

The Naming of Names: Guidelines for Gene Nomenclature in *Marchantia*

John L. Bowman^{1,*}, Takashi Araki², Mario A. Arteaga-Vazquez³, Frederic Berger⁴, Liam Dolan⁵, Jim Haseloff⁶, Kimitsune Ishizaki⁷, Junko Kyojuka⁸, Shih-Shun Lin⁹, Hideki Nagasaki¹⁰, Hirofumi Nakagami¹¹, Keiji Nakajima¹², Yasukazu Nakamura¹⁰, Kyoko Ohashi-Ito¹³, Shinichiro Sawa¹⁴, Masaki Shimamura¹⁵, Roberto Solano¹⁶, Hirokazu Tsukaya^{13,17}, Takashi Ueda¹⁸, Yuichiro Watanabe¹⁹, Katsuyuki T. Yamato²⁰, Sabine Zachgo²¹ and Takayuki Kohchi^{2,*}

¹School of Biological Sciences, Monash University, Melbourne, Victoria 3800, Australia

²Graduate School of Biostudies, Kyoto University, Kyoto, 606-8502 Japan

³University of Veracruz, Institute for Biotechnology and Applied Ecology (INBIOTECA), Avenida de las Culturas Veracruzanas 101, Colonia Emiliano Zapata 91090, Xalapa, Veracruz, México

⁴Gregor Mendel Institute, Dr. Bohrgasse 3, 1030 Vienna, Austria

⁵Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK

⁶Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK

⁷Graduate School of Science, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe, 657-8501 Japan

⁸Graduate School of Life Sciences, Tohoku University, 2-1-1, Katahira, Aoba-ku, Sendai, 980-8577 Japan

⁹Institute of Biotechnology, National Taiwan University, Taiwan

¹⁰National Institute of Genetics, Research Organization of Information and Systems, 1111 Yata, Mishima, 411-8540 Japan

¹¹RIKEN Center for Sustainable Resource Science, Tsurumi-ku, Yokohama, Kanagawa, 230-0045 Japan

¹²Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara, 630-0192 Japan

¹³Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo, 113-0033 Japan

¹⁴Graduate School of Science and Technology, Kumamoto University, Kumamoto, 860-8555 Japan

¹⁵Department of Biological Science, Graduate School of Science, Hiroshima University, Kagami-yama, Higashi Hiroshima, Hiroshima, 739-8526 Japan

¹⁶Departamento de Genética Molecular de Plantas, Centro Nacional de Biotecnología-CSIC, C/ Darwin, 3, Campus Cantoblanco, Universidad Autónoma de Madrid, 28049 Madrid, Spain

¹⁷Okazaki Institute for Integrative Bioscience, National Institute of Natural Sciences, 5-1, Higashiyama, Okazaki, Aichi, 444-8787 Japan

¹⁸Laboratory of Developmental Cell Biology, Department of Biological Sciences, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0033 Japan

¹⁹Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1, Komaba, Meguro, Tokyo, 153-8902 Japan

²⁰Faculty of Biology-Oriented Science and Technology, Kinki University, Nishimitani, Kinokawa, Wakayama, 649-6493 Japan

²¹University of Osnabrück, Botany Department, Barbarastr. 11, D-49076 Osnabrück, Germany

*Corresponding authors: John L. Bowman, E-mail, john.bowman@monash.edu, Fax: +61-3-9905-5613; Takayuki Kohchi, E-mail: tkohchi@lif.kyoto-u.ac.jp, Fax: +81-75-753-6127

(Received August 5, 2015; Accepted November 25, 2015)

While *Marchantia polymorpha* has been utilized as a model system to investigate fundamental biological questions for over almost two centuries, there is renewed interest in *M. polymorpha* as a model genetic organism in the genomics era. Here we outline community guidelines for *M. polymorpha* gene and transgene nomenclature, and we anticipate that these guidelines will promote consistency and reduce both redundancy and confusion in the scientific literature.

Keywords: Gene nomenclature • Liverwort • *Marchantia polymorpha* • Model organism.

