

ESTIMATION OF THE NUMBER OF LETHAL ALLELES IN A PANMITIC POPULATION OF *APIS MELLIFERA* L.

HARRY H. LAIDLAW, JR.¹, F. PIMENTEL GOMES, AND WARWICK E. KERR²

Department of Entomology and Parasitology, University of California, Davis, Department of Mathematics, Universidade de Sao Paulo, Brazil, and Department of Genetics, Universidade de Sao Paulo, Brazil

Received July 5, 1955

BEESKEEPERS have long been familiar with a poor type of brood which due to its scattered appearance is sometimes called "shot brood". An explanation for this character was given by MACKENSEN (1951). MACKENSEN inseminated several queens with semen of their brothers, and by brood viability studies obtained a 1:1 ratio between high (close to 100%) and low (close to 50%) viable brood. He suggested that a series of haploviability homozygous lethals exists so that all females are s_1s_2 , s_2s_3 , s_3s_4 , etc. . . . and the males s_1 , s_2 , s_3 , . . . s_n . Any female s_1s_2 mated with s_3 males would produce highly viable brood and any s_1s_2 female mated either to s_1 or s_2 males would produce brood of 50% mortality.

MACKENSEN suggested further that the inviable combinations (s_1s_1 , s_2s_2 , etc.) could be males and, if so, the sex determination mechanism in *Apis mellifera* L. would be the same as the one in *Habrobracon* except that the alleles are apparently completely lethal when homozygous. This last point of view, that homozygotes may be males, is a question of fact and acceptance of it must await proof that the homozygous dying larvae are male ones. Regardless of whether the series are sex alleles or whether they are only lethal alleles this discovery is outstanding. A queen showing a low degree of brood viability builds such small populations that the colony produces a small amount of honey; and the queen is likely to be superseded, which also lowers the yield.

The studies herewith presented have the scope of estimating the number of these haploviability homozygous lethals in a random mating population of *Apis mellifera* L. so that a better understanding of the population genetics of these bees can be attained.

DEVELOPMENT OF PERTINENT FORMULAE

In undertaking these studies two mathematical problems presented themselves. One was to establish the theoretical frequency of each s allele when the population is in equilibrium, and the other was to find a good estimator for the number of such alleles.

The equilibrium frequencies of the s alleles may be obtained directly from the theory already worked out for selection when the homozygote is at a disadvantage (see WRIGHT 1949, p. 372; LI 1954, p. 260). Suppose the population has n different s alleles, s_1 , s_2 , . . . s_n with frequencies x_1 , x_2 , . . . x_n ($\sum x_i = 1$). If we take the formula for equilibrium given in the above references and assume that each allele has

¹ At present under a Rockefeller Foundation fellowship in the Univ. de Sao Paulo.

² Sequence of authors does not imply seniority.

the same selective disadvantage when homozygous, we are led immediately to the formula, $x_i = 1/n$.

This proves that all alleles in a panmitic population, if all of them have the same adaptive value, reach an equilibrium when all alleles have the same frequency.

This is very easy to understand in a practical way: if one s allele becomes more frequent in one generation it would have a greater chance of uniting with another like itself thus originating a lethal homozygous combination which would decrease its number in the following generation.

If all s alleles have the same frequency, establishing the frequency of one s allele would give us the frequencies of all the others. This conclusion determined our method of work which was designed to detect the frequency of any two alleles taken at random from a random mating population.

The second problem was to find an estimator for the number of s alleles in the population when each queen is inseminated with semen from a single drone. It is shown in the appendix that a satisfactory estimator is given by

$$n^* = 2 \frac{N + 1}{H + 1}$$

where N is the number of colonies and H is the number of colonies where the s allele from the drone is identical with one of the alleles of the queen.

MATERIAL

It is necessary to be careful in choosing a random mating population. The formulae obtained are good only for populations that are random mated for several years, so that they can be said to be in equilibrium. One condition of random mating in honeybees is not to have interference with the raising of new queens.

