

DNA-ID の統計的検証

—実証実験による STR ローカスの相関性の検証—

板倉 征男^{*‡} 橋谷田真樹[†] 長嶋 登志夫[‡] 辻井 重男^{*}

^{*}中央大学 研究開発機構

〒162-8473 新宿区市谷本村町 42-8

[†]東北大学 大学院医学系研究科

〒980-8575 仙台市青葉区星陵町 2-1

[‡]NTT データテクノロジー(株)

〒107-0052 港区赤坂 2-2-12

あらまし

筆者らは DNA の塩基配列の特定領域にある STR (Short Tandem Repeat) 座位 (ローカスと言う) を複数箇所定義しそこで得られる 2~4 個の塩基配列の繰返し回数を一定の順序で並べて個人識別用 ID を生成する方法について実証実験による検証を進めている。特にローカス間の相関関係については、生成した DNA-ID の同値確率に直接影響するので、出来るだけ多くの実データによる検証が必要である。

本論文は実際の人体の DNA を約 500 人の提供者の協力を得て採取分析し、ローカスの多重度と生成した DNA-ID の同値確率の関係及びローカス間の相関関係を検証した実証実験の結果を報告する。

その結果ローカス間の相関はほぼ無視できることが検証でき、提案する DNA-ID は十分実用に供せられることが確認出来た。

キーワード DNA 情報, DNA-ID, STR 繰返し回数, 統計的検証, 同値確率, ローカス間相関

Statistical Verification of DNA-ID

-Verification the correlation among loci by the validation experiment-

Yukio Itakura^{*‡}

Masaki Hashiyada[†]

Toshio Nagashima[‡]

Shigeo Tsuji^{*}

^{*}Chuo University, Research and Development Initiative

42-8 Ichigaya-Honmura-cho, Shinjuku-ku, 162-8473

[†]Tohoku University, Graduate Medical Research Department

2-1 Seiryō-cho, Aoba-ku, Sendai-city, 980-8575

[‡]NTT Data Technology Corporation

2-2-12 Akasaka, Minato-ku, Tokyo, 107-0052

Abstract

We have been proposing the method to generate DNA-ID for personal identification by specifying multiple STR locations (called "loci") and trying validation experiment.

It is very important to verify the correlation among the loci because the matching probability of the DNA-ID directly depends on the correlation function.

Based on the actual more than 500 volunteers data obtained in the validation experiment, we verified the correlation among the loci, the results of which are reported in this paper. We have succeeded in confirming that the DNA-ID that we have been proposing is adequate for use in practical applications.

Keywords: DNA information, DNA-ID, STR repeat count, statistical verification, matching probability, correlation among loci

1. まえがき

本論文は筆者らが先に報告した“個人識別用 DNA 情報に関する実証実験 (その 1, その 2)” に引き続き実証実験の成果について述べる[1][2].

先の報告では, 生成した DNA-ID の同値確率について統計理論的解析及び STR 座位 (ローカス) 間のデータの相関性の解析が不十分であった. 今回はサンプルデータを 350 から 500 以上に増やし, これらの解析をさらに厳密に行った.

即ち各ローカスにおける STR 繰返し数 (アレルという) の分布より, DNA-ID の同値確率を求める理論式を見直し, より正確なものとした. また相関係数の計算方法を見直し, より実態に即した解析を行った. サンプル数自身が増えたので, 全体として前報告より検証の信憑性を上げることが出来た.

2. DNA 情報の採取と DNA-ID の生成

DNA 全塩基配列の内, 遺伝子領域外のところにマイクロサテライトと呼ばれる部分がある[3]. ここは STR (Short Tandem Repeat) と呼ばれるように 2~4 の短い塩基配列のパターンが繰返し並んでいる領域で, その繰返し回数は個人によって異なっている.

図 1 のように A さんの場合 TH01 と名づけられた STR 座位 (ローカスと呼ぶ) で, STR の繰返し回数のペア (STR アレルと呼ぶ) は 5 回と 7 回, B さんの場合 6 回と 6 回となっている.

アレルのペアの一方は父親から, 他方は母親から受継いでいる.

