

IX. A STANDARD FOR MAIZE GENETICS NOMENCLATURE
(From MNL 69:182-184 (1995), as updated Sep 1996; Apr 2000; Apr 2002; Oct 2006.)

PREAMBLE: We wish to have a system that is consistent, compatible with the historical background of maize genetics (insofar as these two goals can be reconciled), is easily understood by plant geneticists working with other species, and forms the basis for the importation of maize data into a general plant genetics database so that the basic knowledge concerning maize genes is available to researchers with other species and vice versa. We believe that this goal is best implemented by the researchers in each species having their own working vocabulary, while the identification of genes that catalyze the same functions in all species should rely on entry into a relational database of the genes' function as an E.C. number (2.4.1.13), trivial name (sucrose synthase), and systematic name (UDPGlucose:D-fructose 2-glucosyltransferase). The situation can be less completely categorized for genes whose products are transcription factors, structural proteins, storage proteins, etc.

If one accepts the premise outlined above that the common ground between species need not reside in the working vocabulary of geneticists using any species as a model system but in the manner in which their data are expressed in the database, then the previously adopted names for maize genes can be retained. It will not be necessary to rename the genes previously named on the basis of the mutant phenotype produced as soon as the function of the nonmutant alleles becomes known, but we should proceed to define more precisely words or terms whose meanings need clarification and to decide how we wish to deal with the new information becoming available.

1. **DEFINITIONS:** The words "locus" and "gene" should not be treated as synonymous. A locus can be defined as "a chromosomal site of variable size at or within which is located a gene, a restriction site, a knob, a breakpoint, an insertion, or other distinguishable feature". This necessitates specifying whether we mean a gene locus or an RFLP locus, etc. We can then define a plant gene as "a DNA sequence of which a segment is regularly or conditionally transcribed at some time in either or both generations of the plant. The DNA is understood to include not only the exons and introns of the structural gene but the cis 5' and 3' regions in which a sequence change can affect gene expression". This treats the gene as a functionally defined entity that is not circumscribed by the transcribed region or other fixed limits.

2. **ANONYMOUS TRANSCRIPTS¹:** For most of the history of genetics, the existence of a gene was recognized when a mutation occurred, and the gene was then named by a word/term that was descriptive of the mutant phenotype. That will continue to be the practice except with isozyme markers, for which the designation will be the enzyme in question, or the instances in which the biochemical lesion responsible for the mutant phenotype is identified before the locus is reported. The loci of these genes have then been placed on chromosome maps in relation to other mapped loci. However, we now have the possibility of recognizing genes in which no mutation has been detected through the construction of cDNA libraries. These anonymous cDNAs are often used as probes in RFLP mapping. When such a probe hybridizes to a single band, it is clear that the RFLP loci circumscribe the transcriptional unit that encodes the message represented by the cDNA, and these RFLP loci with other RFLP loci can be used as the basis for mapping the gene. Mapping a locus in this fashion is encouraged as a means of obtaining maximum coverage of the genome. As long as the locus retains an anonymous status (unknown function or no mutant phenotype), the symbol for the locus should be assigned according to the convention used for RFLP loci (as *umc148*, see Section 8). Further information about the probe and its derivation is best provided in tabular or database form rather than in the symbol itself.

A gene name identifying function for a locus detected with a cloned sequence should be given only when there is unambiguous evidence that this is the site by which that function is encoded. Particular caution should be taken in identifying genes (and their function) from several RFLPs hybridizing to a gene-specific probe from another organism. Until a sequence has been shown to encode the function in question, the gene designation should be that of an RFLP locus (see Section 8).

3. **STANDARD NOMENCLATURE AND SYMBOLS:** The names and symbols that have been used for maize genes should be retained. The name and symbol of a gene locus should be represented with lower-case, italic characters (*defective kernel12*, *dek12*). Note that no hyphen separates the gene name from a numerical suffix, which is a change from previous usage. We use a hyphen in the case of mutant alleles to separate the allele designation from a suffix specifying the particular allele (see Section 5). We advocate strongly that all genes identified in the future be given a three letter symbol. Newly detected maize genes that have been previously identified in other plant species should be named where appropriate (see the last paragraph in Section 2) with reference to the list of generic names compiled by the Commission on Plant Gene Nomenclature.

