International Scholars Journals

Frontiers of Agriculture and Food Technology ISSN 7295-2849 Vol. 8 (5), pp. 001-006, May, 2018. Available online at www.internationalscholarsjournals.org © International Scholars Journals

Review

Developments in the use of Y-chromosome markers in forensic genetics

Neil Leat, Liezle Ehrenreich, Mongi Benjeddou and Sean Davison

Department of Biotechnology University of the Western Cape Bellville 7535 South Africa.

Accepted 09 November, 2017

Sexual assault is a significant problem facing South African society. The South African Police service indicates that between April 2003 and March 2004 over 52 733 rapes were reported. Our research focuses on the development and implementation of genetic identity testing systems for use in sexual assault cases. Genetic identity testing is achieved by examining polymorphic regions of DNA. Typically sets of polymorphisms are examined together to provide a genetic profile. The polymorphic markers most commonly used lie on the autosomal (1-22) chromosomes. While they have an excellent capacity to distinguish between individuals they do have disadvantages. In sexual assault cases it is often difficult to separate the female victim's profile from the rapist's profile. Analysis of Y-chromosome markers overcomes this by generating male specific profiles. In this review we highlight recent developments in the use of Y-chromosome markers in forensic genetics.

Key words: Polymorphism, Y-chromosome, markers, South African.

INTRODUCTION

South African society is faced with high levels of violent crime. Between April of 2003 and March of 2004, over 652 959 acts of violence were reported to the South African Police Service (http://www.saps.gov.za/statistics/reports/crimestats/2004/crime_stats.htm). These reports typically involved: murder, attempted murder, physical assault, assault with the intent to do grievous bodily harm, indecent assault, and rape. Of these reports, 52 733 were of rape, making it one of the most common serious violent crimes in South Africa. The highest per capita incidence of rape occurred in the Northern Cape and the lowest in the Limpopo province (Figure 1). While the figures indicate a high incidence of rape, some organizations suggest that the incidence of sexual assault may be substantially higher. Rape Crisis a non-governmental organization (http://www.rapecrisis.org.za), suggests that police statistics underestimate the incidence of sexual assault for several reasons.

They note that the definition of rape in South African law does not include rape of men, oral rape or rape with objects. In addition, rapes within long-standing relationships are often not viewed as rape and victims appear reluctant to report cases to the police. Despite this, there is agreement that the incidence of sexual violence in South Africa is unacceptably high.

There are a number of factors leading to the high incidence of rape in South Africa. Measures that could address the problem are in most instances beyond the scope of solutions that can be offered by biotechnology. Despite this a contribution can be made in at least one area, forensic genetics. By developing robust DNA based identity testing systems, biotechnology can assist in the identification of perpetrators of violent crime.

The South African Police currently have excellent commercial systems for the analysis of autosomal genetic markers. These systems generate highly specific profiles in both men and women. They are excellent for use on unmixed samples; however their value is compromised when testing mixed samples from sexual assault cases. While systems for the analysis of autosomal genetic markers are well established, systems for the analysis of male specific markers are the focus of a substantial amount of research (Bosch et al., 2002; Butler et al., 2002; Hall and Ballantyne, 2003; Kayer et al., 2004;

*Corresponding author. E-mail: SDavison@uwc.ac.za; Phone: +27 21 9592216; Fax: +21-21-9593505.

