Syst. Zool., 37(2):156–162, 1988

NEI'S MODIFIED GENETIC IDENTITY AND DISTANCE MEASURES AND THEIR SAMPLING VARIANCES*

JÜRGEN TOMIUK[1] AND DAN GRAUR[2,3]

[1]*Lehrstuhl für Populationsgenetik, Institut für Biologie II, Universität Tübingen, Auf der Morgenstelle 28, D-7400 Tübingen, West Germany; and*
[2]*Department of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel*

Abstract.—The conditions under which Nei's (1972) genetic identity measure (I) yields results which are discordant with changes recorded in the gene identities at single loci are defined. We noticed that upon reassessment of allele frequencies, the value of I can in some cases change in the opposite direction of changes recorded in single locus gene identities. This anomaly may affect phylogenetic reconstructions especially when closely related populations and/or rare alleles are involved. We illustrate this problem using two examples, one based on real electrophoretic data from two *Macaca* species, the other based on hypothetical allele frequencies. We propose to use instead of Nei's I, an alternative measure, which we call Nei's modified genetic identity ($\hat{I}$). This measure is based on the arithmetic mean of single locus gene identities. Nei's modified distance ($\hat{D}$) is derived analogically to Nei's D. We present the sampling variances of these modified estimates. [Nei's distance; allele frequencies; philogeny; electrophoresis.]

The most widely used genetic distance estimate is Nei's (1972) D. Its popularity stems from its simplicity and facility of application (Hedrick, 1983; Kimura, 1983). Nei's D has also been shown to have an approximate linear relationship with time of divergence (Nei, 1987). However, during a reassessment of data on the gene diversity at the amylase and hemoglobin loci in two species of macaques (Tomiuk, unpubl.), we noticed that it is possible to record a decrease in the single locus gene identity at one or several loci, and concomitantly record an increase in the overall gene identity across all loci between the two species. This means that despite of a larger divergence at single loci, the total genetic identity may in some cases assume higher values, thus indicating undue genetic similarity.

In this note we assess the conditions under which this type of anomaly may occur. Furthermore, we propose modified versions of Nei's genetic identity and Nei's genetic distance, as suggested but not recommended by Nei in his 1972 paper, and presented again in Hillis (1984). In this note we also calculate the sampling variances of the modified formulae.

DEFINITION

Nei's genetic identity (I) between two populations is defined as:

$$I = J_{12}/(J_1 J_2)^{1/2}, \quad (1)$$

where

$$J_1 = \left(\sum_j^r \sum_i^{l_j} x_{ji}^2\right),$$

$$J_2 = \left(\sum_j^r \sum_i^{l_j} y_{ji}^2\right),$$

$$J_{12} = \left(\sum_j^r \sum_i^{l_j} x_{ji} y_{ji}\right),$$

r is the number of loci analysed in both populations, l_j is the number of alleles at the j-th locus, and x_{ji} and y_{ji} represent the frequencies of the i-th allele at the j-th locus in populations 1 and 2, respectively. Note that in formula (1), J_1, J_2 and J_{12} are defined as sums of allele frequencies, and not as averages over all loci as in Nei (1972). However, this difference is immaterial in

* Dedicated to Prof. K. Wöhrmann on his sixtieth birthday.

[3] To whom all correspondence should be addressed.

the present context since r cancels out in the formula.

THE PROBLEM

In the following we shall consider the reassessment problem. Namely, we shall consider two sets of data: an old set and a new one. The new set may be generated, for instance, by introducing improvements in the resolving power of the electrophoretic techniques involved, and thus usually it is expected that the new set will show more polymorphism than the old one (Graur, 1986). Alternatively, the new set may be generated by increasing the sample size in one of both populations, and in this case the degree of polymorphism may change in either direction. For the sake of simplicity we shall assume in the following that only the data pertaining to one locus are reassessed at a time, and that in each population, at the reassessed locus, there can be a maximum of only two alleles segregating.

Let a_{11} be the old frequency of allele A in population 1, and let a_{21} be the old frequency of this allele in population 2. The old frequency of the alternate alleles in populations 1 and 2 are $a_{12} = 1 - a_{11}$ and $a_{22} = 1 - a_{21}$, respectively. The alternate alleles may or may not be the same in both populations.