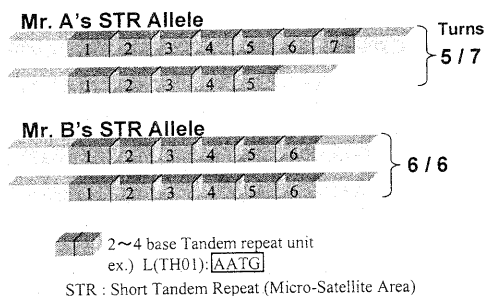


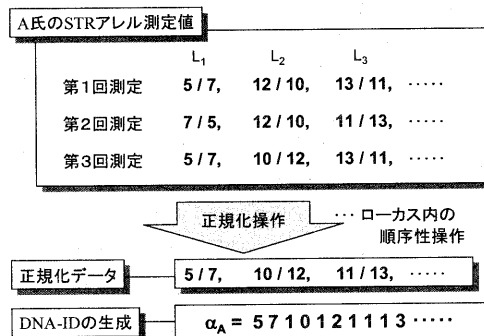
図 1 STR の繰返し数のペア (アレル)

Fig.1 STR repeat number pair (Allele)

アレルの値は, 6~60 回の間で隣り合う 10 前後の回数に分布しており, そのうちメジャーとなるもの即ち 10%以上の出現頻度となるものは, 3~4 種類の回数で分布している. 計測した STR アレルの値は個人差があるので, ローカスを複数個定義し, その順

番でこの値を並べれば個人識別用 ID が得られるというのが提案する DNA-ID の原理である[4][5].

但し, 1 箇所のローカスで計測するアレルのペアは, 5 回/7 回と出る場合と 7 回/5 回と出る場合があるので, ペアの値は必ず小さい方から並べるというルールにする.



注) L1~L3は, ローカス#1~#3を示す.

図 2 DNA-ID の生成方法

Fig.2 DNA-ID Generation Method

こうしないと測定により ID と個人が 1:1 に対応できない (図 2 参照). これをローカス内の順序性操作または正規化操作と呼ぶこととする. 即ち A さんのローカス i におけるアレルのペアを p_i, q_i とすると

$$L_i = p_i \| q_i, p_i \leq q_i \quad (1)$$

と定義する. A さんの DNA-ID : α_A は, 次のようにして生成する.

$$\alpha_A = p_1 \| q_1 \| p_2 \| q_2 \| p_3 \| q_3 \dots\dots \| p_n \| q_n \quad (2)$$

生成された α_A は後述するように一定の同値確率以下で一意性のある個人識別情報となる.

(1)で $p_i \leq q_i$ となるように順序を操作することは, DNA-ID の乱数性に若干悪影響を与えることになるが, その理論的解析は後述する.

上記の DNA-ID 生成方法を使って実際に採取した DNA 情報より, 得られた個人対応の DNA-ID は表 1 のようになる. サンプルデータは N=515 であるが, 最初と最後のサンプルの DNA-ID を例示した.

表中 312 や 332 のように 3 ケタの繰返し回数を示すところがあるが, これは 31 回又は 33 回の 4 塩基配列の繰返しの後に 2 塩基が付加されるパターンを示している. 出現頻度としてはまれなケースであるが再現性は確定的にあるので, このまま直列に並べて ID とすることとしている.

表1 DNA 実採取データから生成した DNA-ID の例

Table 1 Examples of DNA-IDs generated from actual DNA data collected

No	DNA donor	DNA-ID																													
1	A	10	13	10	12	15	20	29	332	16	16	10	11	8	13	10	11	19	23	7	9	8	11	14	17	10	20	10	13	9	13
2	B	12	12	8	9	14	16	31	312	15	17	11	11	8	13	15	15	20	21	9	9	9	11	16	18	11	12	9	14	10	12
3	C	12	13	9	10	13	14	30	30	18	19	11	12	9	12	14	15	20	21	7	7	8	11	17	18	11	12	11	12	9	10
4	D	11	12	9	11	13	17	28	30	15	16	11	11	10	11	13	16	22	22	7	9	8	8	17	18	11	15	9	9	9	11
5	E	10	12	11	12	13	15	30	322	18	18	11	11	8	8	14	16	20	26	6	9	8	11	14	17	12	15	9	13	9	11
6	F	10	10	8	11	15	15	28	322	15	16	10	10	11	12	13	16	25	26	6	9	8	11	16	16	12	17	9	12	10	10
7	G	11	12	8	9	14	16	28	29	15	15	9	11	11	12	10	16	21	23	9	9	8	9	16	17	14	19	9	11	12	13
8	H	10	11	8	12	13	14	30	322	15	16	11	13	10	11	12	12	25	25	6	9	8	12	17	20	5	17	10	11	9	10
9	I	10	12	8	9	13	19	29	30	16	17	13	15	8	12	12	13	20	26	7	93	8	9	14	14	5	15	10	11	9	12
10	J	10	14	8	11	13	15	30	322	16	17	10	12	11	11	14	14	19	25	6	9	8	11	14	17	11	16	9	10	9	9
Reference (Locus no.)		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15															
(Locus name)		CSF1P0	D13S317	D18S51	D21S11	D3S1358	D5S818	D7S820	D8S179	FGA	TH01	TPOX	VWA	PentaE	D16S539	PentaD															

