Letter to the editor

Nomenclature for naming loci, alleles, linkage groups and chromosomes to be used in poultry genome publications and databases

LB Crittenden[1], JJ Bitgood[2], DW Burt[3],

FA Ponce de Leon[4], M Tixier-Boichard[5*]

[1] *Department of Microbiology, Michigan State University, East Lansing, MI 48824;*
[2] *Department of Poultry Science, University of Wisconsin, Madison, WI 53706, USA;*
[3] *Roslin Institute (Edinburgh), Roslin, Midlothian, EH25 9PS, UK;*
[4] *Department of Veterinary and Animal Sciences, 311/Paige Laboratory,*
University of Massachussetts, Amherst, MA 01003, USA;
[5] *Laboratoire de génétique factorielle, Institut national de la recherche agronomique,*
78352 Jouy-en-Josas cedex, France

Summary – The Second International Workshop on Poultry Genome Mapping, meeting at the 1994 Conference of the International Society for Animal Genetics, and the Poultry Committee of the USDA National Animal Genome Research Program have approved a new set of guidelines for poultry gene and allele symbols to replace the guidelines published previously. The new guidelines are a modification of the human gene nomenclature guidelines and are designed to facilitate the naming of loci detected by molecular probes and electronic publication in on-line databases. Authors and journal editors are strongly encouraged to adopt the new nomenclature guidelines. Comments and suggestions are welcome. The complete text of the guidelines can be accessed on the World Wide Web. The home page addresses are: http://www.ri.bbsrc.ac.uk/chickmap/ChickMapHomePage.html for the ChickMap home page, Roslin Institute, UK and http://poultry.mph.msu.edu/ for the Department of Microbiology, Michigan State University, USA. The home pages will contain updates and changes in the guidelines. The home pages also provide access to CHICKBASE, a chicken genome database, which will have a listing of old and new gene symbols. A World Wide Web browser is required for access. The guidelines can also be obtained from LB Crittenden (e-mail: slcritte@facstaff.wisc.edu). The International Poultry Gene Nomenclature Committee comprises the authors of this letter to the editor.

poultry / gene nomenclature

---

* Correspondence and reprints

Résumé – Nomenclature pour désigner les locus, allèles, groupes de liaison et chromosomes des volailles dans les publications et bases de données. *Une nouvelle liste de directives définissant les symboles géniques et alléliques chez les volailles, destinée à remplacer les règles précédemment publiées, a été approuvée par le deuxième séminaire international sur la cartographie du génome aviaire, réuni lors du congrès de la Société internationale de génétique animale en 1994, et par le Comité aux espèces aviaires du programme national de recherche sur les génomes du Département d'agriculture des États-Unis. Les nouvelles directives sont inspirées de celles utilisées pour la nomenclature des gènes chez l'homme et ont été conçues pour faciliter d'une part la désignation des locus révélés par les sondes moléculaires et d'autre part la publication électronique de bases de données en ligne. Les auteurs et les rédacteurs de revues scientifiques sont vivement encouragés à suivre les nouvelles directives. Les commentaires et suggestions sont les bienvenus. Le texte complet des directives est également accessible sur le réseau World Wide Web (W3). Les adresses des pages d'accueils sont : http://www.ri.bbsrc.ac.uk/chickmap/ChickMapHomePage.html pour l'Institut Roslin, Royaume-Uni, et http://poultry.mph.msu.edu pour le département de microbiologie de l'université du Michigan, États-Unis. Les pages d'accueil seront régulièrement mises à jour pour la nomenclature et les directives. Les pages d'accueil permettent aussi d'accéder à la base de données du génome aviaire, Chickbase, qui contiendra une liste des anciens et nouveaux symboles géniques. Un logiciel d'interrogation de W3 est nécessaire. La nomenclature actuellement proposée peut aussi être obtenue auprès de LB Crittenden (adresse électronique : slcritte@facstaff.wisc.edu). Le Comité international pour la nomenclature génique des volailles est constitué par les auteurs de la présente lettre à la rédaction.*

volailles / nomenclature génétique / base de données

INTRODUCTION

These guidelines are based on discussions of the proposal of 23 April 1993 reviewed by the Poultry Species Committee of the National Animal Genome Research Program (NAGRP) at its meeting in Saint Louis on 5 May 1993. A revised version (25 May 1993) was reviewed at an open meeting of the NAGRP Poultry Committee at the 1993 Poultry Science meetings. At the meeting of the NAGRP Poultry Committee on 18 September 1993, an International Nomenclature Committee of Bitgood, Burt, Crittenden, Ponce de Leon and Tixier-Boichard was suggested, with a Resource Panel to help rename classical genes listed by Somes (1988) now being revised by Bitgood. In July 1994 a version dated 11 May 1994 was discussed and generally approved by the Second International Workshop on Poultry Genome Mapping in Prague. The present letter reflects some slight changes suggested by the Workshop participants.

