INHERITED VARIATIONS IN THE SERUM PROTEINS OF CATTLE

O. SMITHIES AND C. G. HICKMAN

Connaught Medical Research Laboratories, Toronto, Ontario and Central Experimental Farm, Ottawa, Ontario

Received July 15, 1957

THE recent demonstration of permanent differences in the serum proteins of normal humans (SMITHIES 1955a, 1955b) which are simply inherited (SMITHIES and WALKER 1955, 1956) suggested the value of a similar investigation of the serum proteins of cattle. We have consequently examined serum samples from over 140 dairy cattle from the Ayrshire and Holstein herds maintained for experimental use at the Central Experimental Farm, Ottawa. The following is a full account of the results of the investigation which was briefly reported by HICKMAN and SMITHIES (1957).

Characteristics of the serum protein types

Five distinct types of serum protein pattern (I to V) have been observed* after electrophoresis in starch gels using the method previously employed for the study of human sera (SMITHIES 1955b). These types are illustrated in Figure 1 which shows a photograph and explanatory diagram of the results obtained when representative examples of the five serum types are compared by electrophoresis on a single starch gel. As can be seen from the figure, the proteins labelled A, B, C, D, and E vary in occurrence and amount from type to type. When the method of two-dimensional electrophoresis (SMITHIES and POULIK 1956) is used to investigate the nature of these proteins they prove to be β-globulins.

The variations in the β-globulins A, B, C, D and E can be generalized as follows:

Protein A is present in the serum types II, IV and V but is absent in Types I and III.

Proteins B and C are present in all five types but in differing proportions. In types II, IV and V (which contain protein A) protein B is present in approximately the same or somewhat greater amounts than protein C, but in types I and III (which lack protein A) protein B is present in considerably less amounts than protein C.

Protein D is clearly demonstrable in sera of types I, II, III and IV. In some animals of serum type V protein D appears to be completely absent, although

* The same five serum types have been observed in cattle by G. C. Ashton (Nature **180**, 917, 1957) in an independent study. Ashton refers to a genetic hypothesis for their control involving five pairs of linked genes.

