samples are within ± 6% of their un-captured counterparts for 9 of the 15 selected loci (Figure 8, panel B). The remaining six loci display captured/un-captured mean heterozygote ratios of 86% or higher. The data do not indicate a relationship between this metric and Identifiler Plus amplicon length as no length-related differences in mean heterozygote ratio are observed for either treatment set.

An analysis was performed to determine if the multiplex Affinity Capture process could non-specifically recover a locus not targeted by the capture probes and thus, introduce artifacts that would result in a CODIS STR call. Based on a 3-loci version of Affinity Capture (amelogenin, FGA and CSF1PO), it was possible to estimate the quantity of a fourth, non-targeted human loci that was non-specifically carried through the process. The mean post-capture quantity of the untargeted D7S820 (mean QPCR size = 156 bp) is 3% of un-captured which is less than that of the three selected loci (amelogenin: 101%, mean QPCR size = 109 bp; FGA: 35%, mean QPCR size = 167.5 bp and CSF1PO: 89%, mean QPCR size =133 bp) (Table 8). This is an over 90% reduction of the untargeted D7S820 component

Affinity Capture Fragment Length Bias

Hybridization based capture methods may be prone to fragment length dependent biases which can negatively impact the recovery of longer template molecules and affect downstream analysis such as conventional CODIS STR PCR and NGS. Results of an experiment comparing the effects of template fragment length on the percent yield of the affinity capture process using sheared human male DNA of two sizes (mean lengths of 300 bp and 800 bp) and 2 loci, amelogenin and FGA are shown in Table 9. For each locus, mean copy number recovery and coefficient of variation (CV) are calculated with respect to the two template fragment lengths and 800 bp / 300 bp ratios of the mean copy recovery are evaluated. The mean and CV values are based on 24 replicate measurements. The QPCR data gauging the relative capture efficiency of

shorter versus longer sheared fragments (mean size of 300 versus 800 bases, respectively) suggests that the longer template is captured at approximately 79% and 71% of the efficiency of the shorter template for the loci surveyed, amelogenin and FGA, respectively (Table 9). However, the standard error of the mean estimate (standard deviation divided by the square root of the number of measurements,) indicates that the mean captured efficiency ratios of these two loci are within 3% to 6% of each other. Therefore, it is reasonable to conclude that the difference between these locus-dependent capture efficiency ratios is not significant. The outcome of this inquiry suggests that a size-dependent capture asymmetry does exist and that additional work should be performed to improve the Affinity Capture for more efficient recovery of longer DNA fragments.

Affinity Capture Mixture Proportion Preservation

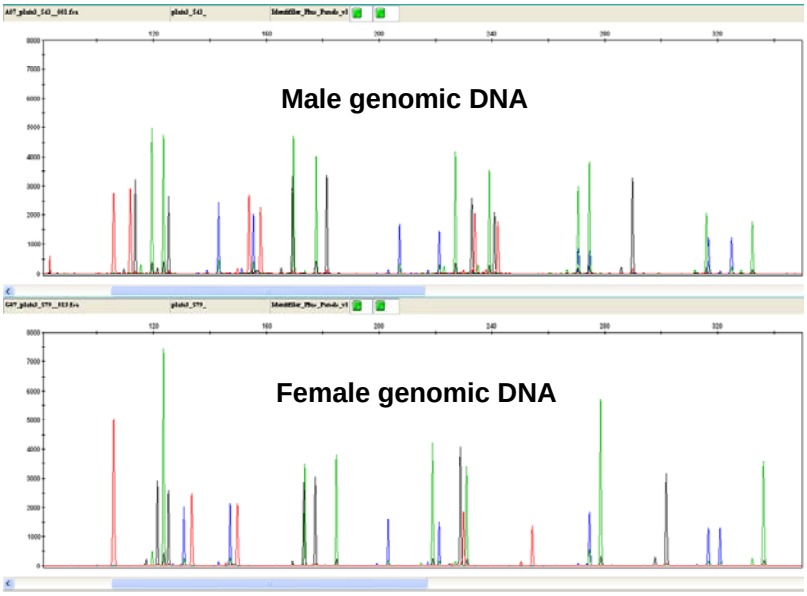
The preservation of human mixtures proportions was evaluated via a series of side-by-side experiments comparing the Identifiler Plus performance on sheared, human genomic DNA samples containing two contributors subjected to capture versus their un-captured counterparts (Figure 9). Initially, five male:female mixture categories (90:10, 70:30, 50:50, 30:70, 10:90) were analyzed. The results were consolidated to three major:minor contributor groups (90:10, 70:30 and 50:50) to increase statistical power (24 versus 12 replicates for each of three categories), based on preliminary analysis indicating no significant difference existing between the mirrored proportions. For each sample, the ratio of the sum of the peak heights for the major contributor alleles (the male is assigned as the major contributor for the 50:50 samples) to the summed height of all peaks present was independently calculated at each locus with non-overlapping male and female alleles. High correlations are present in these ratios across the two treatment groups (captured $r^2 = 0.990$ and un-captured $r^2 = 0.990$) for the three mixture categories and all ten applicable loci (Figure 9, panel A). Additionally, no significant variation was observed in the standard deviations of these ratios across all ten applicable loci for the three mixture categories between the captured and un-captured samples (Figure 9, panel B). Furthermore, there is no discernable locus-specific pattern observed. The 50:50 mixture combination displays a bit more scatter with respect to standard deviation than the two other mixture categories which may be the result of always assigning the male as the major contributor. These results indicate the mixture proportions are preserved by the developed Affinity Capture technique.

Eukaryotic Genome Specificity

The eukaryotic specificity of the multiplex Affinity Capture process was evaluated through a series of experiments introducing bacterial contamination to human genomic DNA samples. Only the amelogenin amplicon was assessed since it is the only locus in the current system at which recovery is nearly complete.

The SYBR Green QPCR results interrogating the bacterial ribosomal 16S locus (28) (Table 10) indicate that there is a mean 94% reduction of prokaryotic DNA in the captured samples. The four mass ratio categories examined displayed similar declines indicating that the decrease in bacterial component appears to be independent of the original human:bacterial genomic DNA quotient. The mean captured to un-captured percentage recovery ratios of human amelogenin template among the four titration groups range from 101.8% to 134.2% (1:5: 102 %, 1:2 : 104%, 1:1: 102%, 1:0: 134%). The recovery rates in excess of 100% reflect the limited resolution of the SYBR Green QPCR assay.

Figures and Tables



Locus	IP Male	SRB Female
AMEL	X,Y	X, X
CSF1PO	10, 12	10, 11
D13S317	11, 14	9, 12
D16S539	10, 11	12, 12
D18S51	14, 14	17, 17
D19S433	12, 15	14, 15
D21S11	30,33.2	29, 33.2
D2S1338	18, 22	23, 23
D3S1358	14, 15	15, 15
D5S818	12, 13	7, 11
D7S820	10, 11	11, 11
D8S1179	13, 16	10, 14
FGA	22, 24	21, 27
TH01	6, 8	7, 9.3
TPOX	9, 11	8, 8
vWA	15, 18	16, 17

Figure 1. Identifiler Plus electrophereograms and profiles of human genomic DNA
Representative un-captured sheared male and female samples with 1500 template copies added to Identifiler Plus PCR. The Identifiler Plus PCR were performed on a GeneAmp 9700 temperature cycler for 29 cycles. Capillary electrophoresis was conducted on a ABI 3100 with POP-6, 50 cm capillary and the results analyzed with GeneMapper 4.0 with LIZ 600 is as the internal size standard.

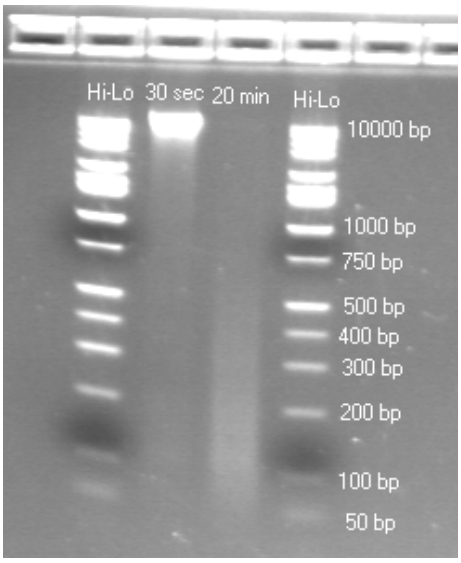


Figure 2. DNAse I digest
Male genomic DNA (Coriell Cell Repository, Camden NJ digested (35 µg DNA / 5 units DNAse). for 30 seconds and 20 minutes by DNase I (Invitrogen by Life Technologies, Carlsbad, Ca, PN 18068-015) to simulate degraded DNA. Approximately 310 ng DNA (as measured by UV absorbance) of each sample is loaded on 2% agarose gel. Mean digested fragment size appears to be 150 bp.