(Remark) Names of donors have been completely erased in accordance with Ethical Guideline.

3. DNA-ID の統計理論的解析

3. 1 ローカス L における同値確率

ローカス L における STR アレル (j, k) がこの組合せで出現する確率を p_{jk} とする。また, j, k の個々の出現確率を p_j , p_k とする。ここで j, k は生成する ID に一意性を持たせるために小さい順に列べる, いわゆる順序性操作を行っていることは前述した。

順序性操作を行うために, (j, k) が出現する確率は p_{jk} だけでなく, (k, j) が出現する確率 p_{kj} も加算されたものとなる。測定では同一人物で (k, j) と出ても, $k > j$ なら (j, k) と書いて (j, k) として扱うからである。

従って p_{jk} は,

$$j \neq k \text{ のとき } (j < k)$$

$$p_{jk} = p_j \cdot p_k + p_k \cdot p_j$$

$$= 2p_j p_k$$

(3)

$$j = k \text{ のとき}$$

$$p_{jk} = p_j \cdot p_j$$

(4)

となる。

任意の 2 人が 1 つのローカスにおいて STR アレルが同値となる確率 p は,

$$j = k \text{ の場合 } \sum_{j=1}^m (p_j \cdot p_j)^2 \quad (5)$$

$$j \neq k \text{ の場合 } \sum_{1 \leq j < k \leq m} (2p_j \cdot p_k)^2 \quad (6)$$

従って p は

$$p = \sum_{j=1}^m p_j^4 + \sum_{1 \leq j < k \leq m} 4(p_j \cdot p_k)^2 \quad (7)$$

となる。

ここで m は j 及び k の上限値で, これまで報告された情報では $m=60$ である [6]。

図 3 は各ローカスにおいて任意の 2 人の STR アレルが同値になる確率 P を示したものである。P の計算は実証実験において採取した $N=515$ の STR アレルの分布データを使って行った。

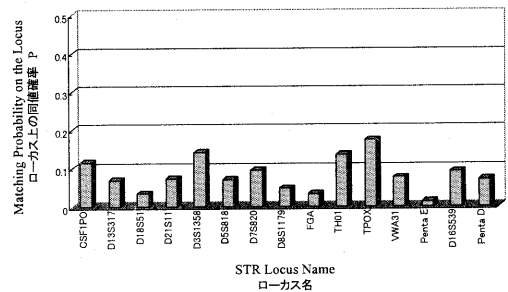


図3 任意の 2 人の STR アレルの同値確率

Fig.3 Matching Probability on the Locus

ローカスによって P の値は多少のバラツキがあるが, 大体 0.2 以下の範囲に入っている。これは STR アレルの出現頻度が均等に分布していることを示している。

3. 2 任意の2人のDNA-IDの同値確率

次に n 個のローカスを多重に重ねた場合の DNA-ID の同値確率 p_n を求める。

i 番目のローカスにおける任意の2人のSTRアレルの同値確率を p_i とする。

ローカスを n 段重ねた時、全STRアレルの配列によって生成したDNA-IDについて任意の2人が同値となる確率 p_n (いわゆるDNA-IDの同値確率) は

$$p_n = \prod_{i=1}^n p_i \quad (8)$$

となる。但しSTRアレル間の相関はないものとする。

実証実験の $N=515$ のデータによるローカスの多重度とDNA-IDの同値数及び同値率の関係は、表2のように段数を重ねるとそれらは次第に小さい値となり、多重度が6以上となると0となった。インターネット公開データから統計理論に基づいて計算した値も、同じ傾向を示している。

