Jpn. J. Human Genet. **29**, 371–380, 1984

A NOTE ON GENE FREQUENCY ESTIMATION IN THE ABO AND ABO-LIKE SYSTEM

Norikazu YASUDA

Division of Genetics, National Institute of Radiological Sciences, Chiba 260, Japan

Summary Based on a random sample of individuals, new conventional formulae for estimating gene frequency in the ABO and the ABO-like system are presented. Applying the formulae on a large number of samples, it is found that the new method is simple and yields estimates as accurate in at least two decimal places as the maximum likelihood solutions. The method however might be fraught with bias when the O phenotype frequency was low or even absent. Degrees of this bias have been discussed in terms of the number of iterative processes by a method of gene counting and of a goodness of fit to the Hardy-Weinberg proportions.

INTRODUCTION

Several authors have already developed formulae for estimating gene frequency in the ABO blood groups (Bernstein, 1925; Wiener *et al.*, 1929; Stevens, 1938; Fisher, 1946; Yasuda and Kimura, 1968). Although recent advents of electronic computer have made it feasible to calculate the ABO gene frequency as the maximum likelihood solution, a simple and biologically sound formula may be still of use. This note based on a method of gene counting (Ceppellini *et al.*, 1955) is an addition of conventionally useful formulae in the sense that the derived estimate is in good agreement with the maximum likelihood solution and that the calculations are manageable by a desk calculator without iteration. In the followings, a random sampling of individuals is assumed.

FORMULAE

The ABO system. Let p, q and r be allelic frequency of A, B and O, respectively. The observed numbers of four phenotypes in population are designated by the phenotype symbols themselves as O, A, B and AB so that the observed number

Received April 8, 1984

of sample is $T=O+A+B+AB$. The new formulae for estimating the ABO gene frequency are

$$\left.\begin{aligned} \hat{p} &= [1/2AB+A+O-\sqrt{O(A+O)}]/T, \\ \hat{q} &= [1/2AB+B+O-\sqrt{O(B+O)}]/T \quad \text{and} \\ \hat{r} &= 1-\hat{p}-\hat{q}. \end{aligned}\right\} \qquad (1)$$

A derivation of the formulae is given in appendix.

The ABO-like system. An extension of the ABO system to include an arbitrary number of codominant alleles has been made, and such a multiple allele locus has been called the ABO-like system (Yasuda and Kimura, 1968). Obviously, one of the most important examples in this genetic system is the HLA polymorphisms.

Suppose that m-1 codominant alleles, A_1, A_2, ..., A_{m-1} and a null allele O have the frequency p_1, p_2,..., p_{m-1} and r, respectively. Let the symbol of phenotypes designate for the observed number. The following formulae are obtained:

$$\left.\begin{aligned} \hat{p}_i &= [1/2H_i+A_i+O-\sqrt{O(A_i+O)}]/T \quad \text{for } i=1,2,\ldots, \text{m-1} \\ \hat{r} &= 1-\hat{p}_1-\hat{p}_2-\ldots-\hat{p}_{m-1} \end{aligned}\right\} \qquad (2)$$

where H_i is the sum of observed numbers of heterozygotes who carry A_i and the one of the other detectable antigens or mathematically, $H_i=\sum_{j\neq i} A_iA_j$.

The A_1A_2BO system. Suppose that p_1, p_2, q and r were allelic frequency of A_1, A_2, B and O, respectively. The new formulae will be

$$\left.\begin{aligned} \hat{p}_1 &= [1/2A_1B+A_1+A_2+O-\sqrt{(A_2+O)(A_1+A_2+O)}]/T \\ \hat{p}_2 &= [1-\sqrt{O/(A_2+O)}](1-\hat{p}_1-\hat{q}) \\ \hat{q} &= [1/2(A_1B+A_2B)+B+O-\sqrt{O(B+O)}]/T \quad \text{and} \\ \hat{r} &= [\sqrt{O/(A_2+O)}](1-\hat{p}_1-\hat{q}). \end{aligned}\right\} \qquad (3)$$

For a derivation, see appendix.