Let us now assume that upon reassessment the new frequencies of allele A in populations 1 and 2 turn out to be x_{11} and x_{21}, respectively. Obviously, the frequencies of the alternate alleles are now $x_{12} = 1 - x_{11}$ in population 1 and $x_{22} = 1 - x_{21}$ in population 2.

In this note we investigate only two possible cases: (a) only one allele is shared by both populations, and (b) both alleles exist in the two populations. In the following we shall treat these two cases separately. Table 1 shows a schematic representation of the allele frequencies at the reassessed locus for the two cases, before and after reassessment.

(a) One Allele is Shared by Both Populations

Upon reassessment of the allele frequencies at one locus, the new genetic identity between the two populations (I_n) derived from formula (1) becomes:

$$I_n = \frac{J_{12} - a_{11}a_{21} + x_{11}x_{21}}{(J_1 - 2a_{11}^2 + 2a_{11} + 2x_{11}^2 - 2x_{11})^{1/2} \cdot (J_2 - 2a_{21}^2 + 2a_{21} + 2x_{21}^2 - 2x_{21})^{1/2}} \quad (2)$$

Differentiating formula (2) with respect to x_{11} we obtain:

$$\frac{\partial I_n}{\partial x_{11}} = \frac{x_{21}(J_1 - 2a_{11}^2 + 2a_{11}) + J_{12} - a_{11}a_{21} - x_{11}(2J_{12} - 2a_{11}a_{21} + x_{21})}{(J_1 - 2a_{11}^2 + 2a_{11} + 2x_{11}^2 - 2x_{11})^{3/2} \cdot (J_2 - 2a_{21}^2 + 2a_{21} + 2x_{21}^2 - 2x_{21})^{1/2}} \quad (3)$$

Therefore, the maximum value of I_n is reached when

$$x_{11} = \frac{x_{21}(J_1 - 2a_{11}^2 + 2a_{11}) + J_{12} - a_{11}a_{21}}{2J_{12} - 2a_{11}a_{21} + x_{21}} \quad (4)$$

Let us now assume that population 1 was considered monomorphic at the reassessed locus ($a_{11} = 1$), and that upon reassessment polymorphism was detected ($x_{11} \neq 1$ and $x_{12} = 1 - x_{11}$). We further assume that the allele frequencies at this locus are unchanged in population 2 ($a_{21} = x_{21}$). If the new allele found in population 1 is not present in population 2 (Table 1, case a), we obtain:

$$I_n = \frac{J_{12} - a_{21} + a_{21}x_{11}}{(J_2)^{1/2} \cdot (J_1 + 2x_{11}^2 - 2x_{11})^{1/2}} \quad (5)$$

I_n will reach its maximum value when $\partial(I_n)/\partial(x_{11}) = 0$, and we can easily see that

$$x_{11} = \frac{a_{21}J_1 - a_{21} + J_{12}}{2J_{12} - a_{21}} \quad (6)$$

Since $0 \leq x_{11} \leq 1$, it follows that $a_{21}J_1 - a_{21} + J_{12} \leq 2J_{21} - a_{21}$ or $a_{21}J_1 - a_{21} + J_{12} \geq 0$. This results in (1) $a_{21} \leq J_{12}/J_1$ or (2) $a_{21} \geq J_{12}/(1 - J_1)$ because in general $J_1 > 1$. In other words, we see that for the biologically meaningful range of $0 \leq x_{11} \leq 1$, I_n reaches an absolute maximum value for $x_{11} \leq 1$ when $0 \leq a_{21} \leq J_{21}/J_1$. When $a_{21} > J_{12}/J_1$, a relative maximum is obtained at $x_{11} = 1$, and the absolute maximum is obtained

when x_{11} is greater than 1. Since gene frequencies cannot exceed 1, I_n will always increase monotonically with x_{11} from 0 to 1.