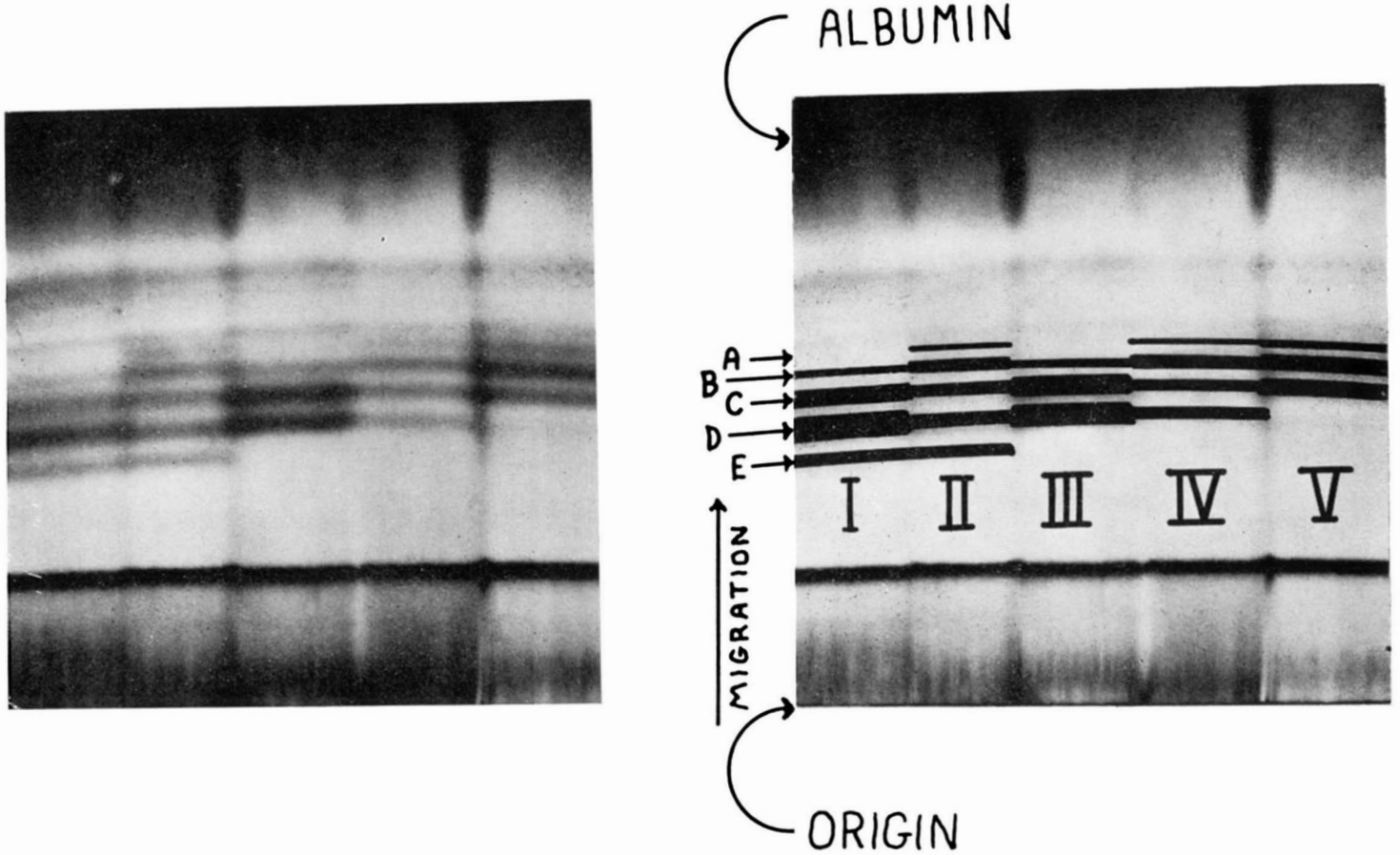

FIGURE 1.—Photograph and explanatory diagram of the results of an experiment in which representative examples of each of the five serum types were compared by electrophoresis in a single starch gel. The part of the gel from the origin to albumin is shown.

more frequently in sera from animals of type V traces of protein D very close to the limits of detectability can be seen. The distinction between types IV and V (which differ chiefly with respect to protein D) has nonetheless presented no serious difficulty in our experience to date, and we regard the two types as being discrete. The presence or absence of protein A is accompanied by characteristic changes in the relative amounts of protein D as well as of proteins B and C. In sera in which protein A is absent (types I and III) protein D occurs in relatively greater amounts than in sera in which protein A is present (types II, IV and V).

Protein E is present in sera of types I and II but is absent in the others.

The serum type of a given animal proves to be a fixed character of the animal within the time of our investigations (over 18 months). We have always found that on retyping an animal the same result is obtained. A pair of monozygotic twins had the same type, IV.

Distribution of serum protein types

Serum samples from all animals in the two available herds which could be grouped into pedigrees (sire, dam and offspring) have been studied. Consequently we assume that the animals tested are random samples of the two populations. More female offspring are kept in the herds and were tested, than male offspring, but we have no evidence for sex-linked differences in the serum protein types.

The distribution of the serum protein types in the two herds (Ayrshire and Holstein) is shown in Table 1, together with the results of a χ^2 test to determine

TABLE 1

The distributions of serum types in the two herds

Herd	Serum types I	II	III	IV	V	Totals
Ayrshire	10	11	10	8	3	42
Holstein	3	10	16	41	32	102
$\chi^2=32.0$ $P<0.00001$						

whether the two herds should be regarded as two samples from a homogeneous population. The P value of $<.00001$ indicates that the distributions of the types in the two herds differ significantly, and that any gene frequency calculations involving the two herds must be made separately. On the other hand, the distribution of the serum types of the parents compared with that of the offspring within either herd shows no significant difference. These comparisons suggest that the β-globulin differences observed are under genetic control.

A hypothesis for the genetic control of the β-globulins

The simplest genetic mechanism consistent with the differences in the β-globulins and the pedigrees which we have observed, requires the existence of a

single locus at which any of three alleles may be present. We assign the symbol β to the locus and designate the three genes β^A, β^E and β^O. The assumed actions of these genes are as follows:

The gene β^A leads to the presence of protein A, and to the occurrence of the proteins B and C in the proportions observed in sera of types II, IV and V. When β^A is absent protein A is absent, and the proteins B and C occur in different proportions, as observed in types I and III.

When gene β^E is present protein E is present, as in types I and II. If gene β^E is absent, protein E is absent.