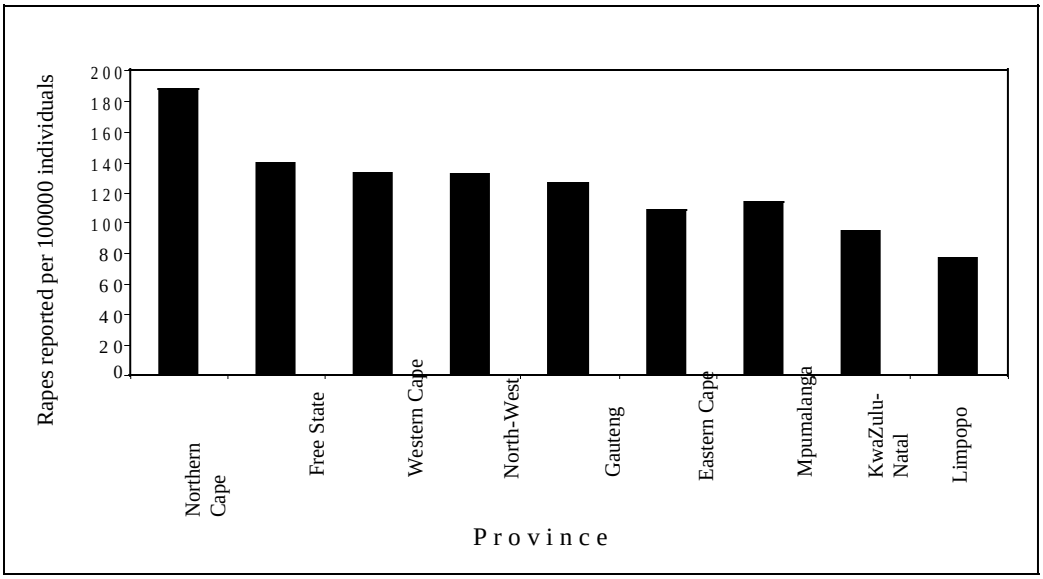

Figure 1. Number of rapes reported to the South African Police per 100000 of the population between April 2003 and March 2004. Statistics obtained from the South African Police (http://www.saps.gov.za/statistics/reports/crimestats/2004/_pdf/crimes/Rape. pdf).

Redd et al., 2002; Schoske et al., 2003; Hanson and Ballantyne, 2004). The primary objective of our research is to exploit Y-chromosome sequence information to develop male specific genetic typing systems for use in sexual assault cases in South Africa.

GENETIC MARKERS USED IN FORENSIC IDENTITY TESTING

Forensic identity testing relies on the analysis of genetic elements which vary from person to person. A number of polymorphic genetic elements could theoretically be used in human identity testing. These include Single Nucleotide Polymorphisms and micro-satellites. Single Nucleotide Polymorphsims (SNPs) typically involve simple single nucleotide substitutions. It has been suggested that on average there are 8.33 SNPs per 10Kb of human genome sequence (Zhao et al., 2003). Despite the large number of SNPs available, they mutate relatively slowly by comparison to microsatellite loci. As a result, they have a modest capacity to discriminate between individuals unless a substantial number are analyzed simultaneously. Micro-satellites or Short Tandem Repeats (STRs) are loci containing tandemly repeated elements typically between 2 and 8 bp in size. A number of STR loci are highly polymorphic and are therefore useful in tests aimed at discriminating between individuals. The repeated elements and the STR loci themselves are of a manageable size and are amenable to analysis using PCR based techniques. Modern forensic genetics relies almost exclusively on STR loci with repeated elements of 4 bp in size.

STR markers suitable for forensic analysis occur on both the autosomal “non-sex” chromosomes 1-22 and the sex chromosomes X and Y. Research on the use of STR markers for identification purposes has focused on markers on the autosomal chromosomes. The most obvious reason for this is the fact that they can be used for the identification of individuals from both sexes. They also have an excellent capacity to discriminate between individuals. Despite their advantages autosomal markers have significant limitations when used to analyze mixed male/female stains. In cases of sexual assault there is a high probability of having to analyse samples containing a mixture of male and female DNA. Samples of this nature produce mixed profiles, which the forensic scientist must attempt to separate. This is particularly difficult if the signal for the female’s profile is stronger than that of the male. In these cases the use of Y-chromosome STR markers provides an attractive alternative, since analysis of these markers will produce only a male profile.

