A PROPOSAL FOR A UNIFORM NOMENCLATURE IN BACTERIAL GENETICS

M. DEMEREC, E. A. ADELBERG, A. J. CLARK, AND PHILIP E. HARTMAN[1]

Received December 29, 1965

A proposal for a descriptive and convenient system of genetic nomenclature for bacteria was drafted by the staff and a number of visitors at Cold Spring Harbor in the summer of 1958 (DEMEREC 1958). The proposal had as its basis a system developed by DEMEREC (1956) which largely adhered to previous genetic conventions yet avoided the complications that have developed in the genetic descriptions of some organisms. At conferences held during the summers of 1962 and 1963 the proposal was critically reviewed and revised in accordance with the increased number of genetic markers available, with usage in other areas (e.g., protein chemistry), with suitability for computer analysis, and with interim developments in bacterial genetics (DEMEREC 1963).

The current proposal is an outgrowth of its predecessors, developed by the present authors in consultation with colleagues in other laboratories and in other countries. The basic system has proven convenient to use in the laboratory and has greatly facilitated understanding and communication among many laboratories in the intervening years; increasing use of the system also speaks for its practicality. Thus this proposal does not intend to present a rigid, "official," frozen system of nomenclature. The system is bound to evolve as knowledge advances in the future. The present communication is aimed at making widely available the proposal as developed to date. Comments, suggestions, and additions are welcome.

The aims of the present proposal are: uniformity; a unique designation for each strain; convenience for typing, editing, printing, record-keeping, and information retrieval; and adaptability, simplicity, clarity, and comprehension by workers in all areas of biology; adaptability to new developments in the forseeable future. The proposal takes the form of a set of guiding principles for dealing with categories where usage can be clearly defined; application to specific situations is left to each individual worker. The standardized system of genetic symbols is designed to serve the following purposes: (1) To distinguish clearly between

[1] The addresses of the authors are: DR. M. DEMEREC. (Deceased April 12, 1966.) Formerly of C. W. Post College, Greenvale, New York 11548. A portion of this system of nomenclature was developed at Brookhaven National Laboratory under the auspices of the U. S. Atomic Energy Commission. DR. E. A. ADELBERG, Department of Microbiology, Yale University, New Haven, Connecticut 06510. DR. A. J. CLARK, Department of Molecular Biology, University of California 94720. DR. PHILIP E. HARTMAN, Department of Biology, The Johns Hopkins University, Baltimore, Maryland 21218.

Editor's note: Because the conventions proposed in this communication have been used in a number of recent papers in GENETICS, (e.g. TAYLOR and THOMAN 1964; SANDERSON and DEMEREC 1965) it is felt that publication of an explicit statement of the rules that have guided these authors would serve a useful purpose. Publication of this proposal in GENETICS does not imply official endorsement by the Genetics Society of America, nor that the recommendations will be obligatory upon authors in the Journal. It is hoped, however, that they may serve to provide a useful guideline for authors.

symbols representing the *genotype* of a bacterial strain, and abbreviations of words which describe *phenotypic properties.* (2) To provide a uniform set of symbols for genetic loci, mutant alleles and mutation sites. These symbols have been designed so as to be readily translatable into computer language, in order that it will be possible to program computers to store the complete genotype of any strain. (3) To provide a system for designating and describing strains that will facilitate both recognition and record-keeeping.

The proposed nomenclature is simple and manageable. It avoids the use of unnecessary commas, semicolons, colons, superscripts, subscripts, or Greek letters.

I. *Genotype Symbols*

Bacterial genetic studies begin with the isolation of a strain from nature or the selection of a prototype strain from an existing culture collection. This strain is arbitrarily designated as wild type; genotype symbols are then devised to designate its genetic determinants, as well as to designate *differences* between the genetic determinants of the wild type and those of derived strains. A set of symbols that describes all the known genetic differences between a derived strain and wild type is used to designate the genotype of the derived strain. A derived strain may differ genetically from wild type in either of two respects: it may carry one or more mutant loci; it may have gained or lost one or more plasmids or episomes. Systems for symbolizing each class of genetic changes will be discussed separately.

A. *Mutant loci.* (1) *General principles:* The terms "locus" and "gene" will be used interchangeably to refer to a specific sequence of nucleotides governing the sequence of amino acids in a specific polypeptide (or the sequence of nucleotides in a specific RNA molecule). Nucleotide sequences which themselves may not be transcribed, but which govern the punctuation or regulation of transcription, are also referred to as "loci." Minor changes in the nucleotide sequence of a locus (substitutions, small deletions, or insertions) may occur by mutation; the different forms of a locus brought about by such mutations are called *alleles.* The problems involved in recognizing mutation sites and loci will be discussed in section I,D. The following system is proposed for designating loci, alleles, and mutation sites.

Locus symbols in current usage are listed in Appendix A.

Recommendation 1: *Each locus of a given wild-type strain is designated by a three-letter, lower-case, italicized symbol.*

The existence of a locus is recognized genetically by the occurrence of a mutation within it. In many cases, a symbol must be invented long before the polypeptide corresponding to the locus in question has been identified; the investigator may only be aware of a gross phenotypic change produced by the mutation. It is thus a common practice to choose three letters which recall this gross phenotypic change. For example, the symbol *ara* was first coined to refer to the loci in which mutations occur that affect the response of the cell to arabinose as a carbon and energy source.

