Person identification using DNA

Mark van Cuijk

July 14, 2009

Abstract

twenty-five years after the discovery was done that DNA could be used for identification purposes, a lot has changed in the DNA identification area. Up to today, these technologies have mostly been used in forensic applications. This paper shows two identification methods known today (Short Tandem Repeats and Single-Nucleotide Polymorphisms) and discusses the performance of DNA-based identification. Finally, the paper contains a discussion about the question whether DNA-based identification can be used in an automatic deployment in the near future.

1 Introduction

The science of using DNA (deoxyribonucleic acid) for identifying persons started in September 1984, when Sir Alec Jeffreys unexpectedly noticed similarities and differences in the DNA of his technician, compared to her family [8].

In the twenty-five years that followed, a vast amount of research has been done in the area of DNA-based person identification. Although the most widespread identification methods in use in forensic applications today are based upon the discovery of Sir Alex Jeffreys, new identifying properties are still discovered and new methods are created.

This paper focusses on two methods well-known today: identification based on Short Tandem Repeats and Single-Nucleotide Polymorphisms. The paper briefly explains the methods to distinguish individuals using these STR and SNP properties and discusses the performance of STR-based identification.

1.1 DNA

To allow readers with a computer science background to understand the remainder of this paper, this section provides a brief and simplified introduction to DNA and describes the key concepts that are relevant for the understanding of this paper. DNA — as it occures in organic cells — consists of two strands that spiral around each other in the shape of a double helix and that are build up from structural units, called nucleotides. A nucleotide contains a base, which can be one out of four types: adenine, cytosine, guanine or thymine, abbreviated as A, C, G and T, respectively. A base has a 3' and a 5' side that can be connected together; this way a form of directionality exists in a DNA strand. The two strands in DNA are structurally positioned in an opposed direction and are hold together by hydrogen bonds between the bases, forming a base pair. Figure 1 shows a graphical representation of a section of two DNA strands with twelve base pairs.

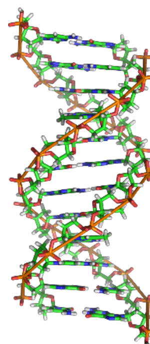

Figure 1: Two DNA strands in a double helix form.

A base pair is always created with complementary bases on both strands: adenine on one strand pairs with thymine on the other and cytosine pairs with guanine. The result of this is that all information that is encoded on a DNA strands has a complementary copy on the other strand. The hydrogen bonds between the two bases can be broken using a mechanical force or high temperature and be rejoined later on. This important feature of DNA allows easy replication, a process that creates a copy of the DNA, such that a cell can divide itself to allow an organism to grow. The double helix can be unfolded, while breaking the two strands apart. An enzyme called DNA polymerase reconstructs the complementary DNA sequence along each strand,

which is then bound to the original strand and refolded into a double helix.

The string of bases that are connected together in a strand is called a DNA sequence. A single strand can have a length of several millions of bases; the position of interesting bases in the sequence — such as the location of a specific gene, like the one that determines the color of the iris of a person — is called a locus. The process of determining the locus for a particular biological trait is called gene mapping, which is out of the scope of this paper. In current DNA-based identification methods, loci in non-coding regions of the DNA strand are used; regions that are made up of bases that do not affect the functioning or appearance of an organism.

1.2 Overview

After this introduction, section 2 describes the PCR method used to amplify the DNA in such a way that a measurement can be made from an extremely small amount of DNA sample, followed by section 3 on Short Tandem Repeats and Single-Nucleotide Polymorphisms, two concepts that allow DNA sequences of individuals to be distinguished. Section 4 contains a discussion about the performance of DNA-based identification and compares it to the performance of other biometric modalities. A discussion on the question whether DNA will be an automatic procedure in the near feature is given in section 5.

2 PCR

As described in section 1.1, DNA has the natural ability to be easily replicated, which is in fact an important feature that is required during the cell division process that occurs in every organism to grow. This section describes how this feature is exploited to amplify a sample of DNA using a technique called Polymerase Chain Reaction.