When designating homozygous genotypes with two or more unlinked genes, the genes are separated by semicolons, e.g., *a1;a2;c1;c2;r*. If linked, the genes are separated by spaces, e.g., *C1 sh1 bz1 Wx1*. Heterozygous genotypes should be written with a slash separating the sets of linked genes, e.g., *C1 Bz1/c1 bz1*. If the genes are unlinked, the proper designation is *Sh2/sh2; Bt2/bt2*.

4. **LOCI WITH THE SAME GENE NAME:** Where we have more than one nonallelic mutant with the same gene name, the earlier recommendation was that the first one to receive that name should not have a numerical suffix but the second has 2 as a suffix. Thus we have *shrunken* (*sh*), *shrunken2* (*sh2*), and *shrunken4* (*sh4*) mutants. Geneticists outside the maize community are apt to misinterpret this convention. We recommend that we be consistent and write *shrunken1* or *sh1* and advocate that even if a new locus is identified and given a unique name, it be designated as 1. This has the definite advantage in maintaining databases and indices that no retrospective correction would be necessary if a second gene locus receives the same designation.

5. **ALLELIC DESIGNATIONS²:** Where a mutant allele is recessive, it should be designated by an italicized symbol (lower case) as *dek12*, which is the same as the symbol of the locus. Since it is unlikely that any two mutant or nonmutant alleles in a highly polymorphic

species such as maize have identical sequences, maize geneticists are encouraged to specify the particular allele with which they are working (see in this Section, Alleles of Independent Mutational Origin and Designation of Nonmutant Alleles). The symbol for dominant, nonmutant (i.e., conditioning a normal phenotype) alleles will be the same italicized three letter symbol as the mutant alleles but with the first letter capitalized (*Dek12*). The symbol of the gene product should not be italicized and should be written with all letters capitalized (e.g., ADH1). The name of the gene product (alcohol dehydrogenase) should neither be capitalized nor italicized.

When the mutant alleles of a gene are dominant, the first letter of the mutant symbol is capitalized. The nonmutant symbol has all the letters lower case. For example, the *corn grass1* (*cg1*) gene locus has several dominant mutant (*Cg1*) alleles as well as nonmutant (*cg1*) alleles. The reference mutant allele is designated as *Cg1-R* or *-1*.

Codominant alleles such as isozymes where the variants are functional and distinguished from each other by electrophoretic mobility, should be designated by symbols with the first letter capitalized and identified by allelic specifications as *Pgm2-5* or *Pgm2-7*.

ALLELES OF INDEPENDENT MUTATIONAL ORIGIN: The unambiguous designation of mutant alleles that have arisen as independent mutational events is increasingly important. It is generally understood that a gene symbol followed by a hyphen plus a letter or number(s) specifies a particular recessive allele at that gene locus. We have referred to the mutation by which the gene was identified as the reference allele; e.g., *bz1-Ref* or *bz1-R*. It is equally appropriate to refer to that allele as *bz1-1*. The mutations in any gene that were identified subsequently have been categorized in various idiosyncratic ways. Alleles that have arisen by independent mutational events have been designated by letters, numbers, a letter plus numbers, the name of the inbred in which the mutation occurred, and sometimes all of these applied to a group of alleles at a gene locus. While all of these designations served the purpose of indicating that these alleles had independent mutational origins, there is a clear advantage to greater standardization. As in the 1973 Nomenclature Standard, it is recommended that new alleles be identified by a laboratory number that might indicate the year of isolation as *sh2-6801*. This has the definite advantage that two laboratories are unlikely to designate two new mutations of the same gene by the same number. However, if two laboratories are targeting the same locus in mutagenesis experiments, they should consult before naming their new alleles to avoid giving the same designation to different alleles. Also recommended is the convention of referring to a new mutation of a given phenotype by a provisional designation as bt-lab number until it is ascertained whether the mutant is a new allele of a known gene or identifies a previously unidentified gene. In the first instance, the proper gene symbol (*bt1* or *sh2*) replaces *bt*, but the lab number is retained (e.g., *bt1-8711*). In the second instance (a previously unidentified locus), a new gene name and symbol would be selected, and this mutant would become the reference allele (*-R* or *-1*).

When mutant alleles are referred to in the generic sense without specification of their origin, a hyphen without further designation (e.g., *bz1-*, *dek12-*) is desirable to make it clear that one is referring to an allele or alleles, not the gene locus.

