Atti Accademia Nazionale dei Lincei

Classe Scienze Fisiche Matematiche Naturali

Rendiconti

Luciano Terrenato, Erna van Loghem, Luigi Bernini, Silvana Augusta Santachiara-Benerecetti, Guido Modiano, Carlo Santolamazza, Rosaria Scozzari, Laura Ulizzi, Maria Beretta

Preliminary data on the genetic heterogeneity among the Sardinian isolates

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Serie 8, Vol. **51** (1971), n.3-4, p. 249–253.

Accademia Nazionale dei Lincei

<http://www.bdim.eu/item?id=RLINA_1971_8_51_3-4_249_0>

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Accademia Nazionale dei Lincei, 1971.

Genetica. — *Preliminary data on the genetic heterogeneity among the Sardinian isolates* (*). Nota (**) di LUCIANO TERRENATO, ERNA VAN LOGHEM, LUIGI BERNINI, SILVANA AUGUSTA SANTACHIARA-BENERECETTI, GUIDO MODIANO, CARLO SANTOLAMAZZA, ROSARIA SCOZZARI, LAURA ULIZZI e MARIA BERETTA, presentata dal Socio G. MONTALENTI.

RIASSUNTO. — Nell'ambito di un programma di ricerche popolazionistiche in Sardegna, sono stati esaminati, in circa 700 soggetti appartenenti a 10 diversi paesi di pianura, 13 marcatori genetici. L'isolamento genetico della Sardegna nel suo complesso rispetto al resto dell'Europa è stato ancora una volta dimostrato sia attraverso la conferma di peculiarità già note, sia con la scoperta di nuove diversità.

È stato poi possibile ottenere una stima dell'effetto della deriva genetica nei 10 diversi isolati, mettendo in evidenza la sua correlazione con il grado di isolamento dei vari paesi esaminati.

Sardinia, with a population of about two million inhabitants, is well known to human geneticists for the studies carried out on the effect of malaria on Thalassaemia and Glucose-6-phosphate dehydrogenase deficiency frequencies. Approximately two thirds of the population of the island are distributed in rather large villages of about five thousand inhabitants each.

A longterm project aiming at a detailed description of the genetic make-up of this population is being performed by the Genetic Institute of the University of Rome, the Central Laboratory of the Netherlands Red Cross Blood Transfusion Service, the Human Genetics Department of the University of Leyden and the International Laboratory of Genetics and Biophysics, Pavia section.

The present report deals with the results obtained with a sample of about 700 individuals, which represents one third of the whole project. The sample is made up of boys and girls, aged from 13 to 16, attending the secondary schools of ten villages scattered in the lowlands of the island. Only the subjects with both parents or with the four grandparents born in Sardinia were considered. The others, amounting to only the 7.2 per cent, were discarded.

Since each subsample was made up of approximately 70 individuals and the population size of each village was about 5000, the ten subsamples represent about 1 per cent of each isolate. The degree of isolation, calculated as the proportion of subjects with both parents born in the same village, is quite different from one village to another, ranging from 20 to 80 per cent (Table I).

(*) Supported by a Grant of the Consiglio Nazionale delle Ricerche – International Biological Program (IBP).

The abstract of this paper was presented at the 4th International Congress of Human Genetics, Paris 1971.

(**) Pervenuta all'Accademia il 15 ottobre 1971.

TABLE I.

A list of the 10 isolates under examination with some relevant characteristics.

ISOLATES	Sub-sample size	Altitude (meters)	Gene frequencies			Percent degree of isolation (*)	Genetic distance (**)
			Gd–	*β–th*	*th*		
Villaputzu	66	13	.32	.03	.12	75.4	.0014
Muravera	76	12	.16	.06	.16	46.3	.0022
Sanvito	68	13	.18	.05	.16	76.1	.0051
Pula	79	15	.28	.04	.13	39.1	.0010
Santadi	72	135	.21	.08	.18	66.7	.0029
Ittiri	63	430	.09	.04	.13	79.4	.0024
Sorso	84	136	.05	.05	.13	75.0	.0023
Ozieri	93	340	.09	.03	.06	36.6	.0007
Thiesi	52	465	.00	.04	.06	48.1	.0020
Tempio Pausania .	106	566	.17	.05	.07	21.7	.0007

(*) See text.
(**) Calculated according to Cavalli-Sforza *et al.* (1969).