In the not too distant past, mankind began to travel the world in search of discovery and, when he or she decided that a geological feature required to be identified from the surrounding landscape, a name was bestowed upon it. However, more often than not they were not the first to do so, and so began the multiplicity of names for singular features, which in turn can

lead to confusion of identity and questions of priority. In the age of genetics and now genomics, mankind travels along the chromosomes of their own and another organism's genomes and bestows names upon landmarks therein. Again, often they are not the first, and the multiplicity of names is one cause for confusion in the literature. While some multiplicity is unavoidable, such as when the same gene is identified independently in different genetic screens, our goal is to minimize duplicity of names.

Since *Marchantia polymorpha* is not the first species to have been examined genetically or to have its genome sequenced, we can build upon the lessons of the past and adapt nomenclatural guidelines to suit an organism in the genomic era. Thus, the guidelines proposed closely follow those of *Arabidopsis thaliana* (Meinke et al. 1998), whose guidelines (formulated by the Arabidopsis Nomenclature Committee at the 1987 Third International Arabidopsis Meeting in East Lansing, MI, USA) were based on those previously produced for *Saccharomyces cerevisiae* (Cherry 1998). In the case of *M. polymorpha*, these guidelines have been

expanded to include types of genes, such as microRNAs, that have been discovered since. While the genome assembly in mid 2015 is not sufficiently complete to assign gene names based on chromosomal location, we anticipate such 'gene ID' names to be formally introduced in the future. Finally, the complexity and sophistication of transgenes have greatly expanded in the past couple of decades and, while a one-size-fits-all nomenclature may not incorporate the full extent of their variety, we provide guidelines for configurations commonly used at present.

With the sequencing of genomes across the breadth of life, including the land plants and their algal relatives, reverse genetics approaches to uncover gene function and trace the evolutionary history of genes has become pervasive. This has led to gene names from one taxon often being adopted for related genes in another taxon. While this promotes cross-referencing and comparisons across taxa, there are risks if appropriate phylogenetic analyses are not performed. We propose that adopting gene names across taxa should only be done in the knowledge that the genes in question are orthologs, otherwise such practice should be avoided. To facilitate both intrageneric and intergeneric comparisons of gene function, we propose that gene names in *M. polymorpha* are prefixed with 'Mp' such that the origin of the gene is immediately recognizable. This prefix may be omitted if, in the context in which the gene is described, it is obvious that the *M. polymorpha* gene is being referenced—for example, in manuscripts or sections of manuscripts where genes from species other than *M. polymorpha* are not mentioned. Finally, in anticipation that other *Marchantia* species may become subjects of genetic and genomic investigations, we provide guidelines for naming of genes in other species of the genus.

To facilitate the sharing, canalization and choosing of unique gene names in *M. polymorpha*, a wiki has been established at the National Institute of Genetics at Mishima, Shizuoka, Japan: <http://marchantia.info/nomenclature/>.

Guidelines for Gene Names in *Marchantia polymorpha* (Table 1)

1. The wild-type allele is written in upper case letters and italics.
2. The mutant allele is written in lower case letters and italics. The origin of the mutant allele, whether via homologous recombination knockout, forward genetics mutagenesis or targeted via genome editing, is irrelevant.
3. The wild-type and mutant names, when written in full, should be the same, except for case.
4. Abbreviated gene symbols should have 2–8 letters and may be followed by a number if there are multiple loci with the same name. In the case of genes encoding microRNAs, the locus designation is a lowercase letter rather than a number.
5. Different mutant alleles of the same gene are distinguished by a number (allele designation) following a hyphen. If only a single

Table 1 Guidelines for gene names in *Marchantia polymorpha*

Gene full name	MpALPHABETICA; with a locus number, MpALPHABETICA2
Gene symbol	MpABC, MpABC2, MpDEF2
Mutant allele	Mpabc
Specific mutant allele	Mpabc-1, Mpabc-2 ^{ko} , Mpabc-3 ^{ge} , Mpabc-4 ^D , Mpabc-5 ^{GOF}
Protein	MpABC
Phenotype	MpAbc ⁺ , MpAbc ⁻
MicroRNA gene	MpMIR160, MpMIR166a, MpMIR166b
MicroRNA	MpmiR160, MpmiR166
Allele in another accession	MpABC ^{Kit-2}

mutant allele is known, it is best to designate it as '-1' to circumvent future changes in nomenclature when additional alleles are discovered. Optional superscript descriptors, such as 'ko' for homologous recombination-mediated knockout or 'ge' for endonuclease-generated genome editing [e.g. CRISPR-CAS9 (clustered regularly interspaced short palindromic repeats/CRISPR associated proteins) or TALEN (transcription activator-like effector nuclease)], may be added.

