Tohoku J. Exp. Med., 2004, **204**, 109-117

Development of Biometric DNA Ink for Authentication Security

MASAKI HASHIYADA

Division of Forensic Medicine, Department of Public Health and Forensic Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan

HASHIYADA, M. *Development of Biometric DNA Ink for Authentication Security.* Tohoku J. Exp. Med., 2004, **204** (2), 109-117 —— Among the various types of biometric personal identification systems, DNA provides the most reliable personal identification. It is intrinsically digital and unchangeable while the person is alive, and even after his/her death. Increasing the number of DNA loci examined can enhance the power of discrimination. This report describes the development of DNA ink, which contains synthetic DNA mixed with printing inks. Single-stranded DNA fragments encoding a personalized set of short tandem repeats (STR) were synthesized. The sequence was defined as follows. First, a decimal DNA personal identification (DNA-ID) was established based on the number of STRs in the locus. Next, this DNA-ID was encrypted using a binary, 160-bit algorithm, using a hashing function to protect privacy. Since this function is irreversible, no one can recover the original information from the encrypted code. Finally, the bit series generated above is transformed into base sequences, and double-stranded DNA fragments are amplified by the polymerase chain reaction (PCR) to protect against physical attacks. Synthesized DNA was detected successfully after samples printed in DNA ink were subjected to several resistance tests used to assess the stability of printing inks. Endurance test results showed that this DNA ink would be suitable for practical use as a printing ink and was resistant to 40 hours of ultraviolet exposure, performance commensurate with that of photogravure ink. ——— biometrics personal authentication; DNA personal identification; DNA ink; short tandem repeat (STR); synthetic DNA

Received April 28, 2004; revision accepted for publication July 20, 2004.

Address for reprints: Masaki Hashiyada, Division of Forensic Medicine, Department of Public Health and Forensic Medicine, Tohoku University Graduate School of Medicine, Sendai 980-8575, Japan.

e-mail: hashiyad@forensic.med.tohoku.ac.jp

A short tandem repeat (STR) is a series of repeated sequences, each 2 to 6 bases long. It is one of the most popular markers for forensic analysis. Our forensic laboratory uses 17 STR loci for paternity tests. Japanese police formally introduced 10 STR loci for crime investigation in August 2003. STRs have 6 to 19 alleles per locus (Hashiyada et al. 2003). Clearly, the analysis of more STR loci lowers the average matching probability.

Because of the high discrimination power provided by STRs, it is possible to use them for biometric personal authentication systems. Many kinds of identity certificates, such as passports, are required during one's lifetime for registered bonds and many types of licenses. However, it is difficult to completely protect these printed materials from being copied. As previously reported (Itakura 2002), DNA information using STR loci is one of the most reliable personal identification markers. If personal DNA information were added into ink, the documents printed in this DNA ink would constitute a novel biometric personal authentication system.

The biggest problem in using STRs is the time between the extraction from materials and the evaluation of alleles. Five to six hours are needed to perform DNA typing using the polymerase chain reaction (PCR). Because of improvements in typing kits, the number of STR loci that can be analyzed at once has increased, but the detection time of each locus has not changed over the last couple of years. If DNA personal identification (DNA-ID) systems were to be utilized for verification of individuals in airports, banks, building security systems, or government organizations, the detection time needs to be dramatically shortened to several minutes. On the other hand, if a DNA-ID were to be used for authentication of rare or expensive goods, time would be unimportant. Therefore, DNA ink could be useful for some commercial applications, even given the current technology. Indeed, a company already advertises a DNA identification technique (http://www.dnatecaus.com/). However, the details of the analysis procedure are not yet available, and any method that uses an individual's DNA directly cannot guarantee the privacy of that individual.

In this report, I outline the development of biometric ink containing DNA whose sequence is based on personal STR information.

Materials amd Methods

STR analysis

Genomic DNA was extracted from a buccal swab using SDS-proteinase K and Chelex 100 (Walsh et al. 1991) treatment followed by phenol/chloroform extraction. We investigated 16 STR loci using the GenePrint® PowerPlex™ 16 System (Promega, Madison, WI, USA) (Promega Corporation 2000). Electrophoresis was carried out on an ABI 310 Genetic Analyzer and the alleles were determined using GeneScan™ 3.7 software (Applied Biosystems, Foster City, CA, USA).

