

Recommendations regarding *Neurospora* genetic nomenclature

D. D. Perkins - Department of Biological Sciences, Stanford University, Stanford CA 94305-5020

Proposals are made that concern (1) distinguishing (in gene symbols) between locus numbers and numbers that describe a product or phenotype, (2) numbering of gene loci, (3) priority of gene names and symbols, and (4) symbolism for heterokaryons.

The four recommendations that follow are being considered for adoption in forthcoming revisions of *N. crassa* maps and of the compendium of markers (Perkins *et al.* 1982 Microbiol. Rev. **46**:426-570). These complement established usage as outlined by Barratt *et al.* 1968 Neurospora Newsl. **8**:23-24 and summarized in the Compendium. I would be glad to have comments.

Recommendation 1. *When the gene name contains a number that is necessary for identifying the product or phenotype, it is proposed that the product-identifying number be included as an integral part of the base symbol, with digits unseparated from the letters by a hyphen.*

Conventions for naming and symbolizing genes in *Neurospora* are derived from usage in *Drosophila* and do not conform with bacterial conventions, which they predate. Different loci with the same base symbol are identified by number. The only digits included in most gene symbols are those that specify the locus number, which is separated from the alphabetical portion of the symbol by a hyphen, as for example *arg-1*, *arg-2*, *arg-3*. No change is proposed in this way of designating loci. Superscripts are used when it is necessary to distinguish different alleles at the same locus, for example *frq¹*, *frq²*, *frq³*; *cnr^R*, *cnr^S*; or *a^{m1}* *a^{m33}*.

The situation becomes complicated when the gene to be named specifies a product requiring a number for its identification. (The histones provide an example.) One way to avoid confusing gene-product numbers with locus-numbers would be to integrate the product number into the unhyphenated base symbol as a digit or digits. For example, chromosomal genes specifying different members of the protein import complex of the mitochondrial outer membrane would be symbolized *mom19*, *mom22*, etc., rather than *mom-19*, *mom-22*. Similarly, *nuo78* would be used rather than *nuo-78* for the nuclear gene specifying the 78 kilodalton subunit of the respiratory chain NADH dehydrogenase (Complex I) in the inner mitochondrial membrane. Histone-specifying genes might be symbolized *hH1*, *hH2*, *hH3*.

In addition to ensuring that numbers describing the product or phenotype not be mistaken for locus numbers, a further advantage of omitting the hyphen in this situation is that doing so makes it possible to use the same base symbol for other loci, as might be desired for members of gene families or sometimes for genes with products that form a functional complex. Genes that specify heatshock protein HSP70 provide an example. Two of these at different loci have been designated *hsps-1* and *hsps-2*, with *s* substituting for the number 70. Under the present proposal,

the symbols would have been *hsp70-1* and *hsp70-2*, which are more explicit and which would leave the way open for unambiguously symbolizing a gene such as *hsp60*.

Recommendation 2. *Hyphens should not be used in gene symbols except when followed by locus numbers. Strain identification numbers should not be converted into locus numbers. Different loci bearing the same gene symbol should preferably be numbered in serial order beginning with one.*

This recommendation reaffirms what is already normal practice. According to convention, arbitrary strain numbers and isolation numbers should not be set off from the base symbol by a hyphen and should not be adopted as locus numbers at the time mutations are mapped and new loci are defined. If the isolation number of an unmapped mutant is to be displayed, the number is normally placed in parentheses following the base symbol. The presence of a hyphen in the symbols of unmapped mutants could be taken to imply that two mutations are at different loci when in fact the mutants have not been tested for allelism. An example of the confusion that may arise from inappropriate use of hyphens is provided by the history of *aod* symbols referenced in the 1982 Compendium. Four unmapped *alternate oxide deficient* mutants were first designated *aod-1*, -2, -3, and -4. Then the *aod-1*, -2, and -3 mutations proved to be alleles and were all assigned to a locus called *aod-1*, while the fourth, originally designated *aod-4*, was assigned to another locus, called *aod-2*.

