DNA-Based Personal Identification and its Technologies

ASOGAWA Minoru

Abstract

As discussed in the feature pages of this special issue, NEC is developing products for matching fingerprints, palmprints and face matching. In addition, NEC has since a few years ago also been conducting R&D for a "DNA analysis system that is capable of field applications." When this system is used together with existing fingerprint, palmprint and face matching systems, a composite matching system is created that contributes significantly to the realization of a safe and secure society.

Keywords

DNA analysis, PCR, electrophoresis, CODIS

1. Introduction

The identification of individuals by their DNA is based on an analysis of human genes. This technique is capable of extremely accurate evidence because it identifies individuals with an accuracy of one to 100- to 500 trillion. This paper describes the basic principles of the DNA analysis method and the mechanism for a "capability of individual identification at 1/100 to 500 trillion." It also introduces the efforts being made at NEC for the implementation of a "DNA analysis system that is capable of field application."

2. Basis of DNA Analysis

The human cell has a nucleus that contains 23 pairs, or 46 individual chromosomes. Among the 46 chromosomes, 23 are inherited from the father and the remaining 23 are from the mother. The chromosome pairs are numbered in order of size, with the largest being called chromosome 1 etc. [*1] It is not possible to determine which of the chromosomes in each pair is inherited from the father or mother.

Every piece of chromosome is basically composed of two "chain molecules" that are entwined around each other. Each "chain molecule" is composed of four kinds of base pairs that are arranged linearly. The number of linearly arranged base pairs per chromosome pair is about 280 million with chromosome 1 being the largest and the total number of base pairs in the 23 chromosome pairs is about 3 billion. The 3 billion base pairs in a "chain molecule" contain all of the information that the owner of the nucleus needs for being as a living organism. This means that individual information on blood type and genetic disease are also contained in the series of base pairs in the "chain molecules."

Among the information in the 3 billion base pairs, 3% is used for the design of proteins and part of the rest is used as signals for systematically running the life mechanism. The other base pairs are not involved in the life mechanism (according to what has been discovered up to the present).

Human chromosomes contain a large number of repeated sequences called microsatellites. Among the human microsatellites, those that are not involved in the life mechanism are used for human DNA analysis. This means that human DNA analysis information includes the information proper to the individual but does not include that related to the blood type or disease.

Among the various microsatellites, the repeats of 4 (or 5) base pairs are used in human DNA analysis. For example, the parts of human microsatellites (gene loci) called vWA refer to the repeats of 4 base pairs contained in chromosome 12, and it is known that the number of repeats is between 10 and 25. In other words, the vWA gene locus include "10-repeat type," "11-repeat type" up to "25-repeat type" locus. Since the number of chromosomes in chromosome pair 12 is two (one from the father and the other from the mother), each human individual has a set of two genotypes in the vWA gene locus (such as "11-repeat type" and "15-repeat type"). Human individual

[*1] As an exception, chromosome 21 is not larger than chromosome 22.

identification via DNA profiling is performed by analyzing genes in a total of 13 gene loci, including the 12 gene locus as described above plus the sex chromosome.

Next, we will explain the degree of authentication capability obtained by analyzing the gene loci. Let us assume that gene locus A has 10 types, for example from "10-repeat type" to "19-repeat type." If it is also assumed that the frequency of occurrence of these types is identical, the probability that a person contains the "18-repeat type" is 1/10. Since gene locus A of an individual contains a set of two genes, it is always expressed as a combination such as {"10-repeat type," "15-repeat type"}. When all of the combinations in locus A is considered comprehensively, it may seem that there would be a total of 100 combinations, which derive from the calculation of 10 types × 10 types. However, since it is not to distinguish the types inherited from the father and those from the mother, the actual number of combinations is 55, which is almost half of the above calculation. Therefore, the probability that the genotype of a person becomes {"10-repeat type," "15-repeat type"} is equal to $(1/10) \times (1/10) + (1/10) \times (1/10) = 1/50$ because two events of {"10-repeat type," "15-repeat type"} and {"15-repeat type," "10-repeat type"} are indistinguishable. Moreover, the probability that the genotype of a person contains a set of two identical repeat types such as {"11-repeat type," "11-repeat type"} is equal to $(1/10) \times (1/10) = 1/100$. Consequently, the probability of the occurrence of the locus type of gene A of a person is only 1/150 at maximum.