The gene β^O is neutral in its effect as far as proteins A, B, C and E are concerned, in that its presence or absence does not affect the occurrence or relative proportions of these proteins.

Either of the genes β^E and β^O leads to the presence of protein D in the serum, but protein D is absent in the homozygote β^A/β^A. Thus we assume that the gene β^A is unable to lead to the presence of protein D in the serum although either of the genes β^E and β^O can do so.

The genotypes are then as follows (see also Figure 2 in which the variations in the β-globulins are shown schematically for the five serum types together with the postulated genotypes):—

I: β^E/β^E or β^E/β^O
II: β^A/β^E
III: β^O/β^O
IV: β^A/β^O
V: β^A/β^A

Two points may be emphasized in regard to this hypothesis:

(a) The genotypes of animals with sera of the types II to V are fixed. A type I animal may be homozygous (β^E/β^E) or heterozygous (β^E/β^O).

(b) The presence of the gene β^A, which it is suggested cannot bring about the presence of protein D in the serum, is reflected in the amount of protein D in sera of the types II and IV. Sera of these two types show less amounts of protein D than do sera of types I and III. This suggests that a "double dose" of the genes able to cause the presence of protein D (i.e. of the genes β^E and/or β^O) leads to the presence of more of this protein than does either gene singly (i.e. when accompanied by the gene β^A). These observations can be interpreted equally well if the gene β^A is regarded as a recessive inhibitor gene as far as its relationship to protein D synthesis is concerned. In the present genetic treatment these two ways of interpreting the action of gene β^A are equivalent.

Tests of the three allele hypothesis

An overall test of the reasonableness of the proposed genetic control of the serum β-globulins can be made by calculating the gene frequencies for the parents of the two herds (weighting each animal according to the number of its offspring), and from these gene frequencies calculating the expected distributions

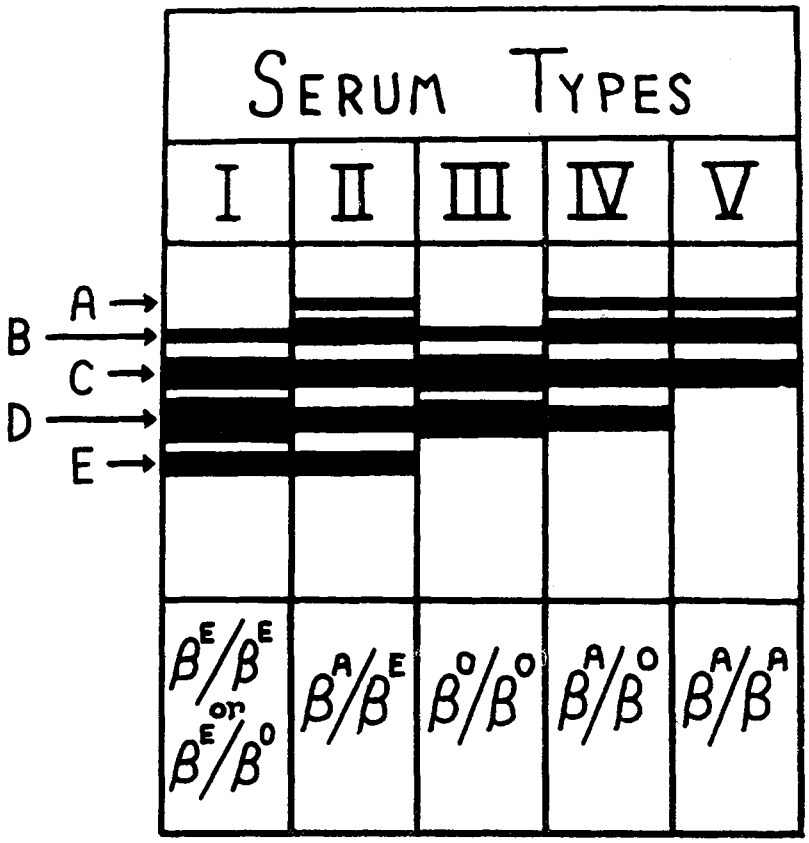

FIGURE 2.—A schematic representation of the variations in the β-globulins for the five serum types. The genotypes postulated are shown under the corresponding serum types.

