

HYPERVARIABLE POLYMORPHIC VNTR LOCI FOR PARENTAGE TESTING AND INDIVIDUAL IDENTIFICATION

Tsuyoshi YOKOI,* Masayuki NATA, Toru ODAIRA, and Kaoru SAGISAKA

*Department of Forensic Medicine, Tohoku University School of Medicine,
Aoba-ku, Sendai 980, Japan*

Summary Three kinds of variable number of tandem repeat DNA probes (VNTR: pYNZ22, pYNH24, and pYNZ2) showing hypervariable polymorphisms were studied. Allelic frequencies and their confidence intervals among Japanese individuals were obtained. Co-dominant segregation of the polymorphism was confirmed in family studies. Two *a priori* probabilities were calculated for each VNTR locus: exclusion probabilities for an alleged father/mother/child trio and for an alleged parent/child duo, and probabilities of matching of genotyped two unrelated individuals or two siblings. Availability as well as highly discriminating polymorphic pattern of VNTR loci makes it potentially very useful for forensic and human genetic purposes.

Key Words hypervariable polymorphism, VNTR loci, parentage testing, individual identification

INTRODUCTION

Many DNA restriction fragment length polymorphisms (RFLPs) have been reported which recognize the presence or absence of particular restriction enzyme sites giving rise to two possible alleles at each locus (Cooper and Schmidtke, 1984). Another class of polymorphisms in hypervariable regions of DNA which resulted from a variable number of tandem repeats (VNTR). VNTR loci are valuable genetic markers for human linkage maps (Nakamura *et al.*, 1988), parentage testing (Jeffreys *et al.*, 1985a; Baird *et al.*, 1986) and individual identification (Jeffreys *et al.*, 1985b; Gill *et al.*, 1985; Giusti *et al.*, 1986). A common feature of these probes is the presence of a short core sequence element (*i.e.*, VNTR or hypervariable repeat, HVR) which is repeated in tandem along the chromosome and provides for the variable polymorphic character the locus (Nakamura *et al.*, 1987). However,

Received October 19, 1989; revised version received January 22, 1990; Accepted February 2, 1990.

* Present address: Division of Analytical Biochemistry, Faculty of Pharmaceutical Sciences, Hokkaido University, Kita-ku, Sapporo 060, Japan.

the analyses of DNA fingerprints with minisatellite probe showed many difficulties assessing which bands are allelic, and undefined chromosomal location provides statistical unsatisfactorily. We chose three kinds of VNTR probes introduced first by Nakamura for forensic studies on Japanese population. The VNTR probes, locating on different chromosome each other, were demonstrated extremely informative for paternity testing and individual identification through statistical analyses.

MATERIALS AND METHODS

Sample analyzed. DNA samples from more than 200 unrelated individuals and 20 two-generation Japanese families living in the north of Japan were analyzed. To prepare DNA from peripheral blood, buffy coat was mixed with several aliquots of 0.2% NaCl containing 10 mM ethylenediamine tetraacetic acid (EDTA) at 4°C. DNA was extracted by the standard technique (Maniatis *et al.*, 1982). The amount of DNA was measured by spectrophotometer scanning between 340 to 260 nm, and calculated using 260 nm absorbance. An optical density (OD) of 1 corresponded to 50 µg/ml DNA. At the same time, the purity of the sample was checked with the ratio between the absorbance at 260 and 280 nm (A_{260}/A_{280}).

DNA probes. VNTR probes (pYNH24, pYNZ22, and pYNZ2) developed originally by Nakamura *et al.* (1987) were kindly supplied from Japanese Cancer Research Resource Bank. The pYNH24 is 2.0 kb of HBV-1 fragment cloned in *AccI* of pUC18, pYNZ22 and pYNZ2 are 1.7 and 2.8 kb of human zeta globin fragment cloned in *BamHI* and *HindIII* site of pBR322, respectively. Chromosome location of pYNH24, pYNZ22, and pYNZ2 are 2q, 17p, and 1p, respectively.

