Retrotransposable Elements: Novel and Sensitive DNA Markers and Their Application in Human Identity

Sudhir Sinha, Ph.D., Gina Murphy, M.S., Hiromi Brown, Ph.D. Andre' Green, Ph.D., Anne Montgomery, M.S., Jonathan Tabak, B.A., Marion Carrol, Ph.D.

InnoGenomics Technologies; New Orleans, LA, USA; www.innogenomics.com

Introduction

Retrotransposable elements (REs) are non-coding genomic sequences of repetitive DNA, consisting of long and short interspersed nuclear elements (LINEs and SINEs), which comprise approximately 40% of the human genome. REs are well characterized, but until now their practical utility for human identity and bio-ancestry testing has been severely limited due to the inherent size difference (>300bp) associated with insertion and null alleles (or INNULs), which has led to difficult multiplexing challenges and poor PCR efficiency. To circumvent the allele size disparity, we have developed a novel "mini-primer" design that reduces the overall amplicon size as well as the difference in amplicon sizes between the two allelic states, insertion or null insertion (Figure 1). The resulting allelic amplicons can now be designed to differ by as little as one base pair with a substantially reduced size (much smaller than STR markers commonly used in forensics), such that degraded and/or low quantity samples can now be typed.

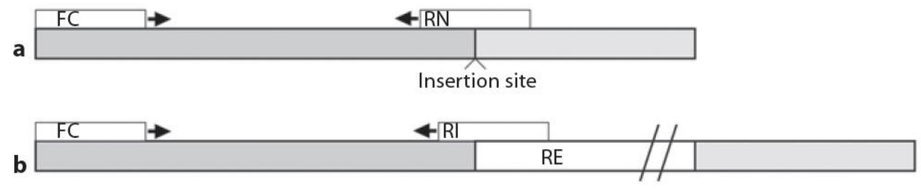


Figure. 1. Novel "mini-primer" design. a) A common forward primer (FC) is used for both insertion and null alleles. A fluorescently labeled 'null-specific' reverse primer (RN) straddles the insertion site of the RE and anneals in the absence of the RE. b) In instances where the RE is present, the annealing site of the reverse primer is disrupted, and the 'insertion-specific' reverse primer (RI) anneals at the site that overlaps with the insertion site and the adjacent portion of the RE.

Utilizing this primer design innovation and other advancements, we have developed highly informative RE-based PCR and Real-Time qPCR assays that offer exceptional sensitivity, reproducibility and robustness, and are compatible with widely used, validated laboratory instrumentation. These include:

InnoQuant® & InnoQuant® HY Kits – robust real-time qPCR assays for simultaneous assessment of human and male DNA quantity, quality and integrity. The use of high copy number retrotransposable element targets provides high sensitivity and reproducibility. Comprehensive and reliable sample assessment enables forensic laboratories to confidently screen samples and make informed workflow decisions, significantly reducing downstream re-processing and improving analysis success rates.

InnoTyper® 21 Kit – a small amplicon (~60-125 bp) DNA typing system for challenging forensic samples that is fully compatible with currently used instrument platforms. This kit contains 20 Retrotransposable Insertion Polymorphism (RIP) markers and Amelogenin, and can provide discriminating results from extremely low-level and/or degraded samples including bone fragments or single hair shafts. The RIP ALU markers in this kit are stable and identical by descent, providing valuable utility for difficult kinship, missing person and mass disaster victim identification cases.

We report here utilization of several of these recently developed technologies to improve analysis efficiency and success rates for highly compromised, degraded as well as trace samples. Preliminary data from the development of a small amplicon, multiplexed primer kit for preparing next-generation sequencing (NGS) libraries that are useful for forensic and bio-ancestral identification from challenging DNA samples is also presented.