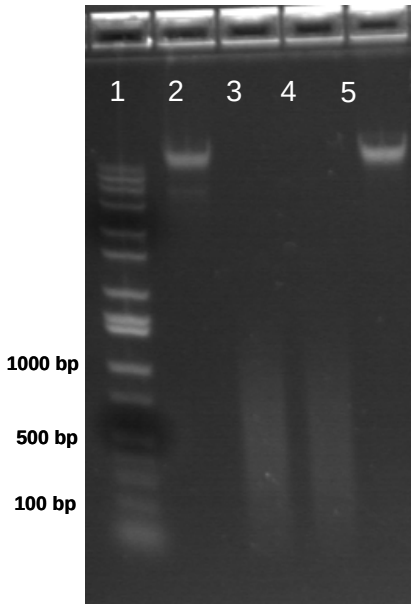


Figure 3. Sheared bacteria and human DNA
Male genomic DNA sample and *Staphylococcus epidermidis* (SE) genomic DNA (American Type Culture Collection, Manassas, VA) mechanically sheared to an estimated mean size of 500 bp by Covaris Inc. (Woburn MA). A 1.2 % agarose gel loaded as follows. Lane 1: Hi-Lo DNA marker (Bionexus Inc, Oakland CA), Lane 2: 300 ng un-sheared SE DNA, Lane 3: 300 ng sheared SE DNA, Lane 4: 300 ng sheared human DNA, Lane 5: 300 ng un-sheared human DNA.

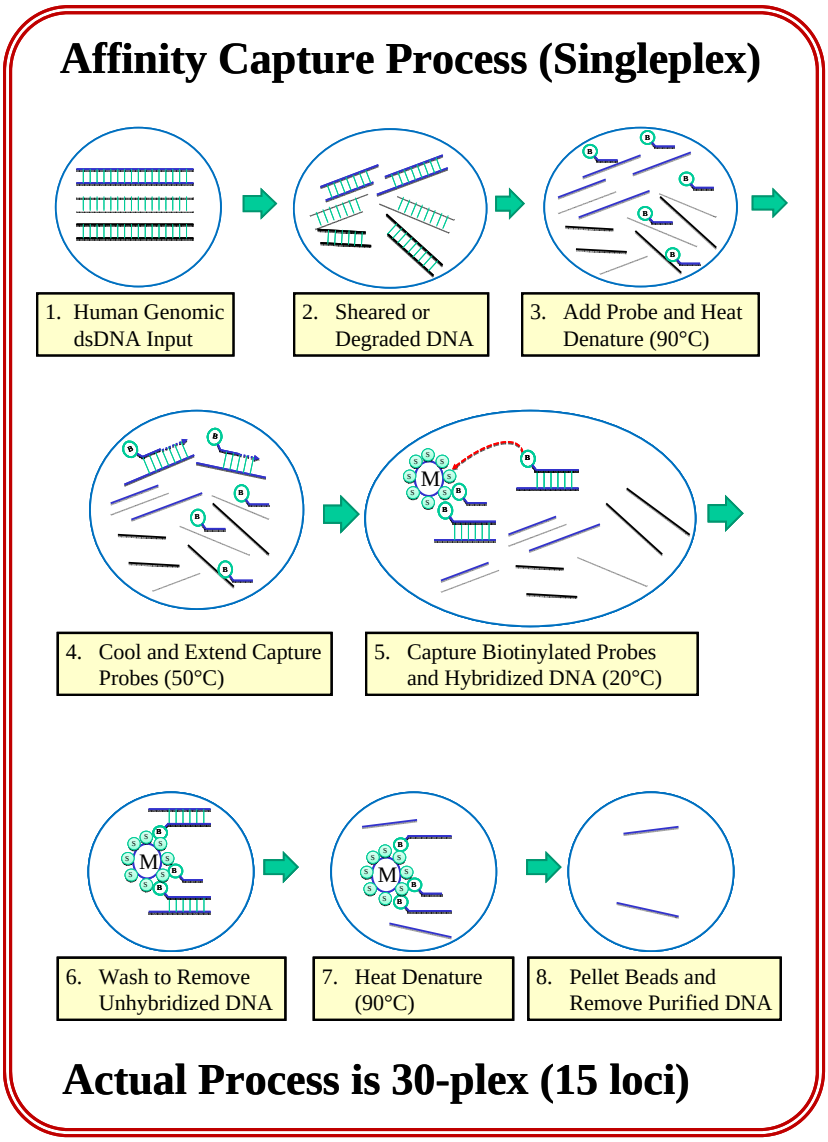
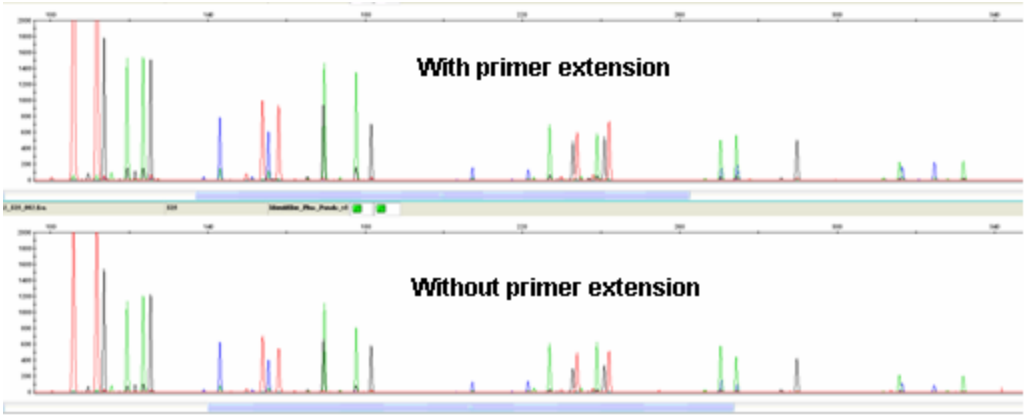


Figure 4. Singleplex capture process
The procedure is as follows: Steps 1 and 2. Double stranded human genomic DNA is either degraded or sheared. Step 3. Biotinylated STR probes are added along with probe extension reagents (AmpliTaq Gold, dNTPs and associated buffer from Applied Biosystems by Life Technologies, Foster City Ca) and heat denatured at 90° (5 min). Step 4. Probe extension occurs at 50°C (10 min.). Step 5. Addition of Dynabeads (Invitrogen by Life Technologies, Carlsbad Ca) to probe/template mixture promotes the capture of the biotinylated probe and hybridized template at 20°C (15 min). Step 6. Bound template is washed at room temperature to remove un-hybridized DNA. Step 7. The probe/template/bead mixture is heat denatured at 90°C (10 min.). Step 8. Beads are pelleted by a nickel plated rare earth cone magnet (SupermagnetMan, Birmingham Al). The singleplex technology was expanded to encompass 15 loci (13 CODIS STR, amelogenin and D2S1338) in a 30-plex reaction.



Locus	Alleles	Mean Peak Height(RFU) With primer extension	Without primer extension	Ratio
AMEL	XY	7904	3640	2.2
CSF1PO	10, 12	342	236	1.5
D13S317	11, 14	1450	1086	1.3
D16S539	10, 11	1112	904	1.2
D18S51	14, 14	639	533	1.2
D19S433	12, 15	3652	2862	1.3
D21S11	30, 33.2	319	279	1.1
D2S1338	18, 22	533	447	1.2
D3S1358	14, 15	2830	2256	1.3
D5S818	12, 13	1888	1192	1.6
D7S820	10, 11	346	326	1.1
D8S1179	13, 16	1671	1029	1.6
FGA	22, 24	1322	1122	1.2
TH01	6, 8	2796	2121	1.3
TPOX	9, 11	920	788	1.2
vWA	15, 18	1579	1248	1.3

Figure 5.Comparison of capture samples with and without primer extension
Multiplex, Affinity Capture process DNAseI treated human male DNA with 1500 template copies added into Identifiler Plus PCR. Mean locus-specific peak heights are based on n=10 replicates for the primer extension and no-primer extension conditions. The peak heights for the two homologous alleles at each locus are combined in each replicate. The ratio is the mean peak height of samples with primer extension to those without primer extension. All Identifiler Plus reactions were performed on a GeneAmp 9700 temperature cycler for 29 cycles. Capillary electrophoresis was conducted on a ABI 3100 with POP-6, 50cm capillary and the results analyzed with GeneMapper 4.0 with LIZ 600 as the internal size standard.