During replication, the two complementary DNA strands are separated and a new strand is constructed along both of them, resulting in two new strands that come into existence. The unbounded DNA strand is surrounded by free nucleosides, molecules that have a structure similar to nucleotides, but have additional phosphate groups. When a nucleoside molecule touches a nucleotide in the DNA strand, hydrogen bonds are automatically created if the the bases are complementary. This results in one long DNA strand with small nucleosides attached to each base.

The nucleosides are connected together by an enzyme called polymerase [2]. The enzyme removes the two additional phosphate groups that a nucleoside has and links the incoming base with the predecessor in the chain. Because polymerase is only able to attach a nucleoside to an already bound nucleotide at the 3' side, a primer that has a 3' end at one side is attached to the DNA strand at the position where replication is to start.

PCR (Polymerase Chain Reaction) [10, 16] is a technique that artificially initiates the replication process in order to amplify a specific piece of DNA in a sample. To perform PCR on a specific locus, the sequence that begins and ends the interesting region must be identified and the complementary of these sequences must be encoded into oligonucleotide primers, such that these can engage hydrogen bonds with the template DNA later on. An oligonucleotide is a short DNA sequence, typically with twenty or fewer bases.

PCR is performed by performing several iterations of a cycle, where each cycle doubles the amount of DNA in the sample, as can be read from table 1. Each cycle consists of three phases:

Denaturation phase For the primers and polymerase to interact with the DNA strands, they must not be joined together. During the denaturation phase, the temperature is set to 94–96°C for a period of 20–30 seconds, such that the hydrogen bonds are broken.

Annealing phase After denaturation, the temperature is lowered to 50–65°C, depending on the primer being used, such that the oligonucleotide primer and polymerase attach to the strand.

Elongation phase During the elongation phase, the temperature is set to optimum activity temperature of the polymerase being used. The polymerase constructs a new strand, linking the newly attached complementary nucleosides together.

Cycle	Templates	Long Products	Short Sequences
0	1	-	-
1	1	1	0
2	1	2	1
3	1	3	4
5	1	5	26
10	1	10	1013
15	1	15	32,752
20	1	20	1,048,555
n	1	n	$(2^n - n - 1)$

Table 1: Number of double strands after 0 to n cycles. This table has been copied from [10].

At the start of the first cycle there are only the two template strands. The two oligonucleotide primers attach to the point where the bases are complementary on the template; one primer on each of the two strands. During the elongation phase, a new strand is constructed until the temperature is raised for the next denaturation step that breaks the hydrogen bonds. At that point there are four strands: the two template strands and the two newly constructed strands, which are called long products. The length of the long products depend on the duration of the elongation phase, the speed of the polymerase and the availability of nucleosides in the primer.

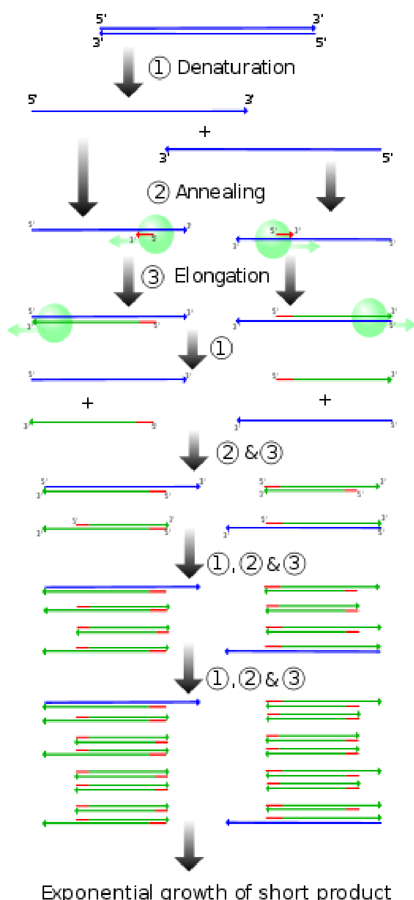

Figure 2: Schematic drawing of four PCR cycles. Blue lines: DNA template; red lines: oligonucleotide primers; green lines: DNA products. The green bubble represents the polymerase. Drawing is taken from [26].

