Theor Appl Genet (1986) 73:223–227

Genetic pollution and number of matings in a black honey bee (*Apis mellifera mellifera*) population

J.-M. Cornuet [1], A. Daoudi [1] and C. Chevalet [2]

[1] Station de Zoologie et d'Apidologie, INRA, F-84140 Montfavet, France
[2] Laboratoire de Génétique Cellulaire, INRA, B.P. 27, F-31326 Castanet Tolosan Cedex, France

Received June 4, 1986; Accepted August 15, 1986
Communicated by A. Robertson

Summary. In the south-east of France, local honey bees possess only the B allele at the MDH locus, whereas the races which are usually imported into this area do not have this allele. The proportion of non-B genes in a sample of drones was used to measure the "genetic pollution" in the local population. Within the course of a breeding scheme of local bees, 99 queens, whose genotypes are BB, were naturally mated between April 25 and June 10, 1985 at la Tave (Gard, France). Twenty daughters-workers of each queen were analysed at the MDH locus. The frequency of the B allele in drones that mated with these queens is estimated by the proportion of workers with genotype BB and the genetic pollution by the cumulated frequency of the other alleles. The sampling variances of these frequencies involve a coefficient which is a function of the average number of drones mated with a queen. This latter parameter is estimated through the maximum likelihood method. In addition to the three well-known alleles, a rare allele (frequency = 0.0055), possibly equivalent to the S1 allele described by Badino et al. (1983), has been found in three different colonies. Cumulating the frequencies of the non-B alleles results in an estimation of the genetic pollution equal to 0.0394 (±0.0071). This low value allows us to proceed to the next step of the selection project. The mean number of drones mated to a queen is 12.4 with a (10.4–19.3) confidence interval at the 90% level.

Key words: *Apis mellifera* – Population genetics – Allelic frequencies – Allozyme variation

Introduction

Among the means at a beekeeper's disposal to improve the production of colonies, one of the most frequently used has been and still is the crossing of geographic races. This practice, however, can lead to a gradual change in the local population's genome. Amongst the drones that mate with the queens of a given area, there are some that come from hybrid or foreign queens. Through natural fecundation, they transmit their "foreign" genes to the local colonies. Whether the consequences are beneficial or detrimental for the concerned populations in the long term is impossible to predict; nevertheless, their genetic evolution will be modified.

Within the course of a breeding program, we estimated the "genetic pollution", defined as the proportion of foreign genome present in the population. For this purpose, we used an allozyme marker, the malate dehydrogenase, whose variability is controlled by a locus with at least 3 alleles, called here A, B and C (Sylvester 1976). The allelic frequencies have been estimated for various races (Cornuet 1982; Sheppard and Berlocher 1984, 1985). It appears that:

1) the local black race (*A. m. mellifera*) only possesses the B allele.
2) the only other races sharing this allele have, as far as we know, never been imported into this area.

Consequently, we shall consider that, in our particular region, the genetic pollution can be estimated by the proportion of males with non B genotypes in the total male population.

Incidentally, the data obtained for this experiment also enables one to jointly estimate the average number of drones mated with a queen through analysing the between colony variability of B genotype frequencies.

Material and methods

1 Sampling

The most direct method to estimate genotypic frequencies in males of a given area would be to analyse a representative sample (e.g. capturing drones in a mating area).

Thanks to a breeding program, we had at our disposal a sample of 99 colonies whose queens were all of the BB genotype and were mated between April 25 and June 10, 1985 in the same location (La Tave, Gard, South-East of France). Twenty workers were sampled from each of the 99 colonies. These were removed from the colony within a few hours before or after emergence in order to be sure that they were daughters of the queen of the colony.

2 Electrophoresis

The genotype at the MDH locus was determined through polyacrylamide gel electrophoresis (Cornuet 1979).