When one gene locus of a sample collected at a crime scene is analyzed, the probability that the type of the gene locus happens to coincide with that of a specific individual is 1/50 at maximum. If 12 such gene loci are analyzed simultaneously, the "probability that the types of all of the 12 gene loci happen to coincide" is $(1/50)^{12}$, or about $1/(2.4 \times 10^{20})$. In fact, however, because the occurrence frequencies of gene repeat counts are deviated and the number of genotypes is no more than 10, the actual probability in the case of 13 gene loci (12 loci + sex chromosome) is around $1/(1 \text{ to } 5 \times 10^{12})$ [3]. This figure is fairly large compared to the world population of 7 billion (7×10^{9}). In other words, it is extremely rare that a person would have the same genotype as the genotype of a sample collected from a crime scene. Therefore, if the genotypes of a person coincides with the sample, it is highly probable that the person is identical to the person who left the sample [*2].

3. DNA Analysis System

In 1998 the FBI of the USA started a system called CODIS (Combined DNA Index System) in order to accumulate the human DNA information that was involved in crimes. Having accumulated about 9 million items of data as of July 2010, the CODIS system is actually contributing to the realization of a safe and secure socity [4]. A similar database is also being built in Japan. The DNA analysis technology is an established technology as may be seen at a system scale of about 9 million items. However, the analysis equipment has a large size (equivalent to a small refrigerator) and it also requires a managed laboratory environment. In addition, it is said that an analysis time of about half a day is necessary. To decrease the size and increase the speed of analysis, a development project for a "DNA analysis system that can be applied in the field" was begun in 2010 under FBI leadership.

At NEC, we also anticipated the need for a "DNA analysis system that could be applied in the field" and we had started its development a few years previously. The development does not aim only at increasing the speed and decreasing the size of the DNA analysis mechanism but also at other advantages considering the situation in the field, such as a reduction in the cost of disposable analysis chips, prevention of contamination an improvement in the ease of use and improvement to the shelf life. The targets for speed increase and size decrease are respectively an analysis time of about 30 minutes and a size equivalent to a suitcase. **Photo 1** shows an attache case-size system developed in FY2007 [1] and a suitcase-size system developed in FY2008 [2][*3]. The 2007 system is capable of

Photo 1 External views of system prototypes.

[*2] The probability of matching increases when the two persons are in a kinship. The genotypesof two persons coincide completely when they are monozygotic twins.

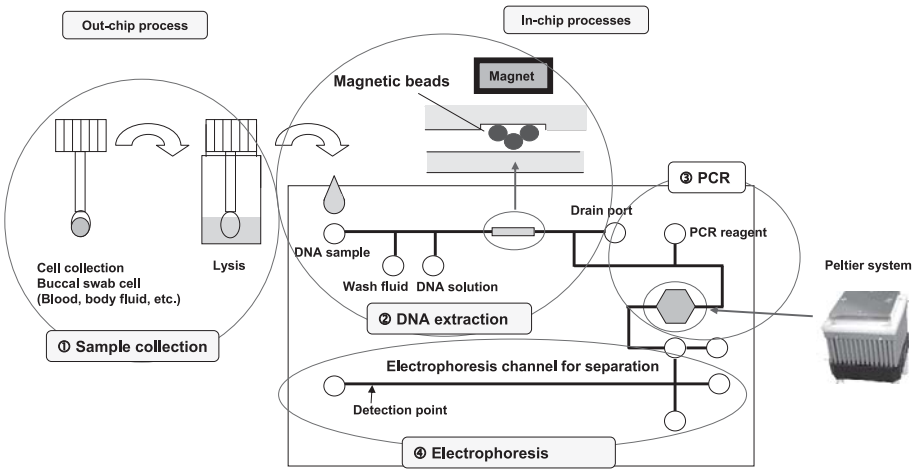

Fig. 1 Process of DNA analysis.

