Génét. Sél. Evol., 1988, **20** (4), 417-424

A test to detect association between autosomal and sex-linked loci in random mating populations : an example with *Drosophila melanogaster*

S. MONT *, A. MOYA * and D. MARINKOVIĆ **

* *University of Valencia, Faculty of Biological Sciences, Laboratory of Genetics, c/ Dr. Moliner 50, 46100 Burjassot, Valencia, Spain*
** *University of Beograd, Faculty of Science, Institute of Zoology, Studentski 16 trg, 11000 Beograd, Yugoslavia*

Summary

A method to detect associations between two autosomal loci and a sex-linked locus is developed. The procedure is the natural extension of HILL's method applied to three autosomal loci in random mating populations. The gametic frequencies are estimated by maximum likelihood. A set of models is developed and tested by means of a likelihood ratio test. The method is applied to data of *Hk, Pgm,* and *6Pgd* loci of *Drosophila melanogaster* located on the second, third, and first chromosomes, respectively. The results of the analyses show the existence of association between the loci *Hk* and *Pgm.*

Key words : enzyme polymorphism, multilocus association, statistical test, Drosophila melanogaster.

Résumé

Un test de détection de l'association entre locus autosomaux et liés au sexe dans des populations panmictiques : un exemple avec Drosophila melanogaster

On a développé une méthode pour détecter des associations entre deux locus autosomaux et un locus lié au sexe. Cette méthode est une extension de celle de HILL s'appliquant à trois locus autosomaux dans des populations panmictiques. Les fréquences gamétiques sont estimées par la méthode du maximum de vraisemblance. Une série de modèles est développée et éprouvée par un test du maximum de vraisemblance. La méthode est appliquée à des données expérimentales concernant les locus *Hk, Pgm* et *6Pgd* chez *Drosophila melanogaster,* situés respectivement sur les 2^{e}, 3^{e} et le 1er chromosomes. Les résultats des analyses montrent l'existence d'une association entre *Hk* et *Pgm.*

Mots clés : polymorphisme enzymatique, association multilocus, test statistique, Drosophila melanogaster.

I. Introduction

The method that we develop here is an extension of HILL's (1975) method for the detection of significant associations between several autosomal loci in random mating diploid populations. We have extended the method to sex linked loci, for a sample of males that have been analyzed by electrophoresis.

Assuming that we are working with a random mating population and that there are no evolutionary forces that can change the equilibrium status of that population, it is possible to get maximum likelihood estimates of male and female gametic (chromosomal) frequencies. In this case we use as data the male (zygotic) frequencies that have been detected by electrophoresis. A likelihood ratio test can be used to detect significant associations between alleles of different loci (linkage disequilibrium).

Due to the source of data, the extended method that we propose in the present paper permits us to estimate both the gametic frequencies coming from males as well as from female parents. It is also possible to test the associations for both sexes.

In the present study we deal with two autosomal and one sex-linked loci. To extend the analysis to more loci is only a quantitative problem.

II. Method

Consider two autosomal loci α and β and a sex-linked locus γ. Each of these three loci has two allelic states (A and a for α, B and b for β and C and c for γ). A sample of N males is taken from the population and examined with regard to different possible genotypes. It is assumed that : (i) there is random mating ; (ii) the zygotic frequencies are the product of gametic frequencies ; and (iii) the different genotypes observed are multinomially distributed.

The expected frequency, for instance, of the genotype $AaBbC$, $\mathrm{E}(AaBbC)$, will be :

$$\mathrm{E}(AaBbC) = ff_{ABC} \cdot fm_{ab} + ff_{AbC} \cdot fm_{aB} + ff_{aBC} \cdot fm_{Ab} + ff_{abC} \cdot fm_{AB}, \quad (1)$$

where, for instance, ff_{ABC} is the frequency of ABC gametes coming from a female parent carrying the A, B, and C alleles, and fm_{ab} is the frequency of ab gametes present in male parents carrying the a and b alleles. According to expression (1), the expected frequencies of each gamete from a given genotype can be determined. For instance, the expected frequency of ABC gametes from $AaBbC$ males is :

$$\mathrm{H}(ABC/AaBbC) = ff_{ABC} \cdot fm_{ab}/\mathrm{E}(AaBbC) = \mathrm{H}(ab/AaBbC), \quad (2)$$

where $\mathrm{E}(AaBbC)$ is calculated according to (1). For the maximum-likelihood estimates of gametic frequencies we have followed HILL's (1975, 1976) « chromosome counting » method. The method is an extension of the « gene counting » method developed by CEPELLINI *et al.* (1955). The gametic frequencies, in this method, are equated to their expectations by means of a series of progressive approaches until convergence is

reached. If ff'_{ABC} is the next estimate of the frequency of ABC gametes and ff_{ABC} the present estimate, then ff'_{ABC} can be deduced from ff_{ABC} by means of :

$$ff'_{ABC} = \Sigma N(x) \cdot H(ABC/x)/2N, \tag{3}$$

where x represents a given genotype, for instance $AaBbC$, $H(ABC/x)$ is calculated according to (2), and $N(x)$ is the observed number of individuals of that genotype. Expression (3) is summed over all genotypes.