3 Statistical analysis

Since all workers are daughters of BB queens, their genotype is either BB (B father) or BX (non B father). Let z_i and n_i denote the number of BB workers and the total number of analysed workers in the i-th colony, respectively. The frequency p_B of the B allele in the drones population may be estimated by:

$$\hat{p}_B = \left(\sum_{i=1}^{I} z_i\right) \Big/ \left(\sum_{i=1}^{I} n_i\right).$$

This estimator is unbiased, but, in order to compute its sampling variance, it is necessary to know the statistical distribution of the number of males mated with a queen. Assuming that this number follows a truncated Poisson distribution with parameter λ, we have (cf. appendix):

$$\mathrm{Var}\,(\hat{p}_B) = p_B\,(1 - p_B) \left(\sum_i n_i + \sum_i n_i\,(n_i - 1)\,C_\lambda\right) \Big/ \left(\sum_i n_i\right)^2$$

where C_λ is a coefficient which depends on the average number of drones mated with a queen:

$$C_\lambda = \sum_{m=1}^{\infty} \frac{e^{-\lambda}}{1 - e^{-\lambda}} \frac{\lambda^m}{m!\,m}.$$

The value of λ may be obtained by analysing the variability of the B allele frequencies z_i/n_i among the I colonies. A simple example will help understand the logic behind this statement. Consider first the case in which each queen mates with exactly one drone: we will obtain only two extreme types of colonies: colonies with only BB workers (B drone) and colonies with only BX workers (X drone). On the other hand, if each queen mates with a very large number of drones, all colonies will exhibit more or less the same proportions of BB and BX workers. In this latter case, the variance of the B allele frequency between colonies is evidently lower than in the former case. As a matter of fact, an analogous computation shows that:

$$E\left(\sum_{i=1}^{I} (z_i - n_i\,\hat{p}_B)^2\right)$$

$$= p_B\,(1 - p_B) \sum_{i=1}^{I} (1 + (n_i - 1)\,C_\lambda)\,n_i \left(1 - n_i \Big/ \sum_{k=1}^{I} n_k\right)^2.$$

Substituting p_B for its estimation $\hat{p}_B$ in the above formula permits the computation of C_λ, and therefore an estimation $\hat{\lambda}$ of the average number of drones mated with a queen.

A more rigorous approach of the joint estimation of both parameters, p_B and λ, based on the maximum likelihood principle is given in the appendix. This method allows one to define confidence intervals for both parameters.

Results

A total of 4 different alleles were found in the 1,980 workers analysed: the 3 classical alleles (A, B and C) and a rare one (D), which may be equivalent to the S1 allele detected by Badino et al. (1983). The total numbers of the various genotypic categories are indicated in Table 1 and the allelic frequencies in males are presented in Table 2. To compute the standard deviations (Table 2) where the value of C_λ is needed, it is necessary to perform the following computations.

The genetic pollution, taken as the cumulative frequency of non B alleles, is estimated to be 0.0394 ($\pm$ 0.0071). The estimation of λ, is obtained through the direct computation of the variance of the B allelic frequencies within colonies (Table 3). Taking the value of

$$p_B^0 = 0.9606 \quad \text{leads to:} \quad \lambda^0 = 12.51\,.$$

Starting with these initial values and solving the maximum likelihood equations by successive iterations leads to the solution

$$\tilde{p}_B = 0.9606\,; \quad \tilde{\lambda} = 12.41\,.$$

The likelihood function gives the a posteriori distribution of these parameters (assuming a uniform

Table 1. Number of the various MDH genotypes of the sampled bees

Genotype	AB	BB	BC	BD	Total
No. of bees	58	1,902	9	11	1,980

Table 2. Allelic frequencies of the drones

Allele	Frequency	Std. deviation
A	0.0293	0.0062
B	0.9606	0.0071
C	0.0045	0.0025
D	0.0055	0.0027

Table 3. Distribution of the numbers of homozygote BB bees sampled in 99 colonies (20 bees were typed in each colony)

