MICROBIOLOGICAL REVIEWS, Dec. 1988, p. 533–535
0146-0749/88/04533-03$02.00/0

Vol. 52, No. 4

New Unified Nomenclature for the Flagellar Genes of *Escherichia coli* and *Salmonella typhimurium*

TETSUO IINO,[1] YOSHIBUMI KOMEDA,[2] KAZUHIRO KUTSUKAKE,[1] ROBERT M. MACNAB,[3]* PHILIP MATSUMURA,[4] JOHN S. PARKINSON,[5] MELVIN I. SIMON,[6] AND SHIGERU YAMAGUCHI[7]

Laboratory of Genetics, Faculty of Science,[1] and Molecular Genetics Research Laboratory,[2] University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, and Laboratory of Biology, Meiji University, 1-9-1 Eifuku, Suginami-ku, Tokyo 168,[7] Japan; Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511[3]; Department of Biological Sciences, University of Illinois at Chicago, Chicago, Illinois 60680[4]; Biology Department, University of Utah, Salt Lake City, Utah 84112[5]; and Division of Biology, California Institute of Technology, Pasadena, California 91125[6]

BACKGROUND

As research into the flagellar gene systems of *Escherichia coli* and *Salmonella typhimurium* has progressed, their complexity has proved to be far greater than originally imagined. Currently, close to 40 genes are known to be involved in the flagellar gene systems of both species (see, e.g., reference 2), making them second only to the ribosomal gene system in size.

Because extensions to the basic symbols (*fla* and, in *E. coli*, *flb*) were assigned as mutants were isolated and subjected to complementation analysis, they bear little relationship to genome order or function. Also, in many cases, what was originally thought to be a single locus has subsequently proved to consist of several genes. As the most extreme example, the original *flaF* locus of *S. typhimurium* is now known to consist of 10 genes. This situation has given rise to the use of a number of gene symbols such as *flaFX*, *flaBII*, and *flaAII.3*, which are awkward and violate the established convention of Demerec et al. (1).

A further complexity is that despite the almost complete homology between the flagellar gene systems of the two species, the symbols used for homologous genes are different, so that it is difficult to recall the homologies or to write in a cogent fashion about studies involving both species. The situation is made even more confusing by the fact that, because the same basic symbol is used, a given extended symbol has two quite different meanings; for example, *flaK* in *E. coli* is the symbol for the structural gene for the hook protein, whereas in *S. typhimurium* it is the symbol for one of the master regulatory genes.

It has been clear for some time to those working on the flagellar genetics of *E. coli* and *S. typhimurium* that the nomenclatures would benefit from major revision, in order to unify them and bring them into compliance with convention. By now, the flagellar systems of both species have been extensively analyzed by classical genetics, molecular genetics, and DNA sequence analysis. The number, order, and homologies of the genes have been largely established. Although it is possible that a few more genes remain to be discovered, it seems likely that the vast majority are known. Now, therefore, seems the appropriate time to make the change.

Rather than patch up the existing nomenclature of gene symbols, we thought it better to abandon entirely the symbols *fla* and *flb* (and special examples such as *H1*), so that the symbol used in a paper in the literature will immediately and unambiguously identify whether it was written under the old or the new nomenclature. We considered the use of a nomenclature system in which the gene symbols would be evocative of specific function but rejected this approach for several reasons, the principal one being that the considerable number of genes of unknown function would have to remain under the old nonunified (and in many instances illegal) nomenclature until their functions were established. We therefore have elected to use a straightforward nomenclature, based on genome order, that can be implemented immediately for all known flagellar genes.

NEW NOMENCLATURE

The base symbol in all cases starts with *fl*, and its third letter is determined by the relevant flagellar region within the chromosome (see reference 2 for a description of these regions). Thus, genes in flagellar region I are assigned the symbol *flg* (for flagellum), those in region II are assigned the symbol *flh*, those in region III are assigned the symbol *fli*, and those in the phase 2 flagellin region of *S. typhimurium* are assigned the symbol *flj*. Genes are given alphabetical extensions (*flgA*, *flgB*, etc.) on the basis of genome order.

The new symbols are given in Table 1, along with the old symbols for both species and an indication of function where that is known.

The nomenclature just described has been incorporated into edition VII of the linkage map of *S. typhimurium* (see preceding article [3]) and will be incorporated into edition VIII of the linkage map of *E. coli*, which will be published in *Microbiological Reviews* in the near future (B. Bachmann, personal communication). We strongly urge all authors to adopt it immediately.

Other Recommendations. (i) In the event that additional genes are discovered, they should be given the next available extension for the appropriate flagellar region, regardless of

* Corresponding author.

