自家不和合性遺伝子座における対立遺伝子の有効数,ヘテロザイゴシティの期待値および遺伝的距離を推定するための統計的方法

誌名	The Japanese journal of genetics
ISSN	0021504X
著者	田嶋, 文生 徳永, 徹 宮下, 直彦
巻/号	69巻3号
掲載ページ	p. 287-295
発行年月	1994年6月

Jpn. J. Genet. (1994) **69**, *pp.* 287–295

Statistical methods for estimating the effective number of alleles, expected heterozygosity and genetic distance in self-incompatibility locus

Fumio TAJIMA[1,*], Tohru TOKUNAGA[2,3] and Naohiko T. MIYASHITA[2]

[1]*Department of Population Genetics, National Institute of Genetics, Mishima, Shizuoka 411*
[2]*Laboratory of Genetics, Faculty of Agriculture, Kyoto University, Kyoto 606*
[3]*Kurume University Fusetsu High-School, Kurume, Fukuoka 830, Japan*

(*Received 14 February 1994*)

ABSTRACT

In order to understand the evolutionary process of self-incompatibility, we must know the genetic variability of the self-incompatibility genes within and between populations. Statistical methods for estimating the effective number of alleles, expected heterozygosity and genetic distance from pollination experiments were developed, which can be applied to both gametophytic and sporophytic self-incompatibility systems. In these methods, bud-pollination, which is necessary for obtaining homozygotes, is not required. Since bud-pollination, which is time-consuming, is not required in the present methods, they might be useful.

1. INTRODUCTION

In many plant species, self-incompatibility is one of the most important mechanisms that prevent self-fertilization and is controlled by a single locus with multiple alleles (S-alleles). Although phylogenetic relationships between S-alleles in various species have been investigated at the DNA level (Anderson et al., 1986; Haring et al., 1990; Ioerger et al., 1990; Clark and Kao, 1991), the attempt to quantify the genetic variation of the self-incompatibility genes in natural populations has not been done. In order to understand the evolutionary process, we must know the genetic variation in a population and the genetic divergence between populations (Fisher, 1930; Dobzhansky, 1970), and we can still obtain such information from pollination experiments (e.g., Levin, 1993; Lawrence et al., 1993; O'Donnell et al., 1993; Lane and Lawrence, 1993).

In this note we present how to estimate the effective number of alleles and expected heterozygosity in a population and the effective number of common alleles and genetic distance between populations from pollination experiments. In the present methods, bud-pollination, which is necessary for obtaining homozy-

* Corresponding author.

gotes, is not required. It should be noted that, as a measure of genetic variation, the effective number of alleles or the expected heterozygosity is more appropriate than the actual number of alleles, since the latter depends on the sample size (Kimura and Crow, 1964; Yokoyama and Hetherington, 1982). Since bud-pollination, which is time-consuming, is not required in the present methods, they might be useful.

In this note we consider both gametophytic and sporophytic self-incompatibilities. In the case of sporophytic self-incompatibility, we consider two cases: (i) alleles are codominant in both pollen and stigma, and (ii) one of alleles is dominant to the other in pollen, whereas alleles are codominant in stigma.

2. MODEL

In this note we consider only the case where crosses are made among heterozygotes randomly sampled from populations, since, under these self-incompatibility systems, the frequency of homozygote is either zero or very low. Namely, we assume the following pollination experiment: (i) crosses are made between random heterozygotes, and (ii) the proportions of successful (compatible) and unsuccessful (incompatible) crosses are obtained.

Within population

Let n be the number of alleles, and assume that the frequency of allele is the same among different alleles. We denote the ith allele by S_i.

When a cross is made between two heterozygotes randomly chosen from a population, there are three possible cases: $S_iS_j \times S_kS_l$, $S_iS_j \times S_iS_k$ and $S_iS_j \times S_iS_j$ where i, j, k and l are mutually different. The probabilities of having these cases can be given by

$$\text{Prob}\{S_iS_j \times S_kS_l\} = \frac{(n-2)(n-3)}{n(n-1)}, \tag{1}$$

$$\text{Prob}\{S_iS_j \times S_iS_k\} = \frac{4(n-2)}{n(n-1)}, \tag{2}$$

$$\text{Prob}\{S_iS_j \times S_iS_j\} = \frac{2}{n(n-1)}. \tag{3}$$

In natural populations, the frequency of allele may not be the same among different alleles. Since we assumed that the frequency of allele is the same among different alleles, we call n the effective number of alleles, which is different from the actual number of alleles unless the frequency of allele is the same among different alleles. Once the effective number of alleles is estimated, we can estimate the expected homozygosity and heterozygosity by

$$H_o = 1/n \quad \text{and} \quad H_e = 1 - 1/n, \tag{4}$$

respectively (Kimura and Crow, 1964). As will be shown in NUMERICAL EXAMPLES, these formulae give estimates which is close to the actual estimates of H_o and H_e.