これ以上の同値確率は、実証実験によって得られた各STR座位におけるアレルの分布データと、インターネットによって公開されているアレル分布データにより計算することが出来る[6]。

表2 ローカスの多重度とDNA-IDの同値数

Table 2 Locus multiplicity and DNA-ID matching probability

N=515			
Locus multiplicity	Number of same values (number of DNA-IDs that are identical)	Matching probability (Experiment data)	Matching probability (Internet data)
1	15185	1.143E-01	1.119E-01
2	2110	1.588E-02	1.535E-02
3	284	2.137E-03	2.127E-03
4	48	3.613E-04	3.811E-04
5	4	3.010E-05	3.609E-05
6~15	0	See Fig.4	See Fig.4

図4は、ローカスの多重度15段までのDNA-ID同値確率を統計理論により計算した結果である。

実証実験に基づく同値確率のSTR段数依存性とインターネット公開データに基づくそれとはよく近似しており、15段で 10^{-15} 程度の値が得られた。

これより n 段の多重STR配列法によるDNA-IDについて、任意の2人が同値となる確率 p_n は、実験式として

$$p_n \doteq 10^{-n} \quad (9)$$

が得られた。いずれも各ローカス間の相関はないとして計算したものである。相関については次節の検証結果から、式(9)のオーダには影響しないことが明

らかになった。

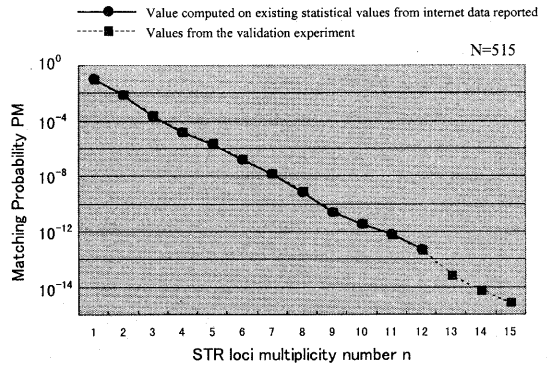


図4 ローカス多重度とDNA-IDの同値率の関係

Fig.4 Relationship between STR loci multiplicity number n and identification matching probability P_M

3. 3 各ローカス間のSTRアレルの相関係数とその影響

STRアレル間の相関がある場合相関係数を ρ とすると、 p_n は、次のように表わせる。

$$p_n = \prod_{i=1}^n (1 + \rho) p_i$$

$$= (1 + \rho)^n \prod_{i=1}^n p_i$$

ここで

$$(1 + \rho)^n = 1 + n\rho + \frac{n(n-1)}{2}\rho^2 + \dots$$

$\rho \ll 1$ なので

$$\doteq 1 + n\rho$$

$$\therefore p_n \doteq (1 + n\rho) \prod_{i=1}^n p_i \quad (10)$$

従ってアレル間に相関 ρ がある場合、これを n 段重ねたときに同値確率に与える影響は $(1 + n\rho)$ 倍程度となる。

ρ は次のようにして求める。

2つのローカスのSTRアレル実測値 (X_i, Y_i) , $i=1, 2, \dots, w$ を母集団 (X, Y) とすると、 X, Y の間の相関係数 ρ は

$$\rho = \text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}X \cdot \text{Var}Y}}$$

$$= \frac{\sum_{i=1}^w (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^w (X_i - \bar{X})^2 \sum_{i=1}^w (Y_i - \bar{Y})^2}} \quad (11)$$

となる [7].

但し, (X, Y) は 2 変量正規分布に従う場合である. 実測データのサンプル数 w が大きくなると, この条件は近似的に満たされる.

$\bar{X}$ は X_i , $\bar{Y}$ は Y_i の平均値である.

また, 上記の相関係数の計算において, STR アレルの実測値 X_i とは, そのローカスにおける STR 繰返し数のペア j, k を並べて

$$X_i = j \| k \ (j \leq k)$$

としている.

一般に ρ の値は, 次のように評価される [8].

.00~.20: 無相関またはほとんど相関が無い

.20~.40: 弱い相関がある.

.40 以上: 相関がある.

