

## VALIDATION EXPERIMENT REPORT ON DNA-ID FOR PERSONAL IDENTIFICATION

**Yukio Itakura<sup>1,3</sup>, Masaki Hashiyada<sup>2</sup>, Toshio Nagashima<sup>3</sup>, Shigeo Tsuji<sup>1</sup>**

<sup>1</sup> Chuo University, Tokyo, Japan

<sup>2</sup> Tohoku University, Sendai, Japan

<sup>3</sup> NTT Data Technology Corporation, Tokyo, Japan



The individual difference in the repeat count of STR (Short Tandem Repeat) bases among all of the base sequences can be used as the unique DNA information for personal identification. We propose “DNA-ID” for personal identification by specifying multiple STR loci and then sequencing the pair of repeat count information thus obtained in a certain order. DNA-ID is generated according to the following procedure:

- Step 1: The STR repeat count at each locus was measured.
- Step 2: The two values obtained from each locus (STR repeat count numbers) were sequenced in ascending order.  
 $L_i: p_i \parallel q_i, p_i \leq q_i$
- Step 3: DNA-ID  $a_A = p_1 \parallel q_1 \parallel p_2 \parallel q_2 \parallel p_3 \parallel q_3 \parallel \dots \parallel p_n \parallel q_n$

The DNA-ID  $a_A$  generated by the procedure step 1~3 becomes personal identification information that is unique at a certain probability. The probability that two persons

A and B will have the same ID value (hereafter referred to as “matching probability”) can be calculated by the statistical data on the frequency of occurrence for the repeat count at each locus and the correlation of them among loci.

As a follow-up to the basic idea, we collected actual human DNA from more than 500 volunteers with their consent and analyzed their samples. Based on the actual data obtained, we verified the relationship between the number of loci required and the matching probability, as well as the correlation among the loci. The relationship between locus multiplicity and matching probability has been established as a validating empirical result.

We could verify that the matching probability becomes exponentially smaller as the number of loci increases and that few correlations exist among the loci. We succeeded in confirming the DNA-ID that our proposition was adequate for use in practical applications.