Between populations

Let us now consider the case where two populations, say populations 1 and 2, are studied. Let n_1 and n_2 be the number of alleles in populations 1 and 2, n_0 be the number of alleles which are common in both populations. When a cross is made between two heterozygotes randomly chosen from two populations, one from population 1 and one from population 2, there are three possible cases as in the case of within-population. The probabilities of having these cases can be given by

$$\text{Prob}\,\{S_iS_j \times S_kS_l\} = 1 - \frac{4n_0}{n_1n_2} + \frac{2n_0(n_0-1)}{n_1(n_1-1)n_2(n_2-1)}, \tag{5}$$

$$\text{Prob}\,\{S_iS_j \times S_iS_k\} = \frac{4n_0}{n_1n_2} - \frac{4n_0(n_0-1)}{n_1(n_1-1)n_2(n_2-1)}, \tag{6}$$

$$\text{Prob}\,\{S_iS_j \times S_iS_j\} = \frac{2n_0(n_0-1)}{n_1(n_1-1)n_2(n_2-1)}. \tag{7}$$

Since we assumed that the frequency of allele is the same among different alleles in each population, we call n_0 the effective number of common alleles. When the frequencies of the ith allele are x_i and y_i in populations 1 and 2, respectively, the genetic distance between populations 1 and 2 is defined as

$$D = -\log_e(J_0/\sqrt{J_1J_2}), \tag{8}$$

where

$$J_0 = \sum_i x_iy_i, \quad J_1 = \sum_i x_i^2 \quad \text{and} \quad J_2 = \sum_i y_i^2 \tag{8a}$$

(Nei, 1972, 1975). Since $x_i = 1/n_1$ and $y_i = 1/n_2$, we have $J_0 = n_0/(n_1n_2)$, $J_1 = 1/n_1$ and $J_2 = 1/n_2$, so that the genetic distance is given by

$$D = -\log_e(n_0/\sqrt{n_1n_2}). \tag{9}$$

This formula can be used for estimating the genetic distance between populations 1 and 2.

3. GAMETOPHYTIC SELF-INCOMPATIBILITY

In gametophytic self-incompatibility system, all pollen is incompatible when crosses are made between S_iS_j and S_iS_j, and a half of the pollen is incompatible when crosses are made between S_iS_j and S_iS_k. On the other hand, all pollen is compatible in crosses $\{S_iS_j \times S_kS_l\}$. These three cases are distinguishable in

pollination experiment.

In the case where heterozygotes are sampled from a population, from (2) and (3) the proportion of incompatible pollen is given by

$$P_{\mathrm{I}} = \frac{1}{2}\frac{4(n-2)}{n(n-1)} + \frac{2}{n(n-1)} = \frac{2}{n}. \tag{10}$$

Therefore, n can be estimated by

$$n = \frac{2}{P_{\mathrm{I}}}, \tag{11}$$

and the expected homozygosity and heterozygosity can be estimated by (4).

When crosses are made between populations 1 and 2, from (6) and (7) the proportion of incompatible pollen is given by

$$P_0 = \frac{2n_0}{n_1 n_2}, \tag{12}$$

where n_1 and n_2 are given by (11). Then, n_0 can be estimated by

$$n_0 = \frac{n_1 n_2 P_0}{2} = \frac{2P_0}{P_1 P_2}, \tag{13}$$

where P_1 and P_2 are the proportions of incompatible pollen in populations 1 and 2, respectively. The genetic distance between populations 1 and 2 can be estimated by (9).

4. SPOROPHYTIC SELF-INCOMPATIBILITY

Model (i)

In this model we assume that alleles are codominant in both pollen and stigma, so that crosses $\{S_iS_j \times S_iS_k\}$ and $\{S_iS_j \times S_iS_j\}$ are incompatible.