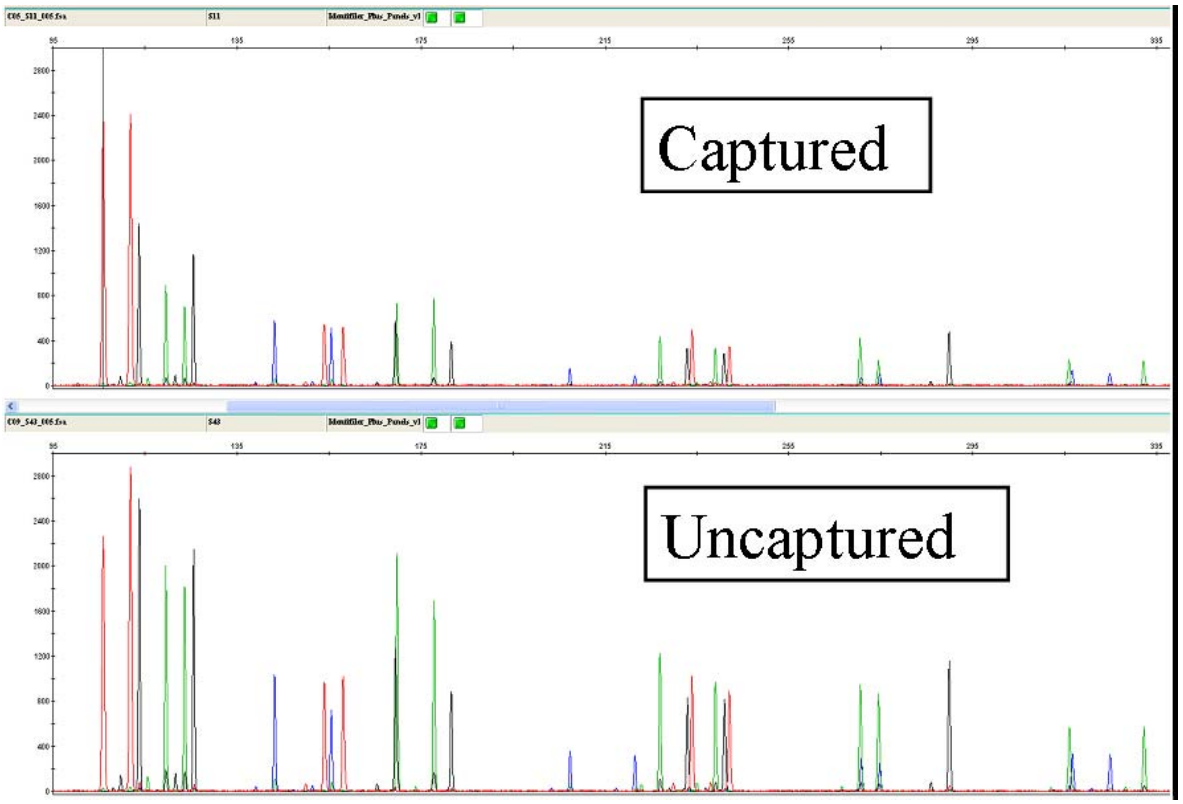


Figure 6. Typical Identifiler Plus electropherograms of captured versus un-captured human male DNA These typical profiles were generated with DNAseI treated DNA at approximately1500 template copies per locus (equivalent to 10 ng of genomic DNA). All Identifiler Plus reactions were performed as previously described (Figure 5).

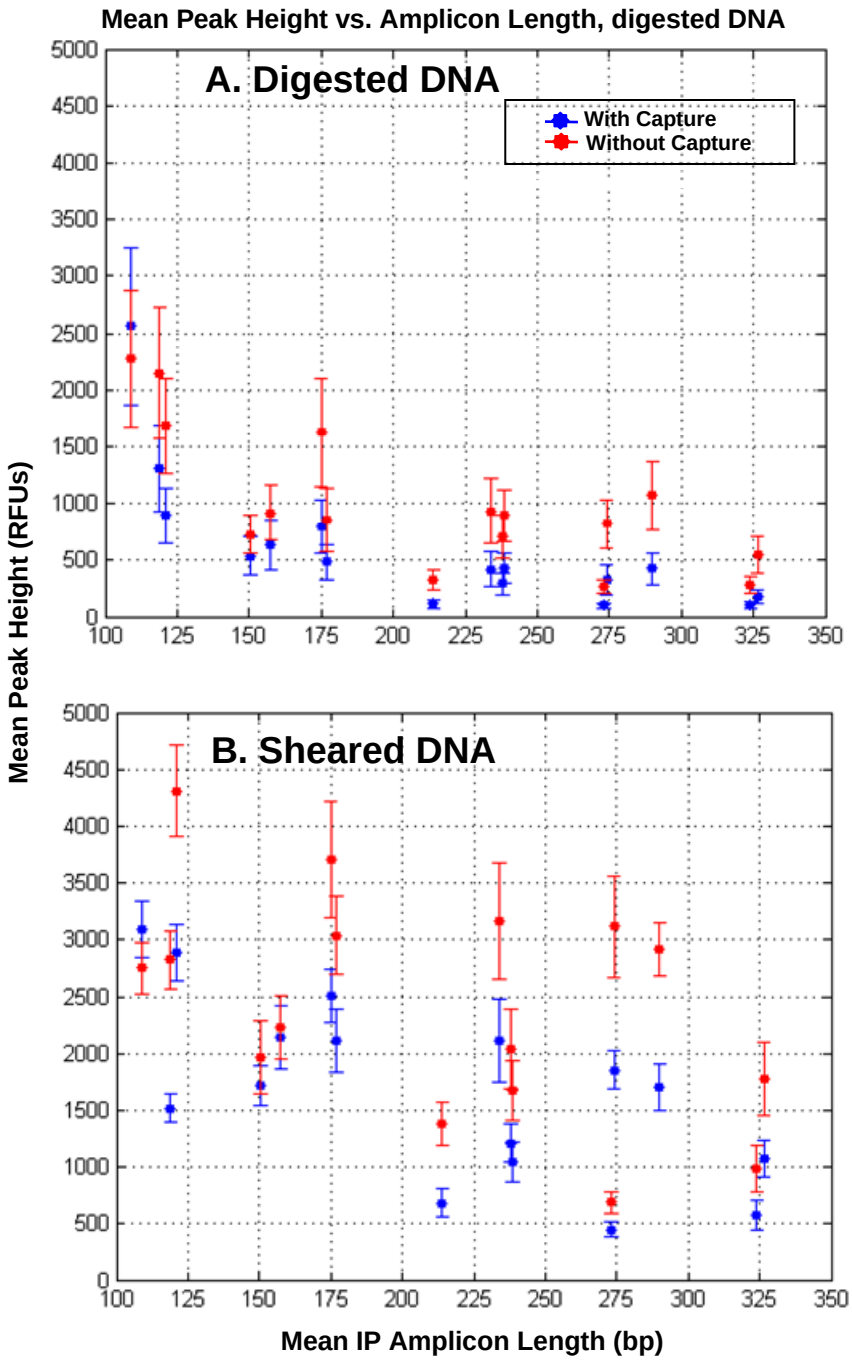


Figure 7. Mean peak height versus mean amplicon length of male genomic DNA
Mean allelic peak heights of captured (blue) and un-captured samples (red) across Identifiler Plus loci for digested (Panel A) and sheared (Panel B) male human DNA. Data are arranged along x axis by mean amplicon length of the 16 Identifiler Plus loci. Each treatment category is represented by n=10 replicates (1500 template copies added to each PCR). The Identifiler Plus reactions were performed as previously described (Figure 5).