This outline is aimed at suggesting nomenclature for use in journal publications and in the international chicken genome database (CHICKBASE) being developed at the Roslin Institute (Edinburgh), UK by D Burt and his colleagues. In certain cases some guidelines for the nomenclature may be different for publication in journals from that in a database, as further information cannot be readily called up. In a database, one can always consult complementary information sources. For

example, footnotes may be used in publications but further fields should be provided in a database.

The intent of these guidelines is to suggest nomenclature for loci that are demonstrated to segregate as Mendelian genes or are physically located to a specific chromosomal region. This version will replace the guidelines published by Somes (1980). When genes are homologous to human genes, the name should be the same as the human gene as listed in the on-line human genome database, GDB, or in the latest 'Catalog of mapped genes' (McAlpine et al, 1993). The name should also be compared with the gene list in CHICKBASE so that duplicate symbols do not occur in the literature. It is also urged that all new gene symbols be submitted to CHICKBASE for review by the Nomenclature Committee for adherence to these guidelines and to avoid duplication of symbols before publication.

Since locus and allele nomenclature will generally follow the human guidelines, the following sections, 'Naming loci and alleles' and 'Genotype terminology', were modified from Shows et al (1987) (p 12–15) to reflect poultry-specific aspects of nomenclature and to use known genes of the chicken as examples. Guidelines for gene nomenclature in ruminants have previously been published in Genetics Selection Evolution (Andresen et al, 1991; Cognosag, 1995).

NAMING LOCI AND ALLELES

Gene symbols

A newly identified locus will be named by the laboratory that first conducts the genetic segregation analysis or assigns a gene to a specific chromosomal location.

Genes are designated by upper-case roman letters or by a combination of upper-case letters and arabic numbers. Since symbols should be short to be useful and should not attempt to indicate all known information about a gene, a total of three characters to designate gene names is optimal; it is recommended that no more than five characters be used except for coded anonymous loci which can have eight. Based on classical genetic guidelines, gene symbols always are either underlined or italicized. Gene symbols need not be italicized in catalogs of known genes. When fragments or synthesized segments of genes are referred to, symbols need not be italicized. New symbols must not duplicate existing gene symbols. Examples: *PO* (polydactyly); *MM7* (micromelia VII); *GPDA* (α-glycerol phosphate dehydrogenase-liver); *HBB* (hemoglobin, β-polypeptide).

The first letter should be the same as that of the name of the gene to facilitate alphabetical listing and grouping.

The initial character should always be a letter. Subsequent characters of the symbol may be other letters or, if necessary, arabic numerals.

All characters in a gene symbol should be written on the same line; thus, no superscripts or subscripts may be used.

No Roman numerals may be used. Roman numerals in previously used symbols should be changed to their arabic equivalents.

Greek letters are not permitted in a gene symbol. All Greek symbols should be changed to letters in the Roman alphabet.

A Greek letter prefixing a gene name must be change to its Roman alphabet equivalent and placed at the end of the gene symbol. This permits alphabetic ordering of the gene in listings with similar properties such as substrate specificities. Examples: *HBA* (α-hemoglobin); *HBB* (β-hemoglobin).

Where gene products of similar function are encoded by different genes, the corresponding loci are designated by arabic numerals placed immediately after the gene symbol, without any space between the letters and numbers used. Example: *PA2*, *PA3* (two loci for pre-albumin). However, single-letter suffixes may be used to designate these different loci only if they exist historically. Example: *ADEA*, *ADEB* (two loci for adenine synthetase).

A final character in the gene symbol may be used to specify a characteristic of the gene. While letters to specify tissue distribution have been used historically, arabic numbers are now preferred as experience has shown that tissue specificity may not be as restricted as described initially.