In the case where heterozygotes are sampled from a population, from (2) and (3) the proportion of incompatible crosses is given by

$$P_{\mathrm{I}} = \frac{4(n-2)}{n(n-1)} + \frac{2}{n(n-1)} = \frac{2(2n-3)}{n(n-1)}. \tag{14}$$

Therefore, n can be estimated by

$$n = \frac{P_{\mathrm{I}} + 4 + \sqrt{P_{\mathrm{I}}^2 - 16P_{\mathrm{I}} + 16}}{2P_{\mathrm{I}}}, \tag{15}$$

and the expected homozygosity and heterozygosity can be estimated by (4).

When crosses are made between populations 1 and 2, from (6) and (7) the proportion of incompatible crosses is given by

$$P_0 = \frac{4n_0}{n_1 n_2} - \frac{2n_0(n_0-1)}{n_1(n_1-1)n_2(n_2-1)}, \tag{16}$$

so that n_0 can be estimated by

$$n_0 = \frac{a - \sqrt{a^2 - bP_0}}{2}, \tag{17}$$

where

$$a = 2(n_1 - 1)(n_2 - 1) + 1 \quad \text{and} \quad b = 2n_1(n_1 - 1)n_2(n_2 - 1), \tag{17a}$$

and n_1 and n_2 can be estimated by (15). The genetic distance between populations 1 and 2 can be estimated by (9).

Model (ii)

In this model we assume that one of alleles is dominant to the other in pollen, whereas alleles are codominant in stigma. [This model also includes the case where one allele is dominant in stigma with codominance in pollen.] Although there are homozygotes in a population, we ignore them since the frequency of homozygote is very low unless the number of alleles is very small. Under this model, crosses $\{S_iS_j \times S_iS_j\}$ are incompatible and crosses $\{S_iS_j \times S_kS_l\}$ are compatible. When pollen from S_iS_j is crossed to S_iS_k stigma, the cross is incompatible if S_i is dominant to S_j, whereas it is compatible if S_j is dominant to S_i. Therefore, a half of crosses $\{S_iS_j \times S_iS_k\}$ are assumed to be incompatible.

In the case where heterozygotes are sampled from a population, from (2) and (3) the proportion of incompatible crosses is given by

$$P_{\mathrm{I}} = \frac{1}{2}\frac{4(n-2)}{n(n-1)} + \frac{2}{n(n-1)} = \frac{2}{n}, \tag{18}$$

so that n can be estimated by

$$n = \frac{2}{P_{\mathrm{I}}}, \tag{19}$$

and the expected homozygosity and heterozygosity can be estimated by (4).

When crosses are made between populations 1 and 2, from (6) and (7) the proportion of incompatible crosses is given by

$$P_0 = \frac{2n_0}{n_1 n_2}, \tag{20}$$

where n_1 and n_2 are given by (19). Then, n_0 can be estimated by

$$n_0 = \frac{n_1 n_2 P_0}{2} = \frac{2P_0}{P_1 P_2}, \tag{21}$$

where P_1 and P_2 are the proportions of incompatible crosses in populations 1 and 2, respectively. The genetic distance between populations 1 and 2 can be estimated by (9).

5. NUMERICAL EXAMPLES

Here we examine two cases, one for gametophytic self-incompatibility and one for sporophytic self-incompatibility.

Gametophytic self-incompatibility

Using bud-pollination, Levin (1993) determined the genotypes of 24 plants sampled from a population of *Phlox drummondii*, which are shown in Table 6 of Levin (1993).

Let us assume that we do not know the genotypes and that all possible crosses (552 crosses) are made among these plants without bud-pollination. Then, we do not expect any case where all pollen is incompatible, whereas we expect that a half of the pollen is incompatible in 46 crosses (Table 1). Therefore, the proportion of incompatible pollens is estimated to be $P_I = 46/(552 \times 2) = 0.0417$, so that we obtain $n = 48.0$ from (11) and $H_o = 0.0208$ and $H_e = 0.9792$ from (4).

From Table 6 of Levin (1993), we can estimate the frequencies of 30 alleles. Using these frequencies, we can estimate the expected homozygosity and heterozygosity and the effective number of alleles, i.e., $H_o = 0.0204$, $H_e = 0.9796$ and $n = 49.0$. These values are close to the estimates obtained by using the present method.