Recommendation 2: *Different loci, any one of which may mutate to produce the same gross phenotypic change, are distinguished from each other by adding an italicized capital letter immediately following the three-letter lower-case symbol.*

When, for example, it was recognized that there are three enzymes in the pathway for arabinose utilization, each controlled by a different locus, the three loci in question were designated *araA*, *araB* and *araD* (ENGLESBERG *et al.* 1962). It is useful, but not essential, to assign the capital letters in the same order as the enzymes in the pathway.

(2) *Application to loci concerned with some commonly observed phenotypes.* (a) *Loci involved in genetic regulation:* Loci concerned with the regulation of polypeptide formation are designated in accordance with the above convention. For example, the locus which regulates the activity of *araA*, *araB* and *araD* has been designated *araC* (ENGLESBERG *et al.* 1962).

Should an author wish to use a symbol which suggests that the locus in question plays a regulatory function, the capital letter might be chosen accordingly. Thus, a "regulator" locus, the product of which represses *arg* loci in the *trans* configuration, might be designated *argR*; if more than one such "regulator" is discovered, they might be symbolized by the sequence *argR*, *argS*, etc. Similarly, an "operator" locus, which can mutate to derepress an *arg* locus in the *cis* configuration, might be designated *argO*. If more than one operator concerned with arginine biosynthetic enzymes are found, they might be assigned the symbols *argO*, *argP*, etc. It is not necessary, however, to use letters having specific connotations, as long as each locus has a unique designation. *araC*, for example, is a perfectly satisfactory symbol for the locus which regulates *araA*, *araB* and *araD*, and has the advantage of not conveying any preconceived ideas of precise gene function.

(b) *Loci governing resistance and sensitivity:* In the absence of knowledge concerning precise mechanisms of resistance or sensitivity, it is customary to choose three letters which recall the deleterious agent. The symbol *str*, for example, was chosen to designate a locus which can mutate to affect sensitivity to streptomycin. When a second locus affecting streptomycin-sensitivity was discovered, the two loci were designated *strA* and *strB* (SANDERSON and DEMEREC 1965). Similarly, the loci within which mutations affect sensitivity to ultra violet light have been designated *uvrA*, *uvrB*, and *uvrC* (HOWARD-FLANDERS *et al.* 1964).

(c) *Suppressor loci:* The change in phenotype produced by a mutation in one locus may be partially or fully reversed by a mutation in a second locus. The second locus is then called a "suppressor locus."

In many cases it has been demonstrated that genetic suppression involves a change at the translation level of protein synthesis, and it is clear that at least some suppressor loci determine the structures of components of the translation machinery (ribosomes, amino acid activating enzymes, transfer RNA's, etc.). In the absence of any direct information, however, it is necessary to invent symbols which avoid unwarranted connotations. The symbol *sup* has been used, fol-

lowed by capital letters which distinguish the loci that have been mapped at different places. Thus, one such set of loci have been designated *supH*, *supL*, *supM*, *supN*, *supO*, *supP* and *supT* (EGGERTSSON and ADELBERG 1965). As more suppressors are mapped, the remaining letters of the alphabet can be used; if these are not sufficient, another set with a symbol such as *spr* may be required.

B. *Mutation sites.*

Recommendation 3: *A mutation site should be designated by placing a serial isolation number after the locus symbol. If it is not known in which of several loci governing related functions the mutation has occurred, a hyphen is used instead of the capital letter.*

For example, GROSS and ENGLESBERG (1959) isolated a large number of *E. coli* B mutants unable to utilize arabinose as carbon and energy source. The mutation sites were provisionally designated *ara-1*, *ara-2*, *ara-3*, etc. Later, 17 of these mutations were mapped, and the enzymes which had been altered were identified. It was then possible to complete the designations by substituting capital letters for the hyphen: e.g., *araB1*, *araA2*, *araC3*, *araA4*, *araC5*, *araB6*, etc. Note that the isolation number is not changed at the time that the locus letter is assigned.

According to this system, the symbol *ara-1* originally designated a specific mutation site on the chromosome of *E. coli* B. The symbol *araB1* still refers to the same site, but now conveys additional information about its location. Furthermore, the symbol *araB1* constitutes the designation of a specific allele: thus, the *araB* locus has a given nucleotide sequence as the *araB1* allele, and a different nucleotide sequence as the *araB6* allele.

Although it is possible that two independent mutations may bring about the exact same base-pair change, the probability of this occurring is low and to prove that it has occurred requires extensive recombination testing. It should thus be assumed, until proved otherwise, that each independent mutation event has occurred at a different site. Even when two mutations are found to occupy identical sites in sensitive recombination tests and to share other properties in common, each mutation still retains its original isolation number.

