Andrea Novelletto

A more generai model for balanced polymorphism

Genetica. — *A more general model for balanced polymorphism.* Nota di Andrea Novelletto (*), presentata (**) dal Socio G. Montalenti.

Riassunto. — Le moderne acquisizioni della biologia molecolare dimostrano che possono esistere molte alterazioni molecolari diverse il cui risultato a livello fenotipico è molto simile in quanto producono un gene « non funzionale ». Ciò fa presupporre che ad ogni generazione la frequenza di questi eventi sia molto alta rispetto alle mutazioni specifiche e unidirezionale in quanto non contrastata dagli eventi specifici « di ritorno ».

L'equazione (3) rappresenta un modello che descrive il polimorfismo bilanciato tenendo conto della continua produzione di nuovi alleli e si verifica che in esso rientrano, come casi particolari, l'equilibrio classico (equazione 1) nonchè l'equilibrio fra selezione e mutazione recessiva dannosa. Il modello si applica alle talassemie, le quali costituiscono un esempio di polimorfismo bilanciato in cui, per i motivi suddetti, si può assumere una frequenza consistente di mutazioni pressochè unidirezionali. Il modello permette di prevedere che le frequenze di equilibrio sono leggermente diverse rispetto al modello classico ma, soprattutto, che esse sono raggiunte in un tempo molto minore (fig. 1 a). Ciò può costituire una giustificazione dell'osservazione che i polimorfismi per le talassemie sono presenti in popolazioni molto più limitate rispetto al polimorfismo per l'emoglobina S.

Introduction

Modern advances in molecular studies are now revealing that certain phenotypes are associated with more than one genotype; sometimes these sets of genotypes are highly heterogenous. Most attention has been paid to those situations in which a given phenotype (either pathologic or not) is attributable to the lack or the deficiency of the product of a gene (the protein or its function); the observation of the molecular heterogeneity of these conditions can be justified by considering that a variety of changes in DNA can result in the loss or the reduction of gene expression. In some cases the frequency of " non functional " alleles of a gene (e.g. i allele of the ABO system, l of the Lewis system, α^{thal} and β^{thal} at the globin loci, Gd^- at G6PD locus etc.) is remarkably high; in all those cases in which the gene product has a well-known role, in order to explain the high frequency of these alleles a specific selective factor has been postulated that favours one of the phenotypes associated with " non functional " alleles; G6PD deficiency and thalassemias (deficiency in the globin chain synthesis) are the most widely known examples in this field, the selective factor being malaria (Haldane, 1949). Thalassemias also represent an example of balanced polymorphism in which the variant allele, quite lethal in the homo-

zygous state, confers on heterozygotes a greater resistance to malaria compared with normal homozygotes, this being a selective advantage in malaric areas.

In the classical model of balanced polymorphism (Fisher, 1922) heterozygotes have a selective advantage compared with both homozygotes. In other words heterozygotes have the maximum fitness (1.0) while homozygote fitnesses are respectively $1 - s$ and $1 - t$. In this situation an equilibrium exists in which two alleles survive in the population with frequencies different from 0. If s and t are the selection coefficients against homozygotes for the allele with relative frequency p and q respectively, at equilibrium we have:

$$(1) \qquad q = \frac{s}{s + t} \quad ; \quad p = \frac{t}{s + t} \cdot$$

In this situation the greater fitness of heterozygotes counterbalances the loss of alleles due to the more selected type of homozygotes. Interestingly this equilibrium is stable in the sense that for infinite populations any perturbation changing the equilibrium frequencies is followed by the re-establishment of the equilibrium. Therefore:

$$(2) \qquad q < \boldsymbol{q} \Rightarrow \Delta q > 0 \quad ; \quad q > \boldsymbol{q} \Rightarrow \Delta q < 0 \,.$$

Thus, starting from the initial situation with a unique allele in the population, the appearance of the other one is followed by the spontaneous approach to equilibrium (Fig. 1 *a*, curve g).

THE MODEL

The classical model does not take into consideration that in natural populations a continuous new production of inactive genes occurs during their reproductive history. Such a model can be applied whenever the recurrent mutation rate is negligible. However it can be argued that, in balanced polymorphisms such as thalassemia, this factor cannot be disgregarded. Two main reasons can be put forward:

1) the great extent of potential molecular damage that can transform a " functional " allele into a " non functional " one leads one to think that the overall frequency of these aspecific events is relatively high and in any case much higher than specific alterations;

2) these events would not be balanced by back mutations, from " non functional " to " functional " alleles, because of their rarity due to the high specificity needed for the proper repairs.

Thus it can be suggested that " non functional " alleles are being accumulated continuously although they are heterogenous at the molecular level. At

each generation the overall frequency of these alleles (q) equals the sum of the pre-existing and the newly formed alleles. The variation of the frequency (Δq) equals the difference between the frequency after the selection process lasting for a generation (1st term of the (3)) and the frequency before such a process (2nd term). Thus:

$$(3) \qquad \Delta q = \frac{2\,q^2\,(1-t) + 2\,pq + 2\,p\mu}{2\,[p^2\,(1-s) + 2\,pq + q^2\,(1-t)]} - q$$

where μ is the overall " production " rate per gamete per generation of alleles, their global frequency being q.

