e-**X**tra*

Review

Guidelines for Genetic Nomenclature and Community Governance for the Model Legume *Medicago truncatula*

Kathryn A. VandenBosch[1] and Julia Frugoli[2]

[1]Plant Biology Department, University of Minnesota, 220 Bio Sci Center, 1445 Gortner Avenue, St. Paul 55108 U.S.A.; [2]Department of Biological Sciences, Clemson University, 132 Long Hall, Clemson, SC 29634 U.S.A.

At the 2nd *Medicago* meeting (a satellite of the 1999 IS-MPMI meeting in Amsterdam), investigators perceived a need for standardization of genetic nomenclature in *Medicago truncatula*, due to the rapid growth of research on this species in the past few years. Establishment of such standards grew out of discussions begun at this meeting and continued electronically throughout the *M. truncatula* community. The proposed standards presented here are the consensus results of those discussions. In addition to standards for gene nomenclature, a method for community governance and a website for cataloging gene names and submitting new ones are presented. The purpose of implementing these guidelines is to help maintain consistency in the literature, to avoid redundancy, to contribute to the accuracy of databases, and, in general, to aid the international collaborations that have made *M. truncatula* a model system for legume biology.

The annual legume species *Medicago truncatula*, or barrel medic, has been chosen by a growing group of investigators as a model genetic organism for the analysis of development during microbial interactions and other aspects of legume biology (Barker et al. 1990; Cook 1999; Cook et al. 1997; Frugoli and Harris 2001; Harrison 2000). Analysis of gene function by mutation and transgenic expression of genes is ongoing in this species, which is also the topic of genomic analysis in the United States (http://www.tigr.org/tdb/mtgi/, http://www.medicago.org, and http://www.noble.org/medicago/) and Europe (http://sequence.toulouse.inra.fr/Mtruncatula.html). The increasing numbers of genes obtained by cloning and those distinguished by mutation and ecotypic variation will necessitate an organized approach to gene nomenclature in this species.

Realization of this fact led to community discussions by participants at the 2nd *Medicago* meeting in Amsterdam in July 1999. An informal committee was formed to develop a

Corresponding author: K. A. VandenBosch; E-mail: vande102@tc.umn.edu

nomenclature system, and documents proposing systems were circulated electronically in the community in the subsequent months. Further dialog at the Plant and Animal Genome meeting in San Diego in January 2000 and 2001 led to the finalization of this document, the result of those discussions.

To aid investigators, editors, and granting agencies who deal with *M. truncatula* genes, we propose the application of guidelines and standards for naming genes identified by sequence or mutation in this species, as well as genes identified through phenotypic variation in natural populations. The guidelines proposed herein for genes identified by sequence are adopted from those spelled out for sequenced genes by the Commission on Plant Gene Nomenclature (http://mbclserver.rutgers.edu/CPGN/Guide.html). Rules for naming genes identified by mutation are adapted from, and consistent with, those established for *Arabidopsis thaliana.* Meinke and Koornneef (1997) have spelled out the *Arabidopsis* rules, and electronic access to this paper and a summary of nomenclature rules may be obtained through The *Arabidopsis* Information Resource (http://www.arabidopsis.org/info/guidelines.html). Though the proposed *Medicago* guidelines were drawn from the same sources as the recently proposed guidelines for *Lotus japonicus* (Stougaard et al. 1999), another model legume, it should be noted that the two sets of guidelines differ in some specifics and in procedures for gene symbol selection.

Community structure.

We propose several steps to implement this plan. First, it is suggested that the investigator who serves as curator of *M. truncatula* gene nomenclature should be designated by a *Medicago* steering committee to act for a fixed term, which may be renewed. This individual will compile a list of gene names and symbols in use and make the list available via a public webpage. The curator will serve as a clearinghouse for registering new gene symbols and will also mediate conflicting uses of gene names or multiple names for the same gene. Procedures for arbitration are suggested below, as are mechanisms for changing the guidelines. Lastly, steps required for publishing a new gene name are also laid out.

