118–119 *Nucleic Acids Research, 2001, Vol. 29, No. 1*

Mendel, a database of nomenclature for sequenced plant genes

Carl A. Price* and Ellen M. Reardon

Commission on Plant Gene Nomenclature, Waksman Institute, Piscataway, NJ 08854-8020, USA

Received August 29, 2000; Revised and Accepted October 16, 2000

ABSTRACT

The Mendel database contains names for plant-wide families of sequenced plant genes. The names have either been approved by the Commission on Plant Gene Nomenclature (CPGN), an organization of the International Society for Plant Molecular Biology (ISPMB), or are identified as provisional or temporary names. Mendel also identifies the corresponding genes in individual species of plants. Mendel can be searched through the mirror sites at Cornell (http://genome.cornell.edu/cgi-bin/WebAce/webace?db=mendel) and Stanford (http://genome-www.stanford.edu/Mendel/). In addition, parts of Mendel can be downloaded from the CPGN Web site (http://mbclserver.rutgers.edu/CPGN/).

INTRODUCTION

Since 1991 the Commission on Plant Gene Nomenclature (CPGN) has been developing a common nomenclature for all sequenced plant genes, nuclear as well as organellar. The Mendel database lists names for plant-wide families of genes (Gene Families) approved by the CPGN (http://mbclserver.rutgers.edu/CPGN/). A downloadable version of Mendel is available on the CPGN Web site. The Mendel mirror sites at Cornell (http://genome.cornell.edu/cgi-bin/WebAce/webace?db=mendel) and Stanford (http://genome-www.stanford.edu/Mendel/) are on the ACeDB platform and have searchable lists in multiple formats.

In the absence of a common nomenclature, individual laboratories tend to adopt idiosyncratic names for the genes that they isolate. For example, the CPGN recently approved the name *GlbS* for one of four families of genes encoding plant globins or hemoglobins, but *GlbS* genes are identified in the literature by seven separate mnemonics, including *Glb*, *L*, *Lb* and *Lbg*. The advantages to scientific communication in having a common genetic language are obvious.

GUIDE TO THE NOMENCLATURE

A guiding principle of the CPGN nomenclature 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 to the same Gene Family. Gene products are considered to be the same when they have the same function and have similar sequences. We find product sequences typically to be ~80–90% similar within a gene family. The identification of individual gene families and the criteria for distinguishing related families are the responsibility of the working groups.

The names of gene families are typically of the form *XyzN* (or *XYZN*). Sets of genes whose products have similar functions but whose sequences contain distinct motifs may be represented by numbers or letters after a shared mnemonic; e.g. *Glb1*, *Glb2*, *GlbC*, *GlbS* for the four families of globin genes. Organellar genes follow the bacterial system with the first letter in lowercase; e.g. *atpA*, *rbcL*, *cox3*. Additional characters required for the mnemonic or to distinguish one gene family from another may be employed; we currently have gene families with four-letter mnemonics; e.g. *Atpv6* (ATPase vacuolar, subunit 6), *Dhps1* (dihydropicolinate synthase), *Lhcb4* (light-harvesting complex, type II, CP29).

Individual genes within a species are further identified by a five-letter abbreviation introduced by SWISS-PROT designating the plant species. To avoid problems with alphabetical searches, it is important that the species identification be treated as a separate field; e.g. *ARAth;Glb1*, not *ARAthGlb1* (or *AtGLB1*).

As multigene families are very common in plants, members of multigene families within a species are designated by a numeral that is also treated as a separate field. Members 1 and 2 of *Glb1* in *Arabidopsis thaliana* are identified by *ARAth;Glb1;1* and *ARAth;Glb1;2* and the 10 members of *GlbS* in *Medicago sativa* are identified by *MEDsa;GlbS;1* … *MEDsa;GlbS;10*.