APPLICATIONS

The ABO system. Fujita *et al.* (1978) have reported extensive data on the ABO blood groups of all Japan. The total number of individuals, T=4,464,349, was classified into O=1,305,924, A=1,725,950, B=988,996 and AB=444,479. The conventional formulae (1) yielded $\hat{p}=0.28313$, $\hat{q}=0.17602$ and $\hat{r}=0.54085$, those of which could be comparable with the maximum likelihood solutions; $p=0.28312$, $q=0.17601$ and $r=0.544087$. In this particular example, the conventional estimates agree in four decimal places to the maximum likelihood solutions. The value of chisquare for a goodness of fit to the Hardy-Weinberg proportions is 1.1761 with one degree of freedom which is not significant.

Now, the following question would be of interest; how many number of iterative processes should it be necessary for arriving at the maximum likelihood so-

lution with allowing a given amount of error when one started at the trial value calculated by the new formulae and then applied a method of gene counting to it? In order to find a practical answer to this question, four hundreds and ninety-two independent subpopulations reported by Fujita *et al.* (1978) have subjected to analysis with an error criterion

$$E=|\hat{p}-p|+|\hat{q}-q|+|\hat{r}-r|<10^{-6}.$$

This infers that both estimates should agree in four decimal places. Results are shown in Table 1 in which the number of iterative processes is distributed in terms of the chisquare for a goodness of fit to the Hardy-Weinberg proportions (the corresponding probability is also given). Mean number of iterative processes and the standard deviation were 4.30 and 0.86, respectively. The mean number reduced to *one* and/or *two* when two decimal places were concerned in the estimates. This suggests that the new formulae (1) are very practical. Table 1 also indicates that there seems no association the number of iterative processes with the value of chisquare for a goodness of fit to the Hardy-Weinberg proportions. Incidentally, of eighteen percent (89/492) populations gathered from prefectural health departments, health centers, and Red Cross blood centers in the whole country, a goodness of fit to the Hardy-Weinberg proportions was significantly deviated at least five percent or higher level.

The ABO-like system. As an extension of the ABO system, HLA-DR antigen

Table 1. Distribution of the iterative number[a] and the chisquare value for a goodness of fit to the Hardy-Weinberg proportions in the ABO blood groups of 492 independent populations in Japan (After Fujita *et al.*, 1978).

Value of chisquare	Probability (%)	Iterative number							
		1	2	3	4	5	6	7	Total
0 - 0.45	100-50	2	13	51	103	15	—	—	184
0.45- 1.07	50-30	—	—	—	51	20	4	—	75
1.07- 1.64	30-20	—	—	—	26	26	3	—	55
1.64- 2.71	21-10	—	—	—	21	34	4	—	59
2.71- 3.84	10-5	—	—	—	8	19	2	1	30
3.84- 5.02	5-2.5	—	—	—	4	18	4	—	26
5.02- 6.63	2.5-1.0	—	—	—	3	11	2	—	16
6.63- 7.88	1.0-0.5	—	—	—	—	8	1	—	9
7.88-10.83	0.5-0.1	—	—	—	2	8	3	—	13
10.83- ∞	0.1-0	—	—	—	4	18	3	—	25
Total		2	13	51	222	177	26	1	492

[a] Error in estimate is less than 10^{-6}. Mean and the standard deviation of iterative numbers are 4.30 and 0.86, respectively.

Table 2. A data assembly and estimation of gene frequency in the HLA-DR system among Japanese (After Baur and Danilovs, 1980).