The bee population studied here belonged to MR. RICARDO GOSSER (Piracicaba, Brazil), a beekeeper who has kept 65 colonies in one apiary for more than 10 years, never rearing queens artificially. The queens mate freely, and within a radius of 10 km (about 6 miles) there are more than 200 other hives. The 65 colonies studied belonged, with 3 exceptions, to the black Portuguese and German races. These 3 exceptions were hybrids in different degrees between black and Italian bees. Practically all 65 colonies showed very good brood with no signs of either 50% or 75% viability. Nine of the 65 colonies showed signs of having superseded the queen shortly before they were used. One of the colonies did not have larvae for grafting in our first collecting of samples because it had a new queen that had not started laying.

METHODS

Queens were reared from as many different colonies of the one apiary as possible, and were inseminated with semen of single drones reared from just one colony. This was accomplished as follows (for details see LAIDLAW 1953):

Four larvae from each colony were "grafted" into artificial queen cups. These grafted cups were put into a "swarm box" for 24 hours and afterwards into a "cell fin-

isher colony" for nine days. On the tenth day the queen cells were transferred to individual screen cages of about $2 \times 3 \times 5$ cm; all cages were put into a frame and placed in a queenless "nursery colony" where the queens emerged and were cared for by the bees until they reached four to ten days old. During this age period each queen was artificially inseminated once with semen of one drone. Before collecting the semen precaution was taken to be certain that the drone had a large amount of semen so that a good insemination could be assured. In the inseminations LAIDLAW equipment and technique (LAIDLAW 1949) were used. After the queens were inseminated the wings were clipped and the queens were put back into the "nursery colony" for one or two days, and were then introduced into queenless nuclei by means of special cages. Two to four days later each queen was released followed by a light

TABLE 1

First and second brood counts and viability percentages. Those with less than 50% are given in parentheses

Hive no.	Number of eggs	Number of worker pupae	Percent of viability	Hive no.	Number of eggs	Number of worker pupae	Percent of viability
D 0-1	507	467	92.11	R 20-2	461	461	100%
R 46	884	396	(44.80)	R 45-1	101	80	79.21
R 68-1	563	408	72.47	R 57-1	550	528	96.00
D 0-2	—	—	100%	R 51-1	438	399	91.10
R 21	1620	1402	86.54	R 24-1	1332	1179	88.51
R 64	2047	944	(46.12)	R 62-1	170	145	85.29
R 40	418	371	88.76	R 9-2	508	475	93.50
R 42	693	572	82.54	R 33-1	186	166	89.25
R 41	1924	945	(49.12)	R 8-2	500	443	88.60
R 59-1	224	167	74.55	R 58-1	1050	1041	99.14
R 7	706	636	90.08	R 56-1	177	150	84.75
R 49-1	178	135	75.84	R 11-2	1671	1431	85.64
R 52	707	543	76.80	R 59-2	600	505	84.17
R 4	—	—	100%	D 2-1	582	563	96.74
R 54-1	1026	973	94.83	R 13-2	1307	602	(46.06)
R 16-1	769	667	86.74	R 28	209	65	(31.10)
R 43	307	147	(47.88)	R 53	371	309	83.29
R 14-1	586	533	90.96	R 35	810	332	(40.99)
R 32	336	316	94.05	R 68	351	279	79.49
R 65	678	643	94.84	R 60	634	520	82.02
R 3	247	180	72.87	R 25-2	643	214	(33.28)
R 55-1	792	691	87.24	R 30	688	632	91.86
R 11-1	489	495	100%	R 15	229	149	65.07
R 9	491	454	92.46	R 34-1	85	76	89.41
R 20-1	559	549	98.21	R 26	278	179	64.39
R 38-1	567	511	90.12	D 0-3	258	187	72.48
R 10	368	306	83.15	R 61	365	154	(42.19)
R 24-2	397	238	59.95	R 6	1056	853	80.78
R 8-1	376	355	94.41	R 29	622	500	80.39
R 48-1	429	412	96.04	R 36	433	343	79.21
R 48-3	296	219	73.99				

spraying of queen and bees with 30% to 50% sugar syrup. All drones used were reared by one colony, colony 39, so they were either s_1 or s_2 as far as their s alleles were concerned. To obtain drones a frame containing sealed drone brood of colony 39 was put into a droneless Italian colony for emergence and the hive entrance was covered with excluder. Even if some indigenous drones occurred in this hive they could be detected by the abdominal color. One to four days after emergence the drones were put into small screen cages, 50 drones per cage, and were left to mature in the same nursery colony as the queens until used. The queens while being inseminated were anesthetised with carbon dioxide. Forty-eight hours after insemination they received an additional CO_2 treatment of 10 minutes duration so they would lay earlier (MACKENSEN 1947). They started laying from 4 to 15 days after insemination.