At this time, an interim *M. truncatula* steering committee has been appointed for community guidance and governance. The interim committee includes two individuals from the United States

and two from Europe, to serve until July 2001. Douglas Cook at the University of California-Davis (drcook@ucdavis.edu) and Maria Harrison at the Noble Foundation (mjharrison@noble.org) have agreed to serve as the U.S. representatives to the interim steering committee. Jean Dénarié of INRA-CNRS in Toulouse, France (denarie@toulouse.inra.fr), and Ton Bisseling of The Agricultural University of Wageningen in The Netherlands (ton.bisseling@mac.mb.wau.nl) have agreed to serve as the European representatives. The committee has appointed Julia Frugoli (jfrugol@clemson.edu) to act as curator for gene nomenclature. The steering committee will also act as a scientific advisory committee for international *Medicago* meetings and will provide leadership for the discussion of scientific priorities and goals for *Medicago* research infrastructure, and as well as coordination of U.S. and E.U. efforts.

Both the composition of the steering committee and the guidelines for gene nomenclature are meant to reflect a community consensus. In so far as is possible, we have consulted a wide variety of individuals while drafting the policy outlined below. The next meeting, the 3rd *Medicago* meeting, will occur in July 2001 in Madison, WI, U.S.A., immediately prior to the 10th International Congress on Molecular Plant-Microbe Interactions (http://www.plantpath.wisc.edu/mpmi/satellite.html). At that meeting, amendments to the gene nomenclature guidelines will be considered. The mission of the steering committee and its term of service, as well as the composition of the committee to succeed the interim committee, will also be discussed.

Nomenclature curation and arbitration.

The steering committee is responsible for appointing an individual from the *M. truncatula* community to act as curator of gene nomenclature. The curator will maintain an up-to-date, alphabetical listing of gene symbols that will be made public through internet access. A website for the gene list has been established (http://www.genome.clemson.edu/affiliated_cugi/medicago/), which is also linked through the *M. truncatula* consortium site (http://www.medicago.org). Submission of gene names and other information will be via submission of a form through the webpage or via an e-mail to the curator (jfrugol@clemson.edu). Gene names and symbols published in the past should be submitted to the curator as soon as possible, so a comprehensive list can be made. A sample of relevant information to be submitted is available at the gene list site. Prior to publication, new gene symbols should be reserved with the curator to assure that duplication will not occur (checklist for publication is below). Reserved symbols must be used in print within a reasonable period of time, as determined by the curator, or they may become available to the community again.

Occasionally, two groups of investigators may choose the same symbol to represent two different genes. In this case, the first group to register the symbol for their gene will have precedence for the use of the symbol. Also, it can be expected that multiple genes identified by mutation, and obtained through independent screens, may turn out to be identical genes. In this case, the symbol for the gene published first will stand. In cases in which conflicts arise over the use of gene symbols, the curator will encourage the participants to resolve the matter by consensus. If the matter cannot be settled by consensus, the curator will consult with the steering committee to resolve the conflict. In publication, if a mutant is known by multiple names, manuscripts should refer to all the names of the gene at one point in the text (abstract, introduction, or methods) with appropriate citation and then use the accepted name alone throughout the rest of the manuscript.

Changes to the *M. truncatula* gene nomenclature guidelines may be made at intervals to accommodate changing needs. Individuals who wish to propose changes to the guidelines should circulate the amendments to members of the research community and post the proposed changes at the *M. truncatula* forum on the World Wide Web (http://www.medicago.org). Proposals to change the guidelines will be acted on at periodic *Medicago* meetings or through electronic interchange.

Format for gene names.

Genes identified by sequence.

Some researchers in the *Medicago* community may wish to give a name to genes identified through sequencing. Investigators desiring to give a name to sequenced genes identified in *M. truncatula* should follow the Commission on Plant Gene Nomenclature (CPGN) guidelines. The Mendel database contains CPGN designations for gene families. The CPGN's guiding principle is that "all genes throughout the plant kingdom that encode the same product will be members of the same gene family, and will therefore be assigned the same Gene Family Number and Gene Family Name." Many gene families have been assigned names and/or numbers, and guidelines have been established for the naming of gene families and their members. The guidelines (http://mbclserver.rutgers.edu/CPGN/Guide.html) and a list of existing names (http://genome-www.stanford.edu/Mendel or its mirror site http://genome.cornell.edu/cgi-bin/WebAce/webace?db=mendel) can be found on the World Wide Web. In brief, names for sequenced genes typically contain a three-letter mnemonic (designating a set of genes whose products have similar function), followed by a character (designating a subtype containing distinct motifs). A capital first letter is used to indicate a nuclear gene, while organelle gene names begin with a lowercase letter, e.g., *His1* for histone H1 or *cyt1* for cytochrome-c oxidase subunit 1.