When only one locus is known that bears a particular name and base symbol, the number -1 is unnecessary and its use is optional. Thus, the only known *inositol* locus is symbolized *inl* rather than *inl-1*.

Locus numbers will ordinarily be assigned sequentially as additional genes with similar phenotypes and the same base symbol are shown to be nonallelic. If these conventions had been followed, *mod-5* would have been called *mod-1* or *mod*, and the first three *blue-light-inducible* loci would have been called *bli-1*, *bli-2*, *bli-3* rather than *bli-3*, *bli-4*, *bli-7*.

Recommendation 3. *Gene names should not be changed without a compelling reason.*

Priority is established by the name that is used for a gene locus when its map location is first published. For example, an auxotrophic mutation tentatively called *ace-6*, which uses acetate or succinate, was shown to be allelic with *suc*, which had been named and mapped long before. Because of priority, the *ace-6* designation was then dropped in favor of *suc* (Kuwana and Okumura 1979 Jpn. J. Genet. **54**:235-244). When *grg-1* and *ccg-1* were found to be alleles, only the latter had been mapped. It was therefore agreed to use the symbol *ccg-1* for the locus. For the same reason, the locus name *eas* was retained in preference to *ccg-2* and *bli-7* when the latter were both shown to be *eas* alleles.

The temptation will always exist to give a more precise name to a gene when new information is obtained, but unless restraint is exercised, synonyms will proliferate and confusion is likely to result. For example, when the linkage group VI gene *Bml* was shown to specify beta-tubulin, a proposal was made to rename the locus *tub-2*, which corresponds to the homologous gene in *Aspergillus* (Orbach *et al.* 1986 Mol. Cell. Biol. **6**:2452-2461). However, the original name

Benomyl-resistant is still descriptive and accurate. It therefore seems wise to avoid change and to continue using *Bml*.

In exceptional situations it may be necessary to change a gene name because the original name was incorrect or misleading. The reason for making the change should then be stated clearly, as was done in changing *met-4* to *cys-10* (Murray 1965 *Genetics* **52**:801-808) and *ol* to *cel* (Henry and Keith 1971 *J. Bacteriol.* **106**:174-182). Rarely, priority may be overridden when belated recognition of allelism would result in replacing an apt name that was widely used and widely published with another, earlier, name that was obscurely published and had been little used. This consideration led Catcheside to propose abandoning *mts* in favor of *cpc-1* (1991 *Fungal Genet. Newsl.* **38**:71). Priority was overridden on another occasion when the original symbols for several amino acid auxotrophs were changed to conform with international biochemical usage (*aspt* to *asp*, *tryp* to *trp*, etc.; Perkins and Barratt 1973 *Neurospora Newsl.* **20**:38).

The temperature-sensitive conditional mutants designated *un* (unknown function) present a special case. When a function or product is identified for one of these, *unknown* is no longer applicable and a new name should be assigned. For example, the temperature-sensitive mutation in strain 55701t, called *un-3*, was shown by Kubelik *et al.* (1991 *Mol. Cell. Biol.* **11**:4022-4035) to be allelic with a temperature-sensitive cytochrome mutant that had been named *cyt-20*. It was proposed therefore to change the name of the gene from *un-3* to *cyt-20*. (Strain 55701t was found also to contain another, separable mutation which confers resistance to ethionine and defines a new locus, *eth-2*, closely linked to *cyt-20*.)

Synonyms will be cross-referenced when the Compendium is revised.

Recommendation 4. *In designating a heterokaryon, it is proposed that genotype symbols for the component nuclei be separated only by a plus sign and that the entire heterokaryon be enclosed as a unit in parentheses.*

This departs from the usage of Barratt *et al.* 1964 *Neurospora Newsl.* **8**:23-24, who proposed that the genotype of each component be enclosed separately in parentheses. If the present recommendation is followed, a heterokaryon of *col-2* with the *a^{ml}* helper strain will be written (*col-2 A + a^{ml} ad-3B cyh-1*) rather than (*col-2 A*) + (*a^{ml} ad-3B cyh-1*). Enclosing both components together rather than each separately is simpler. It also conforms to inclusion of the two nuclear types in the same cytoplasm.