COMMENT AND REVIEWS

A UNIFORM NOTATION FOR THE HUMAN BLOOD GROUPS

E. B. FORD
Genetics Laboratory, University Museum, Oxford

Received 19.xi.54

1. INTRODUCTION

The notation of the human blood groups is at present chaotic. Not only is it out of accord with that adopted elsewhere in genetics, but its usage is inconsistent from one blood group system to another. This adds unnecessary difficulties to the understanding of a subject which is decidedly intricate. Indeed, in the current literature it is often impossible to determine whether a given symbol refers to a gene or to an antigen, and this confused terminology has certainly been a potent factor in preventing many geneticists from including the blood groups within their sphere of interest. Such a situation cannot continue indefinitely, and its revision is overdue. For the complexities of serology are rapidly increasing, while the general theory of polymorphism, developed in other fields of genetics is, as long ago suggested, beginning to have an important influence on the subject (see, for instance, Fisher, 1930 ; Sheppard, 1953).

The unsatisfactory state of the present notation becomes evident on examining the most important and most recent general text book on the subject, the second edition of *Blood Groups in Man* by R. R. Race and R. Sanger, which appeared in August 1954, from which the following facts are extracted. It is not thereby intended to level any adverse criticism at what is deservedly the standard work on the subject, but merely to point out the disadvantages of the conventions which it necessarily adopts, being those in general use.

The symbols employed for the two allelomorphs controlling three of the systems are shown in table 1. The situation in each of these instances

TABLE

Three blood group systems and their genes, in the present notation

System	Genes
Duffy	Fy^a, Fy^b
Kidd	Jk^a, Jk^b
Kell	K, k

is identical, in that both allelomorphs produce a dominant antigen. Clearly, therefore, if those of the first two are to be distinguished by suffixes, so they should be in Kell. The symbols (K, k) indicate in the accepted notation that k does not have a dominant effect, which is incorrect. Moreover, where both allelomorphs produce dominant antigens, large, not small, letters should be used for the suffix to be in accord with genetic usage. The confusion at present introduced will be evident on comparing

such groups as Duffy and Kidd with Lewis. For the gene responsible for the Lewis a-antigen is represented in a similar way (as Le^a) though, unlike the others, its effect is recessive.

The notation adopted for Kell (K, k) is also used for the P groups (P, p) though no antigen produced by p has yet been recognised. Thus we do not know whether it be dominant or recessive (or even, for certain, if it exists). Some distinction should evidently be made between a gene whose action is known and one in which it is merely inferred. The Lutheran groups are, at present, in the same position as the P groups, in the sense that the effect of one only of the two allelomorphs has been recognised by an antibody. The genetic behaviour of the other antigen, if it be produced, is therefore still unknown. Yet the Lutheran genes are represented by a type of notation distinct from that of the P system ; each being designated by a suffix, as Lu^a (though this is dominant in effect) and Lu^b. This is the same as that employed for Duffy and Kidd, though the situation represented is a different one.

Turning now to the MN series, an extraordinary situation is apparent. Not only is this treated in an entirely different way, but it is one which breaks the most fundamental rule of genetic nomenclature. For genes at the same locus are represented by different letters (M and N), there being nothing to show that these are allelomorphic. Further confusion is introduced because these letters, standing alone, are sometimes used for genes and sometimes for antigens.

The same criticism applies to the ABO series, the genes for which, though still presumed allelomorphs, are designated A, B and O. A and B are both dominant to O in their effect, though of this there is no indication in the symbols, in contradiction of genetic usage. Moreover, both genes and antigens are, most confusingly, represented by these letters. The present position is clearly shown by Race and Sanger (1954, table 2, p. 18).