A few days after a queen started laying an empty comb was given to her. One or two days later the comb was taken out, marked, and the eggs were counted with the aid of a hand counter and the counts were recorded. The comb of eggs was then given to a strong two-story hive in which the queen had been confined by a queen excluder in the lower story. A second count was made when the brood was sealed. In almost all cases of 50% viability the counts were repeated once or twice.

Sixty-one queens were reared from the 65 colonies, ten of which were duplications. Fifty-nine of these were tested as described above. Included in the results are two queens (DO-2 and R4 in table 1) that showed 100% viable brood by inspection but actual counts were not made. Figure 1 illustrates the difference in appearance of 50%

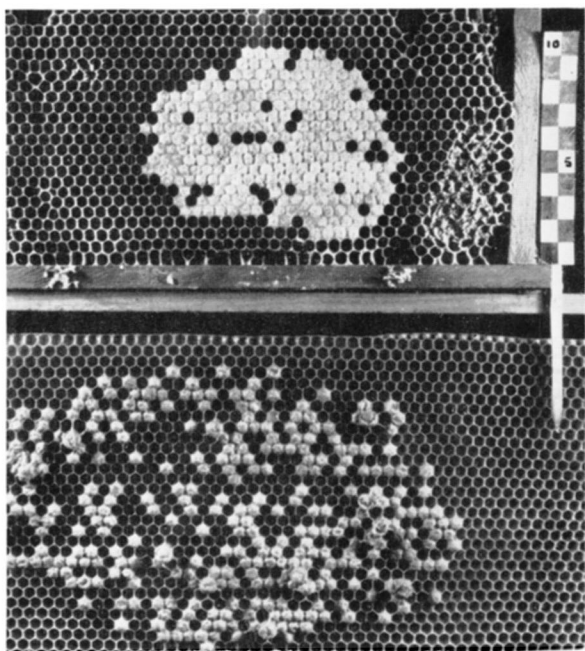


FIGURE 1.—Appearance of 100% viable brood, upper comb, and 50% viable brood, lower comb.

and 100% viability, the upper frame representing 100% viable brood and the bottom one 50% viable brood.

RESULTS AND CONCLUSIONS

The results are summarized in table 1. All percentages of viability smaller than 50% are in parentheses. The figure for R 11-1, where the egg count was smaller than the sealed count, is due to counting mistakes.

A total of 61 queens was analyzed. Nine of these had brood viability ranging from 31.10% to 49.12% and 52 from 59.95% to 100%, table 1.

The bimodal character of the distribution is shown by the graph of figure 2. The only value in the 50% to 60% class is almost 60% and there is little chance that highly viable brood was included with that of low viability.

Applying the formulae developed in the appendix:

$$n^* = \frac{2(N + 1)}{H + 1},$$

$$V(n^*) = \frac{n^2(n - 2)}{2(N + 2)},$$

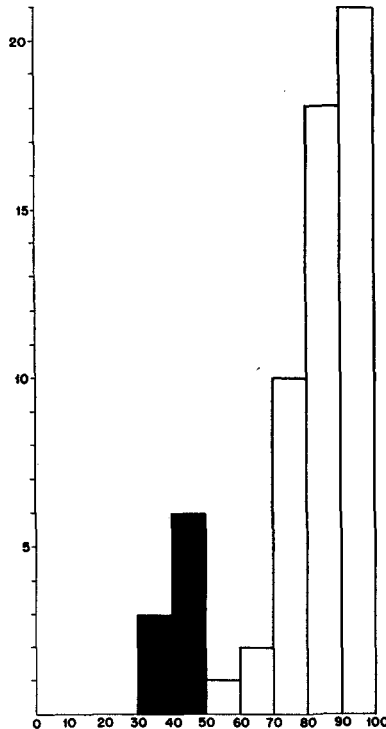


FIGURE 2.—Graph showing the bimodal distribution of different degrees of sterility. The only value in the class 50-60 is 59.9, which is almost in class 60 to 70.

where n = number of different s alleles, N = total number of hives, H = number of hives showing 50% of viability, we have

$$n^* = \frac{2(61 + 1)}{9 + 1} = 12.4,$$

$$s(n^*) = \sqrt{\frac{153.76 \times 10.4}{126}} = \sqrt{12.69} = 3.56.$$

This indicates the number of s alleles in the panmitic population studied is 12.4 with a standard deviation of 3.56.