It is essential that a particular genetic symbol (e.g., *ara-1*) never be utilized on two occasions for two independent mutations. Furthermore, in a useful system of nomenclature, published isolation numbers should not be changed. A few exceptions, however, may occur. For example, a mutation giving rise to valine-resistance in *E. coli* K-12 might receive a particular symbol and isolation number; later, it might be discovered that the mutation actually had occurred in the *ilvB* locus, leading to the formation of an altered, feedback-resistant condensing enzyme elicited by the *ilvB* gene. In this case, the mutation should receive the next available *ilv* isolation number and the change in mutation designation should be noted in the literature.

To avoid duplication of allele numbers, it is urged that geneticists working with the same organism organize a central agency for the assignment of blocks of numbers within each locus. Notices concerning laboratories willing to serve

as clearing-houses for this purpose appear regularly in the Microbial Genetics Bulletin.

C. *Alleles.* The nature of any particular mutational change is not indicated by the genotypic symbol. For example, *araB1* might be a base-pair substitution or a small deletion. In each case, however, the symbol indicates the presence of a unique nucleotide sequence for the locus in question, and thus constitutes the designation of an *allele.*

In bacterial genetics, the practice of using a plus (+) sign to indicate the wild-type allele of a locus has been borrowed from the genetic nomenclature system used for other organisms. Thus, *araB*$^+$ is the wild-type allele of the *araB* locus; it stands for the particular sequence of nucleotides which is found in the *araB* locus of the strain arbitrarily chosen as wild type. Since a locus may have a thousand or more base-pair positions, and since any of four different base-pairs may occupy any one position, the number of possible mutant alleles is very large. It is important that each mutant allele of a particular locus be given a unique designation; e.g., by the use of serial numbers as suffixes. Use of a mutant allele designation is sufficient (e.g., *ara-1*); use of a symbol such as *ara-1*$^-$ is redundant. Superscripts should also be omitted when referring to a particular class of mutants. For example, it is sufficient to speak of all *araB* mutants, or all *ara* mutants, to designate a group of strains all of which have a mutation in the same locus or in the same set of loci.

A deliberate feature of the system recommended above is that the *allele designation conveys no information concerning phenotype.* For example, all alleles of the *strA* locus can be designated simply by the series *strA1*, *strA2*, *strA3*, etc. Some of these alleles may confer on the cell resistance to low levels of streptomycin, some may confer resistance to high levels of streptomycin, and others may make the cell conditionally or absolutely dependent on streptomycin. None of these facts is relevant to the designation of an allele, however, since an allele is defined as a particular sequence of nucleotide pairs. Furthermore, the phenotype associated with a given allele can often be readily altered by mutations at other loci or by changes in the environment. Thus, according to this proposal, superscripts and suffixes such as "R" and "S" for resistance and sensitivity would be rigorously excluded from genotype designations, and should be reserved for use in abbreviations of *phenotype* (see section II).

D. *Recognition of mutation sites and loci.* When a bacterial strain undergoes a genetic change as a result of a single mutation, the site of that mutation may be assigned a genotype symbol even if its map location is unknown. Proof that the observed change reflects a mutation at a single site requires recombinational analysis; until such an analysis is made, the site description remains tentative.

The existence of a mutation site establishes the existence of a locus whose function has been altered by the mutation. When two or more mutations at different sites have altered the same phenotypic property, however, the assignment of locus designation requires further genetic analysis. For example, genetic mapping of a number of mutants derepressed for alkaline phosphatase synthesis revealed that the mutations had occurred at two widely separated regions of the

chromosome. Accordingly, the existence of two loci was inferred, and these were designated *R1pho* and *R2pho* (GAREN and ECHOLS 1962). (To comply with Recommendation 2, these would be changed to *phoR* and *phoS*, respectively.)

When a number of mutation sites all affecting the same phenotypic property are clustered closely together, it is often assumed that the DNA segment within which they are located represents a single functional locus. Without further evidence, however, it is possible that the segment in question includes two or more separate loci having related functions. For example, the early work on mutations affecting the ability to ferment lactose led to the designation of a particular chromosomal region of *E. coli* as the "*lac* locus" (LEDERBERG 1947). Later, this region was found to include at least three loci, governing the formation of beta-galactosidase, beta-galactoside permease, and a repressor regulating the other two loci (JACOB and WOLLMAN 1961).

To prove that two observed mutations are located within the same locus, it is necessary to show that both mutations have affected the amino acid sequence of the same polypeptide. Alternatively, the identity of a locus can be tentatively established by the *cis-trans* test of genetic complementation (JACOB and WOLLMAN 1961; GAREN and GAREN 1963; HELLING and WEINBERG 1963; HAYES 1964; LOPER *et al.*, 1964).

From the foregoing considerations it is clear that extensive genetic analyses are required before the phenotypic differences between two closely related strains isolated from nature can be ascribed to definable genotypes. For example, a particular phage mutation might be found to be suppressed in *E. coli* K-12 but not in *E. coli* B. Such a difference would reflect an unknown number of genotypic differences between the two strains; without further analysis, no assignment of genotype would be possible.

E. *Plasmids and episomes.* Bacteria are host to a variety of genetic elements capable of independent replication. Such elements include plasmids (remaining autonomous) or episomes (capable of alternating between an autonomous state and a state of attachment to the chromosome).

