Apidologie, 1980, **11** (1), 87-93.

NUMBER OF SEX ALLELES IN A SAMPLE OF HONEYBEE COLONIES

Jean-Marie CORNUET (*) and Franck ARIES (**)

Station expérimentale d'Apiculture (*)
Laboratoire de Biométrie (**)
I.N.R.A. Domaine Saint-Paul, 84140 Montfavet (France)

SUMMARY

In order to provide a guide in estimating the number of colonies to be tested at the beginning of breeding work, this paper gives the theoretical distribution of the number of sex alleles in a sample of n queens and m drones drawn at random from a population having a known number k of alleles. It is assumed that the allelic frequencies are equal. If m is fixed at six times the value of n (considering that a queen is inseminated by six drones on average) and for k less than or equal to 20, the computation shows that there is a probability greater than or equal to 0.995 for recovering all the alleles in a sample in which n is equal to k.

A compact brood is one of the first qualities that is sought in honeybee colonies. Among the factors interfering with this trait, there is the possible identity of sex alleles of the workers' parents. When elaborating selection design, it is therefore desirable to take this aspect into account as it has been stressed by several authors (WOYKE 1972, MAUL 1972).

Before any breeding work, there are sex alleles in fixed numbers in bee populations. These numbers are the result of the genetic history of these populations : new alleles occur by mutation or immigration, others are eliminated as a result of genetic drift. Usually, at the first stage of selection, a limited number of colonies is selected to produce the next generation. In this paper, we studied the following question only : how many sex alleles are to be found in such a sample as a function of its size and the number of alleles in the original population (which is assumed to be known)?

RATIONALE

We are looking for the probability distribution of the number of sex alleles in a sample of n queens and m drones (represented by their spermatozoa deposited in the n queens'spermathecae) issued from a population having k alleles.

It is assumed that :

1 - the n queens and m drones are drawn at random from the population;

2 - the sample size is small enough compared with the population size to be considered a sample drawn with replacement;

3 - the sex alleles are in equal frequencies (equal to $1/k$).

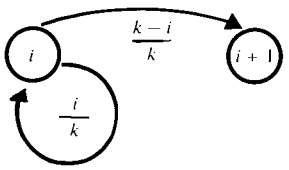
The number of sex alleles in a sample of n queens and m drones drawn at random from a population may be considered as a random variable : $X(n, m)$. The possible values of $X(n, m)$ are the integers between 0 and k.

Let us suppose that, after having drawn n queens and m drones, there are, for instance, i different alleles in the sample (in fact, we are only looking for the probability of such an event). We draw an additional drone. Two mutually exclusive events may then occur :

– either, the allele carried by this drone is identical to one of those already sampled; the probability of such an event is equal to i/k;

– or, this drone carries a new allele; the probability is then equal to $(k - i)/k$.

We may represent the effect of drawing a drone upon the number of sex alleles in the sample by the following scheme :



The circles represent the different " states " of X and the arrows indicate the possible transitions between two states of X with the associated probabilities.

The number of sex alleles obtained through sampling drones may be therefore described by a stochastic process whose $k + 1$ possible states are the integers from 0 to k. The transition probabilities from one state to another are not influenced by the way in which the former state was reached. This process is consequently a Markov chain. Moreover, this Markov chain is discrete (the number of drones is an integer) and homogenous (the transition probabilities do not depend on the number of drones already sampled). As a result, the distribution of $X(n, m)$ is entirely defined by the distribution of $X(n, 0)$ and the transition stochastic matrix Q.

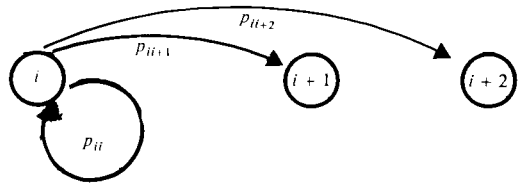
Let us call $z_i(n, m)$ the probability that $x(n, m)$ is equal to i and $Z(n, m)$ the row vector $|z_0(n, m), z_1(n, m), ..., z_i(n, m), ..., z_k(n, m)|$ giving the distribution of $X(n, m)$. We may write :

$$Z(n, m) = Z(n, m - 1) \quad Q = Z(n, 0)\, Q^m \qquad (1)$$

The transition matrix Q is composed of elements q_{ij} $(i, j = 0, 1, 2...k)$ whose value is the conditional probability that $X(n, m + 1)$ is equal to j, knowing that $X(n, m)$ is equal to i. We have already established that if $X(n, m)$ is in state (i), $X(n, m + 1)$ can only be in states (i) or $(i + 1)$. So Q is defined by the following relationships :

$$\begin{aligned} \forall j \quad & \neq i, i + 1 \qquad q_{ij} \quad 0 \\ q_{ii} \quad & = i/k \\ q_{ii+1} & = (k - i)/k \end{aligned}$$