This study was approved by the Ethical Committee of the Tohoku University School of Medicine (receipt number: 2000-103).

Generation of DNA personal ID

STR alleles are defined by the number of repeats they contain. A DNA personal ID α_X (DNA-ID) was generated that included allelic information about 15 of the 16 STR loci analyzed, not including amelogenin. The loci were incorporated in the following sequence, in which the pair of alleles at each locus was arranged in ascending order (Itakura 2002).

($L_I = p_i \| q_i, p_i \leqq q_i$)

α_X = Locus1, Locus2, Locus3, ……, Locus15

= CSF1P0, D13S317, D18S51, D21S11, D3S1358, D5S818, D7S820, D8S1179FGA, TH01, TPOX, vWA, PentaE, D16S539, Penta D

= $p_1 \| q_1 \| p_2 \| q_2 \| \ldots\ldots p_{15} \| q_{15}$

For example, Mr. M had the following alleles at the respective loci: 12/14, 8/11, 13/15, 29/32.2, 15/16, 11/11, 8/8, 12/15, 23/24, 6/7, 8/11, 16/17, 18/20, 12/12, 10/10. The α_X was thus defined as follows.

α_M =12148111315293221516111188121523246781116171820121210 10

When the STR number of an allele had a fractional component, such as allele 32.2 in D21S11, the decimal point was removed, and all of the numbers, including those after the decimal point, were retained.

Establishment of the identification format

Because the α_M contains direct personal STR information, it must be encrypted for privacy protection. This was achieved using a one-way function that also reduced the data length of the DNA-ID (Itakura 2002). This one-way function, the secure hash algorithm-1 (SHA-1), produced an ID with a data length δ_M of 160 bits, according to the following transformation:

$\delta_M = h(\alpha_M)$

DNA ink may be used for group IDs within organizations, as well as for identification of individuals. Half- or quarter-length data could also be considered to reduce costs, though at the expense of precision. In this study, an identifier for practical use that was a quarter of the original 160-bit length was defined as

$\delta_{Mi} = \delta_{M1} \mid \delta_{M2} \mid \delta_{M3} \mid \delta_{M4} \mid$

where δ_{M1}, δ_{M2}, δ_{M3} and δ_{M4}, refer to the identification ID consisting of the first, second, third and fourth 40 bits of δ_M. In this manner, the DNA ink could be used to identify 2^{40}, or about 400 000 products. Each set of 8 data bits is extended by two redundant bits known as the sift and check bits, which serve not only as check bits but also as limiting factors in the latter stages of DNA sequence generation. This limiting factor, a necessity in DNA sequence analysis, acts to exclude five or more repetitions of the same base.

δ_{M1} =1001100110011101011010100101111010 0010 (Extracted 40bits)

δ_{M1} =1001 1001 [10] 1001 1101 [00] 0110 1010 [01] 1001 0111 [00] 1010 0010 [11]

(Shift and check bits show as square brackets with underlines.)

Transformation of binary data into DNA sequence and production of symthetic DNA

Bit series as generated above are transformed into base sequences according to the follow scheme. We called this step the "Encoded Base Array" method.

00=A (adenine), 01=C (cytosine),
10=G (guanine), 11=T (thymine)