The known episomes and plasmids include such elements as the sex factor of *E. coli* K-12, the colicinogenic agents, the so-called "resistance-transfer factors", and a variety of temperate phages such as lambda (λ) and P1. In each case, the element is a DNA structure corresponding to one to two percent of the chromosome in size. As such, it is sufficiently large to contain from 50 to 100 separate loci of average length.

To include information about plasmids and episomes in the genotype of a bacterial strain, the following are needed: (*i*) symbols designating the plasmids and episomes which are present; (*ii*) symbols for the mutant loci and/or mutation sites which they carry.

Recommendation 4: *Plasmids and episomes should be designated by symbols which are clearly distinguishable from symbols used for genetic loci.*

The first letter of a symbol for a plasmid or for an episome is capitalized, the symbol is *not* italicized, and the symbol is placed in parentheses, e.g., (Col E1).

Recommendation 5: *Mutant loci and mutational sites on plasmids and epi-*

somes should be designated by symbols of the same kind as those used for loci and sites on the chromosome.

Recommendation 5 is made with the view that loci on plasmids and episomes are not different in kind from loci on the chromosome, and—as part of the total genetic complement of the bacterial cell—should be symbolized according to the unified system of nomenclature.

The complete genotypic description of strains carrying an episome requires a description of the state of the episome: autonomous or integrated. Such information can best be given in words; e.g.: "RTF is integrated between the chromosomal loci *proA* and *proB*". In the case of strains harboring the sex factor, F, however, a set of symbols is already in use which requires classification. These symbols are F^-, F^+, F′ (F-prime) and Hfr. Different authors have used these symbols to mean different things; for example, "Hfr" has sometimes been used to denote the phenotypic property "high frequency of recombination" (HAYES 1963) and sometimes to mean the genotypic property of integration of sex factor and chromosome (ADELBERG and PITTARD 1965). To avoid further confusion, it is recommended that these four symbols be used as follows:

F^-: The state of lacking the sex factor, F. The criterion of acting as a genetic recipient in conjugation is not sufficient, since strains carrying F can also act as recipients. The criteria of the F^- state include: activity as a genetic recipient; lack of activity as a genetic donor; failure to transmit F; ability to be infected with F; and resistance to male-specific phages.

F^+: The state of harboring an autonomous sex factor which does not carry any genetically recognizable chromosomal fragments.

F′: The state of harboring an autonomous sex factor which carries a genetically recognizable segment of the bacterial chromosome. For example, the strain AB1206 harbors the sex factor F14, in which F DNA has become integrated with a chromosomal fragment bearing such loci as *ilvD*, *metE*, and *argA* (PITTARD *et al.* 1963). Such sex factors attached to chromosomal fragments have been called "substituted sex factors" (HAYES 1964), "F-merogenotes" (CLARK and ADELBERG 1962), or "F-genotes" (RAMAKRISHNAN and ADELBERG 1965). The distinction between the F^+ and F′ states is strictly an operational one, since even wild-type sex factors carry regions of homology with the chromosome (FALKOW and CITARELLA 1965), presumably reflecting incorporated chromosomal fragments (ADELBERG and PITTARD 1965).

Hfr: The state of harboring a sex factor which is integrated with the chromosome. Such a state may or may not confer on the cell the phenotype of a high-frequency genetic donor, since this property depends on the functioning of many loci on the sex factor (and possibly on the chromosome as well). Hfr strains in which the sex factor is defective may be extremely low-frequency donors (CUZIN and JACOB 1965).

Recommendation 6: *The description of a strain carrying an episome should include a statement concerning the state and/or location of the episome. The symbols F^-, F^+, F′, and Hfr should be used only to designate the sex factor states as outlined above, and not to convey information concerning the phenotypic properties of mating activity.*

F. *Changes in genotype symbols.*

Recommendation 7: *Genotype symbols which have already been published and which conform to the system recommended above should not be changed.*

Genotype symbols which do not conform to the above system should be changed accordingly, and the change should be noted when the new symbol is first published.

For example, the set of loci governing the utilization of lactose has been collectively designated as *lac*, but the individual loci have been referred to by the single letters *i*, *o*, *z* and *y* (JACOB *et al.* 1960). Mutant alleles have been designated by symbols such as "i_3^-" "z_4^-", etc. To conform with the standard system proposed here it would, for example, be necessary to change the latter symbols to *lacI3* and *lacZ4*, respectively.

In exceptional cases the italicized capital letter, designating the gene locus, may have to be changed as subsequent tests define the gene-polypeptide chain relationships more precisely. In these cases, the old and the new gene-locus designations are reported and the reasons for the change are stated in the literature. The change should be referred to in use of the new symbol for an appropriate duration thereafter.

II. *Phenotype Abbreviations*

A. *General principles.* The observable properties of a bacterial strain constitute that strain's phenotype. Resistance to a drug such as penicillin, for example, is a phenotypic trait and may reflect any of several diverse genotypes. In publishing a strain description, it is essential that the author make clear whether he is referring to a phenotypic trait or to a genotypic character. In the latter case, a set of symbols such as *penA*, *penB*, *penC*, etc., should be used to designate loci concerned with resistance to penicillin. The phenotype, on the other hand, can best be stated in words: e.g., "penicillin-resistant". In practice, however, there is a justifiable tendency to *abbreviate* what might otherwise be a cumbersome description of phenotype. Thus, the abbreviation "Pen-r10" might be used as an abbreviation of "resistant to 10 units per milliliter of penicillin", provided that the abbreviation is fully explained the first time that it appears in a given paper.

