Int J Leg Med (1991) 104:307–308

International Journal of
Legal Medicine

On the statistics of the "genetic fingerprint"

H. Ritter

Institut für Anthropologie und Humangenetik, Universität Tübingen, Wilhelmstrasse 27, W-7400 Tübingen, Federal Republic of Germany

Received May 29, 1991 / Received in revised form October 17, 1991

Summary. In analogy to the polygene determined morphological features, the DNA-fingerprint is also not suitable for statistical processing. Statements about the individuality are merely speculative. Frequencies of genes cannot be found, since it is impossible to determine which combinations of bands belong to one gene locus. Hence the DNA fingerprint enables the recognition of exclusions from paternity; it does not, however, allow a statistical analysis, no matter which method be employed.

Key words: Statistics – Fingerprint – Probability of band sharing – Individuality

Zusammenfassung. Wie die polygen gesteuerten morphologischen Merkmale ist der „genetische Fingerabdruck" statistisch nicht analysierbar. Angaben über die Individualität sind daher rein spekulativ, da nicht bekannt ist, welche Banden zu einem Locus gehören. Der Fingerprint erlaubt damit lediglich die Erkennung von Ausschlußkonstellationen.

Schlüsselwörter: Statistik – Fingerabdruck – Bandsharing – Individualität

DNA-polymorphisms have been widely used in prenatal diagnosis, in localizing certain genes within the human genome, in analysing a number of cancers, as well as in paternity testing, and in forensic medicine.

These polymorphisms are examined by means of either single locus-probes that can detect only fragments of a single "locus" or multi locus-probes that can detect fragments occurring more than once in the DNA and are distributed over different chromosomes. Since many DNA-fragments per person are detected with the multi locus-probes, the expression "genetic fingerprint" was created (Jeffreys et al. 1985). It is claimed that the genetic fingerprint is suited to prove the genetic identity of a person and hence to be the ideal instrument for paternity testing and forensic medicine.

The probability of accidental bandsharing of two unrelated individuals is given as 2×10^{-8} for the probe/enzyme combination $(CAC)_5$/*Hin*fI (Schäfer et al. 1988) (Table 1).

The question is how reliable are the biostatistics of the fingerprint?

1. Does $\hat{P} = (2pq + q^2)^n$ provide a reliable probability?

The fingerprint represents a polygene system with an unknown number of loci and alleles. In order to test the usefulness of the suggested method of analysis we constructed an artificial print assuming 2 loci A and B with 2 alleles $p = 0.8$ and $q = 0.2$ each.

Table 2 shows an average number of bands of approximately $n = 3$ bands per individual.

So, following Schäfer et al., the probability P of a certain allele is

$$P = (2pq + q^2) = 0.32 + 0.04 = 0.36.$$

Consequently, the probability for a 3-band pattern is then $\hat{P} = 0.36^3 = 0.046$ according to this calculation. As a matter of fact, however, the real value is 0.4352.

Table 1. Variability of DNA fragments in random pairs of unrelated individuals (abridged from Schäfer et al. 1988)

Probe combination $(CAC)_5$/*Hin*fI	
DNA-fragment-range	4–27 kb
Number of individuals	16
Total number of bands	252
Average number of bands/indiv. = n	15.8
$P = (2q - q^2) = (2pq + q^2)$	0.33
$\hat{P} = (2pq + q^2)^n = (2pq + q^2)^{15.8}$	2×10^{-8}
Gene frequency q	0.182

P = mean band frequency = probability of finding an allele in two unrelated individuals

$\hat{P}$ = probability for the whole set of bands

Table 2. Fingerprint with 2 loci A and B with 2 alleles p = 0.8 and q = 0.2 each

Phenotype			B 1	B 2-1	B 2	
	Bands		1	2	1	
A 1	1	e	0.4096	0.2048	0.0256	0.64
		b	2	3	2	
A 2-1	2	e	0.2048	0.1024	0.0128	0.32
		b	3	4	3	
A 2	1	e	0.0256	0.0128	0.0016	0.04
		b	2	3	2	
			0.64	0.32	0.04	1.0

e = expected
b = number of bands

Table 3. Probabilities for the different combinations of bands under the assumption: 2 loci A and B with 2 alleles p and q at each locus

Phenotype			B	B 2-1	B 2
	Bands		1	2	1
A 1	1	e	p^4	$2p^3q$	p^2q^2
		b	2	3	2
A 2-1	2	e	$2p^3q$	$4p^2q^2$	$2pq^3$
		b	3	4	3
A 2	1	e	p^2q^2	$2pq^3$	q^4
		b	2	3	2

e = expected
b = number of bands

Therefore, in a polygene system the formula $\hat{P} = (2q - q^2)^n = (2pq + q^2)^n$ does not apply, as Table 3 shows.

The 3-band pattern has the frequency, as is seen, $4p^3q + 4pq^3$ in the case of different frequencies of p and q, according the example of Table 2, and never $(2pq + q^2)$.

It can be derived from Table 3 that the difference between $\hat{P}$ and the mean number of bands depends on the frequencies of the fragments. The rarer all fragments of a given locus are, the rarer the combinations $2pq \times p^2$ (homozygous for locus 1 and heterozygous for locus 2) and the commoner the combination $2pq \times 2pq$ (heterozygous for both loci).

2. Is there an average band frequency?

Schäfer et al. (1988) assumed, as did Jeffreys et al. (1985), that all identifiable bands of a fingerprint have approximately the same frequency. This presupposition is not fulfilled for reasons of population genetics since in our case:

$P_1 = (2pq + q^2) = 0.36$
$P_2 = (p^2 + 2pq) = 0.96.$

So, according to the reference band, the results are differing, unpredictable, and show unrelatable frequencies that render any further processing futile.

3. Identical alleles in two unrelated individuals?

"The probability P of finding an allele in two unrelated individuals 1 and 2 is $P = (2q - q^2)" = (2pq + q^2)$ (Schäfer et al. 1988). This statement is false.

If the mean frequency for one band $P = (2pq + q^2)$, it follows that:

$P = (2pq + q^2)$ = probability for 1 band of 1 person
$\hat{P} = (2pq + q^2)^n$ = probability for n bands of 1 person, and
$\hat{P}^2 = [(2pq + q^2)^n]^2$ = probability for n bands of 2 persons.

NB: $P = (pq + q^2)$ is better known as "Non-exclusion probability".

References

Jeffreys AJ, Brookfield JFY, Semeonoff R (1985) Possible identification of an immigration test-case using human DNA-fragments. Nature 317:818–819

Schäfer R, Zischler H, Birsner U, Becker A, Epplen JT (1988) Optimized oligonucleotide probes for DNA fingerprinting. Electrophoresis 9:369–374