The genes of the Rhesus system are represented quite differently from those of the MN or the ABO groups, and their notation, which reflects the brilliant deductions of Sir Ronald Fisher (1944), is in accord with standard genetic practice. It seems, however, desirable to indicate an additional feature not generally encountered outside the field of serology : the occurrence of contrasting allelomorphs each of which has dominant effects. Such genes should be distinguished from others in which the action of one member of the pair is as yet unknown, as with F, f. Yet in this instance, it is the smaller letter (f) which has been chosen to represent the gene producing a dominant antigen, though the lower case should be reserved for the recessive.

There are other defects in the present conventions. Thus genes, antigens and antibodies are indiscriminately, though not consistently, placed in italics ; but what has been said is sufficient to indicate the need for reform. It is essential to substitute a uniform notation for the one now in use. This must be in accord with that applied elsewhere in genetics, but it should be extendable to cover the exceptional features of serology and to include new discoveries. The following plan is proposed in an attempt to meet that need. It is not intended to interfere with the present " shorthand " employed for describing the Rhesus phenotypes and genotypes (R_1, R'', Ro, and so forth). This is for specialised use within the field of blood

grouping only, it saves much space when extensive lists are compiled, and it does not compete with normal genetic terminology. It is only necessary to say that, for the purpose of this shorthand, the conventions used in England (see Race and Sanger, 1954) should be adopted, for they have needed no revision. Those of Wiener (1949) should be avoided, for they have been confused by repeated modifications in respect of the phenotypes.

2. THE PROPOSED NEW NOTATION

(i) *Genes*

Genes should always be represented in italics. Antigens and antibodies should not be italicised.

The *locus* is indicated by one capital letter or by a pair of letters of which the first is a capital. Examples : *K* ; *Fy*.

The allelomorph is indicated by the presence or absence of a suffix (attached to the locus-symbol). A capital in the suffix is dominant to a small letter in the suffix. When allelomorphs, each with a capital suffix, are brought together they both excercise their effects. Thus a capital in the suffix represents a gene which exercises its effect whenever present (the effect of one dose does not necessarily equal that of two).

Example :—K^AK^A = Kell+ for A.
K^AK^B = Kell+ for and B.
K^BK^B = Kell+ for

A gene represented without a suffix is dominant in effect to one with a small letter in the suffix, but recessive to one with a capital in the suffix.

Examples :—*Le Le* and Le^aLe are both Lewis negative.
Lu^ALu is Lutheran positive.

A gene without a suffix is one whose antigen has not yet been recognised by an antibody (or perhaps such an antigen does not exist). Examples :— P^A, *P* are at present the genes of the P groups. Le^B, Le^a, *Le* are at present the Lewis genes. Another suffix-letter *B*, *C*, . . . is added as each antigen is discovered. Thus *P* would become P^B if its antigen were detected and found to be dominant, P^b if recessive. The suffix-letter first used is normally *A*, followed by *B*, *C*, . . ., in order of discovery, unless special circumstances make some other suffix-letter desirable (as in the MN group).

When the members of a multiple allelomorph series are distinguished by numbers, a large figure in the suffix represents a gene dominant in effect to one with a small figure. Example : G^{A1}, G^{A_2}, *G*. The first of these is dominant to the second in effect, and both are dominant to *G*, as the foregoing notation indicates. A natural extension of this arrangement covers other instances in which the effect of one gene is dominant to another while both are dominant to a third, as will be seen later in dealing with D^U of the Rhesus series.

(ii) *Antigens*

The initial letter (or letters) for the locus, and the suffix-letter for the allelomorph, become transformed into the discriminators for the antigens (and the antibodies). The way in which they are used prevents any confusion between gene, antigen and antibody, or between any of them and the groups or systems.

An antigen group is shown by the locus-symbol. The suffix-letters follow within brackets and indicate the type of antigen present. Those absent can also be indicated when desired. When the presence of an antigen is dominant, a capital is used, when its presence is recessive a small letter. Presence of the antigen is indicated by a plus, absence by a minus.

Examples :

The Kell antigens are : K(A+), K(A—) ; or, if known, K(A+B—), K(A+B+), K(A—B+). Both these antigens are dominants.