Care in distinguishing between phenotype abbreviations and genotype symbols is all the more urgent in view of the common practice of inventing genotype symbols which are themselves abbreviations (e.g., "*penA*" for a locus which can mutate to produce resistance to penicillin). The phenotype, on the other hand, can and should be described in words. Abbreviations are needed only for the sake of brevity and of clarity in writing. For example, the sentence "A cross was performed between a CSD Met$^-$ strain and an SmR100 Met$^+$ strain" is, perhaps, easier to assimilate than the sentence "A cross was performed between a conditional streptomycin-dependent strain which requires either methionine or streptomycin for growth and a strain which does not require methionine and is resistant to 100 units per milliliter of streptomycin"—*provided that the abbreviations have been clearly defined beforehand.*

To meet these needs, the following recommendation is made:

Recommendation 8: *Phenotypic traits should be described in words, or by the use of abbreviations which are defined the first time they appear in a given paper. The abbreviations should be clearly distinguishable from genotype symbols.*

To conform with Recommendation 8, three-letter lower-case italicized abbreviations should not be used as phenotype abbreviations. When three letters are used, the distinction between genotype symbol and phenotype abbreviation should be emphasized by capitalizing the first letter of the phenotype abbreviation: e.g., "The phenotype Met$^-$ is associated with a mutation in the *metA* locus".

B. *Phenotypic properties involved in mating activity.* The description of mating activity deserves special mention, in view of the problems discussed in connection with the sex factors (Recommendation 6). In many instances, it is necessary to state whether a given strain behaves as a genetic donor or as a genetic recipient, and to indicate the frequency with which it does so. Thus, a strain may be designated as a high-frequency donor, a low-frequency donor, a high-frequency recipient, etc. If such terms require abbreviation, the abbreviations should be carried in accordance with Recommendations 6 and 8; e.g., they should be clearly distinguished from the genotypic symbols F^-, F^+, F' and Hfr.

III. *Descriptions of Strains*

A. *Strain designations.* Every strain must have a unique designation. As is the case of symbols for genetic loci, strain designations should be simple (e.g., free of subscripts, superscripts, Greek letters, etc.) and should be compatible with systems for cataloguing and record-keeping. Accordingly, the following recommendation is made:

Recommendation 9: *Strains should be designated by simple serial numbers. To avoid duplications, different laboratories should use different letter prefixes. Strain designations should not be italicized.*

Phenotypic information should not be included in a strain designation. For example, a designation such as "C600S^R", representing a streptomycin-resistant derivative of strain C600, is undesirable because many different resistant mutants would have the same designation.

Although a strain could be given a unique designation by writing its full or partial genotype (e.g., K-12 (*araB1, metE6, str-17*)), such designations are cumbersome and greatly complicate the jobs of cataloguing and record-keeping. Instead, such a strain should be given a simple serial number, such as JC1234, and its genotype should be described in a table or footnote, or in the text the first time the strain is mentioned. Some examples of prefixes to strain designations are: CL, for *E. coli* London (Stocker laboratory, retained at Stanford); SB, for Salmonella Baltimore (Hartman laboratory); SW, Salmonella Wisconsin (Lederberg laboratory, retained at Stanford), etc. The exact prefixes are unimportant except that each laboratory must be careful not to duplicate a prefix used elsewhere and thus destroy the uniqueness of the strain designation.

In describing an experiment, it is often helpful to stress a relevant phenotypic or genotypic character of a strain. This can be done by supplementing the strain number, rather than replacing it. For example, "A cross was carried out between strain AC100 (carrying *araB1*) and AC101 (carrying *araB6*)."

B. *Changes in strain designation.* When a strain is received from another laboratory, it may be necessary to change its designation for the purposes of local record-keeping. To avoid confusion in publication, however, the following recommendation is made.

Recommendation 10: *Strain designations which have already been published and which conform to Recommendation 9 should not be changed. Strain designations which do not conform to Recommendation 9 should be changed accordingly, and the change should be noted when the new designation is first published.*

C. *Methods for describing strains.*

Recommendation 11: *When a strain is first mentioned in publication its genotype should be described, and relevant phenotypic information should be given. The genotype includes a list of all mutant loci and/or mutation sites, a list of episomes and/or plasmids, and information concerning the state and location of any episome.*

When the genotype includes only one or a few items, it can conveniently be described in the text or in a footnote. When the genotype is long and complex, however, and when there are many strains, it is extremely helpful to tabulate the information. A sample of such a table, describing some hypothetical strains, is given in Appendix B.

IV. *Description of a Cross*

Once strains are clearly defined, crosses can be described by simple reference to the strain designations, for example, AB712 × AB301.

V. *Hybrid Strains*

The system described above for designating mutant loci and mutation sites presents no problems as long as all strains are derived from a single wild type. As discussed under Recommendation 1, a locus is considered mutant if it differs from the corresponding locus in the arbitrarily chosen wild-type strain.