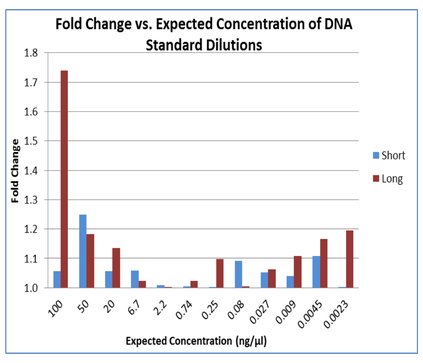
Materials and Methods

- DNA extractions were performed at InnoGenomics using either the BioBasic or QiaAmp DNA Isolation Kits. For the rootless hair shafts, two centimeter hair samples, with follicular tags removed, were cleaned in a series of wash steps before complete digestion and purification using a combination of commercial buffers at Western Carolina University before sending DNA extracts to InnoGenomics for InnoTyper® 21 amplification. For the degraded human remains, organic DNA extraction methods were used according to UNTHSC Center for Human ID protocols for missing persons samples; amplification and analysis also performed at UNTHSC.
- InnoQuant® qPCR quantitation: Primers and TaqMan® probes were designed using two independent intra *retrotransposon* insertion targets and a synthetic target as an IPC. The 2 autosomal targets are: a "short" ALU based target of 80 bp in size, and a "long" target from a separate retrotransposon of 207 bp in size. Real-time PCR reactions were processed on the AB 7500 Real-Time PCR System using Agilent Technologies Brilliant Multiplex QPCR Master Mix as follows: 10 min at 95°C; and 32 cycles of: 15 sec at 95°C, and 2 min at 61°C. Degradation Index was determined by **DI = [short]/[long]**.
- InnoTyper® 21 amplification: Primers were designed for 20 RIP markers and AMEL in a multiplex assay. PCR was performed on an ABI 9700 as follows: 15 min at 95°C; 31 or 32 cycles of: 30 sec at 95°C, 30 sec at 58°C and 1 min at 72°C; one cycle of 60°C for 1 hour. AB 3130 Genetic Analyzer was utilized. Data analysis was performed using GeneMapper® and GeneMapper® *ID/ID-X*.
- Sequencing InnoTyper®21 markers on the Illumina MiSeq®: preliminary data were generated at UNTHSC using 500pg DNA of a known reference sample that was amplified with cold (unlabeled) InnoTyper® 21 primers according to thermal cycler conditions and primer concentrations of the InnoTyper® 21 kit and sequenced on the Illumina MiSeq®. The TruSeq® DNA PCR-Free Sample Preparation Kit (Illumina) was used to prepare the libraries.

Results

Quantitation and Sample Assessment with *InnoQuant®*

Sensitivity:

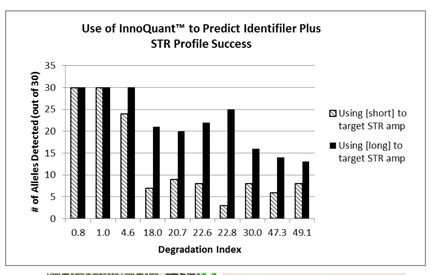


NIST SRM "A" Dilution

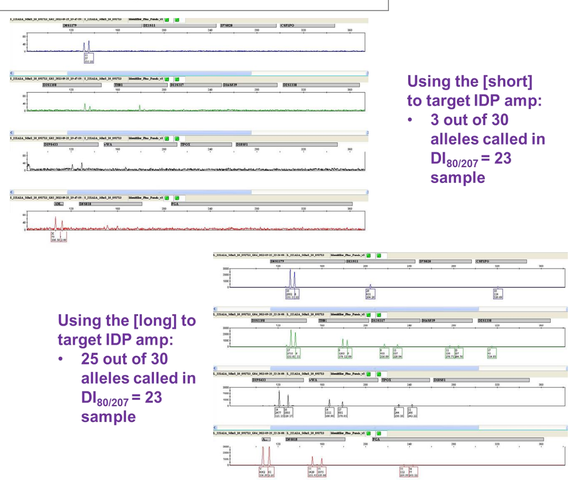
Expected ng/uL	Observed Short Qty	Observed Long Qty	Short Qty Fold change	Long Qty Fold change
5	4.1543	6.8428	1.20	1.37
1.25	1.2454	1.2499	1.00	1.00
0.3125	0.3201	0.2849	1.02	1.10
0.0781	0.0765	0.0608	1.02	1.28
0.0195	0.0202	0.0165	1.04	1.18
0.0049	0.0055	0.0042	1.12	1.17
0.0012	0.0013	0.0010	1.06	1.18
Average Fold Change			1.07	1.18

Highly sensitive and reproducible: less than 20% variation from expected values as low as 1 picogram of DNA. ZERO means ZERO: can stop at quant

Correlation between DI and STR profile success:



As the DI increases, the [long] value can be used to target the STR amplification and thus obtain more alleles in the first attempt.