6. In the near future, or perhaps even at present, the majority of mutant alleles will be endonuclease generated. The implementation of such techniques often generates large numbers of similar, independent alleles. Formal allele designation should only be assigned to those alleles that have been well characterized at the molecular level and that are likely to be maintained. Molecularly identical, but independently derived, alleles are given unique allele numbers.
7. Recessive and dominant alleles are written in the same manner; however, if an allele is dominant to the wild-type allele, a superscript 'D' may be added after the allele designation. If dominance cannot be assessed, given the haploid nature of the *M. polymorpha* gametophyte, but the allele has been shown to be a gain-of-function allele, a superscript GOF can be added to the allele designation.
8. Protein products should be written in upper case letters, no italics.
9. Phenotypes are designated by the gene symbol (no italics) with the first letter capitalized and followed by a superscript + or -. Thus Abc⁺ describes the wild type; Abc⁻ refers to the mutant.
10. MicroRNA nomenclature follows that outlined in Meyers *et al.* (2008) and at www.mirbase.org. The microRNA locus follows the standard gene nomenclature [e.g. see 1 and 2 above]; the microRNA itself is written as 'MpmiRNNN'. However, at miRBase, species are assigned a three-letter code, with *M. polymorpha* designated mpo; thus, a synonym for MpmiR166 at miRBase would be mpo-MIR166.
11. The wild-type allele is that present in the accession whose genome was sequenced, Tak-1. If using another accession in which the allele differs biochemically or biologically from that of Tak-1, it is written with the accession in superscript.
12. Double mutants are written with alleles directly following one another.
13. If the gene name is co-opted from a gene initially identified in another species, the name is preceded by 'Mp' (not italicized) to

designate *Marchantia polymorpha*. In order to use the same name, the *M. polymorpha* gene should be an ortholog, not merely a homolog—orthologs are genes in different species that have descended from a single gene in their last common ancestor (Fitch 1970). Note that this guideline strictly applies only to the alphabetical part of the name and not the numerical if there are multiple loci. While it would be desirable to include the number of the ortholog as well, in cases where there are large differences in gene number between, for example *M. polymorpha* and *A. thaliana*, conservation of gene number may be impractical. Finally, if the numerical designation of an ortholog is retained, and if there are multiple paralogs in *M. polymorpha*, they can be designated with an upper case letter following the number, e.g. MpABC2A and MpABC2B if both are orthologs of the *A. thaliana* gene ABC2. If the gene name originates from studies in *M. polymorpha*, the prefix (Mp) is optional.

- Genes in related *Marchantia* species will have a different prefix (see below).

Guidelines for Transgene Nomenclature in *Marchantia polymorpha* (Table 2)

- Regulatory sequences, colloquially termed promoters, are designated with a subscript 'pro' preceding or following the name of the regulatory sequences.
- Transcriptional fusions are written with a single colon between the name of the regulatory sequence and the name of the coding sequence.
- Translational fusions are written with a '-' between the fused coding sequences, which are listed in the order they appear in the protein. This nomenclature can be extended to include the fusion of more than two coding sequences.
- Specific transgene alleles are written either with a hyphen and a number or with a superscript designation.
- Artificial microRNAs are written with an amiR prefix followed by a hyphen and then the name of gene(s) that are targeted. The original backbone on which the amiR was based is designated in superscript.
- When using a transactivation system (e.g. GAL4/UAS or LhG4/Op) for regulating transgenes a '>>' is written between the name of the regulatory sequences and the name of the coding sequences being regulated.

