On the question of the reliability of silent gene frequencies derived from maximum-likelihood estimates

K. Hummel and S. John

Prof. Dr. K. Hummel, Institut für Medizinische Mikrobiologie und Hygiene der Universität, Hermann-Herder-Str.11, D-7800 Freiburg

On account of the relative high frequency of homozygous people with blood group O or Rh-negative it does not cause any problems to determine frequencies of the silent genes O or d. But this is not the case in systems in which the silent gene is rare. In some cases proof of a man's paternity rests on both he and the child possessing a rare silent gene. In such cases the man will mostly be heterozygous with respect to this gene. In serostatistical terms this means that the triplet in question can give only a negative indication of paternity; moreover, the rarer the gene, the more negative the indication.

In contrast to variant genes, the biostatistical value of silent genes depends on the kind and frequency of the other alleles in the system. Given heterozygosis for the silent gene in child and putative father (pf) and a frequency of 0.01, one will obtain e.g.

in the Fy(a,b,O)system EM = 11.383, and
in the K(1,2,O)system EM = 10.580.

An EM limit for the inclusion or non-inclusion of a rare silent gene in a gene list presupposes that the other allele frequencies in the system are fixed. Alone the problems raised by aliens renders this impracticable. As a solution to the problem we propose the fiction of proof of the silent gene. In this event one would always obtain a positive indication. A frequency of 0.00135 for the silent gene would produce an indication of W = 99.73% and the corresponding predicate "practically proven". Rare silent genes could then be placed in the same category as rare variants: genes with frequencies below the proposed frequency limit of 0.00135 would not be included in gene lists for biostatistical tables, whereas those with frequencies above would (Hummel 1983).

The existence of a rare silent gene is shown by the occurrence of genetically "incompatible" mother-child combinations. The frequency of this gene can be calculated from the frequency of these combinations:

In a case of "incompatibility" let the genotype of the child be A_iO, that of the mother A_jO. Then $a_i \cdot a_j \cdot a_O$ will be the frequency of the "incompatible" mother-child doublet. The probability of such "incompatibilities" is

Advances in Forensic Haemogenetics 2
Edited by W.R. Mayr

$$p = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} a_i \cdot a_j \cdot a_o = a_o \sum_{i=1}^{n} a_i \left[\sum_{\substack{j=1 \\ j \neq i}}^{n} a_j + a_i - a_i \right]$$
$$\approx a_o \sum_{i=1}^{n} \left[a_i (1 - a_i) \right],$$

where n is the number of non-silent genes. The number of "incompatible" mother-child doublets, K, to be expected among N examined doublets is $K = p \cdot N$.
The frequency of the silent gene on the basis of found "incompatible" mother-child doublets will then be

$$a_o = K/N \cdot 1/ \sum_{i=1}^{n} \left[a_i (1 - a_i) \right] ,$$

where a_i (i=1,...,n) is the frequency of non-silent gene in the system.

So far, the existence of silent genes with a frequency greater than 0.00135 has been empirically established - i.e. using "incompatible" mother-child combinations - in the systems Duffy, GPT and Kell-Cellano. Current biostatistical tables take this into account (Hummel 1986). Accordingly, in these systems opposite homozygosity between child and pf no longer results in exclusion of the man from paternity; instead, the result is a high EM value. - One can also estimate the frequency of silent genes using the maximum-likelihood technique. If, using at least 3 phenotype frequencies, one determines not n but n+1 gene frequencies, then, as a rule, the Hardy-Weinberg law of panmixia will be better fulfilled if a silent gene is present than if not. -
Table 1 (see below) compares the results of an empirical survey on silent gene frequencies in Germany[1)] with those obtained from statistical estimates.

As the table shows, there are, for the most part, only minor differences between the empirically determined frequencies of silent genes and the estimates by maximum likelihood. -
Neither the estimates nor the empirical survey produced any indication of the existence of a silent gene of determinable frequency in the systems ADA, AK, C3, C6, F13B, Gc or 6-PGD (not in the table). The existence of silent genes has been proven for the systems acP, Kidd and Bf; however, both the estimated and empirical frequencies are lower than 0.00135.