TABLE II.

Gene frequencies (*) measured on the whole sample.

RED CELL ANTIGENS:	*ABO*	*i*	:	74.58 ± 1.34 (**)
	MN	*M*	:	66.21 ± 1.35 (**)
	Rh	*r*	:	21.68 ± 1.97 (**)
	Kell	*K*	:	4.82 ± 0.62
	P	P_1	:	47.51 ± 1.71
	Duffy	Fy^a	:	37.11 ± 1.56
	Lutheran	Lu^a	:	3.46 ± 0.52
RED CELL ENZYMES:	*Adenosindeaminase*	ADA^1	:	93.83 ± 0.63
	Diaphorase	Dia^2	:	1.25 ± 0.36 (**)
	Acid phosphatase	P^B	:	68.36 ± 1.22
	Phosphoglucomutase	PGM_1^1	:	73.11 ± 1.17
SERUM MARKERS:	*Group component*	Gc^1	:	76.30 ± 1.20
	Immunoglobulin	Inv^b	:	90.33 ± 0.85
	Haptoglobin	Hp^1	:	39.34 ± 1.28

(*) For each marker the frequency (± 1 S.E.) of the most common allele is reported.
(**) These markers are the most useful to show the difference between Sardinian and European populations. The frequencies reported in the Table refer to the alleles most suitable for revealing such difference. In the European population these alleles have about the following frequencies: $i = 60$–70% ; $M = 55$–60% ; $r = 30$–45% ; $Dia^2 = 0.2\%$.

Two types of isolation can be considered: one referring to the whole Sardinian population towards the rest of Europe, and the other concerning the isolation existing among different Sardinian villages.

As far as the Sardinian population on the whole is concerned, the already known peculiarities [1] concerning the ABO, MNS and Rh systems have been confirmed (Table II). Moreover the allele 2 of NADH diaphorase turned out to have a frequency of more than 1 per cent, a very high one, in comparison with those of the other populations so far studied [2].

As far as the isolation among Sardinian villages is concerned, the variance between the gene frequencies found in the ten isolates has been calculated for each of the 13 markers under study. As shown in the fig. 1, these 13 variances are very different from each other at a very high level of significance.

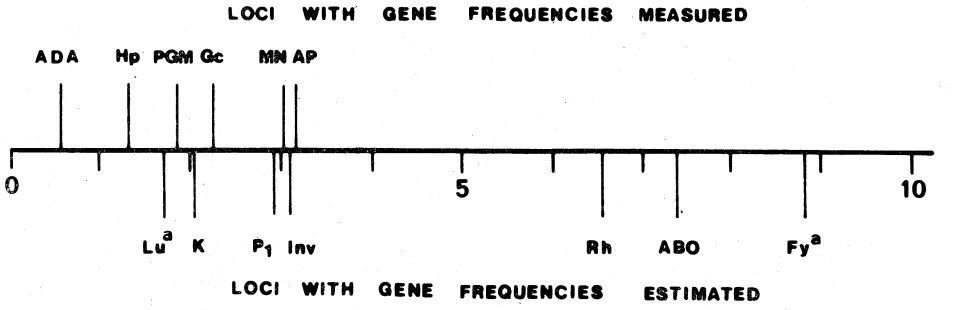

Fig. 1. – A graphical representation of the variances among villages of the 13 genetic markers. For each marker has been calculated the angular transformation of the frequency of its most frequent allele, in each village. Then the variance among villages has been calculated. The 13 variances thus obtained are reported in the figure, using an arbitrary scale. The Bartlett test for the heterogeneity among the 13 variances has given a $\chi^2_{(12df)} = 28.83$ with a $P < .005$.