Labeling of probes. Standard procedures were used for bacterial transformation and plasmid DNA isolation (Maniatis *et al.*, 1982). VNTR probes were labeled with [α - 32 P]dCTP (3,000 µCi/mmol) using the Multiprime DNA labeling system (Amersham, $2-3 \times 10^9$ cpm/µg of probe DNA). Two molecular weight standards were also labeled with the same [α - 32 P]dCTP according to the T4 DNA polymerase labeling method (Hartley and Gregori, 1981, 1×10^6 cpm/µg of marker DNA).

Southern hybridization. Three different restriction enzymes (REs) *HinfI* and *MspI* (Boehringer Mannheim, F.R.G.), *RsaI* (New England Biolab, U.K.) were used for the digestion of DNA specimens (2.5 µg of extracted DNA) and incubated at 37°C for 5-6 hr using 10-12 total units of REs per sample. The digested DNA fragments were loaded on a 24 cm-long horizontal 0.8% agarose gel (*MspI* digested sample probing with pYNH24) or a 1.2% agarose gel (*RsaI* digested sample probing with pYNZ2, *MspI* digested sample probing with pYNZ22). Electrophoresis was carried out in $1 \times$ TBE buffer (pH 8.0) at 1 V/cm of gel for approximately 30 hr. The sizes of polymorphic restriction fragments hybridizing to the probe were determined by comparison to the mobilities of DNA size markers. Two molecular weight standard markers were used for comparison purposes: 1) a 123-base-pair ladder (Bethesda Research Laboratories) and 2) *EcoT* 14 I fragments of lambda DNA (Takara) which reveal the migration of 19.33, 7.74, 6.22, 4.25, 3.47, 2.69, 1.88, 1.49,

0.93, and 0.42 kb, respectively (1 ng of DNA/well for the 123-base-pair ladder and 500 ng of DNA/well for lambda DNA). The loading was in triplicate on each gel, the two outside lanes on each side of the gel, and the two center lanes.

The DNA fragments were then transferred to a nitrocellulose membrane. Prehybridization was carried out at 42°C in a solution containing 50% formamide, 5×SSPE (1×SSPE=0.15 M sodium chloride, 0.01 M sodium dihydrogen phosphate 1 mM EDTA), 0.2 mg/ml salmon sperm DNA, 0.1% SDS, and 5×Denhardt's solution [1×Denhardt's solution=0.2 g/liter Ficoll (Sigma, type 400), 0.2 g/liter polyvinylpyrrolidone (PVP, Kodak), and 0.2 g/liter bovine serum albumin (Sigma, type V)]. The filters were hybridized with probes at 42°C for 12 hr in the same solution as prehybridization except that the concentration of salmon sperm DNA was reduced by half. The filters were washed twice at 50°C for 15 min with 0.1×SSPE, 0.1% SDS, followed with a solution in 0.1×SSPE for few minutes and then air dried. Autoradiography was carried out at -90°C using Fuji X-ray film for one day with two intensifying screens.

Standard paternity testing methods. Paternity testing methods routinely used are taken from standard protocols and include serologic analyses of selected red cell antigen systems (ABO, Rh-Hr, MNSs, P, Duffy and Kidd), red cell enzymes (phosphoglucomutase-1, esterase-D, acid phosphatase, phosphogluconate dehydrogenase), and serum proteins (group-specific component, Km allotype, Gm allotype, α_1 antitrypsin, transferrin and haptoglobin).

Statistical methods. The three VNTR genes were subdivided into 15, 28, and 5 non-overlapping groups of DNA fragments, respectively. Therefore the allele frequency values for VNTR were used directly for calculation. The physical size of the alleles for VNTR was found to vary by increments which were smaller than the size measurement error. The size of the alleles was measured at 100-bp increments (pYNH24 and pYNZ2) (partially 50-bp increment in pYNH24) or 70-bp increments (pYNZ22). To evaluate the precision of the estimates, we calculated 95% confidence intervals, taking into consideration that allelic frequencies of one locus are multinominally distributed. The score of χ^2 (chi-square) was calculated to examine that the population was in Hardy-Weinberg equilibrium.