During the next cycle, four oligonucleotide primers are used: two are attached to the templates — as happend in the cycle before — and two are attached to the long products. Along the templates, two new long products are constructed. However, the construction of the strands along the long products stop when the end of the strand is reached — which is at the position of the primer — and these two new strands are called short sequences. During the next cycles, new long products are constructed along the templates and new short sequences are constructed along each long product, as well as along each short sequence. Figure 2 shows a schematic drawing of the cycles.

3 Identifying properties

In 1980, Wyman and White found a locus in human DNA that has a basepair length that greatly differs among different persons [27]. The same method as used during their discovery has been applied in [17] using oligonucleotide probes. The fact that this region differs in persons is called polymorphism; a highly polymorphic locus is a region where many different sequences are possible. The technique for dividing persons in classes based on the length of such regions is called Restriction Fragment Length Polymorphism (RFLP). When the robustness and sensitivity of these techniques had been improved, the FBI implemented them in 1988 to exonerate one suspect in two rape homocides of young girls [14].

The RFLP involves a slow and cumbersome process and requires large amounts of undegraded sample DNA to be available. New identification methods became available, partly due to the results of the Human Genome Project [23]. These methods are, among other principles, mainly based on Variable Number Tandem Repeats (VNTR, see section 3.1 on STR) and Single-Nucleotide Polymorphisms (SNPs, see section 3.2).

3.1 STR

Human DNA contains regions that comprise several thousand base pairs that are not genes and thus do not contain information that is actively used for the growth of the body. When such a region is made up of a large number of tandomly repeated units, this is called a Variable Number Tandem Repeat (VNTR). The size of a single unit is nearly constant among persons, but a high variability in the number of these units that are tandamly repeated can be found. This is the kind of polymorphism that allows VNTRs to be useful in distinguishing individuals.

Short Tandem Repeats (STRs) are a specific subgroup of VNTRs, so the general principles for using them are the same. However, the repeating unit of STRs are small — defined to be 2 to 7 bases

in length [14] — and the total length of an STR sample is usually less than 500 bases. This small size allows PCR (see section 2) to be applied on a DNA sample, such that very small amounts of DNA can still be used for identification purposes.

The general procedure for creating a profile from a DNA sample is to amplify it using PCR and then place it on a flat gel, which is exposed to an electric field. The speed at which the fragment migrate through the gel depends on the size of the fragments, such that fragments with a small number of repeated units will travel a large distance. The fragments in the gel are transferred to a nylon membrane, which is then flooded with a probe. A probe is a single strand of DNA complementary to tandomly repeated unit, such that hydrogen bonds are formed with the fragments on the membrane. Radioactive atoms are attached to the probe, such that the locations of the fragments will be highlighted on a photographic film that is put in contact with the membrane. An example of a picture that appears on such a film can be found in figure 3.

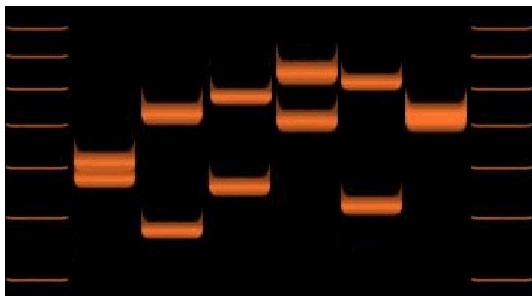

Figure 3: Picture of the photographic film as DNA fragments highlight regions on the film. This film was attached to a membrane with samples from six individuals, such that each sample travels over its own lane. Picture is taken from [24].

Identification using STR is often used in forensic settings. To aid linking individuals to crime scenes, countries maintain databases with DNA profiles of convicted fellons and DNA samples found at crime scenes. In the United States, the nationwide database is called CODIS and is maintained by the Federal Bureau of Investigation. Thirteen loci are identified as the core loci that are stored with each profile in the CODIS database; a list of these loci can be found in table 1 of [15].

Interpol maintains a similar database for cross-nation querying in the European area and they have identified a subset of six loci as minimum input to the database. This set is called the Interpol Standard Set of Loci (ISSOL) and can be found in [18].

3.2 SNP

Certain regions in human DNA contain sequences that largely match between individuals, but have a difference only in a single nucleotide (see figure 4); this situation is called a Single-Nucleotide Polymorphism. Each SNP can split a human population in two disjoint groups: the larger group where individuals with the more common nucleotide are placed and the smaller group where individuals with the more rare nucleotide are placed. Therefore, by adding a new SNP to the comparison, the number of groups (or classes) can be doubled and since thousands single-nucleotide polymorphisms have been identified in human DNA [6] this will exponentially leads to a situation with a large number of classes that each contain only a small number of individuals.