Equilibrium is defined by $\Delta q = 0$.

Besides the two trivial solutions $q = 1$ (fixed allele) and $q = 0$ (possible only when $\mu = 0$) also for the above equation an equilibrium solution exists:

$$(4) \qquad q = \frac{s + \sqrt{s^2 + 4\,\mu\,(t+s)}}{2\,(t+s)} \, .$$

This solution is a generalization of the balanced polymorphism model; in fact it can be seen that by setting $\mu = 0$ eq. (1) is obtained again. Moreover, setting $s = 0$ we obtain the more specific equilibrium between selection and a deleterious recessive mutation (Haldane, 1927): at equilibrium $q = \sqrt{\mu/t}$.

Equation (3) has been studied with the aid of a computer program for computing q at each generation. As inputs for s, t, μ and initial q it is possible to use any value ranging from 0 to 1, obtaining a plot of either q or Δq vs. generations (Fig. 1 a, b).

Intuitively it can be understood that in the present model the frequency of the allele formed by mutation will be higher as compared with the classical model considering the same selection coefficients. As illustrated above, this can be attributed to the mutation pressure that tends to raise the frequency of this allele at each generation. Quantitatively the discrepancy in the equilibrium frequencies between the two models is small; for example, if one assumes, for s and t respectively, figures of the order of 0.1 and 1.0 (those normally cited in the case of thalassemias) it can be seen that the difference becomes appreciable only for relatively high mutation rates: for $\mu = 10^{-4}$ it is still about 10^{-3} while for $\mu = 10^{-3}$ it is about 10^{-2} namely about 10% of the frequency that would be maintained by selection alone. It is worth noting that such a rate of production of new alleles, although exceedingly high for specific mutations, may be close to the actual rate in the case of the summation of several aspecific events (not comprising only mutations in the strict sense). It can be noted that, assuming $4\,\mu\,(t+s) \ll s$ (this condition almost always holds,) eq. (4) can be approximated as: $q = s/(s+t) + \mu/s$. This form leads us to interpret the new equilibrium as the classical one plus the ratio between alleles gained by mutation and lost by selection.
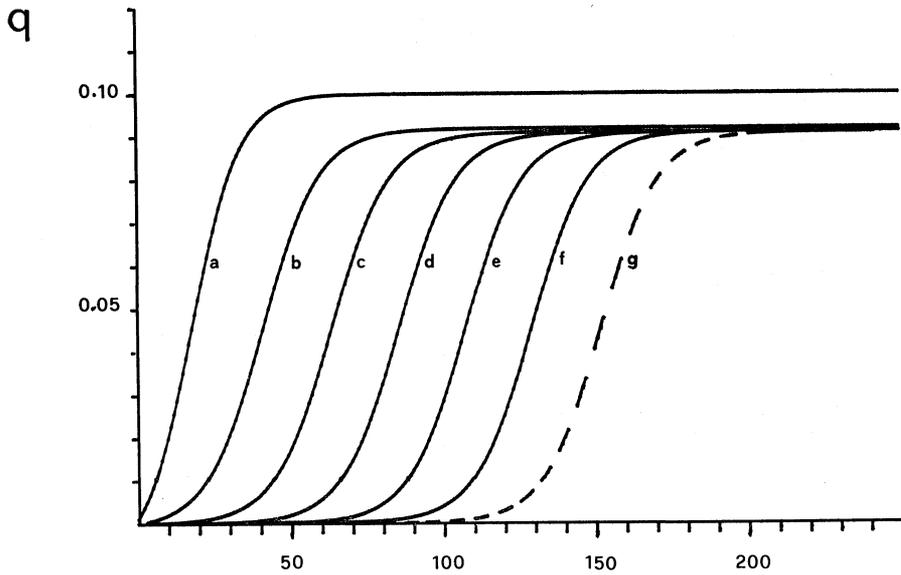
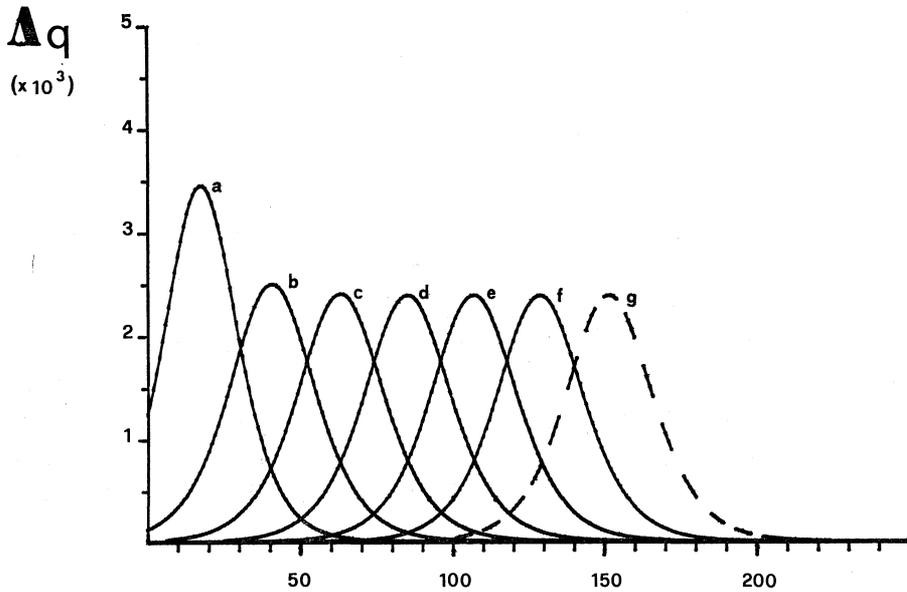Fig. 1 *a*. – Increase of *q* (vertical axis) with generation (horizontal axis).