Y-CHROMOSOME STRUCTURE

Before considering the characteristics of recently discovered Y-STR loci the structure and inheritance patterns of the Y-chromosome will be briefly considered.
The Y-chromosome is one of the smallest chromosomes in the human genome with an overall size of ~60 Mb including approximately 24 Mb of euchromatin and 30 Mb of heterochromatin (Quintana-Murci et al., 2001; Tilford et

al., 2001). At the tip of both arms are pseudo-autosomal regions (PARs) which are homologous to X-chromosome sequences and responsible for correct pairing between the sex chromosomes during meiosis. The majority of the Y-chromosome does not undergo recombination and is always in a haploid state. The non-recombining region of the Y-chromosome (NRY) is inherited intact through paternal lineages unless a mutation occurs at a given locus. This factor highlights one of the extended haplotype key limitations associated with the use of Y-chromosome markers for identity testing. While an autosomal STR profile can be almost unique to an individual, a Y-STR profile is shared by most of a male individual's paternal relatives.

THE DISCOVERY OF NOVEL Y-STR LOCI

As recently as the mid-1990's relatively few Y-STR loci were available for use in forensic studies. A collaborative study was undertaken to assess the suitability of thirteen of the available loci for forensic casework (Kayser et al., 1997). Partly as a result of the impetus from this study an affiliation of forensic laboratories, the Forensic Y User group, was established to regulate the accumulation of population data (http://www.yhrd.org). The Forensic Y User Group coordinates the typing of loci included in either the 'minimal haplotype' or the 'extended haplotype'. The 'minimal haplotype' includes (DYS19, DYS389I/II, DYS390, DYS391, DYS392, DYS393 and the duplicated locus DYS385). The loci of the 'minimal haplotype' are by far the most widely used Y-STR loci in forensic casework. The 'extended haplotype' was initially constituted simply by including the duplicated dinucleotide locus YCA II with the loci of the minimal haplotype. More recently the YCAII locus has been replaced by novel loci considered below.

Prior to the release of large amounts of Y-chromosome sequence, attempts to identify novel Y-STR loci depended on a series of cloning and hybridization steps. In the most recent and almost certainly the last example of this approach a cosmid library was constructed from flow-sorted human Y-chromosomes (White et al., 1999). Probes containing the repeated element $[GATA]_{10}$ or $[TATC]_{10}$ were used to select subclones containing repeated GATA elements. This approach led to the identification of seven loci with the potential for use in human identity testing. At the time this work doubled the number of known tetranucleotide Y-STR loci.

The release of substantial amounts of Y- chromosome sequence allowed for a more straightforward approach to the identification of novel Y-STR-loci. In the first study of its kind. Ayub et al., (2000), surveyed 1.22Mb of Y-chromosome sequence, identifying 18 novel STR sequences. Three of these loci, DYS437, DYS438 and DYS439, have been incorporated into commercial Y-STR typing systems and are widely accepted by forensic scientists. DYS438 and DYS439 have recently been selected as suitable replacements for the YCA II locus of the extended haplotype.

More recently a substantial number of STR loci have been identified from Y-chromosomal sequence. Iida et al. (2001, 2002) characterized five loci DYS441, DYS442, DYS443, DYS444, and DYS445. A more thorough survey was conducted by Redd et al (2002), resulting in the identification of 14 novel Y- STR loci DYS446, DYS447, DYS448, DYS449, DYS450, DYS452, DYS453, DYS454, DYS455, DYS456, DYS458, DYS459, DYS463 and DYS464. By far the most comprehensive analysis has involved a survey of 23 Mb of euchromatic Y-chromosomal sequence (Kayser et al., 2004). Loci were selected with repeated elements between 3 and 6 bp in size. Loci with dinucleotide repeated elements were avoided since they have a propensity to generate PCR stutter artifacts. This approach resulted in the identification of 475 potential Y-STR loci of which 45 had been previously identified. PCR primers were successfully designed for 281 loci *in silico*. Of those, 166 primer-sets generated male-specific amplicons and 139 loci were demonstrated to be polymorphic in a group of eight individuals representing different binary-marker haplogroups. Using this new sequence-based approach, the pool of Y-STR loci available for evolutionary, forensic and paternity testing has increased considerably.