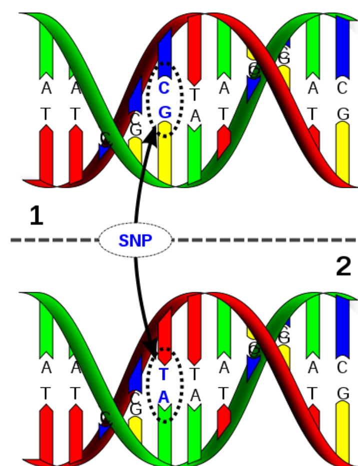

Figure 4: Two double-stranded DNA fragments displaying a polymorphic nucleotide. Picture is taken from [25].

In 1991, SNPs were found in the control region of mitochondrial DNA and a detection process based on PCR (see section 2) and allele-specific matching was described in [19]. The described method has been the basis for several techniques that were used during the 1990s [20, 5, 22]. Several drawbacks of these methods have been solved by [6] in 1999.

In general, the main advantages of SNPs are the numerous amount of markers in the human genome, the many methods of SNP detection that are available and the relative ease of simultaneously amplifying several regions [14]. Although there are few disadvantages, the most important one is the fact that the individual regions are of limited discriminatory value, so a large amount of SNPs are needed to identify an individual in a large population.

4 Performance

The performance of DNA identification largely depends on the number of compared loci. Since every person has only a single DNA sequence, in general it can be stated that different profiles belong to different persons. However, it must be noted that:

- a single person might posses two distinct DNA sequences, an extremely rare disorder called chimerism, and
- mutations in the DNA sequence of a person can occur, resulting in slight variations of DNA sequence in different cells in the body.

Only 30 cases of chimerism are known worldwide, so it is assumed safe to ignore the disorder. The region where a mutation of a DNA sequence occurs can be considered random and therefore — given the extremely small fraction of DNA that is used for identification purposes — have only a very limited effect on the constructed profile [14]. It is therefore often claimed that the False Non-Match Ratio (FNMR) using DNA identification is negligable, such that the False Reject Rate (FRR) is set equal to the error rate of the acquisition and analysis process. In individual cases this has resulted in very awkward situations [1].

When assessing the False Match Rate (FMR), one has to determine the propability that the DNA sequence from two distinct persons result in an identical profile. Several studies have been performed and most of them assume:

- that the used loci are highly polymorphic,
- that no correlation between the used loci exists, and
- a randomly mating population, causing chromosomes to be uniformly distributed among individuals.

The first two assumptions are tested for the 13 CODIS STR core loci (see section 3.1) using population data from African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans and Trinidadians [3]. The test confirmed that the CODIS core loci are highly polymorphic — thus giving the set a large discriminating value — and that no significant correlation between these loci exist. The lack of correlation allows the match probability of all loci to be combined using the product rule.

However, the third assumption is false in general. The majority of humans have parents that share the same race and live in the same geographical region. However, it is shown that the significance of this fact is low, as actual values do not significantly deviate from expected values that can be expected when the assumption holds [3, 9].

For a relatively small dataset, [3] gives the discrimination probability ($P_{i,d}$, $i \in 1..13$) of each of the 13 CODIS loci for several populations. $P_{i,d}$ is the probability that two randomly selected profiles from the dataset can be distinguished by looking at locus i. Therefore, the probability that two randomly selected profiles do match at a locus i is $P_{i,m} = 1 - P_{i,d}$. Under the assumption of non-correlation, the product rule can be applied to determine the probability that two randomly selected distinct profiles from a dataset as $\prod_{i=1}^{13} P_{i,m}$.

Using this method on the data stored in CODIS, table A2 in [14] reveals that the random match probability for Caucasian Americans is 1.738×10^{-15} and for African Americans is 1.092×10^{-15}. For other populations, probabilities are of the same order of magnitude. Taking these numbers as the FMR, they are compared to the performance of other biometric modalities in section 4.1.

