

ENTROPIE DE LA DESCRIPTION D'UNE POPULATION (1)

M. GILLOIS

*Station centrale de Génétique animale,
Centre national de Recherches zootechniques, 78-Jouy-en-Josas*

La relation d'identité et de non-identité entre gènes homologues (M. GILLOIS, 1964) et la relation de dépendance absolue et d'indépendance entre deux gènes homologues (M. GILLOIS, 1965) permettent de mieux définir la description d'une population. L'identité entre gènes homologues entraîne leur dépendance absolue, puisque la connaissance de la nature de l'un nous apporte toute l'information nécessaire pour déterminer la nature de l'autre. L'identité entre gènes homologues est incompatible avec leur indépendance. La non-identité entre gènes homologues est compatible aussi bien avec leur dépendance absolue qu'avec leur indépendance. Deux gènes homologues, non-identiques mais en état de dépendance absolue, sont tels que la connaissance de la nature de l'un nous apporte toute l'information pour déterminer la nature de l'autre ; ces deux gènes homologues, non-identiques pouvant appartenir ou non à la même classe d'isoaction. Par contre, si ces deux gènes homologues, non-identiques, sont en état d'indépendance, la connaissance de la nature de l'un ne modifie pas notre degré d'incertitude quant à la nature de l'autre. Ces remarques ouvrent une voie naturelle pour présenter la description d'une population en utilisant la théorie de l'information.

Toute population, toute génération est décrite par l'énoncé de ses classes d'isoaction et des probabilités qui leur sont attachées, par la description des situations d'identité et des coefficients d'identité, par la connaissance des coefficients d'incertitude zygotique dont nous justifierons plus loin la dénomination. L'ensemble de ces critères traduit en termes condensés un état ou une structure génique. Il ne faut pas confondre cette structure génique avec la structure d'appariement. Seules les populations — même réduites à deux générations successives — peuvent posséder une structure d'appariement : panmixie, homogamie, etc. Décrire la structure génique d'une population c'est pouvoir décider si deux gènes homologues d'un zygote J, pris au hasard parmi tous les zygotes, sont identiques ou non-identiques, si l'un d'eux appartient à une classe d'isoaction particulière, si ces deux gènes homologues, supposés non-identiques, sont absolument dépendants ou indépendants.

(1) Communication faite aux Journées d'Études de la Commission de Génétique de la Fédération européenne de Zootechnie. La Haye, juin 1965.

Il s'agit là de trois événements aléatoires qui nous donnent une certitude plus ou moins grande quant au génotype de ce zygote J.

Étudions l'entropie (1) d'une expérience qui consiste à décider du génotype d'un zygote J pris au hasard dans un groupe ou une génération possédant deux classes d'isoaction A_1 et A_2 , de probabilités *a priori* p et q , et ayant un coefficient de consanguinité f . Soit A le symbole du gène représentant de la classe d'isoaction A_1 et soit a celui du représentant de la classe d'isoaction A_2 . L'expérience a trois issues possibles, les génotypes pouvant être [AA], [Aa], [aa]. Soient P, 2Q, R les probabilités de ces trois issues, en désignant par H l'entropie du système nous avons :

$$H = P \operatorname{colog}_2 P + 2Q \operatorname{colog}_2 2Q + R \operatorname{colog}_2 R.$$

L'entropie sera maximum, et donc l'incertitude la plus grande, pour $P = 2Q = R = 1/3$. H s'annule, si l'une des trois probabilités est égale à un et les deux autres sont égales à zéro ; ce cas est celui de la certitude.

Il est possible d'exprimer P, 2Q, R en fonction des probabilités *a priori* (p , $1 - p = q$, y , $1 - y = y_*$, f , $1 - f = f_*$) (M. GILLOIS, 1964). Ces expressions ne font intervenir aucune hypothèse concernant le mode d'appariement des zygotes de la génération parentale.

$$P = p - pf_*y_*;$$

$$2Q = 2pf_*y_*;$$

$$R = q - pf_*y_*.$$

Les valeurs de p et de f_*y_* qui rendent maximum H, donc l'incertitude, sont :

$$p = 1/2 \quad \text{et} \quad f_*y_* = 1/3.$$

Si la génération, ou le groupe, est tel que $y_* = p = 1/2$, alors $f_* = 2/3$ et le coefficient de consanguinité f égale $1/3$. Ce raisonnement ne permet pas de déterminer les valeurs de f et de y .