of the offspring serum types on the assumption that there exists no selection in favour of any types. A χ^2 test can then be made between the observed and expected distributions. Table 2 shows the results of this test of the hypothesis for the two herds. The gene frequencies are calculated by the maximum likelihood

TABLE 2

Observed and expected distributions of offspring serum types calculated from parent gene frequencies

		Serum types I	II	III	IV	V
	Observed	5	7	4	2	3
Ayrshire	Expected	6.94	4.00	3.08	4.97	2.01
	$\chi^2=3.52$ P>0.3					
	Observed	1	5	8	18	18
Holstein	Expected	2.18	3.23	6.64	21.13	16.82
	$\chi^2=1.54$ P>0.8					

method outlined in the final section of this paper. The results indicate that the hypothesis is consistent with our observations.

A more specific test of the hypothesis can be made by comparing the observed and expected distributions of the serum types in the offspring of the 15 possible matings between five serum types.

Table 3 shows the expected proportions of offspring from the 15 possible matings. Except in the case of matings involving a parent of serum type I the expected proportions are independent of the gene frequencies of the particular herd concerned. The genotype of type I animals cannot be established with certainty

TABLE 3

Expected distributions of offspring serum types

Matings	Offspring serum types				
	I	II	III	IV	V
I × I	$1-\phi^2/4$		$\phi^2/4$		
I × II	1/2	$1/2-\phi/4$		$\phi/4$	
I × III	$1-\phi/2$		$\phi/2$		
I × IV	$1/2-\phi/4$	$1/2-\phi/4$	$\phi/4$	$\phi/4$	
I × V		$1-\phi/2$		$\phi/2$	
II × II	1/4	1/2			1/4
II × III	1/2			1/2	
II × IV	1/4	1/4		1/4	1/4
II × V		1/2			1/2
III × III			1		
III × IV			1/2	1/2	
III × V				1	
IV × IV			1/4	1/2	1/4
IV × V				1/2	1/2
V × V					1

$$\phi = \frac{\text{Frequency } \beta^O}{\text{Frequency } \beta^O + \text{Frequency } \beta^E}$$

in all animals since phenotypically β^E/β^E and β^E/β^O are indistinguishable. In matings involving such animals we use the overall herd gene frequencies to estimate the expected offspring serum types in terms of ϕ

$$\left(\text{where } \phi = \frac{\text{Frequency of } \beta^O}{\text{Frequency of } \beta^O + \text{Frequency of } \beta^E} \right).$$

In some cases the genotypes of type I animals can be determined from their offspring serum types e.g. a type I animal having an offspring of type III or IV must be heterozygous (β^E/β^O). In Table 3 this means of determining the genotype of type I parents is not included since it cannot always be applied.

Table 4 shows a summary of the combined pedigrees of both herds with the observed and expected distributions of serum types of the offspring for all matings studied. In the matings involving a type I animal the expected offspring serum

TABLE 4

Observed/Expected distributions of offspring serum types

Matings	Offspring serum types				
	I	II	III	IV	V
I × I	–/–		–/–		
I × II	2/2.50	3/1.75		0/0.75	
I × III	–/–		–/–		
I × IV	2/1.84	3/1.84	1/1.16	0/1.16	
I × V		1/0.57		0/0.43	
II × II	–/–	–/–			–/–
II × III	2/2.50			3/2.50	
II × IV	0/2.00	3/2.00		2/2.00	3/2.00
II × V		2/2.00			2/2.00
III × III			2/2.00		
III × IV			9/7.00	5/7.00	
III × V				2/2.00	
IV × IV			0/0.50	0/1.00	2/0.50
IV × V				8/7.00	6/7.00
V × V				*2/0*	8/10.00

types are estimated from ϕ for the respective herd unless the genotype of the specific parent can be deduced from its offspring, in which case the deduced genotype is used to calculate the expected offspring serum types.