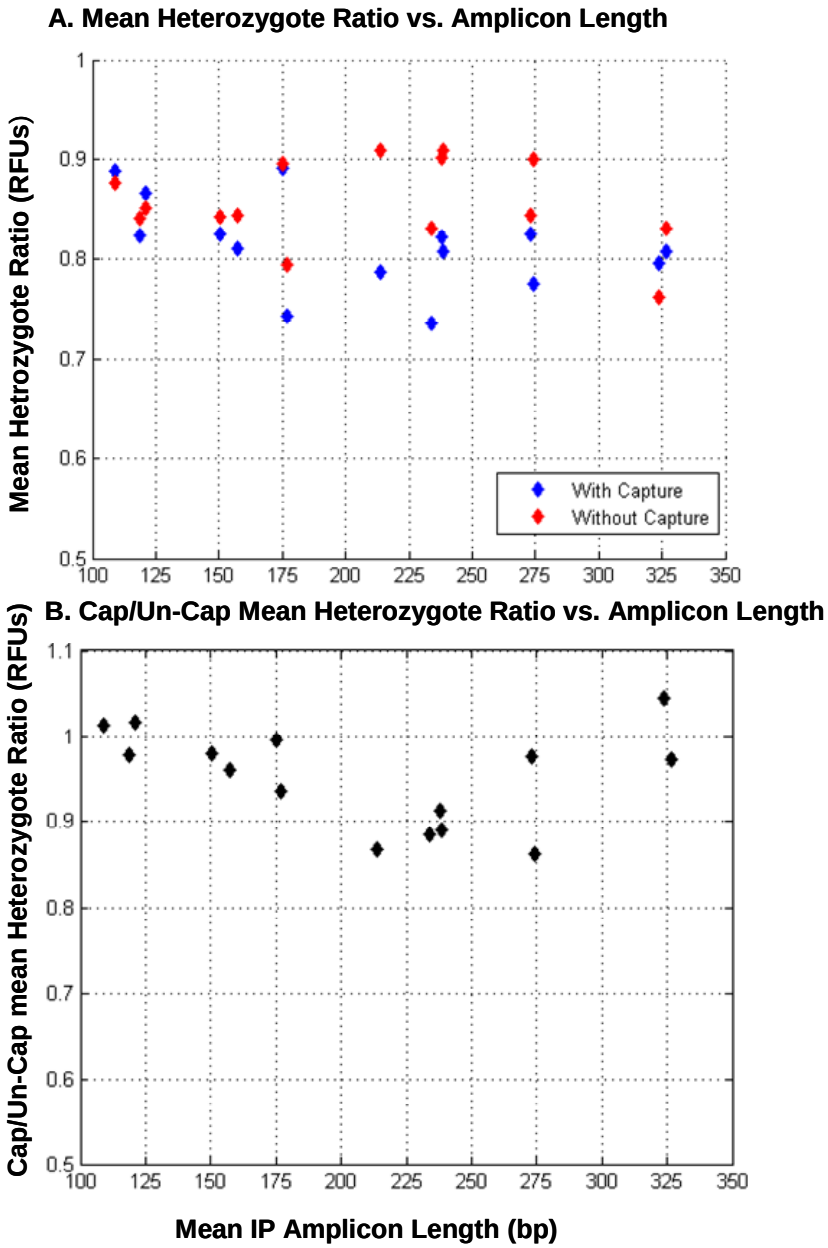


Figure 8. Mean heterozygote ratios versus mean amplicon length of human male genomic DNA digested by DNAse I Panel A: Mean heterozygote ratio of captured (blue) to un-captured (red) samples. Panel B: Ratio of captured/un-captured heterozygote ratios. Each treatment category is represented by n=10 replicates. The Identifiler Plus reactions were performed as previously described (Figure 5).

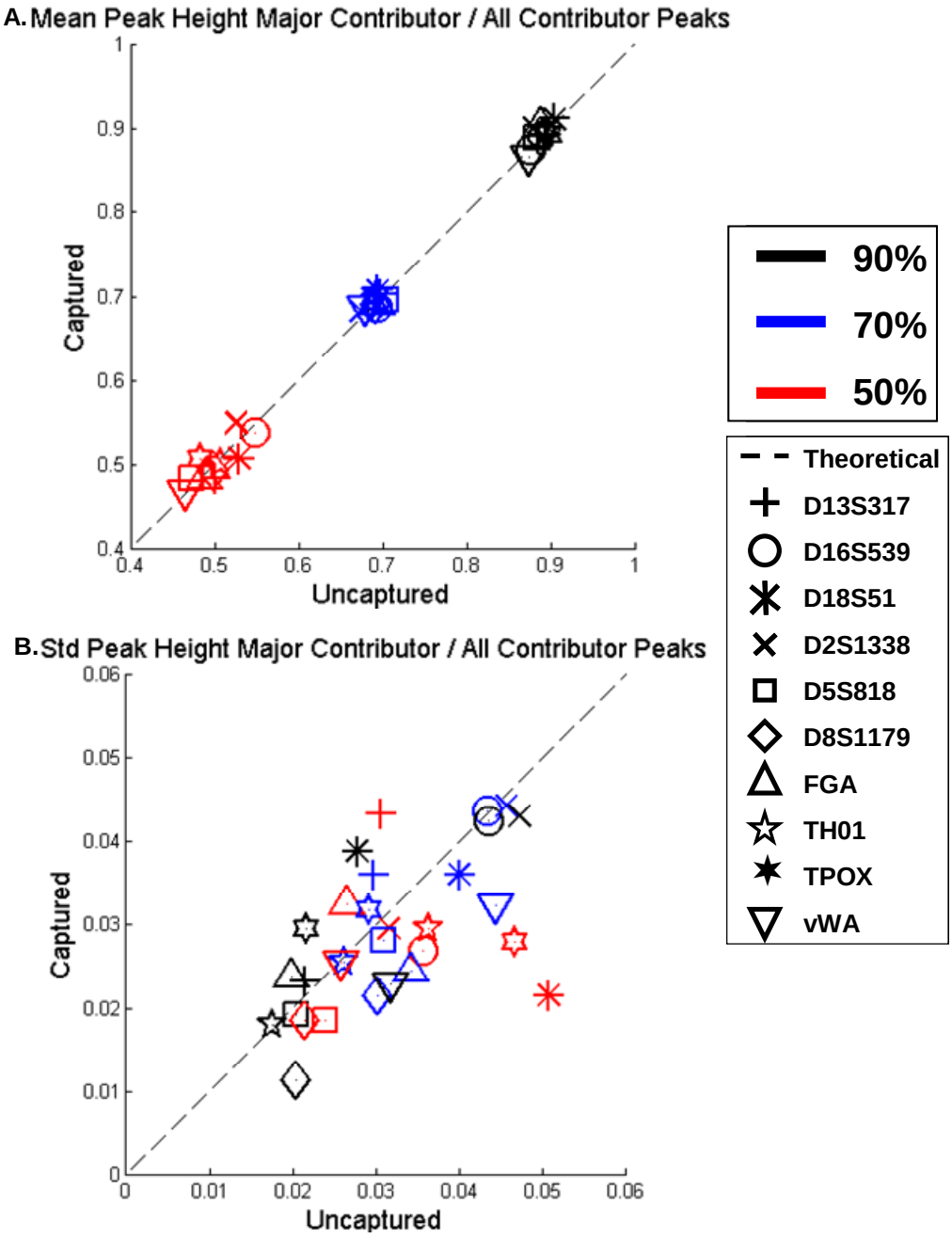


Figure 9. Captured and un-captured mixed human source samples
Mixed sheared (mean size = 800 bp) human DNA samples (male:female) are collapsed into three groups of major contributor/minor contributor mixture proportions: 50:50, 70:30 and 90:10 (1500 template copies added to each Identifiler Plus PCR). Each group is represented by n=24 replicates. Panel A: Comparison of captured and un-captured major contributor/all contributors mean peak height ratios for all 10 non-overlapping loci. Panel B. Comparison of captured and un-captured standard deviation in major contributor/all contributors mean peak height ratios across all 10 loci and 3 mixture categories. The Identifiler Plus reactions were performed as previously described (Figure 5).

Table 1. Capture probes and QPCR primers.