The expected gamete frequency of ABC, ff'_{ABC} is, for example :

$$\begin{aligned} ff'_{ABC} = (&N(AABBC) + N(AABbC) \cdot ff_{ABC} \cdot fm_{Ab}/E(AABbC) + \\ &N(AaBBC) \cdot ff_{ABC} \cdot fm_{aB}/E(AaBBC) + \\ &N(AaBbC) \cdot ff_{ABC} \cdot fm_{ab}/E(AaBbC))/N. \end{aligned} \tag{4}$$

The rest of the gametic frequencies can be obtained in a similar way.

To start the process, we have arbitrarily assigned a value of 0.125 to each female gamete and 0.25 to each male gamete. The new values of the gametic frequencies according to the expression (4) have been taken as the initial values for a new cycle of iteration. As has been demonstrated by ELANDT-JOHNSON (1971), the iterative process converges to the maximum likelihood estimate of gametic frequency. The process can be followed until the required precision is attained.

The gametic counting method also permits us to distinguish between female and male gametic frequencies. Employing these frequencies, and ignoring the constant terms, the log-likelihood functions are :

$$\begin{aligned}
L(\alpha) &= N(AA) \cdot \ln(ff_A \cdot fm_A) + N(Aa) \cdot \ln(ff_A \cdot fm_a + ff_a \cdot fm_A) + \\
&\quad + N(aa) \cdot \ln(ff_a \cdot fm_a) \\
L(\beta) &= N(BB) \cdot \ln(ff_B \cdot fm_B) + N(Bb) \cdot \ln(ff_B \cdot fm_b + ff_b \cdot fm_B) + \\
&\quad + N(bb) \cdot \ln(ff_b \cdot fm_b) \\
L(\gamma) &= N(C) \cdot \ln(ff_C) + N(c) \cdot \ln(ff_c) \\
L(\alpha\beta) &= N(AABB) \cdot \ln(ff_{AB} \cdot fm_{AB}) + \ldots + N(aabb) \cdot \ln(ff_{ab} \cdot fm_{ab}) \\
L(\alpha\gamma) &= N(AAC) \cdot \ln(ff_{AC} \cdot fm_A) + \ldots + N(aac) \cdot \ln(ff_{ac} \cdot fm_a) \\
L(\beta\gamma) &= N(BBC) \cdot \ln(ff_{BC} \cdot fm_B) + \ldots + N(bbc) \cdot \ln(ff_{bc} \cdot fm_b) \\
L^*(\alpha\beta\gamma) &= N(AABBC) \cdot \ln(ff^*_{ABC} \cdot fm_{AB}) + \ldots \\
&\quad + N(aabbc) \cdot \ln(ff^*_{abc} \cdot fm_{ab}) \\
L(\alpha\beta\gamma) &= N(AABBC) \cdot \ln(ff_{ABC} \cdot fm_{AB}) + \ldots \\
&\quad + N(aabbc) \cdot \ln(ff_{abc} \cdot fm_{ab}) \\
L(S) &= N(AABBC) \cdot \ln(N(AABBC)/N) + \ldots + N(aabbc) \cdot \ln(N(aabbc)/N).
\end{aligned}$$

These log-likelihood functions represent different models of association. In the case of $L^*(\alpha\beta\gamma)$ we are considering the model of all the possible pairs of associations, excluding the three-locus association. As there is no explicit formula for the chromosome frequencies under this model, in the standard three dimensional contingency table, an iterative technique has been used to compute these three locus frequencies (ff^*_{ABC}, etc.). The procedure has to be modified for our situation : once the previous pair-wise frequencies have been fitted, find the frequencies, e.g. ff^*_{ABC}, satisfying the model using the iterative technique given by FIENBERG (1970, 1980).

III. Testing the hypothesis

Successive models of dependence among the gene frequencies can be fitted. Given that there is no prior hypothesis on what forces (e.g., selection) are determining specific associations, it seems preferable to follow, as HILL (1975) does, the hierarchy of models of dependence of frequencies considered appropriate for the standard multi-dimensional contingency table. If significant departures from random association are found at any level of the model, the possible forces which could give rise to this can then be studied. The succession of models of association of gene frequencies is as follows :

1) *Association between* α *and* β

$G_{AB} = 2 \cdot (L(\alpha\beta) - L(\alpha) - L(\beta))$.