DISCUSSION

The data summarized in Table 4 indicates very close agreement between the expectations from the three allele hypothesis and the observed pedigree data except for two matings. The offspring of two matings from the Holstein herd are not permitted by the hypothesis. Both contradictions are of the same type in which a type IV offspring is observed from a V × V mating. We therefore considered the possibility that these matings are incorrectly recorded in that one (or both) of the postulated parents is not in fact the parent of the offspring concerned. This explanation of the inconsistency could be tested in one of the matings by determining the red cell antigens of the animals concerned. The result of this test shows that both postulated parents lack the red cell antigen D′ although the offspring has this antigen. Consequently one of these matings can reasonably be excluded from our analysis as being an incorrectly recorded mating. Unfortunately the (male) offspring of serum type IV from the other contradictory mating was sold for beef before it was realized that his blood was particularly valuable. The question therefore arises whether it is reasonable to regard this mating also as being incorrect without being able to prove this with antigen tests. This question can be answered to some extent since the mating concerned

was made by artificial insemination using frozen semen. The probability that the offspring was not from the postulated dam is negligibly small; the most reasonable assumption is that an error was made in selecting the frozen semen. The probability that an error in choice of frozen semen would be detected in the observed manner is given by the sum (over all semen) of the chance of a type IV offspring (for each semen) multiplied by the chance of selecting that semen. At the time of the mating, frozen semen samples from three bulls (a type III, a type IV and a type V) were available. Each must be regarded as having an equal chance of being selected if an error in choice of semen is to be considered. Under these conditions the probability that the error would be detected is 0.5. The coincidence that the second error was detected in the same mating type (V × V) as the other is not particularly unlikely since in the Holstein herd approximately a quarter of all matings were of this type. Consequently we propose to make the assumption that this mating is also incorrectly recorded and can be excluded from our analysis.

A χ^2 test can be made of the observed distribution of offspring serum types in the 71 acceptable matings compared with those expected on the three allele hypothesis—although the small numbers of the expected values in many cases makes the estimate of P so obtained a comparatively crude one. The result is $\chi^2 = 15.02$ with 15 degrees of freedom, corresponding to a value of $P = 0.45$. Since a χ^2 test in this case does not take account of those matings where we expect all offspring to be of one type, this estimate of P is smaller than the true value. Our data thus strongly support the correctness of the three allele hypothesis.

There is a possibility that the animals of serum type I are all heterozygotes (β^E/β^O) and that we have never seen the homozygote (β^E/β^E), which might be a sixth type. Thirteen animals of type I have been observed. Of these five are known from the pedigree data to be heterozygotes. The probability that the remaining eight type I animals (six Ayrshires, two Holsteins) are also heterozygotes is given by $(\phi \text{ Ayr.})^6 \times (\phi \text{ Hol.})^2 = 0.034$. This indicates that it is unlikely that we have not seen an animal homozygous in the gene β^E. Any type I offspring from II × II matings would be homozygotes, but unfortunately this mating has not been observed in the two herds.

A less restrictive hypothesis involving two pairs of genes was considered in the initial stages of this work. It is a possible alternative genetic mechanism not obviously contradicted by our pedigree data—a situation similar to the early interpretation of the genetics of the ABO system in human red cell groups in terms of two gene pairs rather than a three allele system. However we can calculate from our data that such a mechanism is unlikely in the present case. The most striking difference between the two genetic mechanisms is that the two gene pair hypothesis permits certain offspring excluded by the three allele mechanism. No offspring of these particular types has been observed in the matings concerned. The probability that this is a chance observation due to the rather small number of matings involved (18) can be calculated. The calculated value of P is 0.0085,

which indicates that the three allele hypothesis is to be preferred since the probability that our results are explicable by the two gene pair hypothesis is small.

Utility of serum types for parentage tests in cattle breeding

The probability of being able to detect an error in a pedigree is given by the probability of occurrence in the offspring of a serum type other than those possible from the serum types of the postulated parents. For example the probability of being able to say that a given animal is not the offspring of a I × III mating is given by the probability that the animal is not of serum types I and III, which are the permitted offspring serum types for such a mating. If the animal in question can be treated as a random sample from the population, there would, in this case, be a probability of detecting the faulty mating of $(2\,ae + 2\,ao + a^2,)$ using the symbols defined in the final section of the paper.

The probability of detecting errors varies with the different matings, and the matings occur with differing frequencies. However, we can obtain a representative value for the probability of detecting pedigree errors, when the offspring is assumed to be randomly selected from the population, by summing the probabilities for all matings, weighting the probability for each mating by the frequency of its occurrence. The values for this representative probability for the two herds here studied are 0.43 (Ayrshire) and 0.38 (Holstein).