The paternity exclusion (alleged father/mother/child) was calculated according to the calculation form of Komatu (1952). The paternity exclusion (alleged parent/child) and the probability of matching of 2 random persons and 2 siblings was calculated according to Selvin (1980) and Garber and Morris (1983). We calculated 95% confidence intervals of the paternity exclusion and the probability of matching using a bootstrap method (Efron, 1979).

RESULTS

Fifteen different alleles at pYNZ22 probe were found in a sample of 172 unrelated Japanese (Table 1) digested with *MspI*. Samples digested with *HinfI* also showed almost same polymorphisms. As the basic sequence of the repeat (70-bp

Table 1. Allele frequencies and confidence intervals observed by the combination with pYNZ22 probe and *MspI*.

Allele size (kilobases)	Frequency	Confidence 95%	Interval
1.98	0.009	0.000	0.019
1.70	0.020	0.005	0.035
1.49	0.035	0.015	0.054
1.42	0.029	0.011	0.047
1.35	0.058	0.033	0.083
1.28	0.049	0.027	0.072
1.21	0.058	0.033	0.083
1.14	0.064	0.038	0.090
1.07	0.096	0.065	0.127
1.00	0.125	0.090	0.160
0.93	0.108	0.075	0.140
0.86	0.113	0.080	0.147
0.79	0.099	0.067	0.130
0.72	0.096	0.065	0.127
0.65	0.041	0.020	0.062
Total 15 alleles	1.000		

Note. Gene frequencies are based on 172 unrelated Japanese individuals.

sequence) and flanking DNA were reported by Wolff *et al.* (1988), 70-bp increment was adopted to indicate the size fragment between 0.65 and 2.03 kilobase pairs.

Another VNTR probe, pYNH24, was expected to consist of a 30-base pair repeat unit according to the data of Nakamura (1987). Digesting with *MspI*, 28 size classes were adopted ranging from 1.40 to 4.90 kilobase pairs (Table 2) among 216 unrelated Japanese. Figure 1 shows an autoradiogram obtained from analysis of DNA from 10 unrelated individuals probing with pYNH24. We grouped restriction fragments into 100 bp size (partially 50 bp size) for establishing a frequency database. Other REs such as *PstI*, *BamHI*, *HindIII*, and *TaqI* also showed almost the same distribution patterns as *MspI*.

Using pYNZ22 probe, five different size alleles were found in a sample of 173 unrelated Japanese between 1.60 to 2.70 kilobase pairs range (Table 3) examined with *RsaI* digestion (Fig. 1). Relative allelic frequencies (%) for these three VNTR probes are schematically shown in Fig. 2. Relatively small range of allelic frequency was observed on pYNZ22 (0.133–0.006) and pYNH24 (0.106–0.002), while not on pYNZ2 (0.555–0.017). Confidential intervals for each allelic frequency are also indicated in Tables 1 to 3. Heterozygosities of pYNZ22, pYNH24, and pYNZ2 are 0.912, 0.944, and 0.569, respectively.

VNTR probes used in this study were located on different chromosome each

Table 2. Allele frequencies and confidence intervals observed by the combination with pYNH24 probe and *Msp*I.

Allele size (kilobases)	Frequency	Confidence	Interval
			95%
4.90	0.021	0.007	0.034
4.60	0.005	0.000	0.011
4.40	0.009	0.000	0.018
4.30	0.002	0.000	0.007
4.10	0.023	0.001	0.037
3.75	0.007	0.000	0.015
3.60	0.021	0.007	0.034
3.50	0.011	0.001	0.022
3.35	0.025	0.011	0.040
3.20	0.035	0.017	0.052
3.05	0.044	0.025	0.063
2.90	0.023	0.009	0.037
2.75	0.030	0.014	0.046
2.65	0.025	0.011	0.040
2.55	0.049	0.028	0.069
2.45	0.063	0.040	0.085
2.35	0.086	0.059	0.112
2.25	0.086	0.059	0.112
2.20	0.106	0.077	0.136
2.15	0.025	0.011	0.040
2.10	0.063	0.040	0.085
2.05	0.065	0.042	0.088
1.95	0.058	0.036	0.080
1.90	0.019	0.006	0.031
1.85	0.051	0.030	0.072
1.65	0.025	0.011	0.040
1.50	0.014	0.003	0.002
1.40	0.009	0.000	0.018
Total 28 alleles	1.000		