The single locus identity at the j-th locus between two populations (I_{sj}) is defined as

$$I_{sj} = \sum_{i=1}^{n} r_i s_i \Big/ \left(\sum_{i=1}^{n} r_i^2 \sum_{i=1}^{n} s_i^2 \right)^{1/2} \tag{7}$$

where n is the total number of different alleles detected in both populations, and r_i and s_i are the frequencies of the corresponding alleles at the j-th locus in each population.

Let us now consider the new single locus gene identity (I_{sj}) at the reassessed locus. For simplicity the subscript for the locus is omitted. Using the same values as above, i.e., $a_{11} = 1$, $a_{12} = 0$, $a_{21} = x_{21}$ and $a_{22} = 1 - x_{21}$, I_{sj} equals:

$$I_{sj} = \frac{a_{21}x_{11}}{(2a_{21}^2 - 2a_{21} + 1)^{1/2} \cdot (2x_{11}^2 - 2x_{11} + 1)^{1/2}} \tag{8}$$

Differentiating, and setting $\partial I_{sj}/\partial x_{11} = 0$, we obtain $x_{11} = 1$. This means that the maximum value for I_{sj} is always obtained at the point where the reassessed locus in population 1 is monomorphic. Thus, when $a_{21} \leq J_{12}/J_1$, I_n and I_{sj} are not always positively correlated with each other, and it is possible for I_{sj} to decrease while I_n increases. Since the reassessment process involves one locus at the time, an unbiased estimator of genetic divergence should reflect the magnitude and the direction of the changes in the single locus gene identity in the reassessed locus. Because in the calculation of Nei's genetic identity, a mean other than the arithmetic mean was used; I_n does not behave properly in this respect.

Figure 1 illustrates the problem. We see that while the single locus identity increases with the frequency of the allele that is common to both populations over the entire range, the total identity increases only in the range from 0 to J_{12}/J_1 and decreases after this point. The highest discrepancies are for low frequency values of the allele shared by the two populations and for high values of J_{12}. Thus, the problem will be more pronounced in closely related populations than in distantly related ones, and in cases when the shared allele has a lower frequency of occurrence in one of the populations.

TABLE 1. Schematic representation of allele frequencies at the reassessed locus for the two cases, (a) one allele is shared by the two populations, and (b) both alleles are present in the two populations, before and after reassessment.

		Frequency in population 1		Frequency in population 2	
Case	Allele	Before	After	Before	After
(a)	A	a_{11}	x_{11}	a_{21}	x_{21}
	B	a_{12}	x_{12}		
	C			a_{22}	x_{22}
(b)	A	a_{11}	x_{11}	a_{21}	x_{21}
	B	a_{12}	x_{12}	a_{22}	x_{22}

(b) Both Alleles are present in the Two Populations

Let us consider briefly the case where both alleles are present in both populations. Their frequency in each of the populations is different (Table 1, case b). Assuming $x_{21} = a_{21}$, or in other words assuming that the gene frequencies in population 2 remain unchanged after reassessment and the frequencies change in population 1 only, we obtain:

$$I_n = \frac{J_{12} - 2a_{11}a_{21} + a_{11} + 2a_{21}x_{11} - x_{11}}{(J_2)^{1/2}(J_1 - 2a_{11}^2 + 2a_{11} + 2x_{11}^2 - 2x_{11})^{1/2}} \tag{9}$$

In this case, I_n reaches its maximum when

$$x_{11} = \frac{2a_{21}J_1 - J_1 + J_{12} - 4a_{11}^2a_{21} + 2a_{11}a_{21} + 2a_{11}^2 - a_{11}}{2J_{12} - 4a_{11}a_{21} + 2a_{11} + 2a_{21} - 1} \tag{10}$$

while I_{sj} reaches its maximum when $x_{11} = a_{21}$. Both I_n and I_{sj} will reach their maximum values at the same point in only two cases: (1) when $a_{21} = 0.5$ or (2) when $J_{12} = J_1 + 2a_{11}a_{21} - 2a_{11}^2 + a_{11} - a_{21}$. The same discordance between I_n and the single locus genetic identity is observed for most allele frequencies.