Unfortunately, I have been unable to find supportive information to assess the performance of DNA identification based on Single-Nucleotide Polymorphisms (see section 3.2).

The performance numbers for DNA identification are very promising. There are however some remarks that must be made. First of all, it is known that monozygotic (or identical) twins have nearly identical DNA [4]. Monozygotic twins develop when a single egg divides into two separate embryos after fertilization and make up roughly 0.2% of the world population, which makes it an important group when considering any type of DNA-based identification.

Also an interesting issue results from the fact that DNA is inherited from both parents. Each cell in a human contains 46 chromosomes; 23 of which are inherited from the mother, while the other 23 are inherited from the father. The result is that a high degree of resemblance exists between DNA profiles of relatives, and in particular siblings. In siblings, the match probability always involves a factor of 1/4 per locus. However, the current 13 CODIS STR loci are enough to distinguish individuals among relatives and siblings [14].

The fact that the highest degree of resemblance exists between siblings is used in a conservative method to provide an estimate whether a DNA sample belongs to a certain individual (the Sibs methods [14]). The approximate conditional match probability for siblings can be computed and the method then suggests that one can be confident

that the match probability is smaller for any less related individual.

4.1 Comparison with other biometric modalities

Table 2 displays a qualitative comparison among biometric modalities, as presented in [7]. It very clearly shows how DNA can be an interesting biometric modality, given the fact that every person in the world has a DNA profile, these profiles are sufficiently distinctive among different persons and the fact that — despite genomic mutation — these profiles almost do not change during the lifetime of a person. However, collecting DNA samples still is a cumbersome process — compared to e.g. making a photograph of a face — and people may not like the collection process.

Modality	1	2	3	4	5	6	7
DNA	H	H	H	L	H	L	L
Ear	M	M	H	M	M	H	M
Face	H	L	M	H	L	H	H
Facial thermo.	H	H	L	H	M	H	L
Fingerprint	M	H	H	M	H	M	M
Gait	M	L	L	H	L	H	M
Hand geometry	M	M	M	H	M	M	M
Hand vein	M	M	M	M	M	M	L
Iris	H	H	H	M	H	L	L
Keystroke	L	L	L	M	L	M	M
Odor	H	H	H	M	H	M	M
Palmprint	M	H	H	M	H	M	M
Retina	H	H	M	L	H	L	L
Signature	L	L	L	H	L	H	H
Voice	M	L	L	M	L	H	H

Table 2: Qualitative comparison of various biometric technologies based on the perception of the authors of [7]. High, medium and low are denoted by H, M and L, respectively. Colums 1–7 represent: universality (1), distinctiveness (2), permanence (3), collectability (4), performance (5), acceptability (6) and circumvention (7).

Using [12], I've tried to make a comparison between the results obtained in section 4 about the performance of DNA identification techniques and the results found during the FRVT 2006 test about the performance of biometric identification based on facial images and iris scanning technology. It is important to stress that the results from FRVT 2006 are based on practical tests, while the available information about DNA identification is not expressed in numbers of False Accept Rate (FAR) and False Reject Rate (FRR). Table 3 shows numbers from the highest test results from RFVT 2006 and the numbers found in section 4.

Modality	FAR	FRR
DNA	1.0×10^{-15}	negligible
	1.8×10^{-15}	negligible
Iris	1.0×10^{-3}	1.1×10^{-2}
	1.0×10^{-3}	1.4×10^{-2}
Facial	1.0×10^{-3}	8×10^{-3}
	1.0×10^{-3}	1.6×10^{-2}

Table 3: Quantitative comparison of various biometric modalities. The values for DNA are extracted from section 4, the other values are taken from [12]. The two numbers give the upper and lower bounds of the approximate range of values. Note that the values for DNA are not FAR and FRR values from a practical test, but theoretical FMR and FNMR values. Monozygotic twins have been ignored.

It can however be very clear — when judging the performance of DNA-based identification against the FMR/FNMR chart given in figure 3 of [13] — that DNA identification can be positioned at the top left corner of the diagram. In fact, practice shows us that indeed forensic applications are the area where DNA identification technology is mostly deployed and forensics is probably the only application area where DNA technology will have a widespread use in the near future.