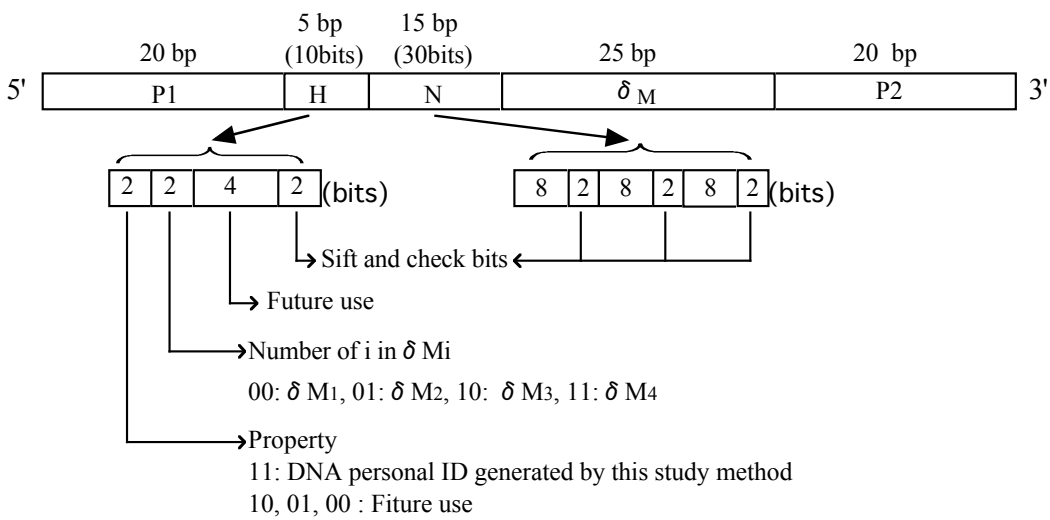


Fig. 1. Sequence structure of the 85-bp single-stranded DNA-ID
P1, P2: Primer sequences are designed so as to not anneal to the human genome.
H: Header
N: Serial number

δ_{M1} =1001 1001 [10] 1001 1101 [00] 0110
1010 [01] 1001 0111 [00] 1010 0010 [11]
=GC GC [G] GC TC [A] CG GG [C] GC
CT [A] GG AG [T]

The identification data format is defined by adding a header (H, 10bits) and a serial number (N, 30 bits) to δ_{M1} (40+10=50 bits). The resulting DNA sequence, consisting of H (5-bp), N (15-bp), and δ_{Mi} (25-bp) would then be flanked by two 20-bp-long primer sequences. This synthetic DNA could then be amplified by PCR, and only those who know the primer sequences would be able to analyze the intervening sequence. Fig. 1 shows the structure of the 85-bp synthetic DNA sequence.

Mixing of synthetic DNA with ink

Single-stranded DNA is more economical to produce than double-stranded DNA, but much less physically stable, so double-stranded PCR-amplified DNA was used for incorporation into the DNA ink. Three mg double-stranded DNA was mixed with 100 ml of ink, which was composed of a colorless transparent pigment to produce an invisible ink and an IR color former added for easy detection of the printed mark. In addition, dummy DNA was also added to make the DNA-ID sequence difficult to analyze by someone who does not know the primer sequences. The manufacturing process for the DNA ink is shown in Fig. 2.