No. of homozygotes BB	20	19	18	17	16	15	14
No. of colonies	67	11	9	4	4	3	1

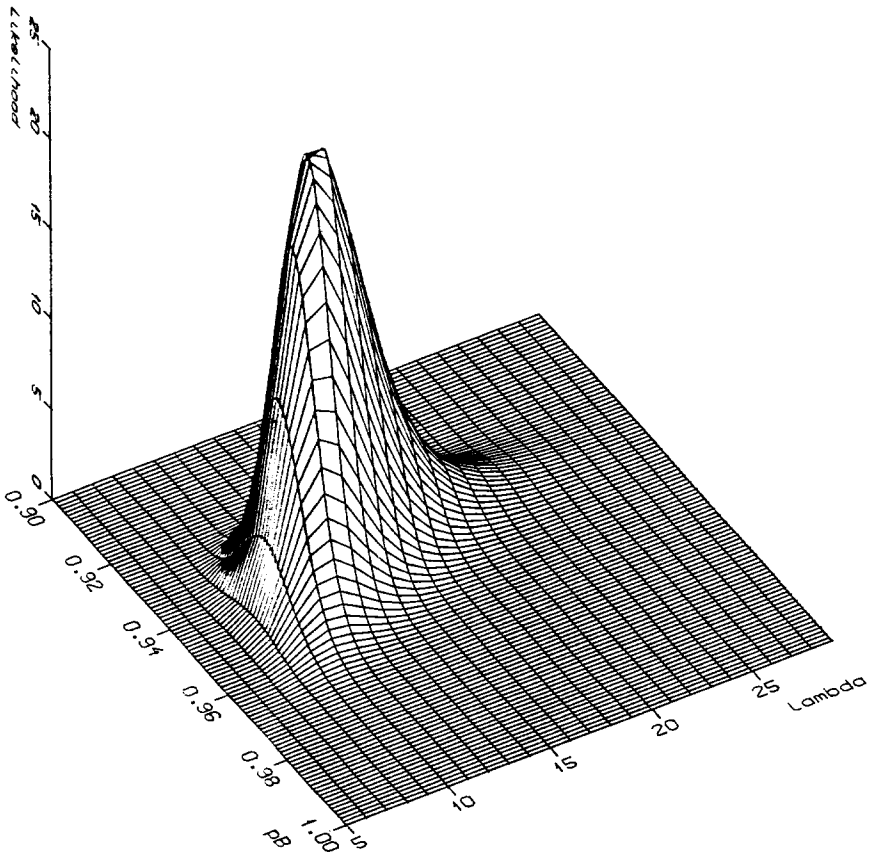

Fig. 1. The likelihood of the sample as a function of the parameters p_B and λ

a priori distribution) and provides joint confidence intervals (Fig. 1):

90% level $0.947 < p_B < 0.971$
$10.4 < \lambda < 19.3$

95% level $0.944 < p_B < 0.972$
$9.8 < \lambda < 20.8$.

Discussion

1 Genetic pollution

The La Tave bee yard was chosen because this area appeared to be isolated from foreign influences. Actually, the low level of genetic pollution found in this experiment confirms this supposition and allows the continuation of the local bee breeding program undertaken by our laboratory staff.

We chose to consider the D allele as a foreign gene, even though we never had found this allele before, neither in the local race nor in any other one. However, since the D allele frequency in our sample is very low (0.0055), the error in estimating the genetic pollution will be necessarily low. On the other hand, the genetic pollution is a factor that we want to reduce to a minimum. Therefore, an overestimation seemed preferable to an underestimation. Whatever it is, the gradual discovery of new alleles at the MDH locus (Badino et al. 1983; Sheppard and Berlocher 1984, 1985) stresses the interest for thorough studies on new populations.

As far as the precision of estimation is concerned, some remarks have to be presented. First, if one is limited by the total number of bees to be analysed, the precision of estimating genic frequencies in males through their daughters increases as the number of sampled colonies becomes higher. Taking the limiting case of one worker per colony, one is sure to have as many males represented as honey bees analysed. Second, if the total number of colonies is limited by experimental conditions, the gain in precision diminishes quickly when the number of analysed workers per colony increases. Third, the precision depends on the average number of matings per queen in the sample of colonies.

2 Mating number

The rationale followed to estimate the average mating number of a queen is analogous to the one presented by Adams et al. (1977) who based their computations on the frequency of diploid males in the brood. Some assumptions had to be formulated.

When workers are sampled in a colony, it is assumed that there is a stochastic independence in the drawing of any two bees. But, particularly in the weeks that follow fecundation, there is an incomplete mixing of sperm inside the spermatheca which increases the probability that two equally aged workers come from the same father. The consequence is the introduction of some genotypic correlation between workers sampled at the same time in the same colony, and an underestimation of the number of drones mated with the queen. In order to prevent, at least partly, these effects, we took the following precautions during sampling: for each colony, 2 samples, of 12 and 8 bees, were taken one month apart. For each sample, bees were taken from different frames and at slightly different ages (a few hours before to a few hours after emergence).