When a gene from *M. truncatula* belongs to a gene family that has been named, the gene should take the family name designation, followed by a numeral to denote the specific gene. If the gene family has not been named, then a suitable name may be proposed to the CPGN registry according to their recommendations. Once a gene family name has been accepted by CPGN, it should be reported to the *M. truncatula* curator. Names that have been proposed to CPGN and are pending approval may be reserved with the *M. truncatula* curator for future use.

Genes identified by mutation.

These guidelines are adopted from the *Arabidopsis* guidelines (Meinke and Koornneef 1997), which may be consulted for additional details and guidelines for special circumstances.

1. Mutant gene symbols should consist of three lowercase letters written in italics or underlined (e.g., *skl*), although symbols longer than three letters may be considered if they improve clarity.

2. Wild-type alleles should consist of the gene symbol in capital letters written in italics or underlined (e.g., *SKL*).

3. Similarly, the full descriptive names of the alleles should be underlined or italicized and then written in lowercase letters for mutants and capital letters for wild type (e.g., *sickle* or *SICKLE*).

Table 1. Websites to check when choosing mutant gene symbols

Names	Website
Previously used *Medicago truncatula* names	http://www.genome.clemson.edu/affiliated_cugi/medicago/genelist.html
Previously used *Arabidopsis thaliana* names	http://mutant.lse.okstate.edu/genepage/genepage.html
Names of sequenced plant gene families	http://genome-www.stanford.edu/Mendel/ http://genome.cornell.edu/cgi-bin/WebAce/webace?db=mendel
Names of *Sinorhizobium* genes	http://sequence.toulouse.inra.fr/meliloti.html

4. Different genes with the same phenotype, which bear the same three-letter symbol, will be distinguished by a number following the three letters (e.g., *dmi1* versus *dmi2*).

5. Different alleles of the same gene will be distinguished by different numbers following the symbol, locus number, and a hyphen (e.g., *dmi1-1* versus *dmi1-2*). If only one allele is known, the hyphen and allele number are unnecessary.

6. The cultivar Jemalong is designated as the reference wild-type genotype because it is a focus of genomic analysis and mutagenic studies. When a variant is discovered in another ecotype, nomenclature should follow the practice used in *Arabidopsis*. Specifically, genes identified from natural variants among ecotypes, such as responses to Nod factors or pathogens, should be named in the same way as mutations. The rationale behind such an approach is that ecotypes are not necessarily entirely genetically uniform or homogeneous and, therefore, variations should be named under a similar system to mutations. When the source ecotype for an allele is other than 'Jemalong', the publication should indicate the ecotype of origin.

7. Protein products of genes should be written in capital letters without italics (e.g., SKL).

Choice of gene symbols for genes identified by mutation.

The *Arabidopsis* guidelines suggest that "…investigators should be free to choose descriptive names that reflect their perspective on the mutant phenotypes and proposed gene functions." Similarly, investigators choosing names for previously unnamed *Medicago* genes are also free to use a symbol of their own choosing within certain parameters.

1. Investigators must avoid redundant use of any three-letter symbol already used for *M. truncatula* genes of distinctly different phenotypes.

2. If new mutants are sufficiently similar to previously described mutants to imply function in a common pathway or process, it may be appropriate to retain a standard three-letter symbol for the phenotype and simply change the locus number.

3. To avoid confusion, investigators are encouraged to use novel three-letter symbols for genes with novel phenotypes, rather than use the same symbol for genes whose mutant phenotypes differ widely (e.g., use of *sym* to denote genes conditioning a wide variety of symbiotic phenotypes). This is particularly the case when the underlying cellular causes of the phenotypes are known and the mechanisms differ among genes. If investigators do choose to use a common symbol such as *sym*, they must use the 'Mt' prefix to differentiate the gene from similarly named genes in other species.