ローカス間の相関係数を調べるには, 個人ごとの STR アレルデータがそろわないと計算できないが, インターネットの公開データもそこまで報告していない. そこで今回約 500 人の DNA 提供者の協力を得て個人ごとの実データを採用し, 分析することによってはじめて表 3 のようなローカス間の STR アレルの相関係数を計算することが出来た. 縦軸と横軸の交点は, 対応する 2 つのローカス間の相関係数を示す. また図 5 にこの相関係数の分布を示す.

相関係数の分布を分析すると, プラス側とマイナス側に平均 -0.0086 標準偏差 0.046 となっており, ほとんど相関が無いことを示している.

多重 STR 配列法による DNA-ID の同値確率は, (10) 式で計算できる. ρ としては標準偏差値をとるのが妥当である.

$\rho=0.046$ とすると $n=15$ 段のときは, $(1+n\rho)=1.69$,

$n=30$ 段のときは $(1+n\rho)=2.38$ となる. いずれにしても p_n のオーダーには影響しない程度であることが判明した.

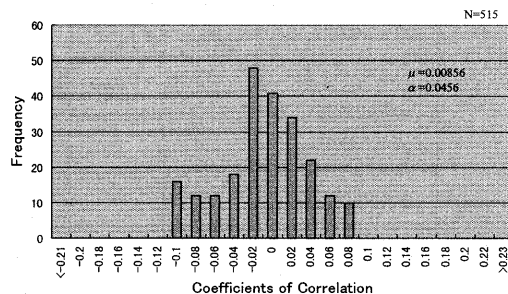


図 5 相関係数の分布

Fig.5 Distribution of correlation coefficients

4. 考察

4. 1 STR 多重度と実用的同値識別率の考察

STR 多重度を n , 生成した DNA-ID を使用する集団の人数を N とすると, N 人の総組合せ同値確率 P (これを実用的同値識別率と呼ぶこととする) は,

$$P \approx \frac{1}{2} N(N-1) P_n \quad (12)$$

となる. ここで P_n は式(9)より $P_n \approx 10^{-n}$ が得られている.

これより n と N の関係を求めると, 図 6 のような結果が得られた.

世界人口の規模をカバーする $N=100$ 億人で本方式による DNA-ID を適用する場合, $n=30$ 段くらいの多重度が必要であり, このときの実用的同値識別率は 100 億分の 1 となる.

表 3 ローカス間の相関係数

Table 3 correlation among Loci

N=515

	CSF1PO	D13S317	D18S51	D21S11	D3S1358	D5S818	D7S820	D8S179	FGA	TH01	TPOX	vWA	Penta E	D16S539	Penta D
CSF1PO	0.00	0.01	-0.10	0.02	-0.05	0.02	0.03	0.05	0.01	-0.01	-0.01	-0.02	-0.02	-0.09	0.01
D13S317	0.01	0.00	-0.02	0.01	0.05	-0.01	0.03	-0.04	-0.10	0.01	0.02	0.04	0.02	-0.01	-0.01
D18S51	-0.10	-0.02	0.00	-0.02	-0.01	-0.02	0.09	-0.02	0.01	-0.08	0.04	0.01	-0.11	-0.07	0.07
D21S11	0.02	0.01	-0.02	0.00	0.01	-0.01	0.00	0.03	0.00	0.08	0.02	0.06	-0.03	0.06	-0.06
D3S1358	-0.05	0.05	-0.01	0.01	0.00	0.08	-0.08	-0.06	-0.04	0.05	-0.04	0.03	-0.03	0.00	-0.01
D5S818	0.02	-0.01	-0.02	-0.01	0.08	0.00	-0.09	0.03	0.05	-0.06	-0.02	0.03	-0.01	-0.04	-0.10
D7S820	0.03	0.03	0.09	0.00	-0.08	-0.09	0.00	-0.03	-0.03	-0.03	-0.04	0.03	-0.02	-0.09	-0.02
D8S179	0.05	-0.04	-0.02	0.03	-0.06	0.03	-0.03	0.00	0.03	-0.03	-0.06	-0.03	-0.10	-0.08	0.01
FGA	0.01	-0.10	0.01	0.00	-0.04	0.05	-0.03	0.03	0.00	-0.07	-0.06	-0.02	0.03	-0.02	-0.04
TH01	-0.01	0.01	-0.08	0.08	0.05	-0.06	-0.03	-0.03	-0.07	0.00	-0.09	0.05	0.01	0.07	0.00
TPOX	-0.01	0.02	0.04	0.02	-0.04	-0.02	-0.04	-0.06	-0.06	-0.09	0.00	0.06	0.00	0.06	0.04
vWA	-0.02	0.04	0.01	0.06	0.03	0.03	0.03	-0.03	-0.02	0.05	0.06	0.00	-0.02	0.00	-0.06
Penta E	-0.02	0.02	-0.11	-0.03	-0.03	-0.01	-0.02	-0.10	0.03	0.01	0.00	-0.02	0.00	0.06	-0.01
D16S539	-0.09	-0.01	-0.07	0.06	0.00	-0.04	-0.09	-0.08	-0.02	0.07	0.06	0.00	0.06	0.00	0.05
Penta D	0.01	-0.01	0.07	-0.06	-0.01	-0.10	-0.02	0.01	-0.04	0.00	0.04	-0.06	-0.01	0.05	0.00