It is also assumed that any male inseminates a queen with an equal number of spermatozoids. Although this assumption seems somehow unrealistic, we had to keep it in order to perform the computations. Conversely, even if it looks as unverifiable as the previous one, the hypothesis of an absence of correlation between the fecundating power of a spermatozoid and its MDH locus genotype seems more reasonable.

Concerning the kind of distribution of the number of matings, the choice of a truncated Poisson law proposed by Adams et al. (1977) seems to be the most logical.

For a more detailed discussion about the estimation methods and the values found by different authors, we refer the reader to the article cited above. We simply observe that our estimation is slightly higher than those previously established under similar latitudes (Taber and Wendel 1958). In any case, the precision of the estimation method is poor. A practical consequence of the former remark would be to consider as insufficient less than 10 drones for artificial insemination of a queen.

Appendix

1 Distribution of the number of BB genotype workers

Let us consider a sample of workers from a given BB queen. Out of the n analysed workers, z have the BB genotype and n−z have another genotype. Since the sample size is very small compared to the whole colony size, sampling approaches one of drawing with replacement. Calling f the probability that a worker of this colony comes from a B father, the distribution of Z, the associated random variable, may be taken as binomial with parameters n and f. Hence, we have the relationship:

$$\text{Prob}\,(Z = z/n) = C_n^z\, f^z\,(1-f)^{n-z}\,. \tag{1}$$

Assuming an equal number of spermatozoids to fecundate an ovule whatever the genotype (B or non B) of the male, it can be written that:

$$f = x/m$$

where x and m are, respectively, the number of B males and the total number of males mated with the queen.

The large number of males present on the mating area allows one to consider the distribution of X, the random variable associated with the number of B males mated with the queen, as binomial with parameters m and p_B (p_B is the B allele frequency in the males population). Then we have

$$\text{Prob}\,(X = x/p_B) = C_m^x\, p_B^x\,(1-p_B)^{m-x}\,. \tag{2}$$

The value of m is the occurrence of an event for which the probability is assumed to follow a truncated Poisson distribution (the existence of workers implies that the queen mated with at least one male) with parameter λ:

$$\text{Prob}\,(M = m/\lambda) = e^{-\lambda}\,\lambda^m/m!\,(1-e^{-\lambda})\,. \tag{3}$$

The distribution of Z is conditional on the values taken on by the two other random variables X and M. By using the corresponding probabilities as weighing factors, we have:

$$\text{Prob}\,(Z = z/n;\, p_B, \lambda) = \sum_{m=1}^{\infty} \frac{e^{-\lambda}}{1-e^{-\lambda}} \frac{\lambda^m}{m!} \sum_{x=0}^{m} C_m^x\, p_B^x\,(1-p_B)^{m-x} \cdot C_n^z \left(\frac{x}{m}\right)^z \left(1-\frac{x}{m}\right)^{n-z}\,. \tag{4}$$

It is easily found that

$$E\,(Z) = n\,p_B$$

$$\text{Var}\,(Z) = n\,p_B\,(1-p_B)\,(1+(n-1)\,C_\lambda)$$

where C_λ is the sum of the series $\sum_{m=1}^{\infty} \frac{e^{-\lambda}}{1-e^{-\lambda}} \frac{\lambda^m}{m!\,m}$.

2 Maximum likelihood estimations

The distribution of the BB workers numbers, z_i, depends on two unknown parameters: p_B, the B allele frequency in males and λ, the average number of males mated with a queen. The likelihood is equal to

$$L = \prod_{i=1}^{l} \text{Prob}\,(Z = z_i/n_i;\, p_B, \lambda)$$

and the corresponding estimations, p_B and λ, are solutions of

$$\frac{\delta \text{Log}\,(L)}{\delta p_B} = 0 \quad \text{and} \quad \frac{\delta \text{Log}\,(L)}{\delta \lambda} = 0\,.$$

