

# TESTING THE HETEROGENEITY OF F VALUES: A SUGGESTION AND A CORRECTION<sup>1</sup>

S. TSAKAS AND C. B. KRIMBAS

*Department of Genetics, Agricultural College of Athens, Greece.*

Manuscript received August 8, 1975

Revised copy received April 30, 1976

## ABSTRACT

The use of a pair of populations removes the problems posed by a structured phylogeny (pointed out by A. ROBERTSON) and permits the application of the original test of LEWONTIN and KRAKAUER. In the case of  $2n$  populations studied a suggestion is made to use only  $n$  pairs in such a way that no population appears in more than one pair. Data on *Dacus* analyzed this way show that at least two out of six genes studied are or have been selected.

TSAKAS and KRIMBAS (1975) used the test devised by LEWONTIN and KRAKAUER (1973), also known as the heterogeneity of  $F$  values test, and especially the variation in space model, on genetic data collected in three populations of the olive fruit fly *Dacus oleae* in Greece. Previously NEVO (1973) has also used the same test on data collected from populations of a rodent.

Recently NEI and MARUYAMA (1975) and especially A. ROBERTSON (1975) have discussed the limitations of this test: the expected variance of the  $F$ 's, under neutrality conditions, is not equal (in a conservative estimation) to  $2\bar{F}^2/n-1$  as LEWONTIN and KRAKAUER suggested but to  $2\bar{F}^2(1/n-1+Vr)$ , where  $Vr$  is the variance of the correlation coefficients between populations, the expected mean of these being equal to zero. This last variance,  $Vr$ , is zero only when all populations derive independently from the original population but different from zero in a structured phylogenetic tree, that is to say, when populations form a definite structure as do the races of the human species. Since this is the general case and the detailed phylogenetic history of the populations is not generally known we should get an independent estimation of  $Vr$  in order to use the test. But this seems to be very difficult, if not practically impossible in most cases.

We suggest an alternative in order to bypass this difficulty: test two populations at a time. This of course put a strong limitation to the power of the test, but in this case  $Vr$  is equal to zero. Whatever is the phylogenetic history of this couple of populations they should ultimately derive from the same population independently by a simple and similar pathway. Then the expected variance under neutrality conditions of the  $F$ 's is equal to  $2\bar{F}^2$ .

Reanalyzing in this way the data of our 1975 publication we get the results shown in Table 1.

<sup>1</sup> Dedicated to the memory of Theodosius Dobzhansky "Tu se'lo mio maestro e il mio autore" Dante (*Inferno*, I, 85).

TABLE 1

Populations	Genes	d.f.	Variance ratio	Significance
Thassos-Corfu	all genes	20	1.302	no
Thassos-Crete	"	21	0.633	no
Corfu-Crete	"	21	2.969	<0.001
"	all except <i>TO</i>	20	2.728	<0.001
"	all except <i>ODH</i>	20	2.743	<0.001
"	all except <i>ADH</i>	20	2.735	<0.001
"	all except <i>APH</i>	20	2.790	<0.001
"	all except <i>EstB</i>	13	1.977	<0.05
"	all except <i>EstA</i>	11	2.850	<0.001

The exception of one gene (*EstB*) from the data makes the test non-significant in the case of the populations of Corfu-Crete. Thus one gene at least should be selected out of six.

In this case, as in the next mentioned below, we should not forget that  $F$ 's estimated from alleles of the same genes have been used. These  $F$ 's are correlated (ALAN ROBERTSON, personal communication) and could inflate the observed variance of the  $F$ 's. It is felt however that this bias is rather unimportant especially since we have removed one degree of freedom for each locus. Furthermore, in order to make the test independent from mutations arising in some of the populations we have excluded from calculations alleles which do not appear in all populations sampled.

TSAKAS *et al.* (in preparation) have studied 12 populations of *Dacus oleae* all over the Mediterranean basin. The same six genes have been studied. From all 66 possible tests of couples of populations, 15 are significant at the level of 0.001. Out of these 15, the heterogeneity of the  $F$ 's could be attributed to one gene for 14 of them and to 2 genes for 1 of them. With  $2n$  populations sampled  $n(2n-1)$  pair tests are possible. However in spite of the fact that any single pair-wise comparison is a valid test, in the collection of all possible ones, the different comparisons are not independent of each other. In order to bypass this difficulty we suggest to use only  $n$  such pair-wise comparisons in such a way that there will be no pair containing a population included in another pair. There is no *a priori* reason to combine in a random way the populations in pairs but it would be preferable to make such combinations before knowing the results. In our case we formed the couples in a random way shown in Table 2.

This result is not a non-representative one. In our case out of 10,395 possible ways to select such six pair out of 12 populations, only 65 (a frequency of 0.0063) would provide non-significant results for all the six pair-wise combinations.

We conclude that LEWONTIN and KRAKAUER test can be used if we compare pairs of populations. Then the variance of the  $F$ 's is equal to  $2\bar{F}^2$ . With  $2n$  populations we suggest to make only  $n$  independent pair-wise comparisons in the way explained above.

TABLE 2

Pair of Population	Significance	Number of genes making heterogeneous the result
Liguria-Madrid	n.s.	
Pyrgos-Cephallonia	<.001	1 gene ( <i>Est A</i> or <i>Est B</i> )
Lesbos-Cyprus	n.s.	
Italy-Split	n.s.	
Syria-Thassos	<.001	1 gene ( <i>ADH</i> )
Corfu-Crete	<.001	1 gene ( <i>Est B</i> )

Thus at least 2 genes (*Est B* and *ADH* making the most conservative choice) out of 6 are or have been selected.

## LITERATURE CITED

- LEWONTIN, R. C. and J. KRAKAUER, 1973 Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**: 175-195.
- NEI, M. and T. MARUYAMA, 1975 LEWONTIN-KRAKAUER test for neutral genes. *Genetics* **80**: 395.
- NEVO, E., 1973 Test of selection and neutrality in natural populations. *Nature* **244**: 573-575.
- ROBERTSON, A., 1975 Remarks on the LEWONTIN-KRAKAUER test. *Genetics* **80**: 396.
- TSAKA, S. and C. B. KRIMBAS, 1975 How many genes are selected in populations of *Dacus oleae*? *Genetics* **79**: 675-679.

Corresponding editor: R. C. LEWONTIN