Table 2 Guidelines for transgene nomenclature in *Marchantia polymorpha*

Promoter	<i>pro</i> MpABC or MpABC _{<i>pro</i>}
Transcriptional fusion	<i>pro</i> MpABC:GFP
Translational fusion	<i>pro</i> MpABC:MpABC-GFP
Transgene allele	<i>pro</i> MpEF1:MpABC-1 or <i>pro</i> MpEF1:MpABC ^{#1}
Artificial microRNA	<i>amiR</i> -MpABC ^{Mpmir160}
Transactivation	<i>pro</i> MpEF1 >> MpABC

A Brief Taxonomic History of *Marchantia polymorpha* (Table 3)

As *M. polymorpha* is a cosmopolitan species, there is significant variation in morphology, and three subspecies that differ in both habitat and morphology are presently recognized (Bischler-Causse and Boisselier-Dubayle 1991, Bischler-Causse 1993). For details of the morphological differences between these subspecies, see Shimamura (2016). As *M. polymorpha* was well known by the Renaissance and possibly much earlier by the Greeks and Romans (Bowman 2016), a long and convoluted taxonomic history exists, and by the late 18th century Schmidel cited nearly 50 synonyms for *M. polymorpha* (Schmidel 1783). Earlier authors, such as Casper Bauhin (Bauhin 1623) and Micheli (Micheli 1729), had already recognized different morphological variants. Perhaps this variation is why Linnaeus, who contributed minimally to liverwort taxonomy and based his nomenclature of cryptogamic plants largely on the previous work of Dillenius (1741), lumped the morphological variants together into a single species 'polymorpha', albeit with three unnamed varieties (Linné and Salvius 1753). Later, Nees von Esenbeck (1838) articulated two distinct varieties, A and B, each with further subvarieties, three of which (var. B *alpestris*, *communis* Aγ *domestica*, and *communis* Aα *aquatica*) corresponded to the three forms identified earlier by Micheli. These varieties were cited as forms by Müller (1906), a practice subsequently followed by Evans (1917). Nees von Esenbeck's names were later adopted by Burgeff when he suggested they should be raised to species rank based on restricted infertility between his European varieties—if female 'aquatica' were crossed with male 'alpestris', the crosses failed, but the reciprocal cross was always successful; crosses between 'polymorpha' and either 'alpestris' or 'aquatica' occurred readily (Burgeff 1943). More recently, Bischler-Causse and Boisselier-Dubayle described three subspecies based on enzyme polymorphism data and suggested that of the three types described by Linnaeus, his α type corresponded to the 'aquatica' variety and that this form should be *M. polymorpha* subsp. *polymorpha* (Bischler-Causse and Boisselier-Dubayle 1991). They could not reconcile Linnaeus' other varieties with enzymatic polymorphism data and thus described two other subspecies, *montivagans* and *ruderalis*, which may correspond to 'alpestris' and 'domesticus' of Nees von Esenbeck. Based on a limited number of isozyme, RFLP (restriction fragment length polymorphism) and RAPD (random amplified polymorphic DNA) markers, the three taxa can be reliably distinguished, with little genetic variation occurring within a taxon (Boisselier-Dubayle et al. 1995). Individual thalli of co-occurring taxa, e.g. *M. polymorpha* subsp. *ruderalis* and *M. polymorpha* subsp. *polymorpha*, living within a few meters of one another did not exhibit any evidence of cross-fertilization—in an exemplar case a strict genetic separation was seen between *M. polymorpha* subsp. *polymorpha* living in contact with the water along a brook and *M. polymorpha* subsp. *ruderalis* just above the water level (Boisselier-Dubayle et al. 1995).