Using the [short] to target IDP amp:
- **3 out of 30 alleles called in $DI_{80/207}$ = 23 sample**

Using the [long] to target IDP amp:
- **25 out of 30 alleles called in $DI_{80/207}$ = 23 sample**

InnoQuant® Results:

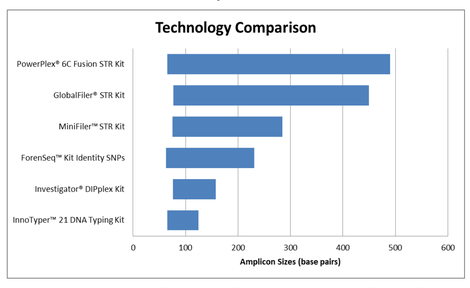
[Short] = 6.68 ng/uL
[Long] = 0.293 ng/uL

DI = 6.68/0.293 = 22.8

Accurate targeting of STR amplifications increases profile recovery.

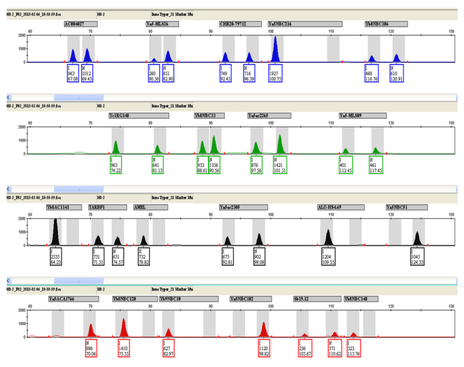
Improved Genotype Recovery from Challenging Samples with *InnoTyper® 21*

A novel primer design allows control of the size of the resultant insertion or null alleles, with amplicon sizes less than 125bp.



Amplicon size smaller than mini-STR kits

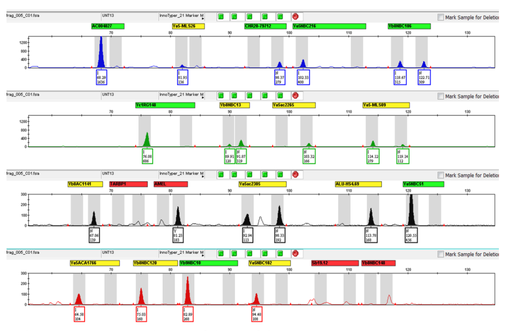
Results with 2 cm rootless hair shafts:



Donor ID	DI	Input DNA (pg)	% recovery	Donor ID	DI	Input DNA (pg)	% recovery
1-1	N/A	0.7	12%	6-1	N/A	1.3	10%
1-2	N/A	0.7	10%	6-2	N/A	1.0	14%
1-3	N/A	1.0	5%	6-3	N/A	1.3	N/A
2-1	7	102.5	74%	7-1	N/A	0.6	14%
2-2	66	56.3	55%	7-2	N/A	1.6	12%
2-3	28	112.1	100%	7-3	N/A	1.6	2%
3-1	N/A	1.5	5%	8-1	19	77.2	90%
3-2	N/A	2.6	0%	8-2	5	67.2	100%
3-3	N/A	1.1	7%	8-3	17	94.0	98%
4-1	16	251.3	100%	9-1	52	106.8	98%
4-2	22	176.6	98%	9-2	N/A	73.8	88%
4-3	17	102.4	98%	9-3	N/A	46.1	60%
5-1	N/A	2.6	19%	10-1	105	101.5	76%
5-2	N/A	5.7	24%	10-2	63	129.1	93%
5-3	N/A	3.9	14%	10-3	114	139.9	76%
Over 50% Genotype Recovery				Less than 50% Genotype Recovery			

~50% of hair shafts produced usable autosomal DNA profiles with InnoTyper® 21

Results with degraded human remains:



Increased allele recovery from degraded samples compared to other commonly used typing methods