Antigen	(A_i)	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	O	H_i	A_i+O	$\sqrt{O(A_i+O)}$	$\hat{p}_i$	p_i	p_{bi}
DRW1	(A_1)	26	25	0	25	2	4	1	8	13	4	—	82	74	59.5986	0.0627	0.0629	0.0631
DRW2	(A_2)	25	109	5	80	6	25	1	23	43	1	—	209	157	86.8101	0.1976	0.1991	0.1998
DRW3	(A_3)	0	5	8	5	3	0	0	4	3	0	—	20	56	51.8459	0.0160	0.0160	0.0160
DRW4	(A_4)	25	80	5	146	9	16	4	28	51	2	—	220	194	96.4987	0.2347	0.2359	0.2345
DRW5	(A_5)	2	6	3	9	8	4	0	4	2	0	—	30	56	51.8459	0.0217	0.0217	0.0217
DRW6	(A_6)	4	25	0	16	4	9	1	18	3	0	—	71	57	52.3068	0.0455	0.0457	0.0463
DRW7	(A_7)	1	1	0	4	0	1	1	0	1	0	—	8	49	48.4974	0.0051	0.0051	0.0051
DRW8	(A_8)	8	23	4	28	4	18	0	14	11	1	—	97	62	54.5527	0.0633	0.0638	0.0649
DRW9	(A_9)	13	43	3	51	2	3	1	11	76	0	—	127	124	77.1492	0.1248	0.1243	0.1223
DRW10	(A_{10})	4	1	0	2	0	0	0	1	0	3	—	8	51	49.4773	0.0063	0.0063	0.0062
DRW-	(O)	—	—	—	—	—	—	—	—	—	—	48				0.2223	0.2192	0.2330
Total												(T=884)				1.0000	1.0000	1.0131

$\hat{p}_i$=estimated by the new conventional formulae (2) in the text. p_i=maximum likelihood solution. p_{bi}=estimated by an extended Bernstein's formulae as commonly used by HLA workers.

phenotypes in the HLA system on Japanese population (Baur and Danilovs, 1980) have been taken for an illustration. Table 2 shows a way of assembling data for the computation of the HLA-DR gene frequency. The row and column indicate the HLA-DR antigens and each entry of the table represents the observed number of phenotype. The diagonal entries are the observed number of individuals who carry a corresponding single antigen detected, while that of heterozygotes is given in the off-diagonal entries in duplicate. For instance, the number of individuals who carry HLA-DRW1 (A_1) and HLA-DRW1/DRW2 (A_1/A_2) are respectively 26 and 25, the number of heterozygotes having HLA-DRW1 (A_1) is $H_1=0+25+2+4+1+8+13+4=82$, and $A_1+O=26+48=74$. Agreement of the estimates with the maximum likelihood solutions is satisfactory at least in two decimal places. Another conventional estimates by the extended Bernstein's formulae, commonly used by HLA-worker, are also given in Table 2 for the sake of comparison. Both conventional methods lead to very similar solutions in this particular example.

A_1A_2BO blood groups. Four hundreds and eighty worldwide samples compiled by Mourant *et al.* (1976) were subjected to the new formulae (3). Having an error criterion

$$E=|\hat{p}_1-p_1|+|\hat{p}_2-p_2|+|\hat{q}-q|+|\hat{f}-r|<4\times10^{-5}$$

which assured four decimal places in accuracy, two populations showed extreme numbers of iterative processes (Table 3). One sample was Amish from Lancaster country of Pennsylvania (McKusick *et al.*, 1967) and comprised $O=23$, $A_1=140$,

Table 3. Distribution of the iterative number[a] and the value of chisquare for a goodness of fit to the Hardy-Weinberg proportions in the A_1A_2BO blood groups of 480 populations in the worldwide (After Table 1.2 of Mourant *et al.*, 1976).