We have yet to define $Z(n, 0)$. Using the same argument, the sampling of queens may also be considered as a stochastic process. Introducing an additional queen may bring 0, 1 or 2 new alleles according to the scheme :



This process is also a discrete and homogenous Markov chain. Calling P the transition matrix, we have :

$$Z(n, 0) = Z(n-1, 0) \qquad P = Z(0, 0)\, P^n \tag{2}$$

Let us establish $P = (p_{ij})\ (i, j = 0, 1, 2 \ldots k)$. Since a queen is heterozygous at the sex locus, there are $k(k-1)/2$ possible genotypes which are encountered with the same probability if the population is at equilibrium. So P is defined by :

$$\forall j \neq i, i+1, i+2 \qquad p_{ij} = 0$$

$$p_{ii} = \frac{\binom{i}{2}}{\binom{k}{2}} = \frac{i(i-1)}{k(k-1)}$$

$$p_{ii+1} = \frac{\binom{i}{1}\binom{k-i}{1}}{\binom{k}{2}} = \frac{2\,i(k-i)}{k(k-1)}$$

$$p_{ii+2} = \frac{\binom{k-i}{2}}{\binom{k}{2}} = \frac{(k-i)(k-i-1)}{k(k-1)}$$

$Z(0, 0)$ is clearly equal to $(1, 0, 0, \ldots, 0)$ (3)

Combining equations (1), (2) and (3), we have

$$\boxed{Z(n, m) = (1, 0, 0, \ldots 0)\, P^n\, Q^m} \tag{4}$$

N.B. : Inverting the two parts of the reasoning, we would have got :

$$Z(n, m) = (1, 0, 0, \ldots, 0)\, Q^m\, P^n \tag{4'}$$

The identity of equations (4) and (4') results from the commutability of the product of matrices P and Q which is easily proved.

Explicit computation of z_i's is complex but expected value and variance of $X(n, m)$ may be expressed in a simple manner :

$$\boxed{\begin{aligned} E[X(n, m)] &= k - \frac{(k-1)^m (k-2)^n}{k^{m+n-1}} \\ V[X(n, m)] &= \frac{(k-1)^m (k-2)^n}{k^{m+n-1}}\left[1 - \frac{(k-1)^m (k-2)^n}{k^{m+n-1}} + \frac{(k-2)^m (k-3)^n}{(k-1)^{m+n-1}}\right] \end{aligned}}$$

TABL. I. — *Theoretical distribution of the number of sex alleles in samples obtained from a population having 12 alleles.*

Sample composition		Distribution of the number of sex alleles in the sample (probability × 1 000)														
Queens	Drones	0	1	2	3	4	5	6	7	8	9	10	11	12	Expected value	Standard deviation
1	6	0	0	0	2	41	219	414	273	51	0	0	0	0	6.067	0.935
2	12	0	0	0	0	0	0	8	65	222	355	261	80	8	9.067	1.101
3	18	0	0	0	0	0	0	0	1	18	114	320	388	158	10.550	0.958
4	24	0	0	0	0	0	0	0	0	1	16	126	413	444	11.283	0.749
5	30	0	0	0	0	0	0	0	0	0	2	35	278	685	11.645	0.557
6	36	0	0	0	0	0	0	0	0	0	0	9	157	834	11.825	0.404
7	42	0	0	0	0	0	0	0	0	0	0	2	82	915	11.913	0.289
8	48	0	0	0	0	0	0	0	0	0	0	0	42	958	11.957	0.205
9	54	0	0	0	0	0	0	0	0	0	0	0	21	979	11.979	0.145
10	60	0	0	0	0	0	0	0	0	0	0	0	10	990	11.990	0.102
11	66	0	0	0	0	0	0	0	0	0	0	0	5	995	11.995	0.072
12	72	0	0	0	0	0	0	0	0	0	0	0	3	997	11.997	0.051
13	78	0	0	0	0	0	0	0	0	0	0	0	1	999	11.999	0.036

RESULTS AND DISCUSSION

Distributions of $X(n, m)$ have been computed for all the values of k (number of alleles in the original population) between 3 and 20. For each k, n varies between 1 and $k + 1$. Finally, considering that a queen is inseminated by 6 drones on average, we have postulated that the number of drones is 6 times greater than the number of queens ($m = 6\ n$).