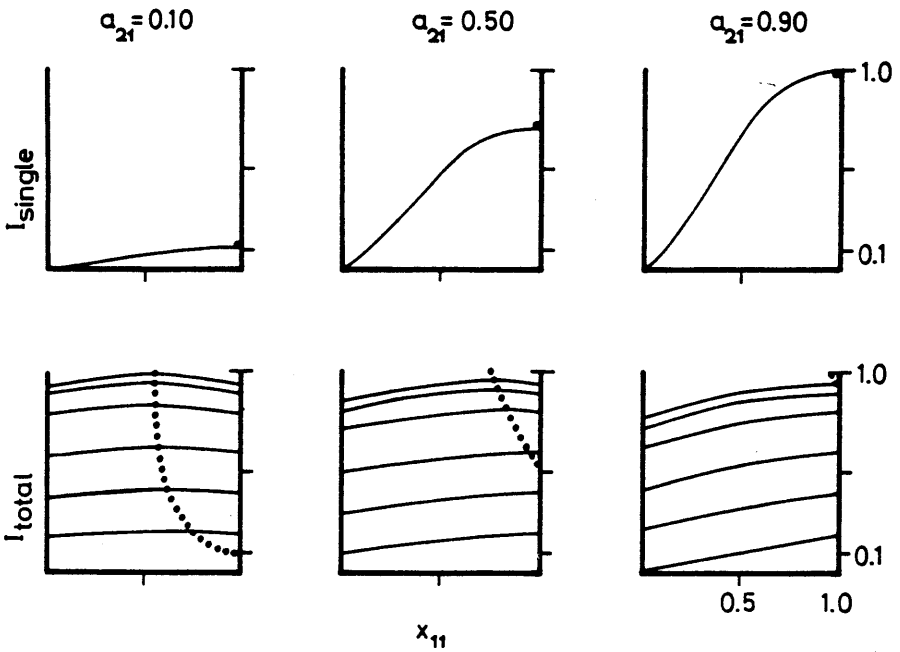

FIG. 1. Changes in single locus gene identity (I_{single}) and total genetic identity (I_{total}) with frequency of reassessed allele (x_{11}). J_1 and J_2 were set in all cases to 5. The lines in the lower figures represent, in ascending order, values of J_{12} of 1, 2, 3, 4, 4.5, and 4.7, respectively (maximum value for J_{12} with the given values for J_1 and J_2 is 5). Circles denote the x_{11} value at which I_{single} and I_{total} reach their respective maxima (see text for details, and Table 1, case a).

NUMERICAL EXAMPLES

We shall provide two numerical examples, one based on empirical data, the other one hypothetical, to illustrate those instances when problems with the application of Nei's genetic identity estimate may arise.

TABLE 2. Genetic variation at the amylase (*Amy*) and hemoglobin (*Hb*) loci in two *Macaca* species.

Source	Locus	Allele frequencies	
		M. fascicularis	*M. mulatta*
Nozawa et al.	*Hb*	0.508	1.000
		0.492	0.000
Tomiuk		0.518	0.658
		0.000	0.342
		0.482	0.000
Nozawa et al.	*Amy*	1.000	1.000
Tomiuk		0.298	0.000
		0.336	0.000
		0.143	0.000
		0.026	0.000
		0.056	0.000
		0.141	0.000
		0.000	0.690
		0.000	0.310

The first example is based on the data of Nozawa et al. (1977), who studied the genetic variation in several *Macaca* species. With respect to the amylase locus (*Amy*), Nozawa et al. found complete lack of variation in both *Macaca fascicularis* and *M. mulatta*, either within or between the species. With respect to hemoglobin (*Hb*), they found lack of variation in all populations of *M. mulatta*, and two alleles in all populations of *M. fascicularis*, one of which, Hb^s, was present in both species. One of us (Tomiuk, unpubl.) reinvestigated the gene frequencies for both the *Amy* and the *Hb* loci in both these species. The relevant gene frequency data from Nozawa et al.'s and Tomiuk's studies are given in Table 2. Data from different geographical populations belonging to the same species from the study of Nozawa et al. were pooled.

The genetic identity calculated from the original set of 29 enzyme and protein loci

TABLE 3. Total (I) and single locus (I_s) gene identities before and after reassessment of two loci, *Amy* and *Hb*, between two *Macaca* species. For data see Nozawa et al. (1977) and Table 1.