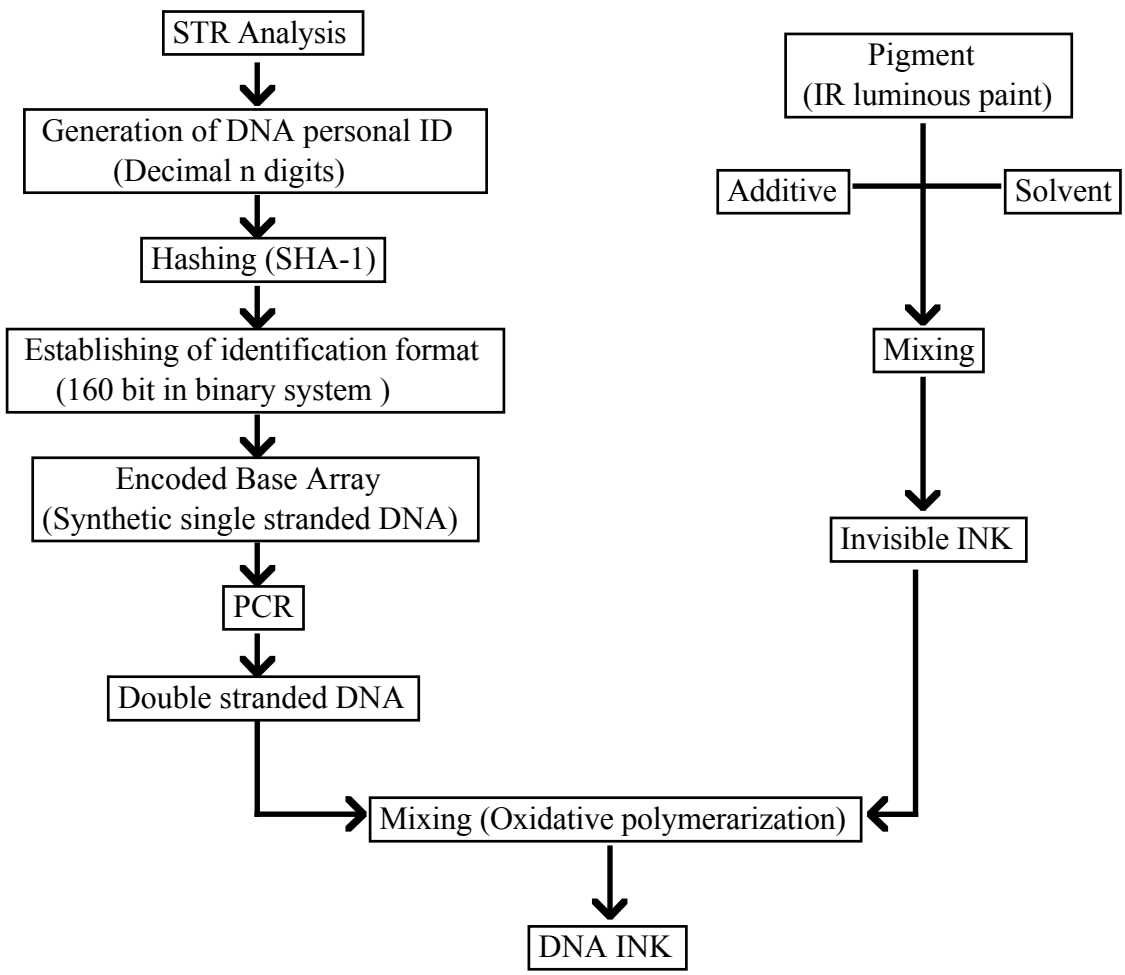


Fig. 2. Synthesis of DNA Ink

TABLE 1. *The list of endurance test used for analysis of printed DNA ink*

Category	Method	Time
Heat	150°C hot air drying machine	10 min.
Acids	20.25°C, 2% sulfuric acid	2 hrs
Alkalis	20.25°C, 1% sodium hydroxide	2 hrs
Paraffin	100.110°C, melting paraffin	10 min.
Soap	20.25°C, 10% soap gel	1 hr
Alcohol	20.25°C, 100% ethanol	24 hrs
UV (ZnO–)	Ci 35A Wether Meter (ATLAS)	1, 3, 6, 10, 20, 40, 60, 80 hrs
UV (ZnO+)	Ci 35A Wether Meter (ATLAS)	1, 3, 6, 10, 20, 40, 60, 80 hrs
UV and water (ZnO–)	Ci 35A Wether Meter (ATLAS), 18 min. shower after 102 min. UV	40, 60, 80 hrs
UV and water (ZnO+)	Ci 35A Wether Meter (ATLAS), 18 min. shower after 102 min. UV	40, 60, 80 hrs
Sunlight	Attached to the sunny window	1 day, 1 week, 1 month

min., minutes; hr(s), hour(s).

Endurance test of DNA ink

Several kinds of resistance tests were used to ascertain the durability of DNA ink for practical use (Table 1). Samples were prepared by full-tone printing on paper using DNA ink and an additional covering of zinc oxide (ZnO) on the surface to enhance resistance to ultraviolet (UV) light, which is thought to be the major case of DNA degradation.

Extraction of DNA. Specimens used for analysis were obtained by cutting 1 cm^2 chips out of the printed sheet after the resistance tests. Then, the specimens were cut into even smaller pieces and put in each 1.5-ml tube containing 500 μl toluene/ methyl isobutyl ketone (4 : 1) and TNE buffer. After mixing for 30 minutes, the aqueous layer was transferred to another tube. Total DNA thus isolated from the surface and interior of the paper was further extracted from this aqueous later by phenol/chloroform followed by ethanol precipitation and dissolved in 20 μl TE buffer.