Thus, a series of mutant loci have beeen designated within strains derived from *E. coli* K-12, another series within strains derived from *E. coli* B, still another within strains derived from *Salmonella typhimurium*, and so on. But what is the genotype of a hybrid strain, arising from a cross between wild-type *E. coli* K-12 and wild-type *E. coli* B? Some of its loci will be derived from one wild type, and some from the other. If K-12 were considered as the reference strain, the loci inherited from B would be mutant, and vice versa. Furthermore, the genotype of the hybrid could not be written until it was known from which parent each locus was derived.

Should it be possible to determine from which parent a particular wild-type locus was derived, a symbol could be devised to convey this information. Most loci, however, are likely to remain unidentified. In some situations, e.g., when many new strains are to be derived from a particular hybrid, it will be best to designate the hybrid itself as a new prototype strain comparable to a wild type.

SUMMARY

Recommendations are made for a convenient system of nomenclature. These specify the manner of symbolizing or designating loci, mutation sites, plasmids and episomes, sex factors, phenotypic traits, and bacterial strains. Symbols are proposed for known genes in *Escherichia coli* and *Salmonella typhimurium*. The system has been employed in a number of recent papers in GENETICS, e.g. by TAYLOR and THOMAN (1964) and SANDERSON and DEMEREC (1965).

LITERATURE CITED

ADELBERG, E. A., and J. PITTARD, 1965 Chromosome transfer in bacterial conjugation. Bacteriol. Rev. **29**: 161–172.

CLARK, A. J., and E. A. ADELBERG, 1962 Bacterial conjugation. Ann. Rev. Microbiol. **16**: 289–319.

CUZIN, F., and F. JACOB, 1965 Analyse génétique fonctionnelle de l'épisome sexuel d'*Escherichia coli* K 12. Compt. Rend. **260**: 2087–2090.

DEMEREC, M., 1956 Terminology and nomenclature. Carnegie Inst. Wash. Publ. **612**: 1–4. —— 1958 Proposal for a uniform nomenclature in bacterial genetics. Microbial Genet. Bull. **16**: 38–41. —— 1963 Proposal for a uniform nomenclature in bacterial genetics. Microbial Genet. Bull. **19**: 30–34.

EGGERTSSON, G., and E. A. ADELBERG, 1965 Map positions and specificities of suppressor mutations in *Escherichia coli* K-12. Genetics **52**: 319–340.

ENGLESBERG, E., R. L. ANDERSON, R. WEINBERG, N. LEE, P. HOFFEE, G. HUTTENHAUER, and H. BOYER, 1962 L-arabinose-sensitive, L-ribulose 5-phosphate 4-epimerase-deficient mutants of *Escherichia coli*. J. Bacteriol. **84**: 137–146.

FALKOW, S., and R. V. CITARELLA, 1965 Molecular homology of F-merogenote DNA. J. Mol. Biol. **12**: 138–151.

GAREN, A., and H. ECHOLS, 1962 Genetic control of induction of alkaline phosphatase synthesis in *E. coli*. Proc. Natl. Acad. Sci. U.S. **48**: 1398–1402.

GAREN, A., and S. GAREN, 1963 Complementation *in vivo* between structural mutants of alkaline phosphatase from *Escherichia coli*. J. Mol. Biol. **7**: 13–22.

GROSS, J., and E. ENGLESBERG, 1959 Determination of the order of mutational sites governing L-arabinose utilization in *Escherichia coli* B/r by transduction with phage P1bt. Virology **9**: 314–331.

HARTMAN, P. E., and S. R. SUSKIND, 1964 *Gene Action*. Prentice-Hall, Englewood Cliffs, N.J. Chapter 5, Complementation.

HAYES, W., 1963 The mechanism of genetic recombination in *Escherichia coli*. Cold Spring Harbor Symp. Quant. Biol. **18**: 75–93. —— 1964 *The Genetics of Bacteria and Their Viruses*. Wiley, New York.

HELLING, R. B., and R. WEINBERG, 1963 Complementation studies of arabinose genes in *Escherichia coli*. Genetics **48**: 1392–1410.

HOWARD-FLANDERS, P., E. SIMSON, and L. THERIOT, 1964 The excision of thymine dimers from DNA, filament formation and sensitivity to ultraviolet light in *Escherichia coli* K-12. Mutation Res. **1**: 219–226.

JACOB, F., D. PERRIN, C. SANCHEZ, and J. MONOD, 1960 L'óperon: groupe de gènes à expression coordonnée par un opérateur. Compt. Rend. **250**: 1727–1729.

JACOB, F., and E. L. WOLLMAN, 1961 *Sexuality and the Genetics of Bacteria.* Academic Press, New York.

LEDERBERG, J., 1947 Gene recombination and linked segregations in *Escherichia coli.* Genetics **32**: 505–525.

LOPER, J. C., M. GRABNAR, R. C. STAHL, Z. HARTMAN, and P. E. HARTMAN, 1964 Genes and proteins involved in histidine biosynthesis in Salmonella. Brookhaven Symp. Biol. **17**: 15–50.