	Amy I_s	*Hb* I_s	Total I
Nozawa et al.	1.000	0.718	0.933
Reassessment:			
1	0.000	0.718	0.916
2	0.000	0.650	0.918

is 0.933. In the first step of the reassessment process, we substituted the data for the *Amy* locus, and kept all the other loci unchanged. The single locus identity, I_{sj}, decreased from 1 to 0, and as expected the total genetic identity, I_n, decreased too from 0.933 to 0.916. In the next step of the reassessment, we substituted the hemoglobin data. Again, the single locus identity decreased from 0.718 to 0.650, a decrease of about 10%. However, we now observe an increase in the total gene identity across loci of about 0.2% (from I = 0.916 to I = 0.918). The single locus and total gene identities before and after reassessment are given in Table 3.

Let us now consider the implications of this sort of effect on the construction of phylogenetic trees. In the following we shall use a hypothetical case to make a point. Consider species A, B, and C. The allele frequencies at six polymorphic loci are given in Table 4. From this table we calculate the gene identities prior to reassessment. These are designated "old" I's in Table 5. Because of identical values of I, it cannot be determined whether species A and B or species A and C are genetically more similar. Table 5 also gives the single locus gene identity for locus 6 ("old" I_6). We now proceed to either collect more data or to refine the resolution of the technique. The new "findings" in regard to locus 6 are also listed in Table 4 ("new" I_6). After recalculation we find out that the total gene identity for the two pairs of species (I_{AB} and I_{AC}) changed exactly in the opposite direction from the changes recorded in the single locus gene identities (Table 5).

The described phenomenon occurs be-

TABLE 4. Hypothetical genetic variation at six polymorphic loci in three species.

Locus	Species A	Species B	Species C
1	0.60	—	—
	0.40	0.35	0.45
	—	0.65	0.55
2	0.50	—	—
	0.50	0.50	0.50
	—	0.50	0.50
3	0.20	—	—
	0.80	1.00	1.00
4	0.40	—	—
	0.60	0.55	0.55
	—	0.45	0.45
5	0.60	—	—
	0.40	0.45	0.35
	—	0.55	0.65
6 "old"	0.90	—	—
	0.10	0.70	0.70
	—	0.30	0.30
6 "new"	0.90	—	—
	0.10	0.65	0.90
	—	0.35	0.10

cause of the assumptions in Nei's (1972) model. Nei assumed that the effective sizes of the two populations are equal, and that they are in a state of equilibrium between mutation, selection and random genetic drift. The probability of substitution is further assumed to be constant either per year or per generation. Thus, the denominator terms, J_1 and J_2, estimate the equilibrium amount of homozygosity under this model. The cross-product, J_{12}, is scaled down relative to J_1 and J_2. The expected degree of homozygosity at particular loci can vary from time to time, but its expectation remains constant. On the other hand, the expectation of gene identity between two populations decreases as time goes on. If

TABLE 5. Total (I) and single locus (I_6) gene identities before and after reassessment between three hypothetical species. For data see Table 4.

Parameter	Old value	New value	% Change
$I_{6(AB)}$	0.102	0.097	−4.9
$I_{6(AC)}$	0.102	0.110	+7.8
$I_{6(BC)}$	1.000	0.927	−7.3
$I_{(AB)}$	0.492	0.493	+0.2
$I_{(AC)}$	0.492	0.482	−2.0
$I_{(BC)}$	0.994	0.979	−1.5

the number of investigated loci is limited, a reassessed locus may decrease the denominator, $(J_1J_2)^{1/2}$, by more than it decreases the nominator. The net effect will be to increase I_n even though the gene frequencies have become further apart.