In practice the conditions of detection of error are usually more restricted than the situation considered above. One parent is frequently known without error, and the choice of the other parent is between two animals. Thus if two cows, one of serum type I and the other of serum type II, freshen concurrently and unattended after mating to the same bull of type I, the question would be which is the correct dam for each calf. The probability that the calves can be assigned to their actual dams is given by one minus the probability that both calves are of a type (or types) possible from both matings. In this example a type I offspring is possible from both matings so that the probability of correct assignment is given (see Table 3) by $1-(1-\phi^2/4) \times \frac{1}{2}$, that is by $\frac{1}{2} + \phi^2/8$. Since the value of ϕ cannot exceed unity, the probability of correct parent assignment in this example lies between 0.5 and 0.625. This procedure can be applied with the use of Table 3 to any similar situation.

We suggest that the determinations of the new serum types may prove of value in cases of questionable parentage in cattle as an economical screening test to precede the expensive and laborious determinations of red cell antigens.

Gene frequency calculations

The following is an outline of the method of estimating gene frequencies by a maximum likelihood treatment in the present three allele system with five phenotypes.

Let the system be represented as follows:

Phenotypes	I	II	III	IV	V
Genotypes	β^E/β^E or β^E/β^O	β^A/β^E	β^O/β^O	β^A/β^O	β^A/β^A
Number of animals	n_{I}	n_{II}	n_{III}	n_{IV}	n_{V}

Let a, e and o be the frequencies of the genes β^A, β^E and β^O respectively, and let the total number of animals be N.

The maximum likelihood estimate of a (a^*) is given by a simple gene count, since all three genotypes involving β^A can be recognized.
Hence

$$a^* = (2n_{\mathrm{V}} + n_{\mathrm{IV}} + n_{\mathrm{II}})/2N \tag{1}$$

and

$$\sigma_{a^*}^2 = a^*(1 - a^*)/2N \tag{2}$$

The derivations of these expressions are given in standard texts (e.g. NEEL and SCHULL 1954).

The likelihood (L) for the system is given by

$$L = \frac{N!}{n_{\mathrm{I}}!n_{\mathrm{II}}!n_{\mathrm{III}}!n_{\mathrm{IV}}!n_{\mathrm{V}}!}(e^2 + 2\,eo)^{n_{\mathrm{I}}}(2\,ae)^{n_{\mathrm{II}}}(o^2)^{n_{\mathrm{III}}}(2\,ao)^{n_{\mathrm{IV}}}(a^2)^{n_{\mathrm{V}}} \tag{3}$$

Since $a + e + o = 1$ then the estimation of a^* leaves the ratio of e to o as the single remaining variable. Let $e/o = \rho$. The maximum likelihood estimate (a^*) of a is independent of ρ. We can therefore express the likelihood in terms of a^* and ρ, and treat a^* as invariable with respect to ρ.

But $a^* + e + o = 1$ and $e/o = \rho$

$$\text{therefore } e = \frac{\rho\,(1 - a^*)}{(1 + \rho)} \text{ and } o = \frac{(1 - a^*)}{(1 + \rho)}.$$

Substituting these values in the likelihood expression (3) and taking logarithms we have

$$\log L = \log k + n_{\mathrm{I}}\log\frac{\rho^2(1 - a^*)^2 + 2\rho(1 - a^*)^2}{(1 + \rho)^2} + n_{\mathrm{II}}\log\frac{2a^*\rho(1 - a^*)}{(1 + \rho)} + 2n_{\mathrm{III}}\log\frac{(1 - a^*)}{(1 + \rho)} + n_{\mathrm{IV}}\log\frac{2\,a^*(1 - a^*)}{(1 + \rho)} + 2n_{\mathrm{V}}\log a^*$$

or

$$\log L = \log k' + (n_{\mathrm{I}} + n_{\mathrm{II}})\log\rho - (2n_{\mathrm{I}} + n_{\mathrm{II}} + 2n_{\mathrm{III}} + n_{\mathrm{IV}})\log(1 + \rho) + n_{\mathrm{I}}\log(2 + \rho)$$

$$\text{Therefore } \frac{d(\log L)}{d\rho} = \frac{n_{\mathrm{I}} + n_{\mathrm{II}}}{\rho} - \frac{2n_{\mathrm{I}} + n_{\mathrm{II}} + 2n_{\mathrm{III}} + n_{\mathrm{IV}}}{(1 + \rho)} + \frac{n_{\mathrm{I}}}{(2 + \rho)} \tag{4}$$