It is worth pointing out that, as expected on obvious statistical grounds, the highest heterogeneity among villages was found for those markers whose gene frequencies are estimated as the square root of the recessive homozygous phenotype frequency. However, besides this heterogeneity depending upon the method of the phenotypic ascertainment, also biologically relevant causes appear to play a role. In fact a considerable variation was observed even with loci studied by the same method: for example the *AP* and *MN* loci vs. the others and the *Rh*, *ABO* and *Fy* loci vs. the others.

Two main biological causes for variation may be considered: namely selection and genetic drift.

With no *a priori* hypothesis a possible approach in order to search for selective factors consists in obtaining the correlation matrix for all the variables, that is the 13 genetic markers under examination, the frequencies of Thalassaemia and Glucose-6-phosphate dehydrogenese deficiency and the altitude; then in studying the distribution of all the correlation coefficients (fig. 2).

All the correlations that are significantly different from zero are worth studying in detail and this is one of the goals of the next part of the project. It should be emphasized that the reliability of this approach is strongly supported by our finding that all the previously known correlations [3] have been in fact picked out or very nearly so, namely altitude vs. Thalassaemia and Glucose–6–phosphate dehydrogenase definiency and Thalassaemia vs. Glucose–6–phosphate dehydrogenase deficiency. The possible explanation for the others, if they turn out to be true, is at the moment only a matter of speculation.

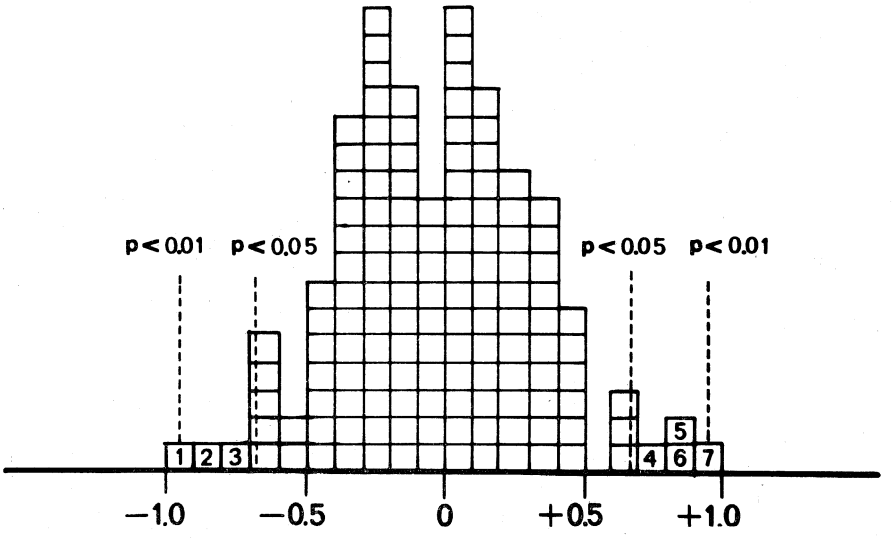

Fig. 2. – Distribution of the 136 correlation coefficients (expressed by the 'z' transformation of 'r') existing between the variabilities in the 10 different villages of the following 17 parameters: *i*) the 13 genetic markers under examination; *ii*) the *Gd–*, β–*th* and *th* frequencies; *iii*) the altitude of the villages. Only the correlations which turned out to be significant are labelled, i.e. (1) alt. vs. *th*, (2) *ADA* vs. *ABO*, (3) alt. vs. *Gd–*, (4) *AP* vs. *Inv*, (5) *Rh* vs. Lu^a, (6) β–*th* vs. *th*, (7) *ABO* vs. Fy^a. The already known correlation between *th* and *Gd–* frequencies has been found on the borderline of the 5 % significance level ($r = 0.448$; $z = 0.485$).

Finally an estimate of the effect of the isolation has been obtained for each of the ten villages, as follows:

First, for each of the 13 markers the genetic distance has been calculated, according to Cavalli–Sforza *et al.* 1969 [4], by using the gene frequencies in the village and those of the whole sample, which have been considered a fair estimate of the overall Sardinian lowland gene frequencies. Secondly, the weighted averages of these 13 genetic distances have been calculated.