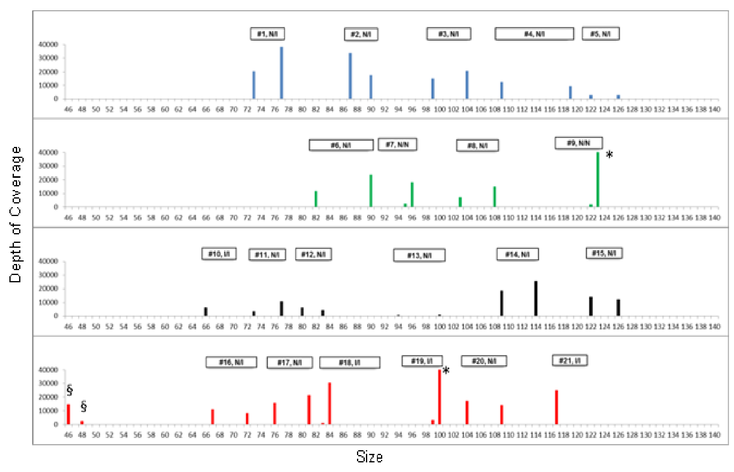
Above: partial profile of human remains which previously yielded no interpretable results with STR and mitochondrial DNA profiling methods

Expanded Utility Using Massively Parallel Sequencing (MPS)

InnoTyper RE markers can be combined with ancestry informative ALU RE markers, all of small amplicon size < 100 bp, to provide additional information. Such an MPS-based targeted-capture multiplex panel can include ~100 small amplicon RE markers to provide:

- increased discrimination power for Human ID purposes
- determination of ancestry within continental and sub-continental population groups
- accurate differentiation of complex mixtures
- effective typing of poor quality (i.e. highly degraded, low-level) samples

ALU REs are exceptional markers for tracking identity by descent and inferring biogeographical origin due to their high copy number, known ancestral state, and genomic stability.



Above: Preliminary results showing a mock electropherogram of Illumina MiSeq® reads for an InnoTyper®21 profile. The expected alleles as given by CE analysis are given for each marker (N,I = no insert, insert allele, respectively). The marker labels (boxes) span the expected sizes of both alleles. An * indicates an allele for which DoC exceeded 40,000 (Markers 9, 19). An § indicates spurious read sizes generated for Markers 16 (left bar) and 21 (right bar).

Conclusions

- Quantification & Sample Assessment Kit InnoQuant® uses high copy number retrotransposable element (RE) targets to obtain highly sensitive and reproducible results, and provides additional information (Degradation Index) prior to PCR amplification that can significantly reduce downstream re-processing and enable DNA analysts to make improved processing decisions.
- ALU RE based typing kit InnoTyper® 21 demonstrates the ability to obtain informative DNA profiles from extremely challenging samples, such as skeletal remains, historical remains and cut hair shafts that have failed to produce informative STR data. InnoTyper® 21 is also significantly more discriminating and less labor-intensive than mtDNA sequencing, which is currently the genotyping standard for heavily degraded and/or low level samples.
- A highly informative small amplicon (50-90 bp) panel of ALU RE markers can be created for Massively Parallel Sequencing (MPS) platforms that is extremely well suited for heavily degraded, low-level and mixed samples, and provides powerful identity and biogeographical ancestry information in a single analysis to preserve scarce biological evidence.

Literature Cited

1. G.M. Pineda et al. *Development and validation of InnoQuant, a sensitive human DNA quantitation and degradation assessment method for forensic samples using high copy number mobile elements Alu and SVA*. Forensic Science International: Genetics 13 (2014) 224–235.
2. S.K. Sinha et al. *Development of a Novel and Sensitive DNA Analysis Multiplex Based on INNUL Markers for Highly Degraded Forensic DNA Samples*. Oral presentation at the 66th Annual Meeting of the AAFS. Seattle, WA.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1230352. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. For additional information, contact: ssinha@innogenomics.com.

We would also like to acknowledge Dr. Arthur Eisenberg, Dixie Peters, Dr. Bobby LaRue, Dr. Bruce Budowle and Jonathan King from the University of North Texas Health Science Center (UNTHSC), as well as Dr. Mark Wilson and Dr. Kelly Grisedale from Western Carolina University for significant contributions to this work.