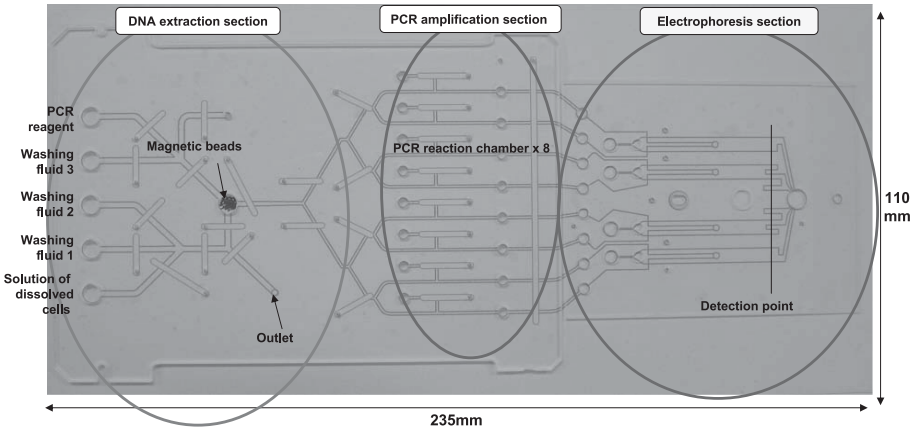

Fig. 2 Outline of chip fabrication.

analyzing one or two gene loci while, as detailed later, the 2008 and later systems are capable of analyzing 8 to 16 loci thanks to the development of new optics.

The system is roughly composed of four sections as shown in **Fig. 1** and **Fig. 2** , and performs analysis in the following process.

1) The sample is collected by harvesting intra-cheek mucosa cells from inside the mouth using a cotton swab, etc. The swab is put in a small special container and the cells are destroyed with the reagent sealed in advanced in the container to prepare a solution in which DNA is eluted.

2) The DNA extraction process uses commercially available magnetic beads. The surfaces of the magnetic beads have a high affinity to DNA so the DNA is caught by them when the beads are mixed with the DNA solution. Next, the beads are collected and washed. The beads are magnetic so that they can be secured easily using the magnet. Thus, only the DNA is transferred to the PCR reaction basins in the next process.

3) The PCR process amplifies the DNA by using a reagent that amplifies only the specific gene loci. As shown in Fig. 2, DNA is divided and injected together with PCR reagents

*3 These systems were developed jointly with Aida Engineering, Ltd.

into the eight PCR reaction chambers prepared for the analyses of eight loci. In fact, two kinds of reagents can be used with each PCR chamber so that up to 16 (= 2 × 8) kinds of gene loci can be analyzed at a time. As shown in **Photo 2** , a metallic member controlled by Peltier devices is installed immediately below the PCR reaction chambers in order to perform a temperature change cycle of 98°C → 58°C → 72°C → 98°C. Each temperature change cycle can double only the "DNA in the repeats of gene loci." This DNA amplification is an artificial version of the DNA replication that occurs during cell division of living organisms. Repeating the temperature change cycles for about 30 times makes it possible to amplify the DNA exponentially. Up to the present time, we have confirmed that the DNA can be amplified to an amount that is suitable for analysis in about 30 minutes. With the previous system, the temperature change was slow because a metallic member of a size capable of processing 96 samples simultaneously was used as shown in **Fig. 3** . The new system uses an ultracompact metallic member of low thermal capacity in order to increase the processing speed.

4) In the next process, the length of the "DNA in the repeats of each amplified gene locus" is measured in order to identify the number of repeats. To measure this distance, DNA is introduced in a very thin (100 μm) channel filled with polymer and subjected to the application of a high voltage. Since the DNA is negatively charged, the application of the voltage causes it to migrate toward the positive electrode. This electrophoresis causes the short DNA to arrive at the detection point in a short time and the long DNA to arrive more slowly, so the DNA length can be measured by measuring the arrival time. When the DNA length is known, the number of repeats in the gene locus can be identified. This process can enable a complete analysis in about 5 minutes.

We have developed a new spectroscopic component with a size of about 1/10th that of traditional devices of this kind. For the size reduction, we fabricated a new aspherical lens (**Photo 3**), which has also made it possible to reduce light loss.

By performing the above steps in sequence, this process enables field DNA analysis in the short time frame of about 30 minutes.