If the name of a gene contains a character or property for which there is a recognized abbreviation, the abbreviation should be used, for example, the single-letter abbreviation for amino acids used in aminoacyl residues or approved biochemical abbreviations such as *GLC* for glucose and *GSH* for glutathione.

Allele symbols

Alleles will be named by the laboratory that first conducts the segregation analysis defining that allele.

The allele symbol should be limited to four characters, with an optimum of three characters. Only capital letters or arabic numerals in any order should be used.

Allele designations are written on the same line as gene symbols. In order to keep the gene and allele designations separated but together, a new character, the asterisk, has been introduced. Advantages of the asterisk are many. The asterisk is convenient, universal, and does not convey past genetic meaning such as the dash, space, or comma. The asterisk preceding a symbol indicates that it is an allele of a gene. Likewise, an asterisk following a symbol indicates that it is a gene. After the gene and allele symbols have been identified, the allele symbol preceded by an asterisk can be used separately in text. There should be no spaces between gene, asterisk, and allele, and the entire symbol should be underlined or italicized. For example: *OV*A*, *OV*B* (for alleles at the ovalbumin locus); *EAA*1*, *EAA*7* (for haplotypes of the blood group A system).

The allele symbol may convey additional information. The first allele in a series may be designated A or 1. The symbol may convey a morphological characteristic, biochemical property, cellular location, control property, or, ultimately, the amino-acid nucleotide substitution (ie, *HBB*6V*). No normal plus (+) or variant minus (−) symbol, Roman numeral, or Greek symbol should be used. If the name of a geographic location is used in designating an allele, it should be limited to no more than a four-character symbol. If an allele lacks function, this is indicated by an *O* (capital letter *O*). For optimal usage, allele symbols should be brief and need not summarize all information known about their genetic specificity.

If the information regarding the genetic specificity is too complex to be conveyed conveniently in a symbol (eg, kinetic properties, amino-acid substitutions, or

subcellular localization), alleles may be designated by letter or number and the information conveyed in tables.

Dominance, recessiveness, and wild type, as these terms have been used for classical genes, are not addressed in Shows et al (1987) presumably because these terms describe the phenotype and not the genotype. We suggest that no symbols denoting dominance or recessiveness be used in the allele symbol, but that tables of genes contain a column stating the dominance relationships of the alleles observed. Difficulty with dominance arises with multiple alleles. We suggest that new allele symbols for currently named genes retain a letter that corresponds to the phenotype observed or use a new one to serve that purpose. Although we prefer to stay away from the wild-type designation, in some cases it may be useful to use N for the normal allele. For example, the W locus could have $^*Y$ and $^*W$ alleles for yellow and white skin. In contrast, the sex-linked DW locus could have alleles $^*N$ for normal and $^*D$ for dwarf as well as the currently used $^*B$ and $^*M$ alleles.

Printing gene and allele symbols

Gene and allele symbols are underlined in manuscripts and italicized in print. Italics need not be used in catalogs. It may be convenient in manuscripts, computer printouts and in printed text to designate a gene symbol by following it with an asterisk (eg, EAA^*). When only allele symbols are displayed, they can be preceded by an asterisk. For example, for EAA^*1, the allele is printed as ^{*1}.

GENOTYPE TERMINOLOGY

Single loci

Heterozygote for alleles at the EAA locus:

$$\frac{EAA^*1}{EAA^*2} \text{ or } EAA^*1/EAA^*2 \text{ or } EAA^*1/^*2$$

Genotypes for sex-linked traits distinguish between males and females. At the dwarf locus (DW), genotypes for heterozygous males and hemizygous females follow a similar pattern:

Males:

$$\frac{DW^*N}{DW^*D} \text{ or } DW^*N/DW^*D \text{ or } DW^*N/^*D$$

Females:

$$\frac{DW^*N}{W} \text{ or } DW^*N/W$$

(The W identifies the female and maintains the diploid nature of the symbol.)

Linkage and phase

Horizontal lines or slashes separate alleles and indicate chromosome location.