Value of chisquare	Probability (%)	Iterative number														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
0 - 1.38	100-50	7	12	38	93	58	12	1	1	—	—	—	—	—	—	222
1.38- 2.40	50-30	—	—	9	32	24	8	1	—	—	—	—	—	—	—	74
2.40- 3.21	30-20	—	—	3	12	22	6	—	—	—	—	—	—	—	—	43
3.21- 4.60	20-10	—	1	3	18	17	9	2	—	—	—	—	1	—	—	51
4.60- 5.99	10-5	—	—	—	13	15	2	1	2	—	—	—	—	—	—	33
5.99- 7.37	5-2.5	—	—	1	1	7	2	1	1	—	—	—	—	—	—	13
7.37- 9.21	2.5-1.0	—	—	1	2	4	3	—	—	—	—	—	—	—	—	10
9.21-10.59	1.0-0.5	—	—	1	1	6	1	—	—	—	—	—	—	—	—	9
10.59-13.81	0.5-0.1	—	—	—	4	4	1	—	—	—	—	—	—	—	1	10
13.81- ∞	0.1-0	—	1	3	4	4	3	—	—	—	—	—	—	—	—	15
Total		7	14	59	180	161	47	6	4	—	—	—	1	—	1	480

[a] Error in estimate is less than 4×10^{-5}. Mean and the standard deviation of iterative numbers are 4.4 and 3.9, respectively.

$A_2=19$, $B=10$, $A_1B=15$ and $A_2B=8$ so that $T=215$. The formulae (3) yielded $\hat{p}_1=0.47474$, $\hat{p}_2=0.11606$, $\hat{q}=0.07884$ and $\hat{r}=0.33036$ so that $\chi^2=4.81$ with two degrees of freedom. By applying a method of gene counting (Yasuda and Kimura, 1968), in which a formula for A_2 was in error and should read as $p_2=[A_2+A_2B+A_2h_{A2}+A_1h_{A12}]/2T$ (h_{A12} is the probability that A_1 individual carries A_2 allele), twelve iterations were required for arriving at the maximum likelihood estimates; $p_1=0.47302$, $p_2=0.13473$, $q=0.07936$ and $r=0.31288$ with $\chi^2=3.54$. The other sample from India (Jalpainquri Toto) (Chaudhuri *et al.*, 1962) resulted 14 cycles of iteration. Here, we have $O=3$, $A_1=27$, $A_2=6$, $B=61$, $A_1B=11$ and $A_2B=7$ so that $T=115$. The conventional estimates by (3) were $\hat{p}_1=0.20435$, $\hat{p}_2=0.11892$, $\hat{q}=0.51429$ and $\hat{r}=0.16244$ with $\chi^2=17.80$ while the maximum likelihood solutions were $p_1=0.19106$, $p_2=0.08244$, $q=0.47049$ and $r=0.25601$ with $\chi^2=12.15$ which was highly significant. A large number of itereative processes here would be due to a low number of O phenotype, the point to be discussed later.

Mean and standard deviation of the number of iterative processes in this set of worldwide samples are 4.4 and 3.9, respectively. Again, no association was found between the number of iterative processes and the value for a goodness of fit to the Hardy-Weinberg proportions.

DISCUSSION

For all practical purposes, any conventional formula for estimating gene frequency should be manageable with a desk calculator and yields a biologically as well as statistically sound solution. The formulae presented in this paper must meet these conditions. However, during applications of the new formulae on random samples of ABO or ABO-like system, it was revealed that the frequency of O (or null) phenotype affected critically the accuracy of the estimates. That is, lower the frequency of O phenotype is, the large number of iterations is required. By any means, the O gene frequency could not be estimated from population data of the ABO system when no O individual was observed. Retabulating Table 1 according to the number of O phenotype in each population, a trend of negative association between the number of iterations and the observed number of O phenotype was noted (Table 4). The trend was enhanced in the Duffy blood groups where two codominant alleles, Fy^a and Fy^b, and a silent allele, Fy, corresponded to the allele A, B and O in the ABO blood groups, and the frequency of Fy(a−b−) phenotype is usually absent but is common in some groups (Race and Sanger, 1975). In Table 5 a result of analysis on twenty samples presented in Table 8.3.2 of Mourant *et al.* (1976) was tabulated in an ascending order of Fy(a−b−) proportion. The negative association is now very clear. The population, which was the lowest frequency of the silent phenotype in Table 5 (Bjarnason *et al.*, 1968), consisted of Fy(a−b−)=1, Fy(a+b−)=436, Fy(a−b+)=548 and Fy(a+b+)=703 so that $T=1,688$, and yielded by the formulae (1) $Fy^a=0.45474$, $Fy^b=0.51959$