Amplification of DNA. Two μl of each extract was used as a template for PCR. Amplification was carried out in a 25 μl total reaction volume containing 2.5 units of Taq DNA polymerase (HotStarTaq™ DNA polymerase, QUIAGEN, Hilden, Germany), 0.4 μM of each of the primers (P1, P2), 2.5 μl 10×PCR buffer, and 200 μM of each nucleotide (QUIAGEN). PCR cycling conditions were 95°C for 15 minutes followed by 35 cycles at 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 1 minute, and then 72°C for 10 minutes in a GeneAmp PCR system 9700 (Applied Biosystems). PCR products were separated in a 4% agarose gel to confirm amplification of the synthetic DNA fragment by size.

Cloning and sequencing. After PCR, the PCR products were purified using a MiniElute™ PCR Purification Kit (QUIAGEN) and subcloned using a QIAGEN PCR Cloning Kit (QUIAGEN). Plasmid DNA obtained from clones was prepared automatically by a MagExtractor MFX-2000 (TOYOBO, OSAKA). Sequencing of the plasmid DNA was performed using a BigDye Terminator ver.3.0 Cycle Sequencing Kit (Applied Biosystems) and analyzed using a 310 Genetic Analyzer (Applied Biosystems).

RESULTS

When used in the ink, single-stranded DNA failed to amplify after either a one-month expo-

sure to sunlight or exposure to acid. In other tests without exposure to sunlight, amplification was successful. Thus, double-stranded DNA was used to increase the durability of the biometric ink. The printed samples containing this DNA were covered with zinc mono-oxide (ZnO) and tested by exposure to the same UV conditions. As expected, durability improved and amplification was successful after 40 hours of exposure to UV, both alone and in conjunction with a water resistance test, in which a cycle comprising a 18-min. water shower and a 102-min. UV exposure was repeated 20 times to make up 40 hours of total UV exposure. These results showed that the DNA ink had similar UV durability to commonly used photogravure inks. Tables 2 and 3 show the results of the durability tests for single- and double-stranded DNA. All samples in which no DNA was detected were subjected to a second round of PCR using 1-μl aliquots of the first PCR product as template, but in no cases was any relevant DNA detected.

Sequence analysis is required to determine whether or not the amplified PCR products were truly derived from the DNA ink. In this study, direct sequencing of the amplified PCR products using primers designed to anneal to the flanking PCR primer sites was unsuccessful. The DNA fragments were so short that peaks corresponding to the nucleotide bases immediately downstream of the primers were hopelessly blurred; even when the sequence was analyzed in both the forward and reverse directions, unreadable bases remained. However, accurate sequence analysis was successfully achieved by subcloning the fragment into a carrier plasmid, raising them in bacterial cells, and isolating and sequencing the resulting amplified plasmids.

Discussion

Development of biometric authentication technologies has progressed rapidly in the last few years. Personal authentication devices based on unique patterns of fingerprints, the iris, or subcutaneous veins in the wrist have been commercialized for personal use, such as in the verification of a user logging in at a computer terminal. Face recognition, voice dynamics, and handwriting analysis are also being used in criminal investigations. All of these methods of verification

Table 2. *The results of a durability test performed without UV exposure*

	Single strand DNA	Double strand DNA
Heat	D	N.T.
Acid	N.D.	N.T.
Alkalis	D	N.T.
Paraffin	D	N.T.
Soap	D	N.T.
Alcohol	D	N.T.

		Single strand DNA	Double strand DNA
Sunlight exposure (ZnO–)	1 day	D	N.T.
	1 week	D	D
	1 month	N.D.	N.D.
Sunlight exposure (ZnO+)	1 day	N.T.	N.T.
	1 week	D	D
	1 month	N.D.	D

D, Detected; N.D., Not Detected; N.T., Not Tested.