Note. Gene frequencies are based on 216 unrelated Japanese individuals.

other as shown in Table 4. The expected and observed frequencies of homozygotes and heterozygotes are summarized. The observed frequencies were very close to the expected ones. In Table 5, exclusion probability is shown when the both parent and the child are genotyped (AF/M/C) or when one parent and the child (P/C) are genotyped. The high scores of exclusion probability (pYNZ22, 0.826; pYNH24, 0.887) means the distinguished usefulness in parentage test and narrow

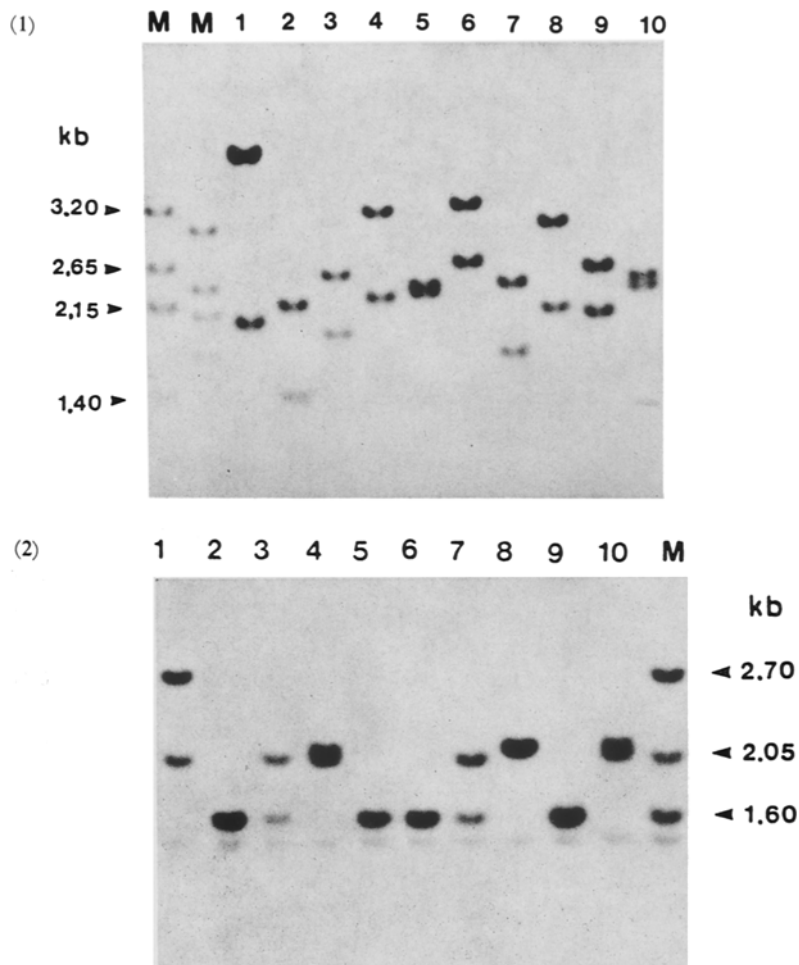


Fig. 1. Autoradiogram obtained from Southern blot analysis. DNA was collected from 10 unrelated individuals, digested with *Msp*I and probed with pYNH24 (1), digested with *Rsa*I, probed with pYNZ2 (2). The arrowheads indicate the size of bands. Fragment length was indicated in kb. The size was determined according to the method indicated in the text.

confidence intervals were observed (pYNZ22, $\pm 0.85\%$; pYNH24, $\pm 0.72\%$). Twenty cases of two generation families were investigated with VNTR probes, whose paternity relations were confirmed with the standard paternity testing method mentioned in materials and method. Every band of a child can be traced to either one of the parents. No mutations were observed. The cumulative exclusion probability was estimated to be 0.986.