The maximum likelihood estimate of ρ (ρ^*) is obtained by equating $\frac{d(\log L)}{d\rho}$ to zero and solving for ρ^*. The result is:

$$\rho^* = -1 + n_{II}/D + \sqrt{1 + (4n_I + 2n_{II})/D + (n_{II}/D)^2} \quad (5)$$

where $D = 4\,n_{III} + 2n_{IV}$

From ρ^* we obtain: $e^* = \frac{\rho^*(1 - a^*)}{(1 + \rho^*)}$ (6)

and $o^* = \frac{(1 - a^*)}{(1 + \rho^*)}$ (7)

The variance of ρ^*, and hence of e^* and o^*, is obtained by differentiating equation (4) with respect to ρ and substituting the expected values for n_I, n_{II} etc. in terms of ρ^*. We obtain

$$\sigma_{\rho^*}^2 = \frac{\rho^*(2 + \rho^*)(1 + \rho^*)^2}{2N(1 - a^*)(2 + a^*\rho^*)}$$

Now it can be shown that $\sigma_{e^*}^2 = \sigma_{o^*}^2 = \sigma_{\rho^*}^2 \left(\frac{de^*}{d\rho^*}\right)^2$

Hence

$$\sigma_{e^*}^2 = \sigma_{o^*}^2 = \frac{(1 - a^*)\,\rho^*(2 + \rho^*)}{2N(1 + \rho^*)^2(2 + a^*\rho^*)} \quad (8)$$

The following gene frequency estimates were obtained using the expressions (1), (2), (5), (6), (7) and (8) for the Ayrshire and Holstein herds considered in the present study.

Ayrshire	$a^* = 0.298 \pm .050$
$N = 42$	$e^* = 0.279 \pm .049$
	$o^* = 0.423 \pm .049$
Holstein	$a^* = 0.564 \pm .035$
$N = 102$	$e^* = 0.065 \pm .017$
	$o^* = 0.371 \pm .017$

SUMMARY

(1) Variations, detected by electrophoresis in starch gels, in the serum proteins of cattle from two herds (Ayrshire and Holstein) enable animals to be grouped into five serum types (I to V). The distributions of the serum types in the two herds differ significantly.

(2) Five proteins, which are shown to be β-globulins, vary in occurrence and amounts in these serum types.

(3) The serum type of an animal appears to be a fixed characteristic of the animal.

(4) A possible genetic mechanism involving three alleles (β^A, β^E and β^O) is suggested for the control of the variable β-globulins.

(5) A comparison of the observed distributions of the serum types of the offspring of 71 matings in the two herds with the distributions expected from the three allele hypothesis shows that the available data strongly support the correctness of this genetic hypothesis.

(6) The possible utility of the serum types for parentage tests in cattle breeding is considered.

(7) A maximum likelihood method for the calculation of gene frequencies in a three allele system with five phenotypes is outlined.

ACKNOWLEDGEMENTS

We thank DR. NORMA FORD WALKER for helpful discussion of this work, MR. D. B. W. REID for guidance in the statistical calculations, and DRS. JAMES F. CROW and NEWTON E. MORTON for drawing our attention to the possible interpretation of the genetic data in terms of multiple alleles.

LITERATURE CITED

HICKMAN, C. G., and O. SMITHIES, 1957 Evidence for inherited differences in the serum proteins of cattle. (Abstr.) Proc. Genetics Soc. of Canada **2**: 39.

NEEL, J. V. N., and W. J. SCHULL, 1954 *Human Heredity*. Pp. 187–188. University of Chicago Press. Chicago.

SMITHIES, O., 1955a Grouped variations in the occurrence of new protein components in normal human serum. Nature **175**: 307.

1955b Zone electrophoresis in starch gels; group variations in the serum proteins of normal human adults. Biochem. J. **61**: 629–641.

SMITHIES, O., and M. D. POULIK, 1956 Two-dimensional electrophoresis of serum proteins. Nature **177**: 1033.

SMITHIES, O., and N. F. WALKER, 1955 Genetic control of some serum proteins in normal humans. Nature **176**: 1265–1266.

1956 Notation for serum-protein groups and the genes controlling their inheritance. Nature **178**: 694–695.