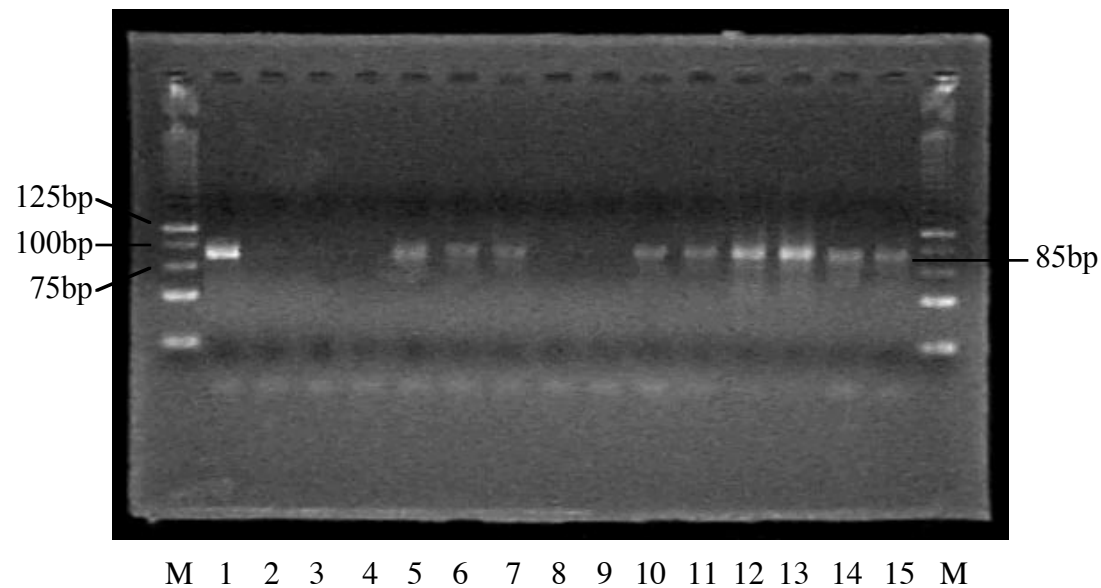
TABLE 3. *The results of a durability test including UV exposure*

	Hours	Double strand DNA
UV (ZnO–)	40	N.D.
	60	N.D.
	80	N.D.
UV(ZnO+)	40	D
	60	N.D.
	80	N.D.
UV and water (ZnO+)	40	D
	60	N.D.
	80	N.D.

D, Detected; N.D., Not Detected.

match analog patterns or feature-point comparisons. If some characteristic landmarks of an individual have been previously registered in the system, the time to verify will be very short. However, because they lack absolute accuracy, these systems are presently limited to local application, and have not yet achieved any universal standards. Among the various types of biometric information, the DNA-ID presented here should prove the most reliable for personal identification. DNA information is intrinsically digital, and does not change either during a person's life or after his death. The discerning power of this data can be enhanced without limit by increasing the number of STR loci analyzed.

Of course, DNA analysis and synthesis involves high capital cost, to buy and maintain equipment as well as purchase the necessary. This problem of the cost may prove to be an eco-



Lane 1 : Positive control
Lane 2 : Negative control
Lane 3 : UV-resistance 60 hr, ZnO–
Lane 4 : UV-resistance 40 hr, ZnO–
Lane 5 : Sunlight exposure 1 month, ZnO+
Lane 6 : Sunlight exposure 1 week, ZnO+
Lane 7 : Sunlight exposure 1 day, ZnO+
Lane 8, 9 : UV-resistance 60 hr, ZnO+
Lane 10, 11 : UV-resistance 40 hr, ZnO+
Lane 12, 13: UV-resistance 20 hrs, ZnO+
Lane 14, 15 : Untested printing samples

Fig. 3. First PCR products after agarose-gel electrophoresis
ZnO+, covered with ZnO; ZnO–, uncovered with ZnO.