The key technology in this system is that for the transfer of the solutions between processes. Silicone rubber films in which patterns are embedded are layered in order to implement the solution transport channels and the solution action control valves.

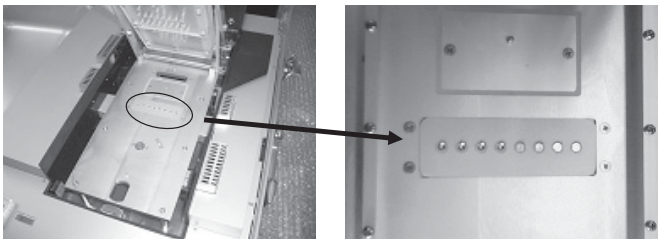

Photo 2 Heat transmission pins in PCR chamber section.

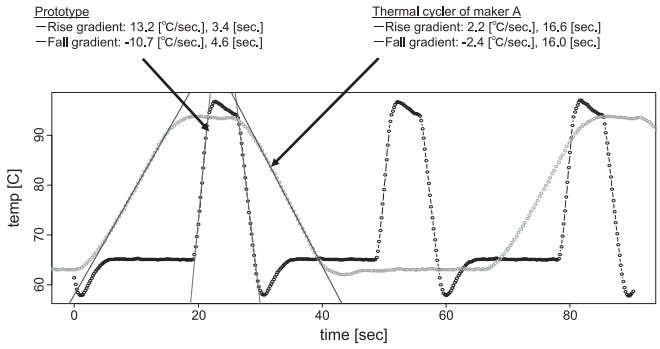

Fig. 3 Comparison of the temperature cycle rate between the existing PCR system and the prototype PCR system.

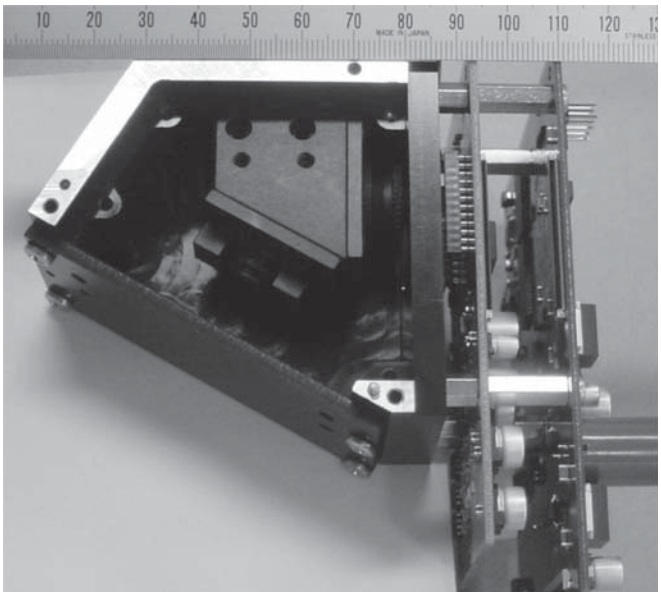

Photo 3 Newly developed spectroscopic device.

DNA-Based Personal Identification and its Technologies

4. Conclusion

Although DNA analysis is expected to form the ultimate means of personal identification, the large size of the requisite analysis systems and the need for managed laboratory environments have made it impossible to effectively perform analyses in the field. In addition, the required analysis time of about half a day has made it unsuitable for prompt criminal investigations.

At NEC, we began development of a "DNA analysis system that can be applied in the field" a few years ago. When a composite matching system can be created by combining this system with fingerprint/palmprint/face matching systems, it is expected to contribute greatly in making society safer and more secure than ever in the future.

References

1) M. Asogawa, M. Sugisawa, et al. "Development of portable STR analysis system aiming on-site screening," Proceedings of International Symposium on Human Identification, (2007).
2) M. Asogawa, M. Sugisawa, et al. "Development of portable and rapid human DNA analysis system aiming on-site screening," MicroTAS 2008, vol. 1, pp.1072-1074. (2008).
3) John M. Butler, "Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers," Academic Press, 2001.
4) http://www.fbi.gov/hq/lab/codis/clickmap.htm

Author's Profile

ASOGAWA Minoru
Expert
2nd Government and Public Solutions Division
Government and Public Solutions Operations Unit