NEI'S MODIFIED GENETIC IDENTITY

Nei (1972) chose the mean values J_1, J_2 and J_{12} as the basis for his calculation of the total gene identity. Nei and Roychoudhury (1974) recognized the fact that the estimate of I is only "asymptotically unbiased", however, they chose it for its mathematical simplicity. Nei (1972) stated that it is also possible to compute the arithmetic mean of the single locus gene identities rather than I. Hillis (1984) discussed the properties of both these estimates, and stated that the normalized genetic identity defined by Nei is distorted by shared and unshared polymorphisms. In this note we identified an additional problem with I. Hillis (1984) suggested the use of the arithmetic mean of the single locus identities ($\hat{I}$) defined as:

$$\hat{I} = 1/r\left[\sum_{j}^{r}\left(\frac{\sum_{i}^{l_j} x_{ij}y_{ij}}{\left(\sum_{i}^{l_j} x_{ij}^2 \sum_{i}^{l_j} y_{ij}^2\right)^{1/2}}\right)\right] \quad (11)$$

where the parameters are the same as in formula (1).

We find that this estimate, which we call Nei's modified genetic identity, changes concomitantly with changes in the single locus identities, and that if there is a local maximum in the single locus identity, there will also be a maximum at the same point in $\hat{I}$. In analogy with Nei's (1972) genetic distance, Hillis (1984) defined Nei's modified genetic distance ($\hat{D}$) as:

$$\hat{D} = -\ln \hat{I}. \quad (12)$$

In the Appendix we present the sampling variances of $\hat{I}$ and $\hat{D}$, including computational details. The sampling variance of $\hat{D}$ is given in formula (6) in the Appendix. The sampling variance of $\hat{I}$ is:

$$V(\hat{I}) = \hat{I}^2V(\hat{D}). \quad (13)$$

We are now in the process of comparing Nei's genetic identity and Nei's modified genetic identity for a large set of allele frequencies from natural populations (Graur and Tomiuk, in prep.).

ACKNOWLEDGMENTS

We thank Drs. K. Wöhrmann, V. Loeschcke, J. W. Archie, J. Felsenstein, D. M. Hillis, and J. L. Rogers for critical comments on the manuscript. Dan Graur was supported in part by a fellowship from the Alexander von Humboldt Foundation and by grants from the Hertz Foundations and the Foundation for Basic Research of Tel Aviv University. Jürgen Tomiuk was supported by a grant of the Deutsche Forschungsgemeinschaft.

REFERENCES

GRAUR, D. 1986. The evolution of electrophoretic mobility of proteins. J. Theor. Biol., 118:443–469.

HEDRICK, P. W. 1983. Genetics of populations. Science Books International, Boston.

HILLIS, D. M. 1984. Misuse and modification of Nei's genetic distance. Syst. Zool., 33:238–240.

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

NEI, M. 1972. Genetic distances between populations. Amer. Nat., 106:283–292.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

NEI, M., AND A. K. ROYCHOUDHURY. 1974. Sampling variances of heterozygosity and genetic distances. Genetics, 76:379–390.

NOZAWA, K., T. SHOTAKE, Y. OHKURA, AND Y. TANABE. 1977. Genetic variation within and between species of Asian macaques. Japan. J. Genet., 52:15–30.

Received 2 February 87; accepted 8 March 88

APPENDIX

The procedures for deriving the sampling variances if $\hat{I}$ and $\hat{D}$ are listed below.

Definitions

r	Number of loci
l_k	Number of alleles at the k-th locus
$jx_k = \sum_{i=1}^{l_k} x_i^2$	Degree of homozygosity at the k-th locus, k = (1, 2, . . . , r), in species X
$jy_k = \sum_{i=1}^{l_k} y_i^2$	Degree of homozygosity at the k-th locus, k = (1, 2, . . . , r), in species Y

$jxy_k = \sum_{i=1}^{l_k} x_i y_i$	Identity of two genes chosen from species X and Y at the k-th locus
$I_k = \frac{jxy_k}{jx_k \cdot jy_k}$	Single locus genetic identity
$\hat{I} = \frac{1}{r}\sum_{i=1}^{r} I_k$	Modified Nei's genetic identity
I	Nei's (1972) genetic identity
$\hat{D} = -\ln \hat{I}$	Modified Nei's genetic distance
D	Nei's (1972) genetic distance

Assumptions

The degrees of homozygosity and gene identity in populations X and Y are assumed to be independent of the locus. That is:

$$\text{cov}(jx_k, jy_s) = 0 \text{ if } k \neq s$$
$$\text{cov}(jx_k, jxy_s) = 0 \text{ if } k \neq s$$
$$\text{cov}(jy_k, jxy_s) = 0 \text{ if } k \neq s$$

where k and s are any two loci.