Fig. 1 *b*. – Variation of *q* (Δq) (vertical axis) with generations (horizontal axis).

Mutation rates and initial values of *q* defined as follows:

a) $\mu = 10^{-3}$, $q = 10^{-3}$.   *b*) $\mu = 10^{-4}$, $q = 10^{-4}$.   *c*) $\mu = 10^{-5}$, $q = 10^{-5}$.

d) $\mu = 10^{-6}$, $q = 10^{-6}$.   *e*) $\mu = 10^{-7}$, $q = 10^{-7}$.   *f*) $\mu = 10^{-8}$, 10^{-8}.

g) $\mu = 0$, $q = 10^{-8}$.

10*

The dynamics of the establishment of equilibrium after the appearance of the first thalassemic allele is even more interesting. Also in the present model eqs. (2) hold true. In the classical model the curve of the increase of the rarest allele is sigmoidal, having a very long phase with vanishing small frequency changes (" lag " phase; Fig. 1 *a*). When the mutation is taken into account the overall shape of the curve does not undergo a substantial change; however the " lag " phase shortens dramatically also when the mutation rate is small (10^{-6}). For an increase of μ of a factor 10 the " log " phase occurs about 25 generations earlier.

DISCUSSION

Going by the present model one might expect that those polymorphisms in which mutations occur continuously in one direction only establish more promptly than those polymorphisms in which an extremely rare mutation only creates the first variant allele. In humans, thalassemias and Hb S polymorphism are representatives of the two types. Although for Hb S a polyphiletic origin of the Asian and the African forms has been demonstrated (Kan and Dozy, 1980), the rate of such a specific mutation as $\beta^A \rightarrow \beta^S$ must be considered very low. Coherently, only those populations that at least once in their history have been numerous, are polymorphic for this trait. On the other hand thalassemia polymorphisms, sharing the same cluster of genes, the same selective factors and very similar selection coefficients, are also present in populations much smaller than the former ones. The suggestion of a heterogeneity of such conditions, initially confined to the distinction between α and β thalassemia, has now been confirmed directly by means of molecular studies. Analyses of the structure of the concerned genes are revealing a large collection of alterations in each of the two gene clusters that can result in these conditions; the same analyses are giving evidence on the vastness of potential alterations, some of which may be still undiscovered.

Considering the high overall rate of these mutations, the present model may represent a quantitative justification of the distribution of these polymorphisms.

However, the possible role of other factors cannot be excluded, including:

1) population size because of stochastic factors that affect the fate of the single mutation making the probability of its survival small (Fisher, 1922; Kimura, 1962);

2) the interaction effect of α and β thalassemias; this effect causes an increase in fitness for those individuals that are homozygotes at one locus and heterozygotes at the other (Kan and Nathan, 1970), with consequences on the joined equilibria.

The simultaneous acting of these factors makes it extremely difficult to trace back a detailed history of the establishment of these polymorphisms.

REFERENCES

FISHER R. A. (1922) – *On the dominant ratio*, « Proc. Roy. Soc. Edinburgh », *42*, 321.

HALDANE J. B. S. (1927) – *A mathematical theory of natural and artificial selection* Part V, « Proc. Cambridge Phil. Soc. », *23*, 838.

HALDANE J. B. S. (1949) – *Disease and evolution*, « Ricerca sci. », *19*, 3.

KAN Y. W. and DOZY A. M. (1980) – *Evolution of the hemoglobin S and C genes in world populations*, « Science », *46*, 341.

KAN Y. W. and NATHAN D. G. (1970) – *Mild thalassemia: the result of interactions of alpha and beta thalassemia genes*, « J. Clin. Invest. », *49*, 635.

KIMURA M. (1962) – *On the probability of fixation of mutant genes in a population*, « Genetics », *47*, 713.