As shown in Table 5, the probability of matching and the confidence interval (95%) were summarized which indicating the usefulness of this analysis for forensic

Table 3. Allele frequencies and confidence intervals observed by the combination with pYNZ2 probe and *RsaI*.

Allele size (kilobases)	Frequency	Confidence	Interval 95%
2.70	0.017	0.004	0.031
2.35	0.009	0.000	0.018
2.15	0.078	0.050	0.106
2.05	0.341	0.291	0.391
1.60	1.555	0.503	0.607
Total 5 alleles	1.000		

Note. Gene frequencies are based on 173 unrelated Japanese individuals.

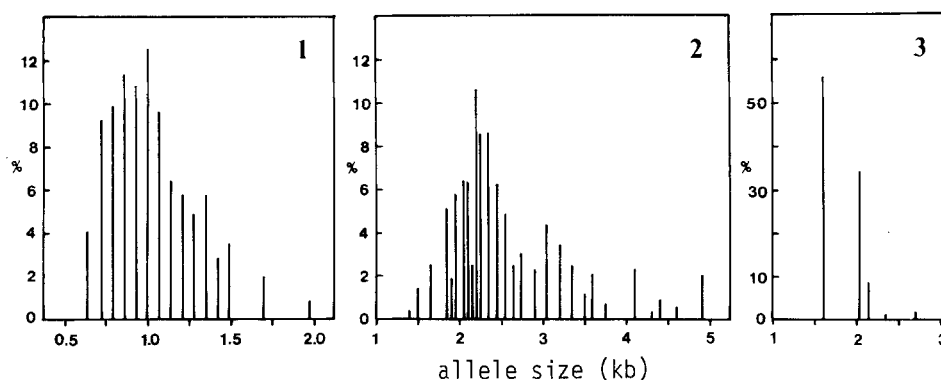


Fig. 2. Relative frequency (in percent) of alleles according to their size (kb) for three VNTR probes analyzed (1, pYNZ22-*MspI*; 2, pYNH24-*MspI*; 3, pYNZ2-*RsaI*).

individualization. From the three analyses, the probability that two randomly selected individuals will have the same genotypes was 2.16×10^{-5} . On the other hand, the probability that randomly chosen two siblings will have the same genotypes was 4.38×10^{-2} .

DISCUSSION

DNA probes are now starting to be used widely due to its hypervariable polymorphisms, especially in forensic individualization (Gill *et al.*, 1985, 1987; Jeffreys *et al.*, 1985b), and in paternity suits (Baird *et al.*, 1986; Jeffreys *et al.*, 1986). Almost all of them are multi-locus probes which recognize several VNTR loci under the conditions of reduced stringency. The single-locus probe is easily handled with under high stringency conditions and appears to be more sensitive and reproducible results than the multi-locus probes. The traditional statistical analyses of paternity and matching probabilities are not suitable for applying to the results of multi-locus probes due to the undefined chromosome location.

Table 4. Chromosomal location of characterized VNTR loci and ex-

Probe	Locus	Chromosome	Enzyme	Number of alleles	Number of individuals
pYNZ22	D17S30	17p	MspI	15	172
pYNH24	D2S44	2q	MspI	28	216
pYNZ2	D1S57	1p	RsaI	5	173

p, significant deviation of observed frequencies from expected frequencies. Homo., homozygotes; dom.

Table 5. Exclusion probabilities and probabilities of matching with confidence intervals calculated in three VNTR DNA systems.