TABLE 1. Old and new symbols for the flagellar genes of *E. coli* and *S. typhimurium*[a]

Old symbol		New symbol	Function
E. coli	*S. typhimurium*		
Region I			
flaU	*flaFI*	*flgA*	Unknown
flbA	*flaFII*	*flgB*	Unknown
flaW	*flaFIII*	*flgC*	Basal-body protein
flaV	*flaFIV*	*flgD*	Basal-body rod modification
flaK	*flaFV*	*flgE*	Hook protein
flaX	*flaFVI*	*flgF*	Basal-body rod protein
flaL	*flaFVII*	*flgG*	Basal-body rod protein
flaY	*flaFVIII*	*flgH*	Basal-body L-ring protein
flaM	*flaFIX*	*flgI*	Basal-body P-ring protein
flaZ	*flaFX*	*flgJ*	Unknown
flaS	*flaW*	*flgK*	Hook-associated protein 1
flaT	*flaU*	*flgL*	Hook-associated protein 3
Region II			
flaH	*flaC*	*flhA*	Unknown
flaG	*flaM*	*flhB*	Unknown
flaI	*flaE*	*flhC*	Regulation of gene expression
flbB	*flaK*	*flhD*	Regulation of gene expression (flagellum-specific σ factor?)
Region III			
flaD	*flaL*	*fliA*	Regulation of late gene expression
–[b]	*nml*	*fliB*	N-methylation of lysine residues in flagellin
hag	*H1*	*fliC*	Flagellin (filament structural protein); phase 1 flagellin gene in *S. typhimurium*
flbC	*flaV*	*fliD*	Hook-associated protein 2
flaN	*flaAI*	*fliE*	Unknown
flaBI	*flaAII.1*	*fliF*	Basal-body M-ring protein
flaBII	*flaAII.2*	*fliG*	Motor switching and energizing
flaBIII	*flaAII.3*	*fliH*	Unknown
flaC	*flaAIII*	*fliI*	Unknown
flaO	*flaS*	*fliJ*	Unknown
flaE	*flaR*	*fliK*	Hook length control
flaAI	*flaQI*	*fliL*	Unknown
flaAII	*flaQII*	*fliM*	Motor switching and energizing
motD	*flaN*	*fliN*	Motor switching and energizing
flbD	*flaP*	*fliO*	Unknown
flaR	*flaB*	*fliP*	Unknown
flaQ	*flaD*	*fliQ*	Unknown
flaP	*flaX*	*fliR*	Unknown
Phase 2 flagellin region			
–[c]	*rh1*	*fljA*	Repressor of phase 1 flagellin gene of *S. typhimurium*
–[c]	*H2*	*fljB*	Flagellin (filament structural protein); phase 2 flagellin gene in *S. typhimurium*
–[c]	*hin*	*hin*[d]	Regulation of flagellin gene expression by site-specific inversion of DNA

[a] For a more extensive discussion of these genes and their functions, including the relevant references, see reference 2.
[b] –, Not established for *E. coli*.
[c] –, Not present in *E. coli*.
[d] The symbol *hin* has been retained for this gene, since it is a member of a family of inversion-stimulated recombinases and follows a nomenclature established for these.

genome order. In the event that what was believed to be a single locus turns out to be multiple, either the original extension should be retained for one of the genes and the next available extensions given to the others, or the original extension should be abandoned and new extensions given to all of these genes. As was mentioned above, we do not expect that there will be many such additions. The practice used in the past—that of giving illegal Roman and Arabic numerical extensions to an alphabetical extension—should be avoided completely.

(ii) We recommend the retention of Fla$^-$ as the abbreviation for nonflagellate phenotype, regardless of the symbol of the gene containing the mutation responsible for that phenotype. Likewise, we recommend the retention of *fla* as a collective symbol for flagellar genes, regardless of region.

(iii) In spoken usage, we recommend that *flg* be pronounced "flag," *flh* be pronounced "fluh" (with the "uh" being the underemphasized, or schwa, vowel as in "locus"), *fli* be pronounced "fly," and *flj* be pronounced "flaj" (with the "j" as in "flagellum").

(iv) In circumstances when the flagellar genes of both *E. coli* and *S. typhimurium* are under discussion and when it is important to distinguish which species contains the gene being referred to, we recommend the use of small capital E or

s as a subscript following the gene symbol, as has been done extensively in the recent past (see, e.g., reference 2). This subscript is not to be regarded as a formal part of the gene symbol.

LITERATURE CITED

1. **Demerec, M., E. A. Adelberg, A. J. Clark, and P. E. Hartman.** 1966. A proposal for a uniform nomenclature in bacterial genetics. Genetics **54**:61–76.
2. **Macnab, R. M.** 1987. Flagella, p. 70–83. *In* F. C. Neidhardt, J. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
3. **Sanderson, K. E., and J. R. Roth.** 1988. Linkage map of *Salmonella typhimurium*, edition VII. Microbiol. Rev. **52**:485–532.