SELECTION OF NOVEL Y-STR LOCI FOR USE IN FORENSIC CASEWORK

It seems plausible that discrimination capacity of Y-STR typing systems could be improved by including novel loci with suitable properties. Important factors in selecting Y-STR loci include their polymorphic content, the ease with which male specific PCR primers can be designed, the repeat-unit size which ultimately determines the extent to which PCR stutter artifacts are generated and copy number on the Y-chromosome. Multi-copy loci generate complex profiles using a single primer set. While this has the advantage of providing a high discrimination capacity for a given primer set, it also means that clear locus-allele relationships cannot be established. In addition, it may be difficult to identify the number of contributors to a mixed stain when using multi-copy loci. Ultimately, it would be desirable to select a set of highly polymorphic, single-copy, male-specific loci which generate minimal stutter artifacts during PCR.

MULTIPLEX DEVELOPMENT

Forensic genetics relies heavily on the use of systems which facilitate the analysis of a number of loci simultaneously. This is typically achieved using multiplex PCR systems, in which two or more target sequences are amplified in the same reaction. Multiplex PCR has been applied to the analysis of deletions, mutations and

Table 1. Recently published Y-STR multiplex typing systems designed to amplify at least nine loci.

Typing system	Number of loci	Minimal haplotype Loci included	Source/Publication
Y-PLEXTM12	12	All	Reliagene Technologies Inc
Mentype Argus® Y-MH	9	All	Biotype AG
PowerPlex Y	12	All	Promega
20-Plex	20	All	Butler et al., 2002
10-Plex	10	DYS19, DYS391, DYS392	Schoske et al., 2003
21-Plex	21	All	Hanson and Ballantyne, 2004
Y Filer	17	All	Applied Biosytems

polymorphisms and has many advantages over uniplex PCR. Key advantages include the fact that a large number of loci can be amplified and analyzed in a relatively short amount of time and smaller amounts of PCR reagents and template DNA are ultimately used.

Systems have been developed where up to 21 markers are efficiently amplified in a single reaction (Table 1). This means a substantial amount of information can be obtained from limited amounts of template DNA.

The development of multiplex PCR systems typically faces a number of challenges. Often the DNA target sequences do not amplify equally and the presence of several primer sets increases the chance of generating non-specific artifacts. This not only makes the PCR less efficient, but also makes analysis of the results more difficult. A substantial amount of optimization is required in order to develop a multiplex PCR system in such a way that all target sequences (and only the target sequences) are consistently and equally amplified.

A number of approaches have been used to develop multiplex PCR typing systems (Butler et al., 2002; Hall and Ballantyne, 2003; Henegariu et al., 1997; Schoske et al., 2003; Wallin et al., 2002). While there are differences in each approach a number of consistent elements exist. Possibly the most important aspect of multiplex development is the design of compatible and specific primer sets. Given that all primer sets must function under a single set of physical conditions it is essential that they are designed to have almost identical melting temperatures. In addition primer sequences should be selected so as to minimize primer-primer interactions. It is also essential that the care is taken to design primers which are specific to the intended locus on the Y-Chromosome. Primers which share a high degree of sequence identity to homologues on the X-chromosome or elsewhere on the genome should be avoided.

Synthesized primer pairs can be tested in uniplex PCR reactions to confirm that male specific amplicons of the expected size are produced. Primer pairs which function well in uniplex reactions can be combined in multiplex reactions and the relative performance of each set monitored. The most common approach to balancing the amount of each product is simply to adjust the concentration of the relevant primer set. Typically the ideal cycling regime and concentration of reagents such as Taq polymerase, KCl, dNTPs and $MgCl_2$ are investigated empirically. Additives such as BSA, glycerol and DMSO have also been used in attempts to improve the performance of multiplex PCR reactions.