A numerical value at an intersection indicates the correlation between two loci.

$\begin{matrix} n(\text{Multiplicity}) \\ N(\text{Population}) \end{matrix}$	15	18	21	24	27	30	Example 適用例
10^5 (10万人)	$\frac{1}{2}10^{-5}$	$\frac{1}{2}10^{-8}$					Small-Middle town 中小都市
10^6 (100万人)	$\frac{1}{2}10^{-3}$	$\frac{1}{2}10^{-6}$	$\frac{1}{2}10^{-9}$				Million population 政令指定都市
10^7 (1000万人)		$\frac{1}{2}10^{-4}$	$\frac{1}{2}10^{-7}$	$\frac{1}{2}10^{-10}$			Tokyo, Osaka 東京, 大阪
10^8 (1億人)			$\frac{1}{2}10^{-5}$	$\frac{1}{2}10^{-8}$	$\frac{1}{2}10^{-11}$		Japan 日本
10^9 (10億人)				$\frac{1}{2}10^{-6}$	$\frac{1}{2}10^{-9}$	$\frac{1}{2}10^{-12}$	China 中国
10^{10} (100億人)					$\frac{1}{2}10^{-7}$	$\frac{1}{2}10^{-10}$	All Over the world 世界人口

図 6 STR 多重度と実用的同値識別率の関係

Fig.6 Relations between STR Loci multiplicity n and Applied ID code matching probability

4. 2 DNA-ID が同値となる可能性とその対策

本論文で提案する多重 STR 配列法による DNA-ID は、多重度 n や適用する人数 N の大小に係わらず次のような場合、同値となるもしくは同値となる可能性がある。

- ① 実用的同値識別率以下の確率で発生する同値の DNA-ID の間
- ② 一卵性双生児の間
- ③ 血液を入れ替えるような医療後の本人血液から採取した DNA と提供者の間
- ④ ガン細胞など DNA 塩基配列に変異を生じた場合

このような場合においても識別を可能とするために、何らかの対策が必要である。方法としては、他の生体情報を併用する場合と、数学的処置を行う場合の 2 通りの手法が考えられる。

他の生体情報を併用する方法としては、ゲノムの中で SNP (Single Nucleotide Polymorphism) といわれる 1 つの塩基配列の変位情報や、細胞核内の DNA とは別の組織にある DNA、例えばミトコンドリア DNA 情報や一卵性双生児間でも異なると言われる指紋などの生体情報を併用して識別する方法である。この場合の問題は、ごく少数の同値 DNA-ID の識別のためにシステム全体に上のようなオプションなしかけを持ち込まねばならぬということである。

一方数学的処置を行う方法とは、本方式で得られた DNA-ID の値 α_A に対して本人の秘密乱数 r_A を加えてハッシュをとった β_A を使うような方式である。

$$\beta_A = h(\alpha_A + r_A)$$

この場合、 α_A を有する本人が確かに r_A という秘密乱数を持っていて β_A を生成できることを、 r_A という秘密情報を示さずにいわゆるゼロ知識証明ができる