The Lewis a-antigens are : Le(a+) and Le(a—). This antigen is recessive and is therefore represented by a suffix-letter in the lower case. The situation for the dominant B antigen can be added if this also is under discussion ; *e.g.* Le(a—B—).

(iii) *Antibodies*

These are indicated by the locus-symbol with the prefix " anti " hyphened to it, followed by the gene-suffix in brackets. Example : Anti-K(A).

(iv) *Systems and groups*

Expressions such as the " Kell system ", the " ABO system ", or the " Rhesus system " are intended to cover the genetic switch-mechanism involved in controlling these systems (whether a pair of allelomorphs, a set of multiple allelomorphs, or several closely linked genes), and the antigens and antibodies associated with them. Consequently such expressions should not be italicised.

A blood " group " represents individuals who are the same in respect of the presence or absence of a given antigen or antigens. Thus the P system contains two groups : those who are positive, P(A+), and those who are negative, P(A—), for the single known antigen concerned. Also the ABO system comprises a number of blood groups : for instance, group B, carrying the antigen G(A—B+), and group O in which the antigen situation is defined by G(A—B—).

3. THE APPLICATION OF THE NEW NOTATION *

(i) *Duffy system*

GENES : Fy^A, Fy^B.

ANTIGENS : Fy(A+), Fy(A—) ; or, if known, Fy(A+B—), Fy(A+B+), Fy(A—B+).

ANTIBODIES : anti-Fy(A), anti-Fy(B).

(ii) *Kell system*

GENES : K^A, K^B.

ANTIGENS : K(A+), K(A—) ; or, if known, K(A+B—), K(A+B+), K(A—B+).

ANTIBODIES : anti-K(A), anti-K(B).

* The order in which these Systems are considered is one of advancing complexity in the light of present knowledge.

(iii) *Kidd system*

GENES : Jk^A, Jk^B.

ANTIGENS : Jk(A+), Jk(A—) ; or, if known, Jk(A+B—), Jk(A+B+), Jk(A—B+).

ANTIBODIES : anti-Jk(A), anti-Jk(B).

(iv) *Lutheran system*

GENES : Lu^A, *Lu*.

ANTIGENS : Lu(A+), Lu(A—).

ANTIBODY : anti-Lu(A).

(No antibody recognising the antigen produced by *Lu* has yet been found.)

(v) *P system*

GENES : P^A, *P*.

ANTIGENS : P(A+), P(A—).

ANTIBODIES : anti-P(A).

(If *P* produces an antigen, no antibody detecting it has yet been found.)

(vi) *Lewis system*

GENES : Le^B, *Le*, Le^a.

ANTIGENS : Le(a+), Le(a—) ; or, if known, Le(a+B—), Le(a—B+), Le(a—B—).

ANTIBODIES : anti-Le(a), anti-Le(B).

(No antibody recognising the gene *Le* by means of its antigen, if this exists, has yet been detected. A combination of both known antigens, Le(a+B+), has never been found in adults.)

Lewis antigens and their genotypes :—

ANTIGENS	GENOTYPES
Le(a+B—)	Le^aLe^a
Le(a—B+)	Le^BLe^B ; Le^BLe ; Le^BLe^a
Le(a—B—)	*LeLe* ; $LeLe^a$

(Le^BLe^a is Le(a—B+) because two doses of the Le^a gene are needed to produce their effect. In this respect, Le^B acts as a dominant over Le^a, just as *Le* does, and as the symbols indicate. It is assumed that Le(B+) is a simple dominant in order to show how the notation would work on that basis, though this group may, in reality, be more complex.)

(vii) *MNL system*

GROUPS : The MN groups.

GENES : Ag^M, Ag^N.

ANTIGENS : Ag(M+N—), Ag(M+N+), Ag(M—N+).

ANTIBODIES : anti-Ag(M), anti-Ag(N).