4. Use of certain types of three-letter symbols should be avoided, including

- symbols already in use for a family of sequenced genes;

- symbols used for genes in another plant species, unless the underlying gene is shown to be orthologous (In the case of orthologous genes, the designation 'Mt' must precede the gene symbol [e.g., Mt*Ein1*] so that it is not confused with the gene from the original plant.); and

- symbols used to designate symbiotic genes or phenotypes of rhizobia (e.g., *fix*, *nod*, *exo*, and *nif*).

Checklist for publication concerning *Medicago truncatula* genes.

1. Choosing the symbol. Choose mutant gene symbols as described previously, making sure that they do not conflict with, and are consistent with, existing symbols. To avoid confusion, there are websites that should be checked to assure that conflicts do not exist (Table 1).

2. Registering the symbol. To facilitate naming, authors should register new gene names/symbols with the gene name curator before publication.

3. Analyzing the phenotype. Phenotypes should be carefully described and compared with existing mutants as much as possible. Whenever possible, more than one mutant allele should be analyzed. Complementation tests should be performed among a collection of similar mutants to determine the number of genes involved and to be sure each putative mutant represents a monogenic trait.

4. Allelism and mapping. An effort should be made to determine a rough map position of each mutation or to perform allelism tests with available mutants having the same or similar phenotype, especially those that map in the same chromosomal region. Recombination data can be used in the case of dominant and gametophytic mutants when complementation tests are not informative. The genetic map of *M. truncatula* will soon be available through the consortium website (http://www.medicago.org) and will be updated regularly.

5. Distributing mutants. Seeds of published mutants must be freely available to other investigators for allelism tests. Without the benefit of a stock center, it is suggested that when feasible, mapping be done first to limit the number of allelism tests required and, thus, not to burden labs unduly with requests for many stocks or large numbers of seed of a single stock. When a stock center is created, we suggest that editors require deposition of seed in the stock center at the time of publication.

ACKNOWLEDGMENTS

LITERATURE CITED

Barker, D. G., Bianchi, S., London, F., Dattee, Y., Duc, G., Esdsad, S., Flament, P., Gallusci, P., Genier, G., Guy, P., Muel, X., Tournerur, J., Denarie, J., and Huguet, T. 1990. *Medicago truncatula*, a model plant for studying the molecular genetics of *Rhizobium*–legume symbiosis. Plant Mol. Biol. 8:40-49.

Cook, D. 1999. *Medicago truncatula*—A model in the making! Curr. Opin. Plant Biol. 2:301-304.

Cook, D., VandenBosch, K. A., de Bruijn, F., and Huguet, T. 1997. Model legumes get the nod. Plant Cell 9:275-281.

Frugoli, J., and Harris, J. 2001. *Medicago truncatula* on the move. Plant Cell 13:458-463.

Harrison, M. 2000. Molecular genetics of model legumes. Trends Plant Sci. 5:414-415.

Meinke, D., and Koornneef, M. 1997. Community standards for *Arabi-dopsis* genetics. Plant J. 12:247-253.

Stougaard, J., Szczyglowski, K., de Bruijn, F. J., and Parniske, M. 1999. Genetic nomenclature guidelines for the model legume *Lotus japonicus*. Trends Plant Sci. 4:300-301.

AUTHOR-RECOMMENDED INTERNET RESOURCES

Arabidopsis genetics, mutant genes: http://mutant.lse.okstate.edu/genepage/genepage.html

Medicago truncatula consortium: http://www.medicago.org

M. truncatula gene nomenclature: http://www.genome.clemson.edu/affiliated_cugi/medicago/

M. truncatula gene nomenclature, gene list: http://www.genome.clemson.edu/affiliated_cugi/medicago/genelist.html

Mendel plant gene nomenclature database: http://genome-www.stanford.edu/Mendel/

Names of sequenced plant gene families: http://genome.cornell.edu/cgi-bin/WebAce/webace?db=mendel

Sinorhizobium meliloti strain 1021 gemone project: http://sequence.toulouse.inra.fr/meliloti.html