It is convenient to transform the double summation in formula (4) using y = m − x so that:

$$\sum_{m=1}^{\infty} \sum_{x=0}^{m} = \sum\sum_{x,y}'$$

where $\sum\sum_{x,y}'$ is the summation over all values (x, y) such that $x \geqslant 0$ and $y \geqslant 0$ but $(x, y) \neq (0, 0)$. Formula (4) then becomes:

$$\text{Prob}\,(Z = z/n;\, p_B, \lambda) = \sum\sum_{x,y}' \frac{e^{-\lambda}}{1-e^{-\lambda}} \frac{\lambda^{x+y}}{x!\,y!}\, p_B^x\,(1-p_B)^y \cdot C_n^z \left(\frac{x}{x+y}\right)^z \left(\frac{y}{x+y}\right)^{n-z}\,. \tag{5}$$

With the following notations:

$$P_i = \text{Prob}\,(Z = z_i/n_i;\, p_B, \lambda)$$

$$PX_i = \sum_{x,y}{}' \frac{e^{-\lambda}}{1-e^{-\lambda}} \frac{\lambda^{x+y}}{x!\,y!} p_B^x (1-p_B)^y \cdot C_{n_i}^{z_i} \left(\frac{x}{x+y}\right)^{z_i} \left(\frac{y}{x+y}\right)^{n_i - z_i} x$$

$$PY_i = \sum_{x,y}{}' \frac{e^{-\lambda}}{1-e^{-\lambda}} \frac{\lambda^{x+y}}{x!\,y!} p_B^x (1-p_B)^y \cdot C_{n_i}^{z_i} \left(\frac{x}{x+y}\right)^{z_i} \left(\frac{y}{x+y}\right)^{n_i - z_i} y$$

the maximum likelihood equations become:

$$\frac{\delta \text{Log}\,(L)}{\delta p_B} = \frac{1}{p_B} \sum_{i=1}^{I} \frac{PX_i}{P_i} - \frac{1}{1-p_B} \sum_{i=1}^{I} \frac{PY_i}{P_i} = 0 \qquad (6a)$$

$$\frac{\delta \text{Log}\,(L)}{\delta \lambda} = \frac{1}{\lambda} \sum_{i=1}^{I} \frac{PX_i + PY_i}{P_i} - \frac{I}{1-e^{-\lambda}} = 0\,. \qquad (6b)$$

This system is solved by successive iterations. Starting from a pair (p_B, λ), we get a new estimation (p'_B, λ') through the equations:

$$p'_B = \frac{\sum_{i=1}^{I} (PX_i\,(p_B, \lambda)/P_i\,(p_B, \lambda))}{\sum_{i=1}^{I} (PX_i\,(p_B, \lambda) + PY_i\,(p_B, \lambda))/P_i\,(p_B, \lambda)} \qquad (7a)$$

$$\frac{\lambda'}{1-e^{-\lambda'}} = \frac{1}{I} \sum_{i=1}^{I} (PX_i\,(p_B, \lambda) + PY_i\,(p_B, \lambda))/P_i\,(p_B, \lambda)\,. \qquad (7b)$$

From a Bayesian point of view, the likelihood function represents the a posteriori probability distribution of parameters p_B and λ, taking into account the observed data (z_i) if a uniform a priori distribution of these parameters is assumed. The computation of this function allows one then to establish confidence intervals.

References

Adams J, Rothman ED, Kerr WE, Paulino ZL (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

Badino G, Celebrano G, Manino A (1983) Population structure and *Mdh-1* locus variation in *Apis mellifera ligustica.* J Hered 74:443–446

Cornuet JM (1979) The MDH system in the honeybees (*Apis mellifera* L.) of Guadaloupe. J Hered 70:223–224

Cornuet JM (1982) The MDH polymorphism in some west Mediterranean honeybee populations. In: IUSSI Congress. Boulder Colo (USA), pp 5–6

Sheppard WS, Berlocher SH (1984) Enzyme polymorphism in *Apis mellifera* from Norway. J Apic Res 23:64–69

Sheppard WS, Berlocher SH (1985) New allozyme variability in Italian honey bees. J Hered 76:45–48

Sylvester MA (1976) Allozyme variation in honeybees (*Apis mellifera* L.) PhD Thesis, University of California (USA)

Taber S, Wendel J (1958) Concerning the number of times queen bees mate. J Econ Entomol 51:786–789