しくみが必要となる。

4. 3 倫理的考察及び対策

本提案による DNA-ID は、DNA 情報のうち人体の構造や病因に関与しない、いわゆる遺伝子領域以外の部分でマイクロサテライトといわれる STR 情報を用いるとはいえ、採取時に遺伝子領域の情報も分析することができるため、十分なプライバシー保護対策が必要である。

その 1 つは、前節で述べたような数学的演算を行って本来の DNA 情報である α_A はそのまま扱うことはせず、変換した β_A を使う方法である。

その 2 つは同様にして α_A を暗号鍵に組込んで秘密鍵及び公開鍵のペアとして ID 機能と等価な識別機能を間接的に持たせる方法である [4]。

いずれにしても、他人の DNA 情報を毛髪などから盗むことは防止できないので、新たなプライバシー保護の視点に立った DNA 情報の取扱いに関する倫理法の制定が必要であり、運用面においても十分な配慮を講じなければならない。

4. 4 分析時間の課題

DNA-ID の大きな課題は、生体情報採取後の分析時間とコストの抜本的なブレークスルーが必要なことである。分析時間は数年前の一晚から昨年は 3 時間で測定可能な機材が商品化された。研究としては 30~50 分の報告が話題になっているが、オンサイトチェックを行うには十数秒のオーダの分析時間短縮が必要で、そのためには数年の期間を要すものと考えられる。コストについては米国における親子鑑定の商用ベースの料金が 200 ドルと言われているが、ID として実用するには 2 ケタ以上のコストダウンが必要である。

5. まとめ

本論文では多重STR配列法によるDNA-IDの生成方法を提案し、IDとしての同値確率について統計理論的な解析を行うとともに、約500人の提供者のDNA情報を実際に採取分析し、実証実験による理論の裏付けを行った。

その結果、各ローカス間のSTRアレルの相関は無視できる、即ち独立した確率変数として扱って問題ないこと、各ローカスにおけるSTRのアレルの分布は、インターネットで公開されているデータと実証実験のデータがほぼ一致し、統計的理論値の信憑性が得られたこと、またこれらによりDNA-IDの実用的同値識別率を得るのに必要なSTRローカスの多重度の関係を統計理論的に解明することができた。

次にある確率で存在しうるDNA-IDの同値者の識別や、個人情報としてのプライバシー保護の方策を検討した。

これらの方法により、DNA-IDは究極の個人識別用IDとして十分適用可能なことが検証できた。

6. 謝辞

本研究に当たり、東北大学医学系研究科社会医学講座法医学分野の舟山真人教授に、実証実験に関する全面的協力をいただきました。

また中央大学理工学部情報工学科の今井桂子教授に統計的理論面からのご教示をいただきました。合せて深謝いたします。

文 献

- [1] 板倉征男, 橋谷田真樹, 長嶋登志夫, 舟山真人, “個人識別用DNA情報に関する実証実験報告(その1),”電子情報通信学会, 信学技報, ISEC2001-12, May. 2001.
- [2] 板倉征男, 橋谷田真樹, 長嶋登志夫, 辻井重男, “個人識別用DNA情報に関する実証実験報告(その2),”電子情報通信学会, 信学技報, ISEC2001-13, May. 2001.
- [3] T. A. Brown 著, 村松正實訳, “ゲノム,”メディカル・サイエンス・インターナショナル, pp.154, 2000.
- [4] 辻井重男, 板倉征男, 山口浩, 北沢敦, 齋藤真也, 笠原正雄, “生体情報が秘密鍵に埋め込まれた構造を有する公開鍵暗号方式,”信学技報, SCIS2000, D07, Jan. 2000.
- [5] 板倉征男, 長嶋登志夫, 辻井重男, “個人識別用DNA情報の統計的検証,”情報処理学会, CSS-2000 シンポジウム, pp.121-126, Oct. 2000.
- [6] Wolfgang Huckenbeck, Kathy Kuntze, Hans-Georg

Scheil, “The Distribution of the Human DNA-PCR Polymorphisms,” A Cooperation Project of Institute of Forensic Medicine Institute of Human Genetics and Anthropology Heinrich-Heine-University Düsseldorf, Germany.

[7] 竹内啓, “現象と行動のなかの統計数理,”新曜社, pp.60-62, 1980.

[8] 岩淵千明, “あなたもできるデータの処理と解析,”福村出版, pp.118, 1997.