Name/Locus	Sequence	Type	Organism
5'Bio-F-Amelogenin-106/112	/5Biosg/TTTTTTTTTTTTTTTCCCTGGGCTCTGTAAAGAATAGTG	Capture Probe	Human/Primate
5'Bio-R-Amelogenin-106/112	/5Biosg/TTTTTTTTTTTTTTTATCAGAGCTTAAACTGGGAAGCTG	Capture Probe	Human/Primate
5'Bio-F-CSF1PO	/5Biosg/TTTTTTTTTTTTTTTACAGTAACTGCCTTCATAGATAG	Capture Probe	Human/Primate
5'Bio-R-CSF1PO	/5Biosg/TTTTTTTTTTTTTTTGTGTCAGACCCTGTTCTAAGTA	Capture Probe	Human/Primate
5'Bio-F-FGA	/5Biosg/TTTTTTTTTTTTTTTAAATAAAATTAGGCATATTTACAAGC	Capture Probe	Human/Primate
5'Bio-R-FGA	/5Biosg/TTTTTTTTTTTTTTTGCTGAGTGATTTGTCTGTAATTG	Capture Probe	Human/Primate
5'Bio-F-D7S820	/5Biosg/TTTTTTTTTTTTTTTGAACACTTGTCATAGTTTAGAACGAAC	Capture Probe	Human/Primate
5'Bio-R-D7S820	/5Biosg/TTTTTTTTTTTTTTTTCATTGACAGAATTGCACCA	Capture Probe	Human/Primate
5'Bio-F-D13S317	/5Biosg/TTTTTTTTTTTTTTTTCTGACCCATCTAACGCCTA	Capture Probe	Human/Primate
5'Bio-R-D13S317	/5Biosg/TTTTTTTTTTTTTTTTCAGACAGAAAGATAGATAGATGATTGA	Capture Probe	Human/Primate
5'Bio-F-D16S539	/5Biosg/TTTTTTTTTTTTTTTATACAGACAGACAGACAGGTG	Capture Probe	Human/Primate
5'Bio-R-D16S539	/5Biosg/TTTTTTTTTTTTTTTGCATGTATCTATCATCCATCTCT	Capture Probe	Human/Primate
5'Bio-F-D18S51	/5Biosg/TTTTTTTTTTTTTTTTGAGTGACAAATTGAGACCTT	Capture Probe	Human/Primate
5'Bio-R-D18S51	/5Biosg/TTTTTTTTTTTTTTTTGTCTTACAATAACAGTTGCTACTATT	Capture Probe	Human/Primate
5'Bio-F-D21S11	/5Biosg/TTTTTTTTTTTTTTTTATTCCCCAAGTGAATTGC	Capture Probe	Human/Primate
5'Bio-R-D21S11	/5Biosg/TTTTTTTTTTTTTTTTGGTAGATAGACTGGATAGATAGACGA	Capture Probe	Human/Primate
5'Bio-F-D2S1338	/5Biosg/TTTTTTTTTTTTTTTTGGAAACAGAAATGGCTTGG	Capture Probe	Human/Primate
5'Bio-R-D2S1338	/5Biosg/TTTTTTTTTTTTTTTTGATTGCAGGAGGGAAGGAAG	Capture Probe	Human/Primate
5'Bio-F-TH01	/5Biosg/TTTTTTTTTTTTTTTTCCTGTTCCTCCCTTATTTCCC	Capture Probe	Human/Primate
5'Bio-R-TH01	/5Biosg/TTTTTTTTTTTTTTTTGGGAACACAGACTCCATGGTG	Capture Probe	Human/Primate
5'Bio-F-TPOX	/5Biosg/TTTTTTTTTTTTTTTTCTTAGGGAACCCTCACTGAATG	Capture Probe	Human/Primate
5'Bio-R-TPOX	/5Biosg/TTTTTTTTTTTTTTTTGTCCTTGTCAGCGTTTATTTGC	Capture Probe	Human/Primate
5'Bio-F-vWA	/5Biosg/TTTTTTTTTTTTTTTTAATAATCAGTATGTGACTTGGATTGA	Capture Probe	Human/Primate
5'Bio-R-vWA	/5Biosg/TTTTTTTTTTTTTTTTATAGGATGGATGGATAGATGGA	Capture Probe	Human/Primate
5'Bio-F-D3S1358	/5Biosg/TTTTTTTTTTTTTTTTCAGAGCAAGACCCTGTCTCAT	Capture Probe	Human/Primate
5'Bio-R-D3S1358	/5Biosg/TTTTTTTTTTTTTTTTCAACAGAGGCTTGCATGTAT	Capture Probe	Human/Primate
5'Bio-F-D5S818	/5Biosg/TTTTTTTTTTTTTTTGGGTGATTTTCCTCTTTGGT	Capture Probe	Human/Primate
5'Bio-R-D5S818	/5Biosg/TTTTTTTTTTTTTTTAACATTTGTATCTTTATCTGTATCCTTAT	Capture Probe	Human/Primate
5'Bio-F-D8S1179	/5Biosg/TTTTTTTTTTTTTTTTTTGTATTTCATGTGTACATTCGTATC	Capture Probe	Human/Primate
5'Bio-R-D8S1179	/5Biosg/TTTTTTTTTTTTTTTTACCTATCCTGTAGATTATTTTCACTGTG	Capture Probe	Human/Primate
F-Amelogenin-106/112	CCCTGGGCTCTGTAAAGAATAGTG	QPCR primer	Human/Primate
R-Amelogenin-106/112	ATCAGAGCTTAAACTGGGAAGCTG	QPCR primer	Human/Primate
F-CSF1PO	ACAGTAACTGCCTTCATAGATAG	QPCR primer	Human/Primate
R-CSF1PO	GTGTCAGACCCTGTTCTAAGTA	QPCR primer	Human/Primate
F-FGA	AAATAAAATTAGGCATATTTACAAGC	QPCR primer	Human/Primate
R-FGA	GCTGAGTGATTTGTCTGTAATTG	QPCR primer	Human/Primate
F-D7S820	GAACACTTGTCATAGTTTAGAACGAAC	QPCR primer	Human/Primate
R-D7S820	TCATTGACAGAATTGCACCA	QPCR primer	Human/Primate
F-D13S317	TCTGACCCATCTAACGCCTA	QPCR primer	Human/Primate
R-D13S317	CAGACAGAAAGATAGATAGATGATTGA	QPCR primer	Human/Primate
F-D16S539	ATACAGACAGACAGACAGGTG	QPCR primer	Human/Primate
R-D16S539	GCATGTATCTATCATCCATCTCT	QPCR primer	Human/Primate
F-D18S51	TGAGTGACAAATTGAGACCTT	QPCR primer	Human/Primate
R-D18S51	GTCTTACAATAACAGTTGCTACTATT	QPCR primer	Human/Primate
F-D21S11	ATTCCCCAAGTGAATTGC	QPCR primer	Human/Primate
R-D21S11	GGTAGATAGACTGGATAGATAGACGA	QPCR primer	Human/Primate
F-D2S1338	TGGAAACAGAAATGGCTTGG	QPCR primer	Human/Primate
R-D2S1338	GATTGCAGGAGGGAAGGAAG	QPCR primer	Human/Primate
F-TH01	CCTGTTCCTCCCTTATTTCCC	QPCR primer	Human/Primate
R-TH01	GGGAACACAGACTCCATGGTG	QPCR primer	Human/Primate
F-TPOX	CTTAGGGAACCCTCACTGAATG	QPCR primer	Human/Primate
R-TPOX	GTCCTTGTCAGCGTTTATTTGC	QPCR primer	Human/Primate
F-vWA	AATAATCAGTATGTGACTTGGATTGA	QPCR primer	Human/Primate
R-vWA	ATAGGATGGATGGATAGATGGA	QPCR primer	Human/Primate
F-D3S1358	CAGAGCAAGACCCTGTCTCAT	QPCR primer	Human/Primate
R-D3S1358	TCAACAGAGGCTTGCATGTAT	QPCR primer	Human/Primate
F-D5S818	GGGTGATTTTCCTCTTTGGT	QPCR primer	Human/Primate
R-D5S818	AACATTTGTATCTTTATCTGTATCCTTATTTAT	QPCR primer	Human/Primate
F-D8S1179	TTTGTATTTCATGTGTACATTCGTATC	QPCR primer	Human/Primate
R-D8S1179	ACCTATCCTGTAGATTATTTTCACTGTG	QPCR primer	Human/Primate
F-Bacteria Ribosome S	TAGAACACCGATGGCGAAGGC	QPCR primer	Bacteria
R-Bacteria Ribosome S	TCGTGGACTACCAGGGTATCTA	QPCR primer	Bacteria

The human amelogenin sequences are reported in Sullivan et al. 1993. MiniSTR sequences are referenced in Butler et al 2003 and the sequence information for the bacterial 16 S ribosomal region is taken from Ecker et al 2005.

Table 2. Locus drop out (LDO) and allele drop out (ADO) results of Identifiler Plus template titration on captured versus un-captured DNA sources.