For example, table 1 gives the theoretical distributions in the case of a population with 12 sex alleles. We see that the probability of drawing the 12 alleles is equal to 0.99 with a sample of 10 queens (inseminated by a total of 60 drones). With 13 queens (and 78 drones), this probability reaches 0.999.

With a ratio of 6 drones to 1 queen, and for $k = 3, 4...20$, this computation shows that with a number of queens equal to the number of sex alleles in the population, the probability of collecting all the alleles in one sample is at least 0.995.

Of course, this result holds only if the three assumptions postulated above are correct.

Firstly, queens must be drawn at random from the population. That is, they must not be more closely related than would be expected in a random sample of the population. In effect, this amounts to excluding samples of sister queens. Natural insemination of queens ensures that the sample of drones is selected at random from the whole available population.

Secondly, the above result shows that a very small sample is sufficient for collecting all sex alleles, which fulfils the second condition.

The third condition is usually assumed in studies of honeybee sex alleles (LAIDLAW and *al.*, 1956; KERR, 1967; WOYKE, 1976; ADAMS and *al.*, 1977). This amounts to assuming equilibrium in an incompatibility system.

Our formulae appear to have general application in estimating the number of sex alleles in a random sample of honeybee colonies. Combined with estimates of the number of sex alleles in panmictic populations (see references already cited), they may provide a guide in estimating the number of colonies to be tested at the beginning of any breeding work. On the other hand, their exactness can not be guaranteed after the first stage of selection since sampling is no longer at random.

In other respects, these formulae may be of some interest for the study of colonization of virgin areas by honeybees. For instance, in Kangaroo Island, a new population of bees was started with six colonies (WOYKE, 1976). If these were derived from a population with 12 sex alleles and if the three above conditions were fulfilled, table 1 shows that there would be a high probability (0.834) that all 12 alleles were carried by these colonies. Of course, we do not know the actuel number in the original population, so further speculation is not possible.

Received for publication in October 1979.

RÉSUMÉ

LE NOMBRE D'ALLÈLES SEXUELS DANS UN ÉCHANTILLON DE COLONIES D'ABEILLES

Plusieurs auteurs ont souligné l'intérêt de tenir compte des allèles sexuels dans tout travail de sélection. La question étudiée dans cet article est la suivante : quelle est la distribution théorique du nombre d'allèles sexuels représentés dans un échantillon de n reines et m mâles provenant d'une population contenant un nombre k (supposé connu) d'allèles.

Le tirage des mâles de l'échantillon est assimilé à un processus aléatoire de même que le tirage des reines. Ces deux processus sont du type chaîne de Markov discrète et homogène. La distribution cherchée se déduit donc directement de la distribution initiale (pour $n = m = 0$) une fois précisées les deux matrices de transitions notées P pour les reines et Q pour les mâles selon la formule :

$$Z(n, m) = (1, 0, 0...0)\, P^n\, Q^m$$

où $Z(n, m)$ est le vecteur ligne des probabilités des $k + 1$ valeurs possibles du nombre d'allèles sexuels de l'échantillon. L'espérance et la variance de cette variable s'écrivent :

$$E = k - \frac{(k-1)^m\,(k-2)^n}{k^{m+n-1}}$$

$$V = \frac{(k-1)^m\,(k-2)^n}{k^{m+n-1}} \left[1 - \frac{(k-1)^m\,(k-2)^n}{k^{m+n-1}} + \frac{(k-2)^m\,(k-3)^n}{(k-1)^{m+n-1}}\right]$$

L'application de ces formules pour k compris entre 3 et 20 et pour un nombre de mâles égal à 6 fois celui des reines ($m = 6\ n$) montre qu'avec un nombre de reines égal au nombre d'allèles dans la population, la probabilité d'avoir tous les allèles représentés dans l'échantillon est toujours au moins égale à 0,995.

Ceci n'est vrai qu'à trois conditions :

1 - reines et mâles sont tirés au hasard dans la population;

2 - la taille de l'échantillon est petite devant celle de la population;

3 - les fréquences géniques sont supposées égales.