The original locus-symbol for this gene was *R* (Ford, 1942), chosen with reference to the rabbit serum used in the preparation of the

antibodies. At that date the Rhesus group had but recently been discovered. However, *R* has subsequently been so much used in connection with Rhesus that it is felt unwise to retain it for the locus of the MN group. The suggestion due to Stern (1949) that *M* be employed for the locus-symbol has been rejected owing to the confusion likely to arise in respect of group MN if $M^N M^N$ were to represent the genotype of group N. Accordingly, *Ag* has been selected as the locus-symbol of the MN group (suggested by Strandskov, 1948.)

GROUPS : The L groups.

GENES : L^A, L^B.

ANTIGENS : L(A+), L(A—), or, if known, L(A+B—), L(A+B+), L(A—B+).

ANTIBODIES : anti-L(A), anti-L(B).

(The *L* and *Ag* loci are extremely closely linked. *L* was formerly known as *S*, a letter which is preoccupied by the *Secretor* gene.)

(viii) *ABO system*

GENES : G^A, G^B, *G*. Two forms of G^A exist, G^{A1} and G^{A_2}.

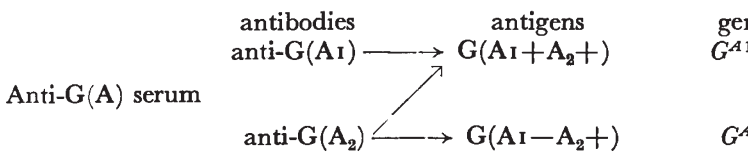
ANTIGENS :

GROUPS	ANTIGENS
O	G(A—B—).
A	G(A+B—).
A1	G(A1+A_2+B—).
A_2	G(A1—A_2+B—).
B	G(A—B+).
AB	G(A+B+).
A1B	G(A1+A_2+B+).
A_2B	G(A1—A_2+B+).

ANTIBODIES : anti-G(A), anti-G(B).

anti-G(A) is composed of an antibody, anti-G(A_2), which reacts both with cells carrying the antigens G(A1) and G(A_2) and one which reacts only with those carrying G(A1). The fact that anti-G(A) is a mixture of two antibodies, one reacting with cells of group A1 only, and the other reacting with cells both of groups A1 and A_2, has been demonstrated serologically. It is most easily explained on the assumption that the antigen possessed by persons of group A1 is itself a mixture of two antigens, as indicated in the notation.

The serological reactions obtained between the antigens and antibodies of groups A1 and A_2 can thus be represented in the following way (agglutination is indicated by an arrow) :



(*Note.*—It was at one time thought that antibodies, then known as " anti-O " and " anti-H ", recognised an antigen produced by *G*. Further work has cast doubt upon this relationship.)

The need for locus-symbols for the blood groups was originally indicated twelve years ago (Ford, 1942). *G*, chosen at that date for the locus of the ABO series, has been retained.

(ix) *Rhesus system*

GENES : C^A, C^B, C^W ; D^A, D ; E^A, E^B ; F^A, F.

(*Note.*—The supposed antigen produced by the allelomorph of D^A does not seem fully established. Therefore the corresponding gene should for the present be written D, not D^B. No antigen produced by F has yet been detected.)

ANTIGENS : (Examples only) C(A+B+), D(A+), E(A—B+), F(A+) could represent the antigens of one donor.
Individual antigens could be stated as C(W+) or E(A—).

ANTIBODIES : (Examples only) anti-C(A), anti-C(B), anti-C(W), . . .

Other Rhesus genes are known, such as D^{U_1}, C^U, C^X, . . . D^{U_1} has been studied in some detail. It is recognisable in the genotypes D^{U_1} D^{U_1} and D^{U_1} D, not in D^{U_1} D^{A1}, as indicated by the introduction of numbers into the notation. This is a natural extension of the system used for groups A_1 and A_2 of the ABO series : a large figure in the suffix represents a gene dominant in effect to one with a small figure, so that D^{A1} is dominant to D^{U_1}, and both are dominant to D (because a capital in the suffix indicates dominance over no letter in the suffix). Numbers should therefore be added when D^{U_1} is being discussed, otherwise the number can be omitted from D^A.