Table 4. Distribution of the iterative number,[a] the number of O phenotype in the sample, and the value of chisquare for a goodness of fit to the Hardy-Weinberg proportions in the ABO blood groups of 492 independent populations in Japan (After Fujita *et al.*, 1978).

No. of O phenotype	Iterative number							
	1	2	3	4	5	6	7	Total
1- 10	—	—	—	1	—	1	1	3
11- 20	—	—	—	2	1	4	—	7
21- 50	—	—	—	—	2 (2)*	3	—	5 (2)
51- 100	—	1	—	7	5	4 (2)	—	17 (2)
101- 200	—	—	3	12	13 (1)	1 (2)	—	29 (3)
201- 500	—	1	6	32 (1)	44 (10)	— (5)	—	83 (16)
501-1,000	—	1	5	36	28 (17)	— (3)	—	70 (20)
1,001-2,000	1	6	10	52	14 (17)	— (1)	—	83 (18)
2,001-5,000	—	3	13	47 (4)	7 (12)	—	—	70 (16)
5,001-10^4	1	—	8	14 (5)	— (4)	—	—	23 (9)
10^4-∞	—	1	6	6 (3)	—	—	—	13 (3)
Total	2	13	51	209 (13)	114 (63)	13 (13)	1	403 (89)

[a] Error in estimate is less than 10^{-6}. * Number in parentheses is the number of populations which have shown a statistical significance at 5% or higher level for a goodness of fit to the Hardy-Weinberg proportions.

Table 5. The number of iterative processes, frequency of silent phenotype, and the value of chisquare for a goodness of fit to the Hardy-Weinberg proportions in the Duffy blood groups (After Table 8.3.2 of Mourant *et al.*, 1976).

Sample	Fy(a−b−)/T	(%)	no.[a]	χ^2(df=1)	Sample	Fy(a−b−)/T	(%)	no.[a]	χ^2(df=1)
1	1/1,688	0.06	46	8.71*	11	169/496	34.07	4	0.08
2	1/128	0.78	26	0.65	12	34/95	35.79	4	0.05
3	4/107	3.74	14	5.22*	13	142/236	60.17	3	1.57
4	4/99	4.07	12	0.16	14	108/179	60.33	4	5.69*
5	4/94	4.25	13	3.97*	15	59/97	60.82	3	0.75
6	4/77	5.19	12	4.26*	16	46/75	61.33	4	4.01*
7	2/27	7.40	10	1.11	17	85/125	68.00	3	0.09
8	11/88	12.50	8	3.04	18	226/236	95.76	2	163.67*
9	30/200	15.00	7	1.04	19	249/260	95.76	2	99.96*
10	58/174	33.33	5	2.52	20#	1,149/1,162	98.88	1	0.02

T=sample size. %=proportion of Fy(a−b−) phenotype in the sample. no.=number of iterative processes. χ^2=value of chisquare for a goodness of fit to the Hardy-Weinberg proportions. [a] Error in estimate is less than 10^{-6}. # The phenotype Fy(a+b+) was absent. * Significant at 5% or higher level.