Treatment	ID+ [1] input copies	Number of replicates	Estimated template size (bp)	LDO [3]	LDO range [4]	ADO	ADO range [5]
Captured_digested	1500	10	150 [2]	0	0	2	0 to 1
Uncaptured_digested	1500	10	150	0	0	0	0
Captured_digested	150	12	150	81	6 to13	29	1 to 4
Uncaptured_digested	150	12	150	31	0 to 5	30	1 to 5
Captured_digested	15	12	150	179	14 to 15	1	0 to 1
Uncaptured_digested	15	12	150	172	12 to 15	7	0 to 3
Captured_sheared	1500	12	800	0	0	0	0
Uncaptured _sheared	1500	12	800	0	0	0	0

[1]ID+: Identifiler Plus

[2] This mean was estimated from from gel of Figure 2

[3]LDO and ADO events are out of a possible 150 and 180 for 10 and 12 replicates, respectively

[4]LDO range: minumum and maximum number of loci affected per replicate

[5]ADO range: minumum and maximum number of loci affected per replicate

Table 3. Impact of capture probe concentration on SYBR Green QPCR.

Probe Concentration	Mean% Captured Amelogenin	CSFIPO	FGA	D7S820
100nM	53	83	34	57
50nM	53	75	34	61
25nM	54	68	46	63
12.5nM	31	32	25	31

Table 4. Mean peak height analysis (by allele) for captured versus un-captured samples using DNAse I digested human male DNA as the starting material.

		Digested (mean size 150 bp)											
		Captured						UnCaptured					
Locus	Size	N 1	N 2	Mean 1	Mean 2	CV 1[1]	CV 2	N 1	N 2	Mean 1	Mean 2	CV 1	CV 2
AMEL	109	10	10	2397.5	2720.2	0.25	0.28	12	12	2128.6	2420.1	0.27	0.25
D19S433[2]	119	10	10	1435.3	1190	0.27	0.30	12	12	2338.5	1957.8	0.26	0.25
D3S1358	121	10	10	934.9	853.5	0.26	0.30	12	12	1803.6	1566.5	0.22	0.27
D8S1179	150.5	10	10	582.5	491.4	0.32	0.28	12	12	786.75	673.83	0.24	0.20
D5S818	157.5	10	10	693.2	568.8	0.35	0.32	12	12	981.17	857.33	0.25	0.27
TH01	175	10	10	838.4	753.9	0.30	0.29	12	12	1650.3	1598	0.28	0.31
vWA	177	10	10	556.1	412.2	0.28	0.30	12	12	947.25	756.33	0.29	0.33
D21S11	214	9	10	126.44	110.1	0.27	0.44	12	12	329.75	326.42	0.30	0.28
D13S317	234	10	10	485.5	356.8	0.32	0.35	12	12	989.25	877	0.35	0.25
TPOX	238	10	10	277.4	307.8	0.26	0.35	12	12	719.5	705.33	0.29	0.25
FGA	238.5	10	10	463.3	398.4	0.31	0.33	12	12	933.58	853.08	0.26	0.26
D7S820	273	10	9	104.5	101.56	0.18	0.31	12	12	277.75	258.83	0.22	0.25
D16S539	274	10	10	360.2	309.1	0.38	0.41	12	12	822.92	814.92	0.25	0.29
D18S51[3]	290	10	N/A	425.3	N/A	0.33	N/A	12	N/A	1067.9	N/A	0.28	N/A
CSF1PO	324	10	10	120.4	96.8	0.31	0.30	12	12	289.25	280.5	0.27	0.25
D2S1338	326.5	10	10	186.7	173.9	0.30	0.35	12	12	560.67	536.25	0.29	0.29

[1]CV: Coefficient of variation is calculated by the division of the standard deviation by the mean peak height

[2]non-specific recovery of D19S433 is most likely due to the high degree of sequence homology to the miniSTR probes

[3]Sample is homozygote at this locus

Indicates Allelic Drop out

Table 5. Mean peak height analysis (by allele) for captured versus un-captured samples using mechanically sheared human male DNA as the starting source material.

		Sheared (mean size 800bp) Captured						UnCaptured					
Locus	Size	N 1	N 2	Mean 1	Mean 2	CV 1[1]	CV 2	N 1	N 2	Mean 1	Mean 2	CV 1	CV 2
AMEL	109	6	6	2974	3209.3	0.09	0.05	5	5	2779.4	2721.8	0.07	0.10
D19S433[2]	119	6	6	1535.3	1502.2	0.10	0.06	5	5	2871.2	2780.8	0.12	0.06
D3S1358	121	6	6	2916.2	2854.3	0.04	0.12	5	5	4305	4313.2	0.11	0.09
D8S1179	150.5	6	6	1738.3	1697.7	0.12	0.10	5	5	2114	1816.2	0.18	0.10
D5S818	157.5	6	6	2235.7	2051.5	0.15	0.09	5	5	2307.4	2142.8	0.12	0.13
TH01	175	6	6	2544.7	2472.7	0.08	0.11	5	5	3840.8	3579.6	0.18	0.08
vWA	177	6	6	2250.8	1969	0.08	0.15	5	5	3181.2	2889.6	0.07	0.14
D21S11	214	6	6	741.5	629.17	0.16	0.19	5	5	1433.8	1327	0.17	0.09
D13S317	234	6	6	2292	1926.8	0.13	0.18	5	5	3397.8	2936.6	0.15	0.15
TPOX	238	6	6	1237	1181.2	0.14	0.15	5	5	2136.2	1941.2	0.15	0.19
FGA	238.5	6	6	1070	1013	0.16	0.19	5	5	1758.2	1584	0.13	0.18
D7S820	273	6	6	469	432.17	0.18	0.12	5	5	708.6	667.6	0.13	0.15
D16S539	274	6	6	1951	1754.2	0.09	0.05	5	5	3159.8	3071.8	0.12	0.18
D18S51[3]	290	6	N/A	1701.5	N/A	0.12	N/A	5	N/A	2917	N/A	0.08	N/A
CSF1PO	324	6	6	618.17	534.83	0.16	0.28	5	5	1039.8	933.4	0.14	0.27
D2S1338	326.5	6	6	1159	977	0.14	0.11	5	5	1898.2	1664.8	0.20	0.14

[1]CV: Coefficient of variation is calculated by the division of the standard deviation by the mean peak height

[2]non-specific recovery of D19S433 is most likely due to the high degree of sequence homology to the miniSTR probes

[3]Sample is homozygote at this locus

Table 6. Heterozygote allele mean peak height ratios of captured versus un-captured DNA samples from digested and sheared DNA sources.

		Digested (mean size 150 bp)						Sheared (mean size 800bp)					
		Captured			Uncaptured			Captured			Uncaptured		
Locus	Size	N	Mean	CV 1[1]	N	Mean	CV	N	Mean	CV	N	Mean	CV
AMEL	109	10	0.89	0.06	12	0.88	0.08	6	0.92	0.07	5	0.95	0.03
D19S433[2]	119	10	0.82	0.11	12	0.84	0.08	6	0.90	0.05	5	0.91	0.07
D3S1358	121	10	0.87	0.07	12	0.85	0.11	6	0.92	0.07	5	0.91	0.06
D8S1179	150.5	10	0.83	0.11	12	0.84	0.08	6	0.93	0.04	5	0.85	0.09
D5S818	157.5	10	0.81	0.13	12	0.84	0.10	6	0.92	0.07	5	0.91	0.07
TH01	175	10	0.89	0.13	12	0.90	0.11	6	0.95	0.05	5	0.90	0.05
vWA	177	10	0.74	0.13	12	0.79	0.14	6	0.87	0.08	5	0.90	0.07
D21S11	214	9	0.76	0.19	12	0.91	0.08	6	0.85	0.06	5	0.90	0.09
D13S317	234	10	0.74	0.17	12	0.83	0.12	6	0.84	0.13	5	0.87	0.11
TPOX	238	10	0.82	0.08	12	0.90	0.08	6	0.89	0.08	5	0.85	0.04
FGA	238.5	10	0.81	0.15	12	0.91	0.05	6	0.94	0.05	5	0.85	0.15
D7S820	273	9	0.81	0.20	12	0.84	0.13	6	0.85	0.11	5	0.91	0.08
D16S539	274	10	0.78	0.19	12	0.90	0.10	6	0.90	0.08	5	0.88	0.11
D18S51[3]	290	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
CSF1PO	324	10	0.80	0.21	12	0.76	0.19	6	0.82	0.15	5	0.88	0.12
D2S1338	326.5	10	0.81	0.11	12	0.83	0.12	6	0.83	0.12	5	0.83	0.15

[1]CV: Coefficient of variation is calculated by the division of the standard deviation by the mean peak height

[2]non-specific recovery of D19S433 is most likely due to the high degree of sequence homology to the miniSTR probes

[3]Sample is homozygote at this locus

Indicates Allelic Drop out

Table 7. Mean Peak height summary and % of samples tested that report an unassigned/ spurious peak migrating at or near a known STR allele in a CE electropherogram utilizing source DNA from a DNAseI digested sample (mean size 150 bp) and processed through Affinity Capture.