DESIGNATION OF NONMUTANT ALLELES: Since it is now apparent that in a species as polymorphic as maize, nonmutant alleles from different sources are apt to have a number of sequence differences one from the other, and these differences can be reflected in gene action (nonmutant isoalleles), it is desirable to specify the nonmutant allele being investigated or used as a control. Incorporating the name of the inbred as part of the allelic designation, *Bz1-W22*, is an appropriate method of doing this. However, mutant alleles should not be designated by the inbred in which they arose (e.g., *bz1-W22*) to avoid confusion with the progenitor allele. Also, there may eventually be numerous mutant alleles of a particular gene isolated in that inbred if a researcher uses that inbred in a mutagenesis experiment. A particular nonmutant allele may be found in an exotic race or other accession that is not an inbred. A unique designator (e.g., a PI number or Bolivia #) should be part of the allelic designation.

RFLPs AND RAPDs AS ALLELES: The presence or absence of a restriction site or a primer-amplifiable sequence at a particular locus represent Mendelian alternatives. They fall under the broadest definition of an allele, and it is appropriate to refer to these alternatives as alleles as has already been done in some reports.

6. **NAMING DELETIONS:** When it is clear that a mutation results from a deletion that has removed all or part of two gene loci, it would be appropriate to indicate this in the following manner. For *an1-6923*, this would be *def(an1..bz2)-6923*, and for *sh-bz-X2*, *def(bz1..sh1)-X2*. When molecular evidence indicates that a deletion has removed all of the structural portion of a gene as is true of *wx1-C34*, it should be indicated in the same manner; i.e., *def(wx1)-C34*.

7. **MUTATIONS RESULTING FROM TRANSPOSABLE ELEMENT INSERTIONS:** There is one further point concerning allelic specification. Maize in particular has many mutable alleles resulting from the insertion of a transposable element. These have been designated by the mutant symbol, a hyphen, a lower case "m", and an isolation number; e.g., *wx-m1*. When the transposable element insertion [*Ac*, *Ds*, *Spm(En)*, *dSpm(l)*, *Mu1..MuX*, etc.] is known, it is suggested that this be indicated by a double colon following the allele as *wx-m1::Ds1*. Since a maize stock may have more than one transposable element family active at the same time, firm genetic and/or molecular evidence is necessary to ascribe mutability to a particular transposable element family. Further, mutable alleles generate both stable nonmutant and stable mutant alleles when the transposable element excises from the gene locus. Since the mutant derivatives are certain to differ in sequence from the nonmutant progenitor allele around the site of the transposable element insertion and the nonmutant derivatives are very likely to differ at that site, researchers should be certain to indicate the origin of such alleles in their reports. One means of doing this is to indicate such an origin by an apostrophe following the locus symbol as *Bz1'-7801* or *bz1'-8905*. The specifics of its origin including the transposable element involved could then be included in the text and entered in the Maize Genome Database. Since transpositions of a transposable element from a site within a gene often insert in locations where they have no phenotypic effect but can be useful markers, it is desirable to have a standard to refer to such insertions. Designate them as RFLP's would be designated (see Section 8), but follow the institutional symbol and number with a double colon and the symbol of the transposable element (e.g., *dnap2094::Ac*).

8. **NAMING RFLPs AND RAPDS:** In naming RFLPs and RAPDs, use a lower case three or four letter code designating the originating university or company followed by a laboratory number (no space between the code and the number). When the probe used is a cDNA or a subclone of a gene, the gene symbol should be added in parentheses after the RFLP locus designation, as *umc000(a1)*. Since a probe not infrequently recognizes RFLPs on two or more chromosomes, these should be designated by the same institutional code, number, and probe followed immediately by A, or B, or C. In so far as possible, the locus with the strongest hybridization should be designated A and the more weakly hybridizing loci be designated B, C etc. in descending order of signal strength.

9. **CHROMOSOME REARRANGEMENTS:** The conventions for dealing with chromosomal rearrangements are well established and adequate for the purpose. To designate particular reciprocal translocations as T1-2a or T1-9(4995) etc. with the breakpoints noted parenthetically or in a table of supporting information is explicit and sufficient. Additional information (the fact that the translocation stock is homozygous for *wx1*) can be incorporated by prefacing the translocation number with the gene symbol as the Coop does in its stock lists (e.g., *wx1* T1-9c). Translocations with B chromosomes have designations that indicate the arm of the A chromosome involved (L or S) as well as a lower case letter distinguishing that translocation from any others involving that particular chromosome arm, as TB-5Sc. The cytological breakpoint in the A chromosome as well as the loci uncovered when the TB translocation is used as a male parent can be noted in the text or in a table of supplementary information. The designations for inversions (e.g., *Inv9b* again with the breakpoints, 9S.05-L.87, listed in a supporting table) are succinct and convey the necessary information.