5 Automated procedure?

To answer the question whether DNA-based identification technologies may be applied automatically in the near future, it's good to look at what exactly is meant by automatic. A distinction can be made between a fully automatic — unattended or unsupervised — use of a device to perform person identification or a less strict definition, such as allowing the acquisition to be performed by a person, but the analysis be performed automatically by an electronic device.

In fact, when using the latter definition, this is currently possible and is actually the way it happens in some forensic laboratories. A DNA sample is collected by a crime scene investigator and transferred to a laboratory. The sample is then placed in a device that automatically performs the PCR amplification, followed by an analysis step. Modern techniques allow a profile to be constructed within minutes [21].

However, a fully automatic deployment of DNA identification technology, e.g. for access control

purposes, will not be feasible in the near future. Although the performance numbers of the technology look very promising, currently available technology is still unable to distinguish monozygotic twins. Today, DNA identification technology could at best be combined in a multi-model setting with another biometric modality, e.g. fingerprint, to be able to make a clear distinction between twins. However, a current research topic is to investigate the genotypical difference between homozygotic twins [4], so in the future this problem might be solved.

Also, the analysis of a DNA sample still takes minutes to complete, prohibiting deployment of DNA-based identification in environments with requirements for high throughput, like airports or most access control environments. As technology advances, this analysis time might decrease, but it is not clear when throughput figures like for fingerprint matching might be achieved.

Another problem with unsupervised DNA identification is that it can be very easily circumvented. It is not very difficult to obtain small amounts of DNA from an unsuspecting individual. Even in a controlled environment a DNA test had been circumvented: in 1992, John Schneeberger placed a plastic tube with the blood of another man in his arm, such that a comparison of his DNA would not match the DNA samples of the alleged rapist's semen [11].

Given the facts stated in this section, I'd say that — although it might be possible that DNA identification technology will be used more widely in the future — in the near future DNA identification will only be limited to forensic applications and not be fully automatic.

6 Conclusion

A lot has changed in the twenty-five years after Sir Alec Jeffreys discovered that DNA can be used for identification purposes. DNA identification technology has long been used in the forensic area, where it is still used extensively today. This paper discussed two DNA-based identification technologies: Short Tandem Repeats (see section 3.1) and Single-Nucleotide Polymorphisms (see section 3.1).

The performance of STR-based identification techniques have been subject to research and — compared to other biometric modalities — DNA has very promising performance numbers; the False Match Ratio is in the order of 10^{-15}. However, at this moment the technology lacks the ability to distinguish monozygotic twins, but research is going on to find new methods to solve this issue.

In the discussion about whether DNA-based identification can be automated in the near future, two distinct questions have been answered. It has been shown that current technology allows for the automatic generation of a DNA profile, based on a small sample, while the acquisition of this sample must still be performed by a human. The discussion also concluded in the fact that unattended or unsupervised DNA-based identification won't be fully automated in the near future, largely because of throughput of the currently available identification devices and the ease of circumventing the system by presenting DNA from a different individual.

References

[1] Claire Ainsworth. The stranger within. `http://www.newscientist.com/article/mg18024215.100-the-stranger-within.html?full=true`, November 2003.

[2] Jeremy M. Berg, John L. Tymoczko, and Lubert Stryer. *Biochemistry (Chapters 1-34)*. W. H. Freeman, 5th edition, 2002.

[3] Bruce Budowle, Tamyra R. Moretti, Anne L. Baumstark, Debra A. Defenbaugh, and Kathleen M. Keys. Population data on the thirteen codis core short tandem repeat loci in african americans, u.s. caucasians, hispanics, bahamians, jamaicans, and trinidadians. *J. Forensic Sci.*, 44:1277–1286, 1999.

[4] Carl E.G. Bruder et. al. Phenotypically concordant and discordant monozygotic twins display different dna copy-number-variation profiles. *Am. J. Hum. Genet.*, 82:761–771, 2008.

[5] Christian A. Heid, Junko Stevens, Kenneth J. Livak, and P. Mickey Williams. Real time quantitative pcr. *Genome Methods*, 6:986–994, 1996.

[6] W. Mathias Howell, Magnus Jobs, Ulf Gyllensten, and Anthony J. Brookes. Dynamic allele-specific hybridization. *Nature Biotechnology*, 17:87–88, 1999.