Probe/Enzyme	Exclusion probability (AF/M/C)	Confidence interval 95%		Exclusion probability (P/C)	Confidence interval 95%	
pYNZ22/MspI	0.826	0.819	0.833	0.702	0.692	0.713
pYNH24/MspI	0.887	0.881	0.893	0.797	0.787	0.807
pYNZ2/RsaI	0.296	0.276	0.317	0.167	0.153	0.181
Cumulative	0.986	0.985	0.987	0.950	0.946	0.954

Probe/Enzyme	Probability of matching 2 random persons		Confidence interval 95%	
pYNZ22/MspI	1.38×10^{-2}		1.27×10^{-2}	1.49×10^{-2}
pYNH24/MspI	5.95×10^{-3}		5.27×10^{-3}	6.62×10^{-3}
pYNZ2/RsaI	0.263		0.241	0.283
Cumulative	2.16×10^{-5}		1.70×10^{-5}	2.62×10^{-5}

Probe/Enzyme	Probability of matching 2 siblings		Confidence interval 95%	
pYNZ22/MspI	0.296		0.294	0.298
pYNH24/MspI	0.279		0.278	0.281
pYNZ2/RsaI	0.531		0.515	0.547
Cumulative	4.38×10^{-2}		4.24×10^{-2}	4.56×10^{-2}

Note. AF, alleged father; M, mother; C, child; P, parent.

The probes used in this study were single-locus probes. We tested first seven kinds of VNTR probes (pYNZ22, pYNH24, pYNZ2, pMLJ14, pJCZ3.1, pEK-MDA2.1, and pCMM86) for genetic study on Japanese population. The latter four probes were cumbersome due to requirement of high concentration of human placental DNA (1.0–0.5 mg/ml) for blocking non-specific reactions. Although the pMLJ14 probe showed excellent polymorphisms with *RsaI*, the allele sizes

pected and observed frequencies of homozygotes and heterozygotes.

Homo.		Hetero.		χ^2	p (d.f.=1)
Obs.	Exp.	Obs.	Exp.		
17	14.7	155	157.3	0.39	0.50 < p < 0.75
16	12.1	200	203.9	1.33	0.10 < p < 0.25
73	74.6	100	98.4	0.06	0.75 < p < 0.90

Hetero., heterozygotes; Obs., observed frequencies; Exp., expected frequencies; d.f., degree of free-

were too large (showed polymorphisms at over 5 kb region). The former three probes were able to be used without special blocking and two of those gave high score (>90%) of heterozygosity. As each of the three probes lies on a different chromosome, they may be assumed to be statistically independent.

We adopted the indication of fragment size in 100-bp increment (partially 50-bp) in the case of pYNH24 probing. The alleles detected by pYNH24 probe have been grouped into 28 size classes among Japanese population, but these categories did not represent individual alleles. The recognition of the single repeat of the core sequence (30 base pair: Nakamura *et al.*, 1987) will be beyond the capability of agarose gel electrophoresis. This will be one of the reason getting relatively high score of homozygotes in individuals compared with that be expected. This probe has theoretically over 110 repeat units between 1.40 to 4.90, kb region but this way of 100-bp increment indication might be a conservative approach for standardization of the fragment size.

Jeffreys *et al.* (1988) reported high mutation rates of some 'fingerprinting' probes one of which, D1S7, demonstrated a mutation rate of 5.2% per gamete. On the other hand, Wolff *et al.* (1988) mentioned that the mutation rate of pYNZ22 was 0.2%. This kind of mutation is actually negligible for individual identification. Although VNTR loci showed more stable inheritance, it should be confirmed with other VNTR probes locating on the other chromosome in the cases of parental negation.

Recently, one of the VNTR probes, pYNZ22, was applied to analysis with polymerase chain reaction (Horn *et al.*, 1989) in which heterozygosity up to 86% was obtained without hybridization procedure. The polymerase chain reaction method on VNTR loci will be extensively and quickly applied to crime investigations.

Probes are easily available either from the ATCC or from the JCRB gene bank, Japan. Now, it can be used without constraint of patents. The combination of single-locus DNA polymorphisms getting from VNTR loci proved very powerful informations for parentage testing and individual identification which has been pointed out by Giusti *et al.* (1986) and Wainscoat *et al.* (1987). Using hypervariable VNTR loci, the power of parentage testing and individual identification are

extremely increased compared with those obtained from classical methods.

Acknowledgments This study was supported in part by the Chiyoda Mutual Life Foundation and the Ministry of Education, Science and Culture of Japan.