			Affinity Captured samples			Uncaptured samples			
Peak migrates at or near locus	size	Matches allele	Mean peak height (mPH)	CV[1]	% (n=10)	Mean peak height (mPH)	CV	% (n=12)	[2] Ratio
AMEL	109								
D19S433	119	13.2	78.5	0.33	100	186.2	0.25	100	0.42
		14.2	71.0	0.27	90	160.3	0.21	100	0.44
D3S1358	121	OL	75.7	0.44	30	109.6	0.20	42	0.69
		12	52.0	0	10				??
		20	97.3	0.11	40	92.5	0.44	50	1.05
D8S1179	150.5								
D5S818	157.5	OL	BT			53.3	0.08	33	**
TH01	75								
vWA	177	17	76.5	0.29	100	172.1	0.29	100	0.44
D21S11	214								
D13S317	234	OL	BT			60.0	0.13	50	**
TPOX	238	OL	64.5	0.01	20	92.0	0.28	100	0.70
FGA	238.5	OL	BT			79.7	0.08	50	**
D7S820	273								
D16S539	274								
D18S51	290	OL	57.0	0.10	10	79.5	0.26	92	0.72
		20.2	BT			63.8	0.25	67	**
CSF1PO	324								
D2S1338	326.5	OL	BT			51.0	0	8	**

[1]CV is calculated by the division of the standard deviation by the mean peak height
[2] Ratio of mean Peak Heights = affinity captured mean peak height ÷ unprocessed sample mean peak height

BT = below 50 RFU threshold

OL = off ladder
% = number of samples where a peak was reported ÷ total number of samples
?? – minimally detected peak of unknown origin in Affinity Captured sample that has no counterpart in the unprocessed sample
** - no peak height reported for Affinity Captured samples (below RFU threshold)

Table 8. Affinity Capture SYBR Green QPCR results for three targeted alleles versus a non-targeted allele.

Locus	Targeted	Mean Percentage Recovery	Copies with Affinity Capture	Control
Amelogenin	yes	101	780	770
CSF1PO	yes	89	400	450
FGA	yes	35	400	1130
D7S820	no	3.0	19.0	630

Table 9. Impact of fragment length on Affinity Capture SYBR Green QPCR Results.

		Affinity Capture of 800 bp fragments		Affinity Capture of 300 bp fragments	
Locus	Ratio 800/300bp[1]	Mean [2]	CV 3	Mean	CV
Amelogenin	0.79	1100	0.12	1400	0.28
FGA	0.71	1160	0.21	1640	0.20

[1]Ratio is based on mean locus-specific template recovery

[2]Mean locus-specific template recovery is based on 24 replicates for all four treatment categories

[3]CV, coeffcient of variation is standard deviation of template recovery over respective mean value

Table 10. Recovery of eukaryotic specific fragments from an excess of bacterial 16 S RNA , SYBR Green QPCR results.

HM/SE ratio[1]	Locus [2]	Mean percentage recovery [3,4]	Copies with Affinity Capture	Expected copies w/o Affinity Capture[5]
1 to 5	Human amelogenin	102	1.1E+03	1.1E+03
1 to 2	Human amelogenin	104	1.2E+03	1.1E+03
1 to 1	Human amelogenin	102	1.3E+03	1.1E+03
1 to 0	Huamn amelogenin	134	1.5E+03	1.1E+03
1 to 5	Bacterial ribosome S	6	7.3E+05	1.2E+07
1 to 2	Bacterial ribosome S	6	2.7E+05	4.8E+06
1 to 1	Bacterial ribosome S	6	1.4E+05	2.4E+06
1 to 0	Bacterial ribosome S	not applicable	5.9E+01	0.0E+00

[1] HM: human male DNA, SE:Staphylococcus epidermidis

[2]QPCR used primers at different concentrations: 250 nM for bacterial (SE) 1 uM for human

[3] Calculated using the ratio of observed human (HM) or bacterial (SE) amplicon copy number in captured samples over that expected with un-captured counterparts based on the copy yield of a 1:1 un-captured sample (1.11 E+3 and 2.38 E+6 for HM and SE, respectively)

[4]Each titration is represented by 8 (4 capturex 2 QPCR assays = 8)

[5]Expected copies calculated from control results and known DNA input (human or bacterial)

Conclusions and Future Implications

Discussion

The primary objective of this project was to develop and subsequently evaluate a focused affinity bead capture technique to facilitate the downstream CODIS STR analysis of two types of sub-optimal evidentiary samples, degraded and mixed source DNA. In the case of degraded DNA, the motivation for the development of Affinity Capture was to determine if template purification performed by the process might reduce the number of incomplete template copies which may compete and, in some instances distort (via the jumping PCR phenomena (2,3)) the true STR allelic composition. Hence, the overall performance of Affinity Capture was assessed by a head-to-head comparison of captured and un-captured fragmented human DNA (enzymatically digested or mechnically sheared) in a downstream CODIS PCR analysis (Identifiler Plus). Perfomance metrics such as locus-specific peak heights and heterozygote peak ratios were analysized statistically and the relative degree of LDO and ADO was noted. In addition, the Affinity Capture process was examined for fragment size capture bias, locus and eukaryotic specificity, and preservation of mixed human source ratios, which are important considerations with respect to both conventional CODIS PCR and NGS analysis.

Preliminary results suggest that the minimally optimized 15 loci affinity bead capture process is able to capture all targeted loci at detectable levels upon conventional CE based fragment analysis. Noteworthy is the observation that the Identifiler Plus profiles of captured samples display above threshold (50 RFUs) allelic mean peak heights across all targeted loci (Figures 6 and 7). However, the current performance of the captured versus un-captured samples is not quite as robust in terms of mean peak heights and DNA input requirements. There appears to be a distinct decay curve (mean peak height as a function of mean amplicon length) in the captured DNAse I digested samples which is much less pronounced for the sheared DNA Affinity Capture set. This finding suggests that the Affinity Capture process in this initial iteration may not be as effective at recovering the longer versus shorter templates in a highly fragmented DNA population. This conclusion is supported by the outcome of an experiment directly assessing the length bias of the Affinity Capture with DNA samples sheared to a larger, relatively uniform size distribution. The results of this experiment suggests that the shorter fragment (mean size 300 bp) was targeted approximately 20 to 30% more effectively than the longer template (mean size 800 bp) (Table 9). Since the effective recovery of longer fragments ($\geq$ 300 bp) is a requirement of a productive front-end for both conventional CODIS

PCR and NGS analyses, further optimization (see discussion on future research) is needed on the nascent Affinity Capture technology.

Nonetheless, even with the moderate degree of size-dependent template loss exhibited by the sheared samples, and the much larger effect induced by the severe degradation of the digested template (estimated mean size =150 bp as seen in Figure 2), the Affinity Capture process is able to generate Identifiler Plus profiles with extremely low dropout (no LDOs in any of the replicates) and allelic peak height CVs that compare to those of the un-captured samples. The incidence of spurious peaks is lower in the captured versus the un-captured PCR amplifications and signal processing analysis (SAIC proprietary algorithms) indicates that this is due to overall signal reduction in the processed sample with the concomitant result of spurious peak intensity falling below threshold detection cutoffs of the commercial software used for analysis. It might be possible that the reduction in the background signal due to a cleaner captured template could have implications for LCN analysis by allowing for an increase in PCR cycle number but such an analysis has not been completed. It is also very encouraging that mean heterozygote allelic ratios are above 0.7 for all 15 targeted loci for both types of captured samples (digested and sheared) and that this ratio does not appear to be related to the mean locus-specific fragment lengths (Figure 8 and Table 6).