The first assembled genome sequence was generated from Tak-1, an accession of *M. polymorpha* subsp. *ruderalis*. While the

Table 3 Taxonomic history of *Marchantia polymorpha*

Micheli 1729	<i>Marchantia major</i> . . . tab. 1, fig 1	<i>Marchantia capitulo</i> . . . tab. 1, fig. 2	<i>Marchantia foliis</i> . . . tab. 1, fig. 3
Dillenius 1741		<i>Domesticus minor</i> . . . tab. 77, fig. 7	<i>Fontanis major</i> . . . tab. 76, fig. 6
Linné and Salvius 1753		<i>M. polymorpha</i> β,γ	<i>M. polymorpha</i> α
Nees von Esenbeck 1838	<i>M. polymorpha</i> <i>B. alpestris communis</i>	<i>M. polymorpha</i> <i>Ag domestica communis</i>	<i>M. polymorpha</i> <i>Aa aquatica</i>
Burgeff 1943	<i>M. alpestris</i>	<i>M. polymorpha</i>	<i>M. aquatica</i>
Bischler-Causse 1993	<i>M. polymorpha</i> subsp. <i>montivagans</i>	<i>M. polymorpha</i> subsp. <i>ruderalis</i>	<i>M. polymorpha</i> subsp. <i>polymorpha</i>

Table 4 Proposed prefix to *Marchantia* species

Species	spp.	Prefix
<i>M. polymorpha</i>	<i>ruderalis</i>	Mp
<i>M. polymorpha</i>	<i>montivagans</i>	Mpmo
<i>M. polymorpha</i>	<i>polymorpha</i>	Mppo
<i>M. paleacea</i>	–	Mpa
<i>M. foliacea</i>	–	Mfo

species complex is presently considered to consist of three subspecies, future taxonomic revision may be instigated by additional genetic and genomic data. If the species complex is deemed to represent multiple species, it is possible that priority for the name *M. polymorpha* may not go to the common laboratory variety, *M. polymorpha* subsp. *ruderalis*. Thus, to avoid future nomenclatural confusion, we propose that the prefix Mp be applied to the laboratory strain (*M. polymorpha* subsp. *ruderalis*), and that modifications of this ‘moniker’ be applied to the other subspecies: e.g. Mpmo for ‘montivagans’ and Mppo for ‘polymorpha’ in the scheme of Bischler-Causse and Boisselier-Dubayle (Table 4). Alternatively, in the future there may be an requirement for an agreement similar to that for the genus *Acacia*, where retypification allows the Australian clade to conserve the name over priority (Smith and Figueiredo 2011).

The *Marchantia* Genus

As with all model species, some features of *M. polymorpha* are derived relative to related species, and thus tools and resources developed for *M. polymorpha* will be turned towards other *Marchantia* species. For example, an obvious derived feature of intense interest is the lack of mycorrhizal associations in the common laboratory strains of *M. polymorpha* subsp. *ruderalis* (Golenkin 1902, Stahl 1949, Ligrone et al. 2007). Mycorrhizal associations have not been observed in *M. polymorpha* subsp. *polymorpha*, but they are present in at least some accessions of *M. polymorpha* subsp. *montivagans* (Ligrone et al. 2007). As research may proceed rapidly to other *Marchantia* species, we propose that genes in other species within the *Marchantia* genus should have different prefixes,

such as Mpa for *M. paleacea* and Mfo for *M. foliacea*, two species in which mycorrhizal associations have been more recently investigated (Russell and Bulman 2005, Humphreys et al. 2010). Prefixes for other species can be assigned as the need arises.

Additional signatories

Shin-ichi Arimura, Kevin Davies, Feng Chen, Taku Demura, Beverley J. Glover, Ueli Grossniklaus, Takashi Hashimoto, Yuki Hirakawa, Yoko Ikeda, Yutaka Kodama, Nayelli Marsch Martinez, Kenji Matsui, Hiroyasu Motose, Ryuichi Nishihama, Toshiharu Shikanai, Kosaku Takahashi, Mizuki Takenaka, Daisuke Takezawa, and Masaaki Umeda.

Funding

This work was supported by the Australian Research Council (DP130100177) to JLB, Japan Society for the Promotion of Science (JSPS) Bilateral Research Program, and JSPS Grant-in-Aid for Scientific Research on Innovative Area (25113001).

Disclosures

The authors have no conflicts of interest to declare.