Some anti-D(A) sera react with the D(U+) antigen while others do not. The serological basis for this distinction is so far not clear, so that it cannot yet be described precisely in the notation.

4. DISCUSSION

Linked genes are conveniently represented together on the same side of a line. The other side is blank when there is no homologous chromosome, as in the gametes ; thus, $Lu^A Le^a/$. When both homologous chromosomes are present, the line separates the allelomorphs carried respectively in them ; thus, $Lu^A Le^a/LuLe^a$.

If several genes control a polymorphism, evolution tends to produce extremely close linkage between them. Some of the blood groups are determined in this way, by a block of two or more genes which, in general, acts as a single switch-mechanism ; that is to say, a " super-gene " within which crossing-over must be very rare. It is suggested that, when necessary, such close linkage be indicated in the notation by placing the genes concerned together within a bracket, as illustrated by the following examples :

gametes : $(Ag^M L^A\ )/$ or $(D^A C^A E^B F)/$
zygotes : $(Ag^M L^A)/(Ag^N L^A)$ or $(D^A C^A E^B F)/(DC^B E^B F^A)$

Two points of a general nature must be mentioned in conclusion. (1) I had originally considered using large numerals throughout where numbered groups (*e.g.* A1 and A_2) are involved, with the simple statement that the figure 1 is dominant to the figure 2, so obviating the disadvantage of employing numerals both in large and small type. This notation I rejected for the following reason. When it is found that a group can be subdivided into stronger- and weaker-reacting types, and these are given numbers (as with group A), the distinction will have been obtained serologically, and it may be some time before the genetics of the situation are analysed. If it were then proved that the group labelled 2 were in fact dominant to that labelled 1, the numerals would have to be interchanged if the principle that 1 is dominant to 2 had been laid down. This would cause great confusion if the numbers had already been published as, in the circumstances, they probably would have been. It is worth noting also that the large and small case is already a recognised method of indicating dominance, while a statement such that 1 is dominant to 2 has never been so employed.

(2) The antigen situation shown in the foregoing analysis represents, in respect of particular groups, the presence or absence of given antigens upon the red cells : that is to say, the antigen-phenotype of the individual. It may also be necessary to refer to the antigen itself, such as a G(A) antigen in the saliva. In such circumstances, the plus and minus is not relevant and can be omitted. Thus, for instance, one can refer to the Fy(A) antigen or the G(A) antigen.

I am greatly indebted to Professor C. D. Darlington for his encouragement and advice in the preparation of this account. Also to Professor K. Mather for a very useful suggestion. I have received much information and valuable criticism from Dr R. R. Race and Dr R. Sanger who have given freely of their time in helping me The support of such experienced serologists has been a great stimulus in developing this notation. For any drawbacks from which it may suffer I am, however, alone responsible.

5. REFERENCES

FISHER, R. A. 1930. *The Genetical Theory of Natural Selection.* Oxford Press.
FISHER, R. A., cited by RACE, R. R. 1944. *Nature, 153,* 771-772.
FORD, E. B. 1942. 1st ed. *Genetics for Medical Students.* Methuen, London.
RACE, R. R., AND SANGER, R. 1954. 2nd ed. *Blood Groups in Man.* Blackwell Scientific Publications, Oxford.
SHEPPARD, P. M. 1953. *Am. Nat.., 87,* 283-294.
STERN, C. 1949. *Principles of Human Genetics.* Freeman, San Francisco.
STRANDSKOV, H. H. 1948. *J. Hered., 39,* 108-112.
WIENER, A. S. 1949. *Hereditas,* suppl. vol.