This is χ^2 distributed with 1 d.f. under the null hypothesis of no association.

2) *Association between* α *and* γ, *once* α *and* β *have been fitted*

$G_{AC} = 2 \cdot (L(\alpha\gamma) - L(\alpha) - L(\gamma))$.

This is χ^2 distributed with 1 d.f.

3) *Association between* β *and* γ, *once* $\alpha\beta$ *and* $\alpha\gamma$ *have been fitted*

$G_{BC} = 2(L^*(\alpha\beta\gamma) - L(\alpha\beta) - L(\alpha\gamma) + L(\alpha))$,

which is χ^2 distributed with 1 d.f.

4) *Association between* $\alpha\beta$, $\alpha\gamma$, *and* $\beta\gamma$, *once the previous pairs have been fitted*

$G^*_{ABC} = 2(L(\alpha\beta\gamma) - L^*(\alpha\beta\gamma))$,

which is χ^2 distributed with 1 d.f.

5) *Two- and three-locus associations*

$G_{ABC} = 2(L(\alpha\beta\gamma) - L(\alpha) - L(\beta) - L(\gamma))$,

which is the sum of the previous G-tests and is χ^2 distributed with 4 d.f.

6) *Final test*

A final test which may be useful is for a fit to Hardy-Weinberg associations of gametic frequencies, i.e., that genotype frequencies equal the product of gametic frequencies :

$G_{HW} = 2(L(S) - L(\alpha\beta\gamma))$,

which is χ^2 distributed with 6 d.f.

IV. Results and discussion

We use available data on hexokinase (*Hk*, second chromosome), phosphoglucomutase (*Pgm*, third chromosome), and 6-phosphodehydrogenase (*6Pgd*, first chromosome). All the data are from *Drosophila melanogaster* and they are the same experiment as MARINKOVIĆ *et al.* (1987). The flies derived from 200 *Drosophila melanogaster* females were collected from a natural population in Furnace Creek (Mojave Desert, California). They were brought to the laboratory and distributed into several dozen separate cultures (half-pint bottles) with standard *Drosophila* medium. These cultures were subcultured by transferring 20-30 F_1 individuals from each culture to a fresh one. Groups of 10-20 pairs from the F_2 generation (collected from several cultures in such a way that males and females always came from different cultures) were placed in new culture bottles. Three hundred males of the F_3 generation were randomly chosen from the cultures and assayed by starch gel electrophoresis.

In the model α is *Hk*, β is *Pgm*, and γ is *6Pgd*. The more frequent allele at each locus is represented by *A*, *B*, and *C*, respectively, and the set of the less frequent alleles by *a*, *b*, and *c*, respectively. Table 1 shows the observed genotypes of the 295 males that have been electrophoretically analyzed. From these data, the maximum likelihood estimates of gametic frequencies, following the above described « gamete counting » method, have been determined. The corresponding log-likelihood functions

TABLE 1

Observed number of male genotypes for Hk, Pgm, *and* 6Pgd *loci in* Drosophila melanogaster, *with* A/a, B/b, *and* C/c *alleles, respectively*

	C			*c*		
	BB	*Bb*	*bb*	*BB*	*Bb*	*bb*
AA	128	38	4	28	7	1
Aa	57	9	1	8	2	0
aa	9	1	0	2	0	0

TABLE 2

Log-likelihood function values to study the association of Hk, Pgm *and* 6Pgd *loci in* Drosophila melanogaster

Function	Value
$L(\alpha)$	− 216.71
$L(\beta)$	− 173.36
$L(\gamma)$	− 131.02
$L(\alpha\beta)$	− 387.96
$L(\alpha\gamma)$	− 347.49
$L(\beta\gamma)$	− 304.38
$L^*(\alpha\beta\gamma)$	− 518.71
$L(\alpha\beta\gamma)$	− 518.63
$L(S)$	− 516.49

TABLE 3

Maximum likelihood test for the association between Hk, Pgm, *and* 6Pgd *loci of* Drosophila melanogaster