Thus a cumulative and weighted estimate of the genetic distance was available for each of the ten villages. These ten figures turned out to be quite different from each other, ranging from less than 0.001 to about 0.005 (Table I). A possible cause for this variation is, of course, the degree of isolation, which, as discussed before, is very heterogeneus. The correlation between these two

parameters has been therefore measured (fig. 3), and it came out to be rather high, $r = 0.68$, and statistically significant ($P < 0.05$).

Both on the basis of our data, i.e. that the vast majority of the correlation coefficients (fig. 2) turned out to be not statistically different from zero, and

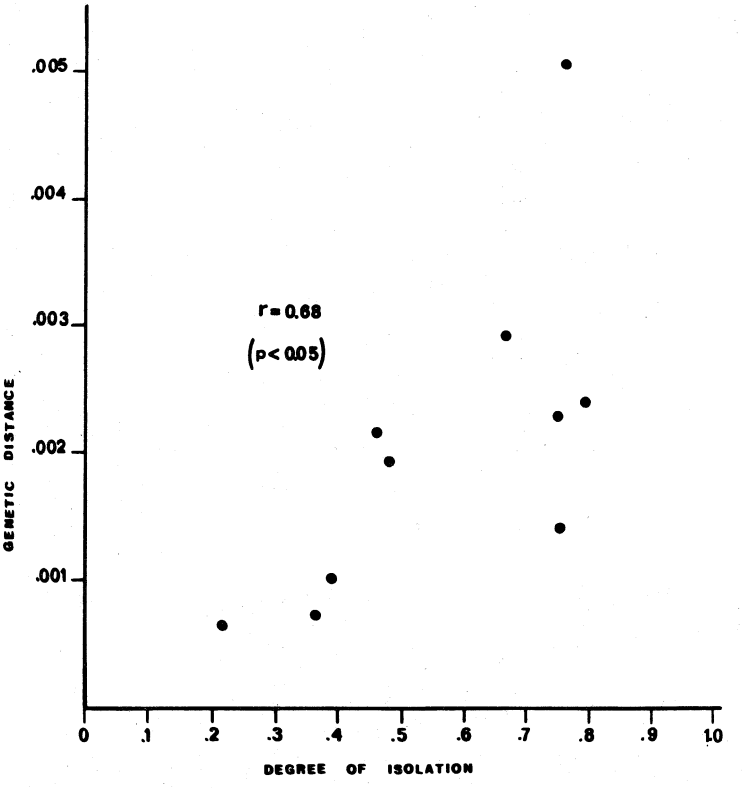

Fig. 3. – Searching for a correlation between the degree of isolation and the genetic distance (the data are reported in Table I).

of *a priori* considerations, taking into account the great ecological similarity among the villages under examination, we can suggest that the main component of the effect of the isolation can be ascribed to the effect of genetic drift. If this is the case, the values shown in fig. 3 are reliable estimates of the effect of drift on villages of about 5000 inhabitants at different degrees of isolation.

REFERENCES

[1] A. E. MOURANT, *The distribution of the Human Blood Groups*, Blackwell Scientific Publications, Oxford (1954).

[2] D. A. HOPKINSON, G. CORNEY, P. J. L. COOK, E. B. ROBSON and H. HARRIS, « Ann. Hum. Genet., London », *34*, 1 (1970).

[3] M. SINISCALCO, L. BERNINI, G. FILIPPI, B. LATTE, P. MEERA KHAN, S. PIOMELLI and M. RATTAZZI, « Bull. W. H. O. », *34*, 379 (1966).

[4] L. L. CAVALLI-SFORZA, L. A. ZONTA, F. NUZZO, L. BERNINI, W. W. W. DE JONG, P. MEERA KHAN, A. K. RAY, L. N. WENT, M. SINISCALCO, L. NIJENHUIS, E. VAN LOGHEM and G. MODIANO, « Amer. J. Hum. Genet. », *21*, 252 (1969).