Table 1. Crosses in which a half of the pollen is incompatible among 24 plants of *Phlox drummondii*, obtained from Table 6 of Levin (1993)

Stigma	Pollen	Stigma	Pollen	Stigma	Pollen
1	6 10 18 19	10	1 19	19	1 9 10 24
2	24	11	3 4	20	12 15 18
3	11 22 23	12	15 20	21	5 17
4	11	13	7	22	3 23
5	21	14	9	23	3 22
6	1 18	15	12 20	24	2 9 19
7	13	17	21		
9	14 19 24	18	1 6 20		

Sporophytic self-incompatibility

Ford and Kay (1985) and Stevens and Kay (1989) estimated the number of alleles as well as their frequencies and dominance relationships in a natural population of *Sinapis arvensis* by using bud-pollination technique. The allele composition of a sample of 35 plants is shown in Table 1 of Stevens and Kay (1989).

Table 2. Incompatible crosses among 35 plants of *Sinapis arvensis*, obtained from Table 1 of Stevens and Kay (1989)

Stigma	Pollen	Stigma	Pollen
2	1 7 8	19	17 34
3	4	20	10
4	3	21	10 22 23
5	6 12 13 14	22	10 12 15 21 23 30 31 32
6	5 13 14 17	23	10 21 22
7	8	24	28
8	7	26	17
10	20 21 22 23	27	18
12	5 15 22 30 31 32	28	16 24
13	5 6 14 17	30	12 15 22 31 32
14	5 6 13 33	31	12 15 22 30 32
15	12 22 30 31 32	32	12 15 22 30 31
16	28	33	14
18	17 27	34	19

Let us now assume that we know neither the genotypes of plants nor the dominance relationships, and that all possible crosses (1190 crosses) are made among these plants without bud-pollination. The results expected under these assumptions are given in Table 2, where only incompatible crosses are shown. There are 80 incompatible cases, so that the proportion of incompatible crosses is $P_{\mathrm{I}} = 80/1190 = 0.0672$.

In order to estimate the effective number of alleles and the expected homozygosity and heterozygosity, we must know the dominance relationships. If we assume that alleles are codominant in both pollen and stigma (Model i), we obtain $n = 59.0$ from (15), and $H_{\mathrm{o}} = 0.0170$ and $H_{\mathrm{e}} = 0.9830$ from (4). On the other hand, if we assume either that one of alleles is dominant in pollen with codominance in stigma or that one allele is dominant in stigma with codominance in pollen (Model ii), we obtain $n = 29.8$ from (19), and $H_{\mathrm{o}} = 0.0336$ and $H_{\mathrm{e}} = 0.9664$ from (4). Thus, the values of estimates depend on the model used.

Stevens and Kay (1989) observed codominance in both pollen and stigma (in 43% of heterozygotes), dominance of one allele in pollen with codominance in stigma (in 50% of heterozygotes), dominance of one allele in stigma with codominance in pollen (in 5% of heterozygotes), and dominance of one allele in both pollen and stigma (in 2% of heterozygotes). This means that the true situation might be located between Models (i) and (ii). If we assume that Models (i) and (ii) are correct each in 50% of the cases, P_{I} is given by

$$P_{\mathrm{I}} = \frac{1}{2}\frac{2(2n-3)}{n(n-1)} + \frac{1}{2}\frac{2}{n} = \frac{3n-4}{n(n-1)}. \tag{22}$$

Then, n can be estimated by

$$n = \frac{P_I + 3 + \sqrt{P_I^2 - 10P_I + 9}}{2P_I}. \tag{23}$$

In this example, we have $n = 44.3$, so that we obtain $H_o = 0.0226$ and $H_e = 0.9774$.

Another way to estimate the effective number of alleles is to use the weighted average of two estimates. In this example, the effective number of alleles is estimated as $n = (59.0 + 29.8)/2 = 44.4$, so that we have $H_o = 0.0225$ and $H_e = 0.9775$. These estimates are essentially the same as the above ones.

Figure 2 of Stevens and Kay (1989) gives the frequency distribution of 35 alleles obtained by using bud-pollination. From this figure we can estimate the expected homozygosity and heterozygosity, and the effective number of alleles, i.e., $H_o = 0.0257$, $H_e = 0.9743$ and $n = 39.0$. These values are not very different from the estimates obtained by using the present method.