ROBERTS and MACKENSEN (1951) calculated that with single drone matings and eight s alleles 25% of the colonies should show 50% viable brood; with $n = 16$ this figure dropped to 12.5%. In the case of double mating and 12 s alleles 2.8% of the colonies should show 50% viability and 27.8% should show 75% viability.

These percentages did not occur in the natural population studied but the reason was evident in the experiments. Practically all nuclei showing 50% viability had the queens superseded so quickly that in some cases second counts could not be taken. This contributes to maintain the number of s alleles as large as possible.

Any new allele that appears in the population by mutation has every chance to become established.

SUMMARY

A series of haploviaible homozygous lethals was discovered by MACKENSEN (1951) in *A. mellifera* L. and in this paper an estimate of the number of such alleles in a panmitic population is given.

It was demonstrated that if there are n alleles in the population equilibrium is attained when all alleles have the same frequency $\left(\frac{1}{n}\right)$.

An estimator for the number of different s alleles was derived. This was found to be $n^* = \frac{2(N + 1)}{H + 1}$, and the variance to be $V(n^*) = \frac{n^2(n - 2)}{2(N + 2)}$ where N = total number of hives and H = number of hives showing 50% viability.

The material used was 65 hives of a random mating population from which 61 samples (virgin queens) were taken. These queens were inseminated with just one drone each from one colony of the same population. Brood viability tests were made revealing that nine queens had low viable brood and 52 high viable brood. Applying the formulae, 12.4 s alleles were found, with a standard deviation of 3.56. Conditions are offered in a bee population to quickly establish any new s allele arising by mutation.

RESUMO

Este trabalho trata da estimação do número de aleles múltiplos, s , em uma população panmítica de *Apis mellifera* L., do tipo de homozigotos letais descoberto por MACKENSEN (1951).

Demonstra-se no trabalho que quando há n aleles na população, o equilíbrio é

atingido quando todos têm a mesma frequência $1/n$. Uma estimativa n^* para o número n de aleles s é dada pela fórmula:

$$n^* = \frac{2(N + 1)}{H + 1}$$

onde N é o número de colmeias do experimento e H é o número de colmeias em que apenas 50% da prole sobrevive. A variância de n^* é dada pela fórmula

$$V(n^*) = \frac{n^2(n - 2)}{2(N + 2)}$$

Um experimento foi feito com 65 colmeias de uma população com fecundação ao acaso há muitos anos. Dessa população, 61 rainhas virgens foram tomadas e inseminadas artificialmente, cada uma com semen de um único zangão. Todos os zangões provieram de uma mesma colmeia. Das 61 rainhas, 52 deram prole com alto índice de sobrevivência (perto de 100%) e 9 com baixo índice de sobrevivência (menos de 50%). A aplicação das fórmulas nos deu $n^* = 12,4$ aleles com desvio padrão = 3,56.

Sugere-se que qualquer novo alele s aparecido por mutação tem tódta facilidade de se estabelecer na população.

ACKNOWLEDGMENTS

Grateful acknowledgment is made to DR. ERICO AMARAL for his personal help and for the means put at our disposal to accomplish this problem and to MR. RICARDO GOSSER for allowing us to use his hives. Also, without the aid from the Rockefeller Foundation and from the Conselho Nacional de Pesquisas (Brazilian Research Council) this work could not have been done.