References

- Bauhin, C. (1623) Pinax Theatri botanici Caspari Bauhini. Ludovici Regis, Basileae Helvet.
- Bischler-Causse, H. (1993) *Marchantia* L. The European and African taxa. J. Cramer, Berlin-Stuttgart.
- Bischler-Causse, H. and Boisselier-Dubayle, M.-C. (1991) Lectotypification of *Marchantia polymorpha* L. *J. Bryol.* 16: 361–365.
- Boisselier-Dubayle, M.-C., Jubier, M.F., Lejeune, B. and Bischler, H. (1995) Genetic variability in the three subspecies of *Marchantia polymorpha* (Hepaticae)—isozymes, RFLP and RAPD markers. *Taxon* 44: 363–376.
- Bowman, J.L. (2016) A brief history of *Marchantia* from Greece to genomics. *Plant Cell Physiol.* 57: In press.
- Burgeff, H. (1943) Genetische Studien an *Marchantia*. Gustav Fischer, Jena.
- Cherry, J.M. (1998) *Saccharomyces cerevisiae*. *Trends Genet.* 14: S10–S11.

- Dillenius, J.J. (1741) *Historia Muscorum: A General History of Land and Water, &c. Mosses and Corals, Containing All the Known Species*. J. Millan, London.
- Evans, A.W. (1917) The American species of *Marchantia*. *Trans. Conn. Acad. Arts Sci.* 21: 201–313.
- Fitch, W.M. (1970) Distinguishing homologous from analogous proteins. *Syst. Biol.* 19: 99–113.
- Golenkin, M. (1902) Die Mycorrhiza-ähnlichen Bildungen der Marchantiaceen. *Flora* 90: 209–220; Taf. XI.
- Humphreys, C.P., Franks, P.J., Rees, M., Bidartondo, M.I., Leake, J.R. and Beerling, D.J. (2010) Mutualistic mycorrhiza-like symbiosis in the most ancient group of land plants. *Nat. Commun.* 1: 103: 1–7.
- Ligrone, R., Carafa, A., Lumini, E., Bianciotto, V., Bonfante, P. and Duckett, J.G. (2007) Glomeromycotean associations in liverworts—a molecular, cellular, and taxonomic analysis. *Amer. J. Bot.* 94: 1756–1777.
- Linné, C.v. and Salvius, L. (1753) *Species plantarum. Impensis Laurentii Salvii, Holmiae* [Stockholm].
- Meinke, D.W., Cherry, J.M. and Koornneef, M. (1998) *Arabidopsis thaliana*. *Trends Genet.* 14: S24–S25.
- Meyers, B.C., Axtell, M.J., Bartel, B., Bartel, D.P., Baulcombe, D., Bowman, J.L., et al. (2008) Criteria for annotation of plant microRNAs. *Plant Cell* 20: 3186–3190.
- Micheli, P.A. (1729) *Nova plantarvm genera. B. Paperinii, Florentiae*.
- Müller, K. (1906) *Die Lebermoose in Dr. L. Rabenhorst's Kryptogamen-Flora von Deutschland, Oesterreich und der Schweiz. Vol. 6*. Eduard Kummer, Leipzig.
- Nees von Esenbeck, C.G. (1838) *Naturgeschichte der europäischen Lebermoose Vol. 4*. Grass, Barth and Comp, Breslau.
- Russell, J. and Bulman, S. (2005) The liverwort *Marchantia foliacea* forms a specialized symbiosis with arbuscular mycorrhizal fungi in the genus *Glomus*. *New Phytol* 165: 567–579.
- Schmidel, C.C. (1783) *Icones Plantarvm Et Analyses Partivm Aeri Incisae Atqve Vivis Coloribvs Insignatae*. Valentin Bischoff, Norimberga.
- Shimamura, M. (2016) *Marchantia polymorpha*: taxonomy, phylogeny and morphology of a model plant. *Plant Cell Physiol.* 57: 230–256.
- Smith, G.F. and Figueiredo, E. (2011) Conserving *Acacia* Mill. with a conserved type: what happened in Melbourne? *Taxon* 60: 1504–1506.
- Stahl, M. (1949) Die Mycorrhiza der Lebermoose mit besonderer Berücksichtigung der thallosen Formen. *Planta* 37: 103–148.