Table 6. A comparison of gene frequencies estimated by three methods in the HLA-A system from an Okinawa study (After Yasuda and Tusji, 1975)

A	$\hat{p}_i$	p_i	p_{bi}
1	0.0346	0.0190	0.0174
2	0.5172	0.4211	0.4128
9	0.3103	0.2643	0.2300
10	0.1379	0.0924	0.1094
—	0.0000	0.2032	0.0000
Total	1.0000	1.0000	0.7696

A=HLA-A antigen. $\hat{p}_i$=estimated by the new formulae (2) in the text. p_i=maximum likelihood estimate. p_{bi}=estimated by an extended Bernstein's formula.

and Fy=0.02567 with $\chi^2=21.52$ while after 46 times of iteration the maximum likelihood solution accurate to four decimal places was $Fy^a=0.43959$, $Fy^b=0.50246$ and Fy=0.05796 with $\chi^2=8.72$. In this particular example, the Fy gene estimated by (1) was only of 44.5 percent that of the solution of maximum likelihood.

In the ABO-like system, the above problem would become rather serious, especially when no null phenotype is observed, but the estimation indicates the existance of null alleles in heterozygous conditions with more than two codominant alleles. Such a typical example in the HLA polymorphisms (Yasuda and Tsuji, 1975) yielded the result shown in Table 6. The estimated gene frequencies for four codominant alleles by the formulae (2) were far different from the maximum likelihood solutions. Rather, in this special example, the commonly used Bernstein's formula gave good estimates for these alleles, except the null allele. It is a matter of conjecture that whether implementation of heterozygous phenotypes in the new formulae, comparable with the Bernstein's formulae in which homozygous phenotypes were only taken into account, made this difference.

The situation seems to be the same even in A_1A_2BO blood groups as already mentioned. The new conventional formulae will give an underestimate for the O (or null) gene frequency when the number of O individuals was small in the sample. Nevertheless, the new formulae provide first order approximations and trial values for further improvement such as the maximum likelihood solutions by a gene counting which has never failed to converge, at least in the ABO and the ABO-like system examined so far.

Acknowledgments The author would like to thank Mr. Kanjiro Fukuhisa for his technical assistance in running computer programs on ACOS-700S installed in the National Institute of Radiological Sciences. For the excellent secretarial assistance the author is grateful to Mrs. Hiroko Itoh.

This study was supported in part by a grant for the Intractable Diseases from Ministry of Health and Welfare, Japan.

REFERENCES

Baur, M.P. and Danilovs, J.A. 1980. Population analysis of HLA-A, B, C, DR and other genetic markers. In *Histocompatibility Testing 1980*, Terasaki, P., ed., UCLA Tissue Typing Laboratory, Los Angeles, pp. 955–993.

Bernstein, F. 1925. Zusammenfassende Betrachtungen über die erblichen Blutstrukturen des Menschen. *Z. Indukt. Abstamm. -u. Vereblehre* **37**: 237–270.

Bjarnason, O., Bjarnason, U., Edwards, J.H., Fridriksson, S., Godber, M.J., Mourant, A.E., Tills, D., and Woodhead, B.G. 1968. Unpublished observatien (Cited in Mourant, A.E., Kopec, A.C., and Domaniewska-Sobczak, K., 1976).

Ceppellini, R., Siniscalco, M., and Smith, C.A.B. 1955. The estimation of gene frequencies in a random mating population. *Ann. Human Genet.* **20**: 97–115.

Chaudhuri, S., Chakuravartti, M.R., Mukherjee, B., Sen, S.N., Ghosh, J., and Maitra, A. 1964. A study of haematological factors, blood groups, anthropometric measurements and genetics of some of the tribal and caste groups of: South India-Kerala, Nilgiris and Andhra Predesh. 2. North eastern India (Indo-Bhutan border). *Totopara. 9th Congr. Int. Soc. Blood Transfus., Mexico 1962. Bibl. Haemat.* **19**: 196–205, 1964 (Cited in Mourant, A.E., Kopec, A.C., and Domaniewska-Sobczak, K., 1976).

Fisher, R.A. 1946. In Dobson, A.M. and Ikin, E.W. 1946. The ABO blood groups in the United Kingdom; frequencies based on a very large population. *J. Pathol. Bact.* **48**: 221–227.

Fujita, Y., Tanimura, M., and Tanaka, K. 1978. The distribution of the ABO blood groups in Japan. *Jpn. J. Human Genet.* **23**: 63–109.