[7] Anil K. Jain, Arun Ross, and Salil Prabhakar. An introduction to biometric recognition. *IEEE Security and Privacy*, 14:4–20, 2004.

[8] Alec J. Jeffreys, Victoria Wilson, and Swee lay Thein. Hypervariable 'minisatellite' regions in human dna. *Nature*, 314:67–73, 1995.

[9] A. M. Lins, K. A. Micka, C. J. Sprecher, J. A. Taylor, J. W. Bacher, D. R. Rabbach, R. A.

Bever, S. D. Creacy, and J. W. Schumm. Development and population study of an eight-locus short tandem repeat (str) multiplex system. *J. Forensic Sci.*, 43:1168–1180, 1998.

[10] Kary B. Mullis, Henry A. Erlich, Norman Arnheim, Glenn T. Horn, Randall K. Saiki, and Stephen J. Scharf. Process for amplifying, detecting, and/or cloning nucleic acid sequences. US Patent Number 4800159, Jan 1989.

[11] CBC News. Sask. doctor sentenced for rape. `http://www.cbc.ca/canada/story/1999/11/26/saskdr991126.html`, November 2000.

[12] P. Jonathon Phillips, W. Todd Scruggs, Alice J. O'Toole, Patrick J. Flynn, Kevin W. Bowyer, Cathy L. Schott, and Matthew Sharpe. Frvt 2006 and ice 2006 large-scale results, March 2007.

[13] Salil Prabhakar, Sharath Pankanti, and Anil K. Jain. Biometric recognition: Security and privacy concerns. *IEEE Security and Privacy*, 1:33–42, 2003.

[14] Janet Reno, Daniel Marcus, Mary Lou Leary, Julie E. Samuels, Julie E. Samuels, and Christopher Asplen. Forensic dna testing, 2000.

[15] Ugo Ricci, Ilaria Sani, Michael Klintschar, Nicoletta Cerri, Francesco De Ferrari, Maria Luisa, and Giovannucci Uzielli. Forensic sciences. identification of forensic samples by using an infrared-based automatic dna sequencer. *Croation Medical Journal*, 44(3):299–305, 2003.

[16] J.J. Rossi, R. Kierzek, T. Huang, P.A. Walker, and K. Itakura. An alternate method for synthesis of double-stranded dna segments. *J. Biol. Chem*, 257:9226–9229, 1982.

[17] Randall K. Saiki, Stephen Scharf, Fred Faloona, Kary B. Mullis, Glenn T. Horn, Henry A. Erlich, and Norman Arnheim. Enzymatic amplification of -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, 230(4732):1350–1354, 1985.

[18] Werner Schuller, Lyn Fereday, and Richard Scheithauer. Interpol handbook on dna data exchange and practice, 2001.

[19] Mark Stoneking, Dennis Hedgecock, Russel G. Higuchi, Linda Vigilant, and Henry A. Erlich. Population variation of human mtdna control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am. J. Hum. Genet.*, 48:370–382, 1991.

[20] Ann-Christine Syvänen. Solid-phase minisequencing as a tool to detect dna polymorphism. *Methods in Moolecular Biology*, 98:291–298, 1998.

[21] F. R. R. Teles and L. P. Fonseca. Trends in dna biosensors. *Talanta*, 22:606–623, 2008.

[22] Sanjay Tyagi and Fred Russel Kramer. Molecular beacons: Probes that fluoresce upon hybridization. *Nature Biotechnology*, 14:303–308, 1996.

[23] Office of Science U.S. Department of Energy. Human genome project information. `http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml`.

[24] Wikipedia. File:d1s80demo.gif. `http://en.wikipedia.org/wiki/File:D1S80Demo.gif`.

[25] Wikipedia. File:dna-snp.svg. `http://en.wikipedia.org/wiki/File:Dna-SNP.svg`.

[26] Wikipedia. File:pcr.svg. `http://en.wikipedia.org/wiki/File:PCR.svg`.

[27] Arlene R. Wyman and Ray White. A highly polymorphic locus in human dna. *Proc. Natl. Acad. Sci. USA*, 77:6754–6758, Nov 1980.