6. DISCUSSION

In this note we present the methods for estimating the effective number of alleles and the expected homozygosity and heterozygosity in a population, and the effective number of common alleles and the genetic distance between populations. Since bud-pollination is not required in the present methods, they might be useful as long as we know the system of self-incompatibility.

Although we considered three models, there might be several other models. For example, Lawrence (1975) presented, in addition to the models we considered, another model which assumes that one allele is dominant in both stigma and pollen. Under this model, high frequency of recessive homozygote is expected (Cope, 1962), so that we cannot use the present method which ignores the effect of homozygotes. In this model, however, we can obtain different information about genetic variation.

Let n_p be the number of phenotypes and assume that the frequency of phenotype is the same among different phenotypes (Cope, 1962). As before, we call n_p the effective number of phenotypes. Since a cross is incompatible only when the cross is made between the same phenotype, the proportion of incompatible crosses is given by

$$P_I = \frac{1}{n_p}, \tag{24}$$

so that n_p can be estimated by

$$n_p = \frac{1}{P_I}. \tag{25}$$

It should be noted, however, that the effective number of phenotypes is not equal to the effective number of alleles, so that we cannot estimate the expected heterozygosity and homozygosity in this case.

We thank Drs. K. Tsunewaki and M. K. Uyenoyama for their valuable suggestions and comments. We also thank two anonymous reviewers for their comments. This research is partly supported by NIG Cooperative Research Program ('93-30). This is contribution no. 1988 from the National Institute of Genetics, Mishima, Shizuoka 411, Japan.

REFERENCES

Anderson, M. A., Cornish, E. C., Mau, S.-L., Williams, E. G., Hoggart, R., Atkinson, A., Bonig, I., Grego, B., Simpson, R., Roche, P. J., Haley, J. D., Penschow, J. D., Niall, H. D., Tregear, G. W., Coghlan, J. P., Crawford, R. J. and Clarke, A. E. (1986). Cloning of cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*. *Nature* **321**, 38–44.

Clark, A. G. and Kao, T.-h. (1991). Excess nonsynonymous substitution at shared polymorphic sites among self-incompatibility alleles of Solanaceae. *Proc. Natl. Acad. Sci. USA* **88**, 9823–9827.

Cope, F. W. (1962). The effects of incompatibility and compatibility on genotype proportions in populations of *Theobroma cacao* L. *Heredity* **17**, 183–195.

Dobzhansky, Th. (1970). *Genetics of the Evolutionary Process*. Columbia University Press, New York.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Ford, M. A. and Kay, Q. O. N. (1985). The genetics of incompatibility in *Sinapis arvensis* L. *Heredity* **54**, 99–102.

Haring, V., Gray, J. E., McClure, B. A., Anderson, M. A. and Clarke, A. E. (1990). Self-incompatibility: a self-recognition system in plants. *Science* **250**, 937–941.

Ioerger, T. R., Clark, A. G. and Kao, T.-h. (1990). Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *Proc. Natl. Acad. Sci. USA* **87**, 9732–9735.

Kimura, M. and Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

Lane, M. D. and Lawrence, M. J. (1993). The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. VII. The number of S-alleles in the species. *Heredity* **71**, 596–602.

Lawrence M. J. (1975). The genetics of self-incompatibility in *Papaver rhoeas*. *Proc. R. Soc. Lond. B* **188**, 275–285.

Lawrence, M. J., Lane, M. D., O'Donnell, S. and Franklin-Tong, V. E. (1993). The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. V. Cross-classification of the S-alleles of samples from three natural populations. *Heredity* **71**, 581–590.

Levin, D. A. (1993). *S*-gene polymorphism in *Phlox drummondii*. *Heredity* **71**, 193–198.

Nei, M. (1972). Genetic distance between populations. *Am. Nat.* **106**, 283–292.

Nei, M. (1975). *Molecular Population Genetics and Evolution*. North-Holland, Amsterdam.

O'Donnell, S., Lane, M. D. and Lawrence, M. J. (1993). The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. VI. Estimation of the overlap between the allelic complements of a pair of populations. *Heredity* **71**, 591–595.

Stevens, J. P. and Kay, Q. O. N. (1989). The number, dominance relationships and frequencies of self-incompatibility alleles in a natural population of *Sinapis arvensis* L. in South Wales. *Heredity* **62**, 199–205.

Yokoyama, S. and Hetherington, L. E. (1982). The expected number of self-incompatibility alleles in finite plant populations. *Heredity* **48**, 299–303.