Ces conditions sont habituellement remplies, ce qui confère au résultat énoncé plus haut une bonne généralité. L'intérêt majeur de nos formules est de servir de guide pour l'estimation du nombre de géniteurs à tester au départ d'une éventuelle sélection.

ZUSAMMENFASSUNG

DIE ANZAHL DER SEXALLELE IN EINER PROBE VON BIENENVÖLKERN

Mehrere Autoren haben betont, wie wichtig es ist, während der ganzen Selektionsarbeit den Sexallelen Beachtung zu schenken. Die Frage, die in dieser Arbeit untersucht wird, ist die folgende : Wie ist die theoretische Verteilung der Zahl der Sexallele in einer Stichprobe aus n Königinnen und m Drohnen, die aus einer Population mit k Allelen (es wird vorausgesetzt, dass diese Zahl bekannt ist) stammen.

Die Entnahme der Drohnen aus der Stichprobe erfolgt nach einem Prozess der zufälligen Auswahl. Ebenso die Entnahme der Königinnen. Diese beiden Prozesse entsprechen dem Typ einer diskreten und homogenen Markovschen Reihe. Die gesuchte Verteilung leitet sich direkt von der Ausgangsverteilung (für $n = m = 0$) ab, sobald die beiden Transitionsmatrizen (P für die Könniginnen und Q für die Drohnen) nach der folgenden Formel bestimmt sind :

$$Z(n, m) = (1, 0, 0...0)\, P^n\, Q^m$$

Dabei ist $Z(n, m)$ der lineare Vektor der Wahrscheinlichkeiten von $k + 1$ möglichen Werte der Anzahl der Sexallelein der Probe. Der Erwartungswert und die Varianz dieser Variablen wird wie folgt beschrieben :

$$E = k - \frac{(k-1)^m (k-2)^n}{k^{m+n-1}}$$

$$V = \frac{(k-1)^m (k-2)^n}{k^{m+n-1}} \left[1 - \frac{(k-1)^m (k-2)^n}{k^{m+n-1}} + \frac{(k-2)^m (k-3)^n}{(k-1)^{m+n-1}} \right]$$

Die Anwendung dieser Formeln für ein k zwischen 3 und 20 und für eine Zahl von Drohnen, der einem Sechsfachen der Zahl an Königinnen ($m = 6\ n$) entspricht, zeigt, dass mit einer Königinnenzahl, die gleich ist mit der Zahl der Sexallele in einer Population, die Wahrscheinlichkeit immer mindestens 0,995 beträgt, alle Sexallele in einer Stichprobe vertreten zu haben.

Das trifft aber nur unter den folgenden drei Bedingungen zu .

1 - Königinnen und Drohnen werden rein zufallsmässig aus der Population entnommen.

2 - Die Grösse der Stichprobe ist klein im Vergleich zur Grösse der Population.

3 - Es wird angenommen, dass die Genfrequenzen gleich sind.

Diese Bedingungen werden gewöhnlich erfüllt sein, so dass man den oben angeführten Resultaten ein hohes Mass an Allgemeingültigkeit zuerkennen kann. Die wesentlichste Bedeutung unserer Formeln liegt darin, dass man sie als Leitlinie für die Schätzung der Zahl der zu prüfenden Ahnen zu Beginn einer geplanten Selektion benutzen kann.

REFERENCES

ADAMS J., ROTHMAN E. D., KERR W. E., PAULINO Z. L., 1977. – Estimation of the number of sex alleles and queen mating from diploid males frequencies in a population of *Apis mellifera*. *Genetics*, **86 :** 583-596.

KERR W. E., 1967. – Multiples alleles and genetic load in bees. *J. apic. Res*, **6** (2) : 61-64.

LAIDLAW H. H., GOMES F. P., KERR W. E., 1956. – Estimation of the number of lethal alleles in a panmictic population of *Apis mellifera L. Genetics*, **41 :** 179-188.

MAUL V., 1972. – Programme d'élevage avec allèles sexuels déterminés. « Symposium international de Lunz am See : Contrôle de l'accouplement et sélection chez l'Abeille mellifère ». 75-79.

WOYKE J., 1972. – Les allèles sexuels et la fécondation contrôlée. « Symposium international de Lunz am See : Contrôle de l'accouplement et sélection chez l'Abeille mellifère ». 69-74.

WOYKE J., 1976. – Population genetic study on sex alleles in the honey bee, using the example of the Kangaroo Island bee sanctuary. *J. apic. Res*, **15** (3/4) : 105-123.