10. **ORGANELLAR GENES:** For chloroplast and mitochondrial genes, we accept for the present the proposals already in place. For chloroplast genes, this is Hallick and Bottomley, *Plant Mol. Biol. Rep.* 1(4):38-43, 1983, as updated at SwissProt <<http://www.expasy.ch/cgi-bin/lists?plastid.txt>> or by the Chloroplast working group for the Commission on Plant Gene Nomenclature. For mitochondrial genes, this is Lonsdale and Leaver, *Ibid.* 6(2):14-21, 1988, updated by the Mitochondrion working group for the Commission on Plant Gene Nomenclature. For brevity's sake, these are not summarized here.

11. **TRANSCRIPTION FACTORS:** (Oct 2006 addition) We define here TFs as proteins that contain a DNA-binding domain and that fall within one of the families described in <http://arabidopsis.med.ohio-state.edu/AtTFDB/>.

There is currently no coherent effort in maize for a rational and organized naming of transcription factors (TFs). The use of GenBank accession numbers, EST names or locus identifiers provides an impractical mechanism, which often leads to ambiguities, for example because of multiple entries in GenBank or of several ESTs for the same protein. Thus, we propose here to create a uniform nomenclature for maize TFs, following the lead from *Arabidopsis*. A similar proposal is being adopted by the TIGR rice annotation group and by the SUCEST-FUN sugarcane annotation group.

Recommendation: GENE PRODUCTS: Each transcription factor will have an organism identifier (*Zm*) to be used only in the context of other organisms, followed by letters that represent the TF family (e.g., MYB, bHLH, HD, bZIP) and by a number that will start with '1'. A similar strategy is currently being applied to other maize gene families (e.g., the kinesins, see Lawrence et al., *J. Cell Biol.* 167:19-22). Since we realize that many TFs are known by their genetic names, this nomenclature will permit the use of synonyms. For example, KNOTTED could be named HD1(KN) (or *ZmHD1(KN)* when being compared to HDs of other species) and C1 would be MYB1(C1) (or *ZmMYB1(C1)*). In addition, whenever possible, we will try to have the numbers provide a historic perspective of which TFs have been first identified. In that regard, since KN and C1 correspond to the founding members of their respective families in maize, they are assigned the number '1'. Prior genetic nomenclature will be incorporated in the database.

GENES: Existing names for genes encoding TFs will not be altered. If necessary, and only as a way to provide coherence with the naming of the gene products, the synonym strategy described above would be used. In that regard, *c1* would continue to be *c1* but could also be cross-referenced as *c1(myb1)*. New genes will be named according to their products. If mutant phenotypes are identified at a later date, gene names derived from mutant phenotypes will be added as synonyms, but the original name will not be changed. As indicated for the gene products, the use of the prefix *Zm* in front of the gene's name will only be used when comparing maize genes with related genes from other species (e.g., *Zm myb1*).

For generating a position for transcription factors, Erich Grotewold served on the Nomenclature Committee in an ad hoc capacity.

12. **CLEARING HOUSE FOR NOMENCLATURE:** We also believe that it is desirable to initiate a clearing house for maize nomenclature so that a researcher wishing to name a recently identified gene can ascertain almost immediately that no one has used the proposed designation and symbol. This clearing house can, in principle, function through the MaizeGDB <<http://www.maizegdb.org/>>, which will be refereed by a cooperator. The same facility could be used to insure that allelic designations are not duplicated or to answer questions concerning nomenclature.

Current Members Include: Hugo Dooner, chair; Tom Brutnell; Vicki Chandler; Curt Hannah; Toby Kellogg; Marty Sachs; Mike Scanlon; Mary Schaeffer (Polacco); Philip Stinard.

Footnotes:

¹ANONYMOUS TRANSCRIPTS: 1996 decision made not to utilize the parenthetic 'gfu' designation for "gene, function unknown". **RATIONALE:** in common usage, the 'gfu' suffix has proven confusing, implying 'known function', especially to researchers from other species. The confusion arises from the practice in RFLP naming to include parenthetic acronyms where sites are detected by probes with an assigned or putative identity with a particular gene product.

²ALLELIC DESIGNATIONS: 1996 decision made to use '-', rather than '+', in designations of non-mutant alleles. **RATIONALE:** use of '+' has met with resistance by journal editors; definition of non-mutant alleles can be a grey area.