Results of the locus specificity capture experiment indicate an 11 to 33 fold enrichment of the targeted versus untargeted loci (Table 7). However, in the 15 loci Affinity Capture experiments, the single untargeted Identifiler Plus D19S433 locus was captured along with the targeted loci and its correct allelic peaks consistently exceeded the Applied Biosystem's default peak height threshold (Tables 4 and 5) at the 1500 template copy level. The unintentional recovery of this locus is most likely due to the high degree of sequence similarity of this region with those of miniSTR probes. The quasi-specificity of these CODIS capture probes will, most likely, not be a hindrance to downstream applications as long as it is not does not interfere with the capture of intended genomic regions or is not too extensive. It would be desirable to assess degree of STR locus specificity against a variety of loci across the human genome via a genomic hybridization array or a NGS analysis. These types of empirical screens will signal any non-relevant regions that are inadvertently hybridized during the capture process.

Another template condition required for accurate CODIS STR profile resolution of mixed human DNA is the consistency of contributor ratios across CODIS loci so that allelic information from each locus can be linked together to generate a complete profile. Consequently, it is critical that the capture process preserve these source proportions from locus to locus. It is encouraging to note that that there is no discernable difference in the degree of source

proportion preservation between captured and equivalent un-captured samples (Figure 9). Thus, it may be concluded that the Affinity Capture process does not introduce additional bias in the analysis of human genomic DNA mixtures with two contributors at the ratios tested.

To prevent prokaryotic data from swamping that generated from human template in downstream analysis using universal or random primers (such as NGS), a capture technology should be sufficiently effective at human DNA enrichment to the point at which the human component prevails. The capture process appears to be specific for the human component as a mean of only 6% of the initial SE template is retained in the post capture sample. These series of experimental assessments involved only the human amelogenin locus, a single species of bacteria, one individual human male DNA and a QPCR assay of coarse resolution and so this should be viewed with cautious optimism until a more complete analysis with other loci and bacterial genomes are queried.

Implications for policy and practice

The data demonstrates that the prototypic Affinity Capture does not interfere with the conventional CODIS STR analysis by introducing artifacts. The post-captured profiles (at the 1500 copy number) are characterized by detectable allelic peak heights, and expected heterozygote balance Additionally, the Affinity Capture process preserves two contributor human mixture ratios and effectively reduces the prokaryotic DNA component of a sample, two critical requirements for both conventional downstream analysis as well as emerging DNA technologies targeted at addressing CODIS STR mixture de-convolution.

Preliminary success has been achieved in the development of a 15 loci multiplex Affinity Capture process. It is apparent that such a technology needs further improvement before it is sensitive or specific enough for routine application in forensic laboratories. At this stage, the process is not as effective at recovering longer versus shorter template molecules so does not appear to confer an advantage over the conventional non-capture treatment with respect to highly fragmented DNA samples, whereas the process is more successful with moderately fragmented materials. How this applies to problematic DNA samples frequently encountered at crime scenes is still unknown. The probe binding efficiency of the longer DNA fragments needs to be increased across all targeted loci before this procedure can be used as a productive front end for conventional CODIS STR profiling as well as NGS mediated mixture de-convolution.

The Affinity Capture was developed as a multiplex to be consistent with current processing of forensic samples and minimize the need to aliquot limited sample followed by

pooling product for further processing. The Affinity Capture process developed requires equipment standard in most crime laboratories. The application of the Affinity Capture as a front-end for an NGS technology does, however, requires the use of technology not yet found in crime laboratories. Affinity Capture combined with NGS will be most useful for DNA samples that are not amenable to conventional methods. Once optimized, this procedure should be applicable to a kit format and can be performed in the pre-amplification area of the local crime laboratory. For any future NGS analysis, a core facility either at the state or federal level could process sub-optimal sample DNA (post extraction and/or capture) along with the reference sample controls.

Implications for future research

The logical next step would be to undergo optimization of the process to increase both sensitivity and specificity as mentioned above. Maximizing the recovery of the longer fragments is especially critical. Other avenues of investigation include: incubation/temperature modifications, adjustment of buffer salinity to encourage probe/ template binding specificity, and the use of a more processive polymerase with a higher rate of nucleotide incorporation to increase the rate of primer extension. These modifications will enable the optimization of signal to noise in the sensitivity/specificity continuum and reduce the required amount of primer and template.

Technical and experimental improvements to the evaluation of the capture process include the use of more sensitive QPCR chemistry such as the TaqMan®[i] 5' nuclease assay with the minor groove binding fluorescent probe and associated hardware (Applied Biosystem's 7500 /7900 Real-Time PCR instruments). In addition, future widespread testing of the capture technology should involve individuals of different ethnicities, more complex mixture combinations (3 or more contributors in various proportions) and a eukaryotic specificity survey against a diverse spectrum of prokaryotic organisms.

Once optimized for the 15 loci multiplex capture of CODIS STR loci, the Affinity Capture process can be extended to encompass more human loci including non-CODIS STRs. Eventually it will be informative to test an array of post-capture human DNA mixtures of varying complexity on a NGS platform. Finally, the affinity bead capture technology may be adapted to other issues related to DNA template purification such as PCR inhibitor removal.

References

1. Lynch M, Endeavour 2003, 27: 93-97.

2. Alaeddini R, Walsh S, Abbas A, Forensic Science International: Genetics 2010, 4: 148-157.

3. Pääbo S, Proc. Natl. Acad. Sci. USA 1989, 86: 1939-1943.

4. Lindahl T, Philos. Trans. R.Soc. Lond. B: Biol. Sci 1996, 351: 1529-1538.

5. Swango K, Timken M, Chong M, Buoncristiani M, Forensic Science International 2006, 158: 14-26.

6. Butler J, Shen Y, McCord B, J Forensic Sci. 2003, 48 (5): 1054-1064.

7. Hill C, Kline M, Coble M, and Butler J, J. Forensic Sci., 2008, 53(1): 73-80.

8. Bill M, Proceedings of the 19th International Symposium on Human Identification Oct. 2008, Hollywood California.

9. Wang T, Xue N, Birdwell D, J Forensic Sci 2006, 6: 1556-4029.

10. Cowell RG, Lauritzen SL , Mortera J, Forensic Science International 2007, 166: 28-24.

11. Lucas C, Wang TW, Birdwelll JD, Proceedings of the 15th International Symposium on Human Identification Oct. 2005, Phoenix Arizona.

12. Butler J, 2001 Academic Press, London.

13. Boutrand L, Egyed B, Füredi S, Mommers N, Mertens G, Vandenberghe A, 2001, Int J Legal Med 114: 295-297.

14. Clayton T, Whitaker J, Sparkes R, Gill P, 1998, Forensic Science International 166: 28-34.

15. Levinson G, Gutman G, Mol. Biol. Evol. 1987, 4: 203-221.

16. Margulies M, et al. Nature. 2005 September 15; 437(7057): 376–380.

17. Garza J, Freimer N, Genome Research 1996, 6: 211-217.

18. http://www.justice.gov/ag/dnapolicybook_solve_crimes.html.

19. Herráez D, and Stoneking M, Analytical Biochemistry 2008, 383: 329-331.

20. Parham N, Picard F, Pytavi R, Gagnon M, Seyrig G, Gagné P, Boissinot M, Bergeron M, Clinical Chemistry 2007, 53: 1570–1576.

21. http://www.illumina.com.

22. http://www.invitrogen.com.

23. Whitney D, Skoletsky J, Moore K, Boynton K, Kann L, Brand R,Syngal S, Lawson M, ShuberA, J. Mol. Diagn. 2004, 6: 386-395.

24. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProus E, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe D, Lander E, Nusbaum C, Nat. Biotechnol. 2009, 27(2): 182-189.

25.Garber K, Nature Biotechnology 2008, 26(10): 1101-1104.

26. Chandler D, Stults R, Anderson K, Cebula S, Schuck B, and Brockman F, Analytical Biochemistry 2000, 283: 241–249.

27. Sullivan K, Mannucci A, Kimpton, C and Gill P, BioTechniques 1993,15: 637-641.

28. Ecker D, Sampath R, Blyn L, Eshoo M, Ivy C, Ecker J, Libby B, Samant V, Sannes-Lowery K, Melton R, Russell K, Freed N, Barrozo C, Drader J, Wu J, Rudnick K, Desai A, Moradi E, Knize D , Robbins D, Hannis J, Harrell P, Massire C, Hall T, Jiang Y, Ranken R,White N, McNeil J, Crooke S, and Hofstadler S , PNAS 2005, 102): 012-8017.

29. Gilder J, Inman K, Shields W, Krane D, Int J Legal Med 2010, 115: 64-69.

Dissemination of Research Findings

Preliminary research results were presented (poster presentation) at the 2010 Annual NIJ Award Conference (Arlington VA). No other project related publications are presently planned.