McKusick, V.A., Bias, W.B., Norum, R.A., and Cross, H.E. 1967. Blood groups in two Amish demes. *Humangenetik* **5**: 36–41 (Cited in Mourant, A.E., Kopec, A.C., and Domaniewska-Sobczak, K., 1976).

Mourant, A.E., Kopec, A.C., and Domaniewska-Sobczak, K. 1976. *The Distribution of the Human Blood Groups and Other Polymorphisms* (2nd ed.), Oxford Univ. Press, London.

Race, R.R. and Sanger, R. 1975. Chapter 10. The Duffy blood groups. In *Blood Groups in Man* (6th ed.), Blackwell Scientific Publications, Oxford, pp. 350–363.

Stevens, W.L. 1938. Estimation of blood-group gene frequencies. *Ann. Eugen.* **8**: 362–375.

Wiener, A.S., Lederer, M., and Polayes, S.H. 1929. Studies in isoheamagglutination. I. Theoretical considerations. *J. Immun. (Balt.)* **16**: 469–482.

Yasuda, N. 1980. A note on gene frequency estimation in the ABO blood group system. *Ann. Rep. Natl. Inst. Radiol. Sci.* NIRS-20: 47–48.

Yasuda, N. and Kimura, M. 1968. A gene-counting method of maximum likelihood for estimating gene frequencies in ABO and ABO-like systems. *Ann. Human Genet., Lond.* **31**: 409–420.

Yasuda, N. and Tsuji, K. 1975. A counting method of maximum likelihood for estimating haplotype frequency in the HLA system. *Jpn. J. Human Genet.* **20**: 1–15.

APPENDIX

A derivation of the new conventional formulae

The ABO system. Using the notation in the text, the method of gene counting yields the followings:

$$\left.\begin{aligned} p &= [AB + A + Ah_A]/2T, \\ q &= [AB + B + Bh_B]/2T \\ \text{and} \quad r &= 1 - p - q. \end{aligned}\right\} \qquad \text{(A1)}$$

In which $h_A=p/(p+2r)$ and $h_B=q/(q+2r)$ (Yasuda and Kimura, 1968). Now, let us consider observable quantities

$$k_A=O/(A+O) \quad \text{and} \quad k_B=O/(B+O).$$

In terms of the expectation k's could be expressed as

$$k_A=Tr^2/[T(p^2+2pr)+Tr^2]=[r/(p+r)]^2$$
$$\text{and } k_B=[r/(q+r)]^2.$$

Then h's in (A1) can be expressed as

$$h_A=(1-\sqrt{k_A})^2/(1-k_A)$$
$$\text{and } h_B=(1-\sqrt{k_B})^2/(1-k_B).$$

Or, in terms of the observed number of phenotype, we have

$$\left.\begin{aligned} Ah_A &= A+2O-2\sqrt{O(A+O)} \\ \text{and} \quad Bh_B &= B+2O-2\sqrt{O(B+O)}. \end{aligned}\right\} \qquad \text{(A2)}$$

Substituting (A2) for (A1), the new formulae (1) in the text can be derived (Yasuda, 1980). The derivation of the formulae (2) in the ABO-like system is essentially the same.

A_1A_2 BO blood groups. For allele A_1 and B, the derivation is the same as that of the ABO system. Then the following simultaneous equations hold and can be solved in terms of p_2 and r;

$$\begin{cases} p_2+r=1-\hat{p}_1-\hat{q} \\ [r/(p_2+r)]^2=O/(A_2+O). \end{cases}$$

Namely,

$$\hat{r} = \sqrt{O/(A_2+O)}\,(1-\hat{p}_1-\hat{q})$$
$$\text{and } \hat{p}_2=[1-\sqrt{O/(A_2+O)}]\,(1-\hat{p}_1-\hat{q}).$$

In principle, the above procedure can be applied to any genetic system and would provide a simple and practical formula for gene frequency estimation.