Source		d.f.	*G*-value
Association of :			
Hk and *Pgm*			
$2 \cdot (L(\alpha\beta) - L(\alpha) - L(\beta))$		1	4.22*
Hk and *6Pgd*			
$2 \cdot (L(\alpha\gamma) - L(\alpha) - L(\gamma))$		1	0.48
Pgm and *6Pgd*			
$2 \cdot (L(\beta\gamma) - L(\beta) - L(\gamma))$		1	0.00
Association of :	Conditional on :		
Hk and *Pgm*	*6Pgd*		
$2 \cdot (L^*(\alpha\beta\gamma) - L(\alpha\gamma) - L(\beta\gamma) + L(\gamma))$		1	4.28*
Hk and *6Pgd*	*Pgm*		
$2 \cdot (L^*(\alpha\beta\gamma) - L(\alpha\beta) - L(\beta\gamma) + L(\beta))$		1	0.54
Pgm and *6Pgd*	*Hk*		
$2 \cdot (L^*(\alpha\beta\gamma) - L(\alpha\beta) - L(\alpha\gamma) + L(\alpha))$		1	0.06
All two-locus associations :			
$2 \cdot (L(\alpha\beta\gamma) - L^*(\alpha\beta\gamma))$		1	0.16
All two-locus and three locus associations :			
$2 \cdot (L(\alpha\beta\gamma) - L(\alpha) - L(\beta) - L(\gamma))$		4	4.92
Hardy-Weinberg equilibrium			
$2 \cdot (L(S) - L(\alpha\beta\gamma))$		6	4.28

*$P < 0.05$

are shown in table 2. Finally, table 3 shows the tests for different associations between *Hk, Pgm,* and *6Pgd.* As this table shows, the test for Hardy-Weinberg equilibrium gives 4.28 with 6 d.f., which indicates no departure and the assumption of random mating in the analysis seems tenable. The same have been obtained by treating separately *Hk* and *Pgm.* On the other hand, there is significant association between *Hk* and *Pgm.* This association is also detected when the corresponding genes are conditioned on the third one, *6Pgd.* In the current case the total two- and three-locus associations are completely determined by the two-locus associations.

MARINKOVIĆ *et al.* (1987) have also detected significant linkage disequilibrium between *Hk* and *Pgm* following the method proposed by P.M. BURROWS, which yields a correlation coefficient that incorporates the departures from Hardy-Weinberg for the sample frequencies at each locus. The method has been developed by WEIR & COCKERHAM (1979) and, as is well known, does not require the assumption of random mating. Both methods, maximum likelihood and correlation, seem to be equally efficient at detecting significant association among *Hk* and *Pgm,* making the result more solid. So, as has been pointed out by other authors (*see* for a review HEDRICK *et al.,* 1978), it is possible to find associations between loci on different chromosomes.

The flies here assayed are the F_3 generation from flies collected in nature. It is possible that linkage disequilibrium may have built up in those few generations. But the flies are randomized samples from numerous cultures with several thousand individuals

in toto. Whether either natural selection or random drift are responsible for the detected association remains unknown to us, and more experimental studies should be done in this line (*see*, for instance, the neutrality tests used by HEDRICK & THOMSON, 1986). It is evident that the sample used is not large enough to study loci with alleles at low frequencies. However, the main objective of the present is to develop a method for detecting associations between loci of different chromosomes, especially among autosomal and sex-linked loci. We have no evidence of significant association between *6Pgd*, sex-linked, and the autosomal loci. Similar results have also been obtained with the above mentioned correlation method (MARINKOVIĆ *et al.*, 1987).

Received January 21, 1987.
Accepted February 10, 1988.

Acknowledgements

The work has been partly supported by grant PB86-0517 of « Dirección General de la Investigación Científica y Técnica de España » to A.M.

References

CEPELLINI R., SINISCALCO M., SMITH C.A.B., 1955. The estimation of gene frequencies in a random-mating population. *Ann. Eugen.*, **20,** 97-115.

ELANDT-JOHNSON R.C., 1971. *Probability models and statistical methods in genetics*. 340 p., Wiley, New York.

FIENBERG S.E., 1970. The analysis of multidimensional contingency tables. *Ecology*, **51,** 419-433.

FIENBERG S.E., 1980. *The analysis of cross-classified categorical data*. 198 p., MIT press, Massachusetts.

HEDRICK P.W., THOMSON G., 1986. A two-locus neutrality : application to humans, *E. coli*, and lodgepole pine. *Genetics*, **112,** 135-156.

HEDRICK P.W., JAIN S., HOLDEN L., 1978. Multilocus systems in evolution. *Evol. Biol.*, **11,** 101-184.

HILL W.G., 1975. Test of association of gene frequencies in randomly mating populations. *Biometrics*, **31,** 881-888.

HILL W.G., 1976. Non-random association of neutral linked genes in finite populations. *In :* KARLIN S., NEVO E. (ed.), *Population genetics and ecology*, 339-376, Academic Press, New York.

MARINKOVIĆ D., TUCIC N., MOYA A., AYALA F.J., 1987. Genetic diversity and linkage disequilibrium in *Drosophila melanogaster* with different rates of development. *Genetics*, **117,** 513-520.

WEIR B.S., COCKERHAM C.C., 1979. Estimation of linkage disequilibrium in randomly mating populations. *Heredity*, **42,** 105-111.