Designations of alleles follow the appropriate procedure for the relevant plant species. Mutant alleles are usually represented in lower case with the allelic designation separated by a hyphen; e.g. a wild-type gene in maize encoding alcohol dehydrogenase is designated *ZEAma;Adh1;1*, and the *C-m* allele is designated *ZEAma;Adh1;1-C-m*.

Temporary designations using the initial letter *Y* can be assigned on request to conserved gene families or ORFs whose functions are not yet known. Examples are the *Ypr* mnemonic for genes encoding *p*athogenesis-*r*elated proteins, the *ycf* mnemonic for *c*hloroplast open-reading *f*rames and *ymf* for *m*itochondrial open-reading *f*rames.

Further details of the CPGN nomenclature are described in (1), and an updated A Guide to Sequenced Plant Genes is posted on the CPGN Web site (http://mbclserver.rutgers.edu/CPGN/Guide.html).

SEARCH TERMS

Mendel may be searched at the mirror sites by the following terms:

*To whom correspondence should be addressed. Tel: +1 760 942 6050; Fax: +1 732 445 5735; Email: cpgn@mbcl.rutgers.edu

- Gene Family: plant-wide families of genes.
- Gene Name: names for genes in individual species of plants.
- Gene Synonyms: alternate terms used for the genes.
- Gene Product: the multiple terms used to describe the gene product.
- DNAseqAC: accession number in nucleotide databases.
- ProtSeqAC: accession number in protein databases.
- Product Family: sets of plant genes sorted on the basis of sequence similarity of the gene products.
- Working Group: the names of scientists responsible for naming various groups of gene families.

SAMPLE LISTINGS IN MENDEL

A typical search of Mendel would be to find the CPGN-approved name knowing the EMBL-GenBank accession number, the gene product, or a name from the literature (Gene Synonym), etc. A search for *Cab9*, for example, yields two genes in different gene families. Clicking on these provides further data. We should note that frustration with the disparate meanings of the *Cab* mnemonic led to the first set of CPGN-approved gene families (2). The two listings for *Cab9* are:

Gene name: *LYCes;Lhcb5;1*

GeneFamily: *Lhcb5*
GeneProduct: light-harvesting complex type I CP29, …
GeneSynonym: *Cab9*
MendelNumber: 308
Species: *Lycopersicon esculentum*
MemberNumber: 1
DNAseqAC: X61287 [EMBL|GenBank]
DNAseq_Description: *L. esculentum Cab9* gene for type I (26 kD) CP29 polypeptide
ProtSeqAc: Q00321 [SwissProt]
SwissProt_Description: CHLOROPHYLL A-B BINDING PROTEIN OF LHCII TYPE I PRECURSOR (CP29).

Gene name: *PISsa;Lhcb1;5*

GeneFamily: *Lhcb1*
GeneProduct: light-harvesting complex type I LHCII, …
GeneSynonym: *Cab9*
MendelNumber: 995
Species: *Pisum sativum*
MemberNumber: 5
DNAseqAC: M86906 [EMBL|GenBank]
DNAseq_Description: Pea (subclone AB9) *Cab9* gene, 3′-end.
ProtSeqAc: Q41007 [SwissProt]
SwissProt_Description: CHLOROPHYLL A/B-BINDING PROTEIN (FRAGMENT).

Most of the items—GeneFamily, … Species—contain internal links to further information or, shown in square brackets, links to external databases; e.g. [SwissProt].

REFERENCES

1. Price,C.A., Reardon,E.M. and Lonsdale,D.M. (1996) A guide to naming sequencing plant genes. *Plant Mol. Biol.*, **30**, 225–227.
2. Jansson,S., Pichersky,E., Bassi,R., Green,B.R., Ikeuchi,M., Melis,A., Simpson,D.J., Spangfort,M., Staehelin,L.A. and Thornber,J.P. (1992) A nomenclature for the genes encoding the chlorophyll *a*/*b*-binding proteins of higher plants. *Plant Mol. Biol. Rept.*, **10**, 242–253.