Multiplex typing systems must be optimized to the point where they meet certain performance standards. There are several governing bodies that ensure that high typing and analysis standards are maintained. Among these is the International Society for Forensic Genetics (ISFG), the Scientific Working Group on DNA Analysis Methods (SWGDAM), and the European DNA Profiling Group (EDNAP). These organizations lay down guidelines for the use and validation of multiplex PCR typing systems. Some common validation exercises include: [1] establishing that the typing system is sensitive and performs consistently using freshly prepared and stored DNA, [2] that identical results are obtained irrespective of the type of tissue from which DNA was extracted, [3] that the systems yield consistent results in several laboratories, and [4] that the system performs well when used to analyze samples similar to those encountered in forensic casework. In this regard efficient typing should be obtained for DNA extracted from body fluids mixed with commonly encountered substances (e.g. dyes, soil, leather, denim). The influence of environmental factors such as temperature, humidity and UV should also be established and the capacity of the system to analyze mixtures of male and female DNA should be examined.

A number of Y-STR multiplex PCR typing systems have been developed using the approaches outlined above (Table 1). While these multiplex sets demonstrate the potential for combining large numbers of primer sets into multiplex reactions they almost all include the loci of the minimal haplotype as core loci. While the loci of the minimal haplotype have had an important role to play it seems reasonable to suggest that with the number of novel Y - STRs available, loci with more suitable properties can be selected. Some of the limitations of the minimal haplotype are illustrated in preliminary studies conducted at the University of the Western Cape (UWC) (Leat et al., 2004).A preliminary study was carried out at UWC to determine the allele and haplotype frequencies for the loci of the minimal haplotype in communities living in the

Cape Town metropolitan area. Samples were collected from 100 English-speaking Caucasian males and 99 Xhosa males. The samples were typed for the loci of the minimal haplotype and allele frequencies and gene diversities values were calculated for each locus. Gene diversity provides a simple indication of the polymorphic content of a locus and is calculated as $1 - \Sigma P_i^2$, where P_i is the allele frequency. Gene diversity values can range between zero and one with higher values reflecting higher polymorphism.

Gene diversity values for the Caucasian English-speaking sample ranged from 0.33 for DYS393 to 0.82 for DYS385. All of the loci studied could potentially be used in forensic casework in this particular subpopulation. By contrast two of the loci, DYS391 and DYS392 were found to have low levels of polymorphism in the Xhosa sample. Similar results were observed for an indigenous population sampled in Maputo, Mozambique (Álves et al., 2003). For DYS391, 93% of the Xhosa and 82% of the Mozambican sample shared the same allele. For DYS392, 96% of the Xhosa and 99% of the Mozambican sample shared the same allele. Low levels of polymorphism were also observed for DYS392 in a population of Central Africa Pygmies where over 96% of the individuals sampled shared the same allele (Kayser et al., 2001). These data suggest that DYS391 and DYS392 may not be ideally suited to forensic casework in several sub-Saharan populations.

Given that several of the loci of the minimal haplotype seemed to be of limited value in discriminating between the Xhosa individuals sampled, an attempt was made to assess the value of novel Y-STR loci. Our approach was similar to that recently reported by Kayer et al. (2004). Available Y-chromosomal sequence was surveyed for STR sequences leading to the identification of 232 trinucleotide, 437 tetranucleotide and 118 penta-nucleotide STR sequences. While many of these loci had already been submitted to the Genome Database (http://www.gdb.org) little data was available on their suitability for forensic casework. These sequences were ranked according to the degree of homology between repeated units and the number of repeat units present. Unlabeled PCR primers were designed for twenty five loci. Male specificity, copy number and polymorphism were assessed for 46 individuals using polyacrylamide gel electrophoresis and silver staining. On the basis of their genetic diversity thirteen loci were selected for further analysis using dye-labeled PCR primers combined into two multiplex PCR reactions. Over 100 Caucasian and 75 Xhosa samples were typed. Gene diversity values, the number of alleles identified and the average stutter was determined for each locus.

This has resulted in the identification of several loci which appear to be highly polymorphic in the subpopulations we have surveyed. We are currently in the process of assessing the polymorphism of these markers in a number of other South African population groups with a view to developing a multiplex PCR systemthat can be used by the police in forensic casework.

REFERENCES

Alves C, Gusmao L, Barbosa J, Amorim A (2003). Evaluating the informative power of Y-STRs: a comparative study using European and new African haplotype data. Forensic Sci. Int. 134: 126-33.