REFERENCES

- Baird, M., Balazs, I., Giusti, A., Miyazaki, L., Nicholas, L., Wexler, K., Kanter, E., Glassberg, J., Allen, F., Rubinstein, P. and Sussmann, L. 1986. Allelic frequency distribution of two highly polymorphic DNA sequences in three ethnic groups and its application to the determination of paternity. *Am. J. Hum. Genet.* **39**: 489-501.
- Cooper, D.N. and Schmidtke, J. 1984. DNA restriction fragment length polymorphisms and heterozygosity in the human genome. *Hum. Genet.* **66**: 1-16.
- Efron, B. 1979. Bootstrap methods: Another look at the jackknife. *Annu. Statist.* **7**: 1-26.
- Garber, R.A. and Morris, J.W. 1983. General equations for the average power of exclusion for genetic systems of n codominant alleles in one-parent and no-parent cases of disputed parentage. In: *Inclusion Probabilities in Parentage Testing*, Walker R.H., ed., American Association of Blood Banks, Arlington, Virginia, pp. 277-280.
- Gill, P., Jeffreys, A.J. and Werrett, D.J. 1985. Forensic application of DNA 'fingerprint.' *Nature* **318**: 577-579.
- Gill, P., Lygo, J.E., Fowler, S.J. and Werrett, D.J. 1987. An evaluation of DNA fingerprinting for forensic purposes. *Electrophoresis* **8**: 38-44.
- Giusti, A., Baird, M., Pasquale, S., Balazs, I. and Glassberg, J. 1986. Application of deoxyribonucleic acid (DNA) polymorphisms to the analysis of DNA recovered from sperm. *J. For. Sci.* **31**: 409-417.
- Horn, G.T., Richards, B. and Klinger, K.W. 1989. Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. *Nucleic Acids Res.* **17**: 2140.
- Jeffreys, A.J., Brookfield, J.E.Y. and Semenov, R. 1985a. Positive identification of an immigration test-case using human DNA fingerprints. *Nature* **317**: 818-819.
- Jeffreys, A.J., Wilson, V. and Thein, S.L. 1985b. Individual-specific 'fingerprints' of human DNA. *Nature* **316**: 76-79.
- Jeffreys, A.J., Wilson, V., Thein, J.L., Weatherall, D.J. and Ponder, B.A.J. 1986. DNA "fingerprints" and segregation analysis of multiple markers in human pedigrees. *Am. J. Hum. Genet.* **39**: 11-24.
- Jeffreys, A.J., Royle, N.J., Wilson, V. and Wong, Z. 1988. Spontaneous mutation rates of non-length alleles at tandem repetitive hypervariable loci in human DNA. *Nature* **332**, 278-281.
- Komatu, Y. 1952. Probability-theoretic investigations on inheritance. VII₂ non-paternity problems. *Proc. Jpn. Acad.* **28**: 105-108.
- Maniatis, T., Fritsch, F.F. and Sambrook, J. 1982. *Molecular Cloning. A Laboratory Manual (1981)*. Cold Spring Harbor Lab., Cold Spring Harbor, New York.
- Nakamura, Y., Leppert, M., O'Connell, P., Wolf, R., Holm, T., Culver, M., Martin, S., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**: 1616-1622.
- Nakamura, Y., Lathrop, M., O'Connell, P., Leppert, M., Barker, D., Lalouel, J.M. and White, R. 1988. A mapped set of DNA marker for human chromosome 17. *Genomics* **2**: 302-309.
- Selvin, S. 1980. Probability of nonpaternity determined by multiple allele codominant systems. *Am. J. Hum. Genet.* **32**: 276-278.
- Wainscoat, J.S., Pilkington, S., Peto, T.E.A., Bell, J.I. and Higgs, D.R. 1987. Allele-specific DNA identity patterns. *Hum. Genet.* **75**: 384-387.
- Wolf, R.K., Nakamura, Y. and White, R. 1988. Molecular characterization of a spontaneously generated new allele at a VNTR locus: No exchange of flanking DNA sequence. *Genomics* **3**: 347-351.