PITTARD, J., J. S. LOUTIT, and E. A. ADELBERG, 1963 Gene transfer by F′ strains of *Escherichia coli* K-12. I. Delay in initiation of chromosome transfer. J. Bacteriol. **85**: 1394–1401.

RAMAKRISHNAN, T., and E. A. ADELBERG, 1965 Regulatory mechanisms in the biosynthesis of isoleucine and valine. III. Map order of the structural genes and operator genes. J. Bacteriol. **89**: 661–664; **90**: 295.

SANDERSON, K. E., and M. DEMEREC, 1965 The linkage map of *Salmonella typhimurium.* Genetics **51**: 897–913.

TAYLOR, A. L., and M. S. THOMAN, 1964 The genetic map of *Escherichia coli* K-12. Genetics **50**: 659–677.

APPENDIX A: LIST OF PROPOSED SYMBOLS

(For data on many of these loci in *E. coli* and *S. typhimurium*, see reviews in TAYLOR and THOMAN [1964] and SANDERSON and DEMEREC [1965].)

acr genes determining response to acridine (resistance or sensitivity)
ade see: *pur*
adl genes determining adonitol utilization
ala genes determining and regulating alanine biosynthesis
ank genes determining and regulating K antigen synthesis
ara genes determining and regulating arabinose utilization
arg genes determining and regulating arginine biosynthesis
aro genes determining and regulating biosynthesis of several aromatic amino acids and aromatic vitamins
ars genes determining sensitivity to arginine
asc genes determining and regulating ascorbate biosynthesis
asn genes determining and regulating asparagine biosynthesis
asp genes determining and regulating aspartic acid biosynthesis
att prophage attachment site
azi genes determining response to azide (resistance or sensitivity)
bio genes determining and regulating biotin biosynthesis
chr genes determining response to chromium (sensitivity or resistance)
cit genes determining and regulating citrate utilization
clb genes determining and regulating cellobiose fermentation
clk genes determining response to colicine K (resistance or sensitivity)
(Col K) bacteriocinogenic for colicine K
cys genes determining and regulating cysteine biosynthesis
cyt genes determining and regulating cytosine biosynthesis
dds genes determining D-serine deaminase and regulating its production
dta genes determining and regulating D-tartrate utilization
dul genes determining and regulating dulcitol utilization
fdp gene(s) determining fructose-1,6-diphosphatase
fim genes determining and regulating fimbriation (piliation)
fla genes determining and regulating flagellation (presence of flagella)
gal genes determining and regulating galactose utilization
gas genes determining gas formation from fermentable sugars
gln genes determining and regulating glutamine biosynthesis

glp genes determining and regulating glycerol and glycerolphosphate utilization
glu genes determining and regulating glutamic acid biosynthesis
gly genes determining and regulating glycine biosynthesis
gra genes determining response to gramicidin (resistance or sensitivity)
gua genes determining and regulating guanine biosynthesis
hag genes determining and regulating the synthesis of flagellar antigens
hcr genes determining host cell reactivation
hem genes determining and regulating heme biosynthesis
his genes determining and regulating histidine biosynthesis
hom genes determining and regulating homoserine biosynthesis
ilv genes determining and regulating isoleucine and valine biosynthesis
inl genes determining and regulating inositol utilization
ita genes determining and regulating isotartrate utilization
lac genes determining and regulating lactose utilization
lam genes determining response to phage lambda (resistance or sensitivity)
leu genes determining and regulating leucine biosynthesis
lev genes determining and regulating levulose utilization
lon genes determining cellular growth into filaments
lys genes determining and regulating lysine biosynthesis
mal genes determining and regulating maltose utilization
man genes determining and regulating mannose utilization
mel genes determining and regulating melibiose utilization
met genes determining and regulating methionine biosynthesis
mlz genes determining and regulating melizitose utilization
mot genes determining the functioning of flagella (i.e. flagella present but bacteria nonmotile)
mtl genes determining and regulating mannitol utilization
mut genes determining functions whose aberration leads to heightened spontaneous mutation rates
nfn genes determining response to nitrofuran (resistance or sensitivity)
nic genes determining and regulating nicotinic acid biosynthesis
nit genes determining and regulating nitrate utilization
nml genes determining the presence of E-N-methyl-lysine in flagellar protein
nol genes determining response to norleucine (resistance or sensitivity)
nov genes determining response to novobiocin (resistance or sensitivity)
(P1) lysogenicity for phage P1
(P22) lysogenicity for phage P22
pab genes determining and regulating *p*-aminobenzoic acid biosynthesis
pan genes determining and regulating pantothenic acid biosynthesis
pdx genes determining and regulating pyridoxine biosynthesis
pen genes determining response to penicillin (resistance or sensitivity)
pgi gene(s) determining phosphoglucosisomerase
phe genes determining and regulating phenylalanine biosynthesis
pho genes determining alkaline phosphatase and regulating its production
pig genes determining pigment formation (pigment of unknown nature)
pmi gene(s) determining phosphomannosisomerase
pmx genes determining response to polymixin (resistance or sensitivity)
pro genes determining and regulating proline biosynthesis
pur genes determining and regulating purine biosynthesis
pyr genes determining and regulating pyrimidine biosynthesis
raf genes determining and regulating raffinose utilization
rbs genes determining and regulating ribose utilization
rec genes affecting genetic recombination
rha genes determining and regulating rhammose utilization
rib genes determining and regulating riboflavin biosynthesis