LITERATURE CITED

- CRAMER, HAROLD, 1946 *Mathematical Methods of Statistics*. Princeton Univ. Press. Princeton.
- KENDALL, M. G., 1948 *The Advanced Theory of Statistics*. Vol. II., 2nd edition, Charles Griffin & Company, London.
- LIDLAW, H. H., 1949 Development of precision instruments for artificial insemination of queen bees. *J. Econ. Entomol.* **42**: 254-261.
- 1953 Beekeeping management for the bee breeder. *Am. Bee J.* **94**: 92-95.
- MACKENSEN, OTTO, 1947 Effect of carbon dioxide on initial oviposition of artificially inseminated and virgin queen bees. *J. Econ. Entomol.* **40**: 344-349.
- 1951 Viability and sex-determination in the honey bee (*Apis mellifera* L.). *Genetics* **36**: 500-509.
- MOOD, A. M., 1950 *Introduction to the Theory of Statistics*. 1st edition. McGraw-Hill Book Company, New York.
- ROBERTS, W. C., and O. MACKENSEN, 1951 Breeding improved honey bees. *Am. Bee J.* **91**: 382-384.

APPENDIX

Statistical aspects of the estimation of the number of alleles

The second problem was to find an estimator for the number of s alleles in the population when each queen is inseminated with semen from just one drone.

While developing the theory, it should be kept in mind that the experiment was carried out with 61 colonies of bees and led to an estimate of about 12 alleles. This knowledge is not essential, of

course, for the derivation of the estimator, but it is important when discussing the degree of approximation of the formulae obtained.

Let us assume that there are n equally frequent s alleles in the population. Then the probability p that the s gene from the drone is identical to one of those from the queen (event A) is $p = \frac{2}{n}$, and the probability of the complementary event (event B) is $q = \frac{n-2}{n}$.

Hence the frequency function $f(x)$ will be $f(x) = p^x q^{1-x}$, $x = 0, 1$, where $x = 1$ denotes event A and $x = 0$ denotes event B. If we now consider N colonies the likelihood will be $L = p^H q^{N-H}$, where $H = \sum x$ is the number of colonies where the s allele from the drone is identical with one of the alleles of the queen.

Let us try to apply the method of maximum likelihood to obtain an estimator for n . It is easier to use $\log L$ instead of L itself, so we set up:

$$\log L = H \log p + (N - H) \log q,$$

$$\frac{d}{dn} \log L = \frac{2}{n} \left(-\frac{H}{2} + \frac{N - H}{n - 2} \right).$$

By equating this derivative to zero we obtain the estimator

$$\hat{n} = \frac{2N}{H}.$$

Let us now study the properties of this estimator. First of all, we are supposed to calculate its expectation $E(\hat{n})$ to see if it is biased. Now we have a difficulty, for the estimator cannot be used for $H = 0$. But we may take this case, the probability of which is q^N , and consider

$$E(\hat{n}) = \frac{2N}{1 - q^N} \sum_{H=1}^N \frac{1}{H} \binom{N}{H} p^H q^{N-H},$$

where $\binom{N}{H}$ denotes the binomial coefficient.

To carry out the summation we can use the identity

$$\frac{1}{H} = \frac{1}{H + 1} + \frac{1}{(H + 1)(H + 2)} + \frac{2}{H(H + 1)(H + 2)},$$

from which we obtain the approximate formula

$$(I) \quad \frac{1}{H} \approx \frac{1}{H + 1} + \frac{1}{(H + 1)(H + 2)}.$$

$$\frac{1}{1 - q^N} \sum_{H=1}^N \frac{1}{H + 1} \binom{N}{H} p^H q^{N-H} = \frac{1}{(1 - q^N)(N + 1)p} [1 - q^{N+1} - (N + 1)pq^N]$$

$$\approx \frac{1}{(N + 1)p} - \frac{q^N}{1 - q^N}.$$

Let us put

$$y = \frac{q^N}{1 - q^N}$$

and assume that $n < \frac{N}{4}$. This condition can be always fulfilled through the increase of the number N of colonies. Then we have

$$q < 1 - \frac{8}{N} \text{ and } y < \frac{1}{\left(1 - \frac{8}{N}\right)^{-N} - 1} < \frac{1}{e^8 - 1} \approx \frac{1}{2980}.$$

This term is negligible, since for $\frac{N}{6} < n < \frac{N}{4}$ we have

$$\frac{1}{(N+1)p} = \frac{n}{2N+2} > \frac{N}{12N+12} \approx \frac{1}{12}.$$

Hence

$$\frac{1}{1-q^N} \sum_{H=1}^N \frac{1}{H+1} \binom{N}{H} p^H q^{N-H} \approx \frac{1}{(N+1)p}.$$