Nous allons déterminer quelles valeurs particulières f_* et y_* doivent simultanément prendre pour rendre l'entropie H maximum lorsque p est considéré comme un paramètre. Nous avons précédemment montré que l'expérience qui consiste à décider du génotype d'un zygote J pris au hasard dans une génération possédant deux classes d'isoaction A_1 et A_2 , peut être décomposée en trois expériences E_1 , E_2 , E_3 . Les expériences E_1 et E_2 sont indépendantes :

$$H(1, 2) = H(1) + H(2).$$

L'expérience E_3 est dépendante des expériences E_1 et E_2 :

$$H = H(1, 2, 3) = H(1) + H(2) + H_{1,2}(3).$$

L'expérience E_1 consiste à décider si les deux gènes homologues du zygote J sont identiques ou non-identiques.

$$H(1) = -f \log f - f_* \log f_*.$$

(1) L'entropie H d'une expérience ayant n_i issues possibles de probabilité p_i est $H = \sum_i p_i \operatorname{colog}_2 p_i$.

L'expérience E_2 consiste à décider de la nature de la classe d'isoaction de l'un de ces deux gènes homologues.

$$H(2) = -p \log p - p_* \log p_*$$

L'expérience E_3 consiste à décider de la nature de la classe d'isoaction d'un gène sachant que les deux gènes homologues sont non-identiques et connaissant la nature de la classe d'isoaction de l'un d'entre eux.

$$H_{1,2}(3) = f_* p [-y \log y - y_* \log y_*] + f_* q [-z \log z - z_* \log z_*].$$

En considérant p comme un paramètre, f, f_*, y, y_*, z, z_* comme des variables, cherchons les valeurs de ces dernières qui rendent l'entropie H maximum. Utilisons la méthode des multiplicateurs de Lagrange en introduisant la fonction auxiliaire telle que :

$$\Psi = H + \mu_2 G_2 + \mu_3 G_3 + \mu_4 G_4.$$

avec

$$G_2 = f + f_* - 1 = 0;$$

$$G_3 = y + y_* - 1 = 0;$$

$$G_4 = z + z_* - 1 = 0.$$

L'ensemble des valeurs des variables, qui rendent l'incertitude H maximum sachant que p est un paramètre est

$$y = y_* = 1/2;$$

$$z = z_* = 1/2;$$

$$f = 1/3; f_* = 2/3.$$

En particulier il vient $f_* y_* = 1/3$, valeur précédemment trouvée quand $p = 1/2$.

Pour $p = 1$ et pour $p = 0$, l'entropie H s'annule et il n'y a que deux cas de certitude.

Puisque l'entropie H s'exprime en fonction de p et de $f_* y_*$ nous appelons la quantité $f_* y_*$ le coefficient d'incertitude zygotique.

La notion de coefficient d'incertitude peut être étendue au cas du multi-allélisme. C'est cette quantité qui est atteinte dans le cas de l'estimation de paramètres de la structure génique d'une population quand l'information apportée est le décompte des génotypes ou des phénotypes.

Reçu pour publication en novembre 1965.

SUMMARY

ENTROPY OF POPULATION DESCRIPTION

Describe the genic structure of population implies to be able to decide :

1° Whether two homologous genes from a zygote I, taken at random among all zygotes, are identical or non-identical.

2° Whether one of them belongs to a particular class of isoaction.

3° And whether these two genes, supposed non-identical, are absolutely dependant or independant.

Here are 3 hazardous events because of which we cannot be sure of the nature of genotype of this zygote I.

This uncertainty can be reckoned by the usual formula :

$$H = \sum_i p_i \log p_i$$

(p_i is the probability of each hazardous event).

An application to the case of two classes of isoaction is described and discussed.

RÉFÉRENCES BIBLIOGRAPHIQUES

- BRILLOUIN L., 1956. *Science and information Theory*. Academic press Inc., N. Y.
GILLOIS M., 1964. *La relation d'identité en génétique*. Thèse Fac. Sciences Paris, 294 p.
-