In the Tf system the single occurrence of a mother-child "incompatibility" in 2,200 doublets leads mathematically to a relatively high frequency of the silent gene. But this frequency is not supported by the maximum-likelihood estimates. Hence, one may assume that the silent gene frequency in the Tf system lies below 0.00135. - In the GLO system, too, the calculations with the phenotypes produced a low silent gene frequency; however, among 12,177 mother-child doublets 10 were incompatible, giving a fre-

[1] For the kind cooperation we are very indebted to the colleagues Prof.Berg, Dr.Fingscheidt, Prof.Goedde, Prof.Hilgermann, PD Dr. Höher, Prof.Jürgens, Dr.Luboldt, Dr.v.Pritzbuer, Prof.Pulverer, Dr.Röhrborn, Dr.Werner, Prof.Wuermeling, Prof.Zang

quency of the silent gene of 0.0016, i.e. just above the limiting value of 0.00135. Nevertheless it seems not justified to include a silent gene in the GLO gene list.

In the systems Hp and Pi the M.-L.-estimates produced silent-gene frequencies of 0.0014 and 0.0065 respectively, i.e. above 0.00135. However, as the empirical survey does not support these values these systems do not have to be extended to include a silent gene either.

This leaves the systems PLG, EsD and PGM_1. In all three both estimates and empirical survey indicate a silent gene with a frequency of more than 0.00135. Hence, in new biostatistical tables these 3 systems should be expanded to include a silent gene, as was done some time ago in the cases of the systems Duffy, GPT and Kell-Cellano.

References

Hummel K (1983) Selection of gene frequency tables. In: Walker RH (ed) Inclusion probabilities in parentage testing. Am Ass Blood Banks, Arlington (Virg) 231-243

Hummel K (1986) Biostatistische Abstammungsbegutachtung - Biostatistical Opinion of Parentage - Ergänzung zu Tabellenband 1 26 polymorphe Systeme des Blutes, Privatdruck, Freiburg

Table 1. Frequency of silent genes in 10 different systems, determined a) by estimates using the maximum-likelihood method, and b) on the basis of the frequencies of "incompatible" mother-child combinations.[1,2].

run.no.	system	f(*O) max-likel.-estimates	no. of persons	chi^2	fg	$\bar{X}$ empir. mother-child	f(*O) determined "incomp."	$\bar{X}$ freq.of "incomp."	N doub-lets	k m.-ch. "incomp."
1	acP(A;B;C)	0.0000	1.800	4.872	1	0.00020	0.00060	0.00160	11.700	4
2	Tf(C1;C2;C3;B;D)	0.0000	1.017	0.00		0.00400	0.02200	0.08190	2.200	1
3	Jk(a;b)	0.0005	6.297	0.00		0.00024	0.00080	0.00198	11.000	4
4	GLO(1;2)	0.0006	5.862	0.00		0.00080	0.00160	0.00300	12.177	10
5	Bf(F;F_1;S;SO,7)	0.0008	1.414	3.69	2	0.00030	0.00090	0.00220	13.217	4
6	Hp(1;2)	0.0014	7.315	0.01		0.00026	0.00080	0.00198	11.000	4
7	PLG(A;B;M;A_V;B_V)	0.0052	1.490	5.17	3	0.00130	0.00390	0.00990	2.200	4
8	Pi(M1;M2;M3;S;Z;V)	0.0065	3.454	15.80	11	0.00004	0.00090	0.00480	2.717	1
9	EsD(1;2;5)	0.0074	1.825	1.96	1	0.00037	0.00138	0.00400	11.600	3
10	PGM_1(1;2)	0.0100	4.756	0.48		0.00150	0.00290	0.00510	11.500	12

[1] In order of increasing frequency of the silent gene (determined by estimates using the maximum-likelihood method)

[2] excluded the systems ADA, AK, C3, C6, F13B, Gc and 6-PGD systems, for which neither estimates nor empirical survey produced indications of a silent gene, as well as the MNSs system which has so far defied methods of simple calculation

VIII. Varia