(Rtf) harboring resistance-transfer-factor
rou genes determining functions whose aberration leads to rough colony morphology or serotype
scr genes determining and regulating sucrose utilization
ser genes determining and regulating serine and glycine biosynthesis
som genes determining and regulating somatic antigen synthesis
sor genes determining and regulating sorbose utilization
srl genes determining and regulating sorbitol utilization
str genes determining response to streptomycin (resistance, sensitivity, or dependence)
suc genes determining and regulating succinic acid utilization
sul genes determining response to sulfonamide (sensitivity or resistance)
tfr genes determining response to phage T4 (resistance or sensitivity)
thi genes determining and regulating thiamine biosynthesis
thr genes determining and regulating threonine biosynthesis
trp genes determining and regulating tryptophan biosynthesis
tna genes determining tryptophanase and regulating its production
ton genes determining response to phage T1 (sensitivity or resistance)
tre genes determining and regulating trehalose utilization
trp genes determining and regulating tryptophan biosynthesis
tsx genes determining response to phage T6 (resistance or sensitivity)
tur genes determining and regulating turanose utilization
tyr genes determining and regulating tyrosine biosynthesis
uvr genes determining repair of ultraviolet radiation damage to DNA
val genes whose aberration leads to valine-resistance
vio genes determining response to viomycin (resistance or sensitivity)
xyl genes determining and regulating xylose utilization

APPENDIX B:

Example of a table illustrating description of some hypothetical strains

		Mutant loci and mutation sites									Relevant phenotype*							Episomes†		
Strain	Source	*lacZ*	*lacY*	*lacI*	*lacO*	*his*	*try*	*str*	*proA*	*proB*	β-gal	Perm	Hist	Tryp	Sm	Prol	M. Act.	λ	P1	F
AC9991	‡	+	+	3	+	1	6	12	2	+	C	C	—	—	R	—	LFD	+	+	1**
AC9992	‡	13	+	3	+	+	6	7	+	+	—	C	+	—	D	+	HFD	+	—	1§
AC9993	‖	+	16	+	4	1	+	+	+	7	C	—	—	+	S	—	Rec	131¶	—	—

* (β-gal) β-galactosidase; (Perm) β-galactoside permease; (Hist) histidine; (Tryp) tryptophan; (Sm) streptomycin; (Prol) proline; (M. Act.) mating activity; (C) constitutive; (—) absent or required; (+) synthesized; (R) resistant; (D) dependent; (S) sensitive; (LFD) low frequency donor; (HFD) high frequency donor; (Rec) recipient.
† (+) wild-type episome present; (—) episome absent; numbers are strain numbers of episomes.
‡ This laboratory.
§ F1 is integrated between *proB* and *lacY;* promotes chromosome transfer in the order *proB, leu, thr, metA* . . . etc.
‖ Obtained from J. C. SUNSET. Formerly designated A3–7/*lac*^{c}.
¶ λ 131 carries the mutant alleles *susA1*, *susB7*.
** F1 is autonomous.

APPENDIX C: SUMMARY OF RECOMMENDATIONS

1. Each locus of a given wild-type strain is designated by a three-letter, lower-case italicized symbol.

2. Different loci, any one of which may mutate to produce the same gross phenotypic change, are distinguished from each other by adding an italicized capital letter immediately following the three-letter lower-case symbol.

3. A mutation site should be designated by placing a serial isolation number after the locus symbol. If it is not known in which of several loci governing related functions the mutation has occurred, the capital letter is replaced by a hyphen.

4. Plasmids and episomes should be designated by symbols which are clearly distinguishable from symbols used for genetic loci.

5. Mutant loci and mutational sites on plasmids and episomes should be designated by symbols of the same kind as those used for loci and sites on the chromosomes.

6. The description of a strain carrying an episome should include a statement concerning the state and/or location of the episome. The symbols F^-, F^+, F' and Hfr should be used only to designate the sex factor states as outlined above, and not to convey information concerning the phenotypic properties of mating activity.

7. Genotype symbols which have already been published and which conform to the system recommended above should not be changed. Genotype symbols which do not conform to the above system should be changed accordingly, and the change should be noted when the new symbol is first published.

8. Phenotypic traits should be described in words, or by the use of abbreviations which are defined the first time they appear in a given paper. The abbreviations should be clearly distinguishable from genotype symbols.

9. Strains should be designated by simple serial numbers. To avoid duplications, different laboratories should use different letter prefixes. Strain designations should not be italicized.

10. Strain designations which have already been published and which conform to Recommendation 9 should not be changed. Strain designations which do not conform to Recommendation 9 should be changed accordingly, and the change should be noted when the new designation is first published.

11. When a strain is first mentioned in publication its genotype should be described, and relevant phenotypic information should be given. The genotype includes a list of all mutant loci and/or mutation sites, a list of episomes and/or plasmids, and information concerning the state and location of any episome.