Similarly we obtain

$$\frac{1}{1-q^N} \sum_{H=1}^N \frac{1}{(H+1)(H+2)} \binom{N}{H} p^H q^{N-H} \approx \frac{1}{(N+1)(N+2)p^2},$$

therefore

$$\begin{aligned} E(\hat{n}) &\approx 2N \left[\frac{1}{(N+1)p} + \frac{1}{(N+1)(N+2)p^2} \right] \\ &\approx n \left[1 + \frac{n-2}{2(N+1)} \right] \end{aligned}$$

It is evident, therefore, that there is a bias

$$b = \frac{n(n-2)}{2(N+1)}.$$

This bias approaches zero when $N \rightarrow \infty$, so that the estimator is asymptotically unbiased. However, for $n = 12$, $N = 60$ the bias is $b \approx 1.0$.

We can try to improve the formula obtained to make it unbiased or, at least, with a smaller bias. We suggest the formula

$$n^* = 2 \frac{N+1}{H+1},$$

which can be used for $H = 0$ and for which we have

$$E(n^*) = n - nq^{N+1}.$$

For $n = 12$, $N = 60$ the bias is now $b = \frac{1}{5634}$, which is negligible.

We shall show now that the estimator n^* is sufficient (KENDALL, vol. II, p. 8; CRAMER, p. 488; Mood, p. 151).

Let x_1, x_2, \dots, x_N be the observed values of x . We have:

$$L(x_1, x_2, \dots, x_N | n) = p^H q^{N-H},$$

where $H = \sum x$ and L denotes the likelihood. Also

$$L_1(n^* | n) = \binom{N}{H} p^H q^{N-H}$$

$$L_2(x_1, x_2, \dots, x_N | n^*) = \frac{L}{L_1} = \frac{1}{\binom{N}{H}}.$$

Since $H = \frac{n^*}{2(N+1)} - 1$, we see that L_2 does not depend on the values of x , but just on the estimate n^* . Hence it is sufficient. The variance of n^* will be given by the formula

$$V(n^*) = E(n^{*2}) - n^2,$$

for the bias is so small that we may take $E(n^*) = n$. But

$$E(n^{*2}) = 4(N+1)^2 \sum_{H=1}^N \frac{1}{(H+1)^2} \binom{N}{H} p^H q^{N-H}.$$

We may use again formula (I) and take

$$\frac{1}{(H+1)^2} \approx \frac{1}{(H+1)(H+2)} + \frac{1}{(H+1)(H+2)(H+3)}.$$

We obtain:

$$\begin{aligned} E(n^{*2}) &= \frac{4(N+1)}{(N+2)p^2} + \frac{4(N+1)}{(N+2)(N+3)p^3} \\ &\approx n^2 \left[1 + \frac{n-2}{2(N+2)} \right], \end{aligned}$$

from which we conclude that

$$V(n^*) \approx \frac{n^2(n-2)}{2(N+2)}.$$

Note that for $n = 2$ the variance is zero, which should be expected, for then event B is impossible, so that we must have $H = N, n^* = 2$.

Let us now study the efficiency of the estimator n^* . It is known (CRAMER, pp. 486-487; KENDALL, pp. 17-20) that for any estimator \hat{n} of n we have

$$V(\hat{n}) \geq \frac{\left(1 + \frac{db}{dn}\right)^2}{-N \sum_x \left(\frac{d^2}{dn^2} \log f\right) f}$$

and that equality holds only for efficient estimators. For n^* we have

$$\begin{aligned} b &= -nq^{N+1}, \\ \frac{db}{dn} &= -q^N \left[q + \frac{2}{n}(N+1) \right]. \end{aligned}$$

For $n = 12, N = 60$, the absolute value of this expression is less than 0.0002. Also

$$\begin{aligned} \frac{d^2}{dn^2} \log f(x) &= \frac{1}{n^2} - \frac{1-x}{(n-2)^2}, \\ -N \sum_{z=0}^1 \left[\frac{1}{n^2} - \frac{1-x}{(n-2)^2} \right] p^z q^{1-z} &= \frac{2N}{n^2(n-2)}, \end{aligned}$$

so that we have very closely

$$V(\hat{n}) \geq \frac{n^2(n-2)}{2N}.$$

This last expression is practically identical to the one obtained for $V(n^*)$, so that n^* is an efficient estimator.