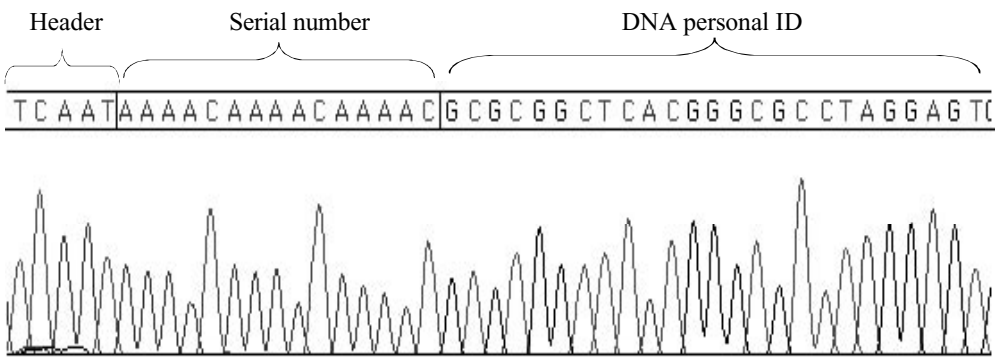


Fig. 4. Results of sequencing analysis by subcloning

nomic obstacle that will prevent its worldwide application. On the other hand, this high cost might actually turn out to be an advantage, because it would be difficult to copy DNA and use it without ample funding. However, like most products, the more popular such DNA techniques become, the lower the unit costs of the apparatus and reagents will become.

Another major problem using DNA materials is how to protect personal privacy. Based on our previous work concerning the generation of personal identifiers for DNA certification (Itakura 2002), we synthesized DNA containing base sequences that provides personal identification information while protecting privacy. The one-way encryption implemented here makes it impossible to recover any original DNA information from the sequence of the final synthetic DNA. Even determination of the final encrypted DNA-ID sequence itself would be virtually impossible and would require knowledge of the primer sequences used to amplify it; trying to determine them *a priori* would be an enormous waste of time and money. Furthermore, the inclusion of immaterial spacer DNA sequences further protects the DNA-ID from amplification, even if a random primer is used. This study solves both the privacy and the decoding problems of such DNA-ID sequences and demonstrates that its use for commercial application is feasible, as exemplified by its inclusion in DNA ink.

DNA ink sources need not to be limited to human beings; genetic information for animals and plants can also be converted to DNA sequences. DNA ink can thus be regarded as a sort of biological memory.

If it is to be used practically, DNA ink must be resistant to UV irradiation. In this study, the durability of the DNA ink approached that of photogravure inks; ideally, it would be made even more stable, allowing its use for long-term authentication of important documents. The greater the amount of DNA incorporated, the better its durability against physical attacks; however, the cost of production would be increased. To keep costs reasonable for commercial use, a better way to protect the DNA in the ink, such as wrapping it in small capsules or suspending it in some other solvent, could be developed. Another target would be to reduce the time needed for DNA analysis. The cost and time required for analysis are advantageous in that they present obstacles for readout and copying; however, situations requiring urgent verification are not currently amenable to these relatively slow DNA ink-based authentication systems. Considerable advances in technology are still needed to meet such requirements.

Acknowledgments

This work was supported in part by the Telecommunication Advancement Organization's project "High-performance security function based on hyperelliptic cryptography and its application in elec-

tronic social systems." The author is grateful to Prof. S. Tsujii (Chuo University) for his extensive support and gives special thanks to Dr. Y. Itakura (NTT Data University Corporation) and Mr. H. Ito (ID Technica Corporation) for their technical assistance and valuable comments. I also thank Prof. M. Funayama (Tohoku University) for reading the manuscript and giving me helpful advice.

References

Hashiyada, M., Itakura, Y., Nagashima, T., Nata, M. & Funayama, M. (2003) Polymorphism of 17 STRs by multiplex analysis in Japanese population. *Forensic Sci. Int.*, **133**, 250-253.

Itakura, Y., Hashiyada, M., Nagashima, T. & Tsuji, S. (2002) Proposal on Personal Identifiers Generated from the STR Information of DNA. *Int. J. Information Security*, **1**, 149-160.

PowerPlex™ 16 Kit User Manual (2000) Promega Corporation.

Walsh, P.S., Metzger, D.A. & Higuchi, R. (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques*, **10**, 506-518.