Derivation of Sampling Variance

Differentiating $\hat{D}$ with respect to jx_k, jy_k and jxy_k we obtain:

$$\frac{\partial \hat{D}}{\partial jx_k} = \frac{I_k}{2r\hat{I}jx_k} \quad (1)$$

$$\frac{\partial \hat{D}}{\partial jy_k} = \frac{I_k}{2r\hat{I}jy_k} \quad (2)$$

$$\frac{\partial \hat{D}}{\partial jxy_k} = \frac{-I_k}{r\hat{I}jxy_k} \quad (3)$$

The intra-locus variance of $\hat{D}$ is by analogy with Nei and Roychoudhury (1974)

$$V_s(\hat{D}) = \sum_{k=1}^{r}\left[\left(\frac{\partial \hat{D}}{\partial jx_k}\right)^2 V(jx_k) + \left(\frac{\partial \hat{D}}{\partial jy_k}\right)^2 V(jy_k) + \left(\frac{\partial \hat{D}}{\partial jxy_k}\right)^2 V(jxy_k) + 2\left(\frac{\partial \hat{D}}{\partial jx_k}\frac{\partial \hat{D}}{\partial jxy_k}\right)\text{cov}(jx_k, jxy_k) + 2\left(\frac{\partial \hat{D}}{\partial jy_k}\frac{\partial \hat{D}}{\partial jxy_k}\right)\text{cov}(jy_k, jxy_k)\right]. \quad (4)$$

From (1), (2), (3) and (4), we obtain:

$$V_s(\hat{D}) = \frac{1}{(2r\hat{I})^2}\sum_{k=1}^{r}\left[\left(\frac{I_k}{jx_k}\right)^2 V(jx_k) + \left(\frac{I_k}{jy_k}\right)^2 V(jy_k) + \left(\frac{2I_k}{jxy_k}\right)^2 V(jxy_k) - \frac{(2I_k)^2\text{cov}(jx_k, jxy_k)}{jx_k jxy_k} - \frac{(2I_k)^2\text{cov}(jy_k, jxy_k)}{jy_k jxy_k}\right]. \quad (5)$$

Rearranging (5), we obtain:

$$V_s(\hat{D}) = \frac{1}{(2r\hat{I})^2}\sum_{k=1}^{r} I_k^2 \left[\frac{V(jx_k)}{jx_k^2} + \frac{V(jy_k)}{jy_k^2} + \frac{4V(jxy_k)}{jxy_k^2} - \frac{4\,\text{cov}(jx_k, jxy_k)}{jx_k jxy_k} - \frac{4\,\text{cov}(jy_k, jxy_k)}{jy_k jxy_k}\right]. \quad (6)$$

Compare (6) to Nei and Roychoudhury's (1974) formula for $V_s(D)$:

$$V_s(D) = \frac{1}{4r^2}\left[\frac{V(Jx)}{Jx^2} + \frac{V(Jy)}{Jy^2} + \frac{4V(Jxy)}{Jxy^2} - \frac{4\,\text{cov}(Jx, Jxy)}{JxJxy} - \frac{4\,\text{cov}(Jy, Jxy)}{JyJxy}\right]. \quad (7)$$

Differentiating $\hat{D}$ with respect to $\hat{I}$, we obtain analogously the total variance $V(\hat{D})$:

$$V(\hat{D}) = \frac{1}{\hat{I}^2}V(\hat{I}) = \frac{\sum_{k=1}^{r}(I_k - \hat{I})^2}{r(r-1)\hat{I}^2}. \quad (8)$$

We can, hence, calculate the inter-locus variance of D as:

$$V(\hat{D}) = V_s(\hat{D}) + V_l(\hat{D}). \quad (9)$$

Rearranging formula (9), we obtain

$$V(\hat{I}) = \hat{I}^2 V(\hat{D}). \quad (10)$$