Ayub Q, Mohyuddin A, Qamar R, Mazhar K, Zerjal T, Mehdi SQ, Tyler-Smith C (2000). Identification and characterization of novel human Y-chromosomal microsatellites from sequence database information. Nucl. Acids Res. 28: e8.

Bosch E, Lee A, Calafell F, Arroyo E, Henneman P, de Knijff P, Jobling M (2002). High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. Forensic Sci. Int. 125: 42-51.

Butler J, Schoske R, Vallone P, Kline M, Redd A, Hammer M (2002). A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. Forensic Sci. Int. 129: 10-24.

Hall A, Ballantyne J (2003) . Strategies for the design and assessment of Y-short tandem repeat multiplexes for forensic use. Forensic Sci. Rev. 15:139-149.

Hanson E, Ballantyne J (2004). A highly discriminating 21 locus Y-STR "megaplex" system designed to augment the minimal haplotype loci for forensic casework. J. Forensic Sci. 49: 40-51.

Henegariu O, Heerema N, Dlouhy S, Vance G, Vogt P (1997). Multiplex PCR: critical parameters and step-by-step protocol. Biotechnology 23: 504-511.

Iida R, Tsubota E, Matsuki T (2001). Identification and characterization of two novel human polymorphic STRs on the Y chromosome. Int. J Legal Med. 115: 54-56.

Iida R, Tsubota E, Sawazaki K, Masuyama M, Matsuki T, Yasuda T, Kishi K (2002). Characterization and haplotype analysis of the polymorphic Y-STRs DYS443, DYS444 and DYS445 in a Japanese population. Int. J. Legal Med. 116: 191-194.

Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Roewer L (1997). Evaluation of Y-chromosomal STRs: a multicenter study. Int. J. Legal Med. 110: 125-33, 141-9.

Kayser M, Kittler R, Erler A, Hedman M, Lee A, Mohyuddin A, Mehdi S, Rosser Z, Stoneking M, Jobling M, Sajantila A, Tyler-Smith C (2004). A comprehensive survey of the human Y-chromosomal microsatellites. Am. J. Hum. Genet. 74: 1183-1197.

Kayser M, Krawczak M, Excoffier L, Dieltjes P, Corach D, Pascali V, Gehrig C, Bernini LF, Jespersen J, Bakker E, Roewer L, de Knijff P (2001). An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am. J. Hum. Genet. 68: 990-1018.

Leat N, Benjeddou M, Davison S (2004). Nine-locus Y-chromosome STR profiling of Caucasian and Xhosa populations from Cape Town South Afr. Forensic Sci. Int. 144:73-5.

Quintana-Murci L, Krausz C, McElreavey K (2001). The human Y chromosome: function, evolution and disease. Forensic Sci. Int. 118: 169-181.

Redd A, Agellon A, Kearney V, Contreras V, Karafet T, Park H, de Knijff P, Butler J, Hammer M (2002). Forensic value of 14 novel STRs on the human Y chromosome. Forensic Sci. Int. 130: 97-111.

Schoske R, Vallone P,, Ruitberg C, Butler J (2003). Multiplex PCR design strategy used for the simultaneous amplification of 10Y-chromosome short tandem repeat (STR) loci. Anal. Bioanal. Chem. 375: 333-343.

Tilford C, Kuroda-Kawaguchi T, Skaletsky H, Rozen S, Brown L, Rosenberg M, McPherson J, Wylle K, Sekhon M, Kucaba T, Waterston R, Page D (2001). A physical map of the human Y chromosome. Nat. 409: 943-945.

Wallin J, Holt C, Lazaruk K, Nguyen T, Walsh P (2002). Constructing universal multiplex PCR systems for comparative genotyping. J. Forensic Sci. 47: 52-65.

White P, Tatum O, Deaven L, Longmire J (1999). New male-specific microsatellite markers from the human Y chromosome. Genomics. 57: 433-437.

Zhao Z, Fu YX, Hewett-Emmett D, Boerwinkle E (2003). Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. Gene. 312: 207-13.