Loci not located on the same chromosome are separated by a semicolon:

$$\frac{I^*N}{I^*DW}; \frac{EAJ^*1}{EAJ^*2} \text{ or } I^*N/I^*DW; EAJ^*1/EAJ^*2 \text{ or } I^*N/^*DW; EAJ^*1/^*2$$

Loci on the same chromosome (linked or syntenic), where the phase is known, are joined by a horizontal line but separated by a space and listed in alphabetical order when the gene order is not known:

$$\frac{EAJ^*1 \ SE^*N}{EAJ^*2 \ SE^*SE}$$

For text, the loci can be printed on a single line, with a space separating genes in phase and a slash indicating different homologs:

$$EAJ^*1 \ SE^*N/EAJ^*2 \ SE^*SE$$

Loci on the same chromosome but with phase not known are separated by a comma:

$$\frac{EAJ^*1, SE^*N}{EAJ^*2, SE^*SE}$$

or printed on a single line with a separating comma:

$$EAJ^*1/EAJ^*2, SE^*N/SE^*SE$$

If the linear order and phase of the genes on the same chromosome are known, they are listed in order from the end of the short arm (p) to the end of the long arm (q) of the chromosome and separated by a space:

$$\frac{EAH^*1 \ SE^*N \ EAJ^*1}{EAH^*2 \ SE^*SE \ EAJ^*2}$$

or $EAH^*1 \ SE^*N \ EAJ^*1/EAH^*2 \ SE^*SE \ EAJ^*2$. The linear order on chromosome 1 is presumed to be pter-EAH-SE-EAJ-cent.

If the gene order on the same chromosome is not known, then the loci are listed on the linear map alphabetically, separated by a comma, and enclosed by parentheses:

$$\text{pter-}SE\text{-}EAJ\text{-}(EV1, O, P)\text{-cent}$$

CLASSICAL LOCI CATALOGUED BY SOMES (1980)

The present standard nomenclature will be converted to the new nomenclature and the new nomenclature will be used for naming any newly identified genes. The new terminology will be much more adaptable for use in computer databases, and will appear as entered on non-graphics screens, except for italics. The Somes nomenclature should be directly convertible to the new nomenclature, in many cases, reported by the Resource Panel appointed by the Nomenclature Committee.

LOCI DETECTED BY DNA PROBES

The use of DNA probes adds another level of nomenclature to the system: the probe name. Probe names cannot be used directly for locus symbols because one probe often detects polymorphisms at more than one locus, and the laboratory probe name may not even reflect the name of the cloned gene and is often long and complex. No attempt to standardize probe names will be made at this time.

Loci detected by anonymous DNA sequences

Such loci have no known physiologic function and can be detected by restriction fragment-length polymorphism (RFLP) techniques using random genomic or cDNA library members as probes, or by polymerase chain reaction (PCR) techniques using arbitrary primers or primers derived from cloned sequences.

Such loci will be named by each laboratory defining them, using a laboratory code of not more than three upper-case letters and sequential arabic numbers of four digits with right justification of the number and preceded by zeros if less than 1 000 (eg, *COM*0099). Expressed genes, such as those detected by cDNA library members, that have no known function shall be followed by an uppercase *E* (eg, *COM*0110*E*). Note that the locus symbols exceed the limit of five characters suggested for named genes. However, allele symbols should be short so that the total symbol can be less than 12 letters or digits.

This system does not contain an embedded chromosome number or other information on the type of probe, as does the human system, since a standard system for naming microchromosomes has not been implemented. However, the advantages are that a unique name can be assigned to the locus by the typing laboratory which does not have to be changed with chromosome assignment, or assigned by the database manager or a committee. However, the locus should be renamed once it is shown to contain coding sequences for a named gene product (see the next section). Further information about each locus will be available in the original publication and in supplementary tables that can be called up in a database.

The locus name can be clarified in publications by adding a code for the type of probe in upper case letters in parentheses. *F* for RFLP, *A* for RAPD, *E* for endogenous viral genes, *M* for microsatellite, and *V* for minisatellites are suggested. These letters will not be considered part of the official name and will not be included in the database, but are optional in journal publication for clarity and should be footnoted.

Loci detected by DNA sequences that represent coding sequences for a known gene product

These loci should be named in uppercase letters and numbers that reflect the name of the gene product. The name should begin with a letter that reflects the first letter of the gene product and numbers should be used when necessary. The general rules for naming loci and alleles should follow Shows et al (1987) as modified above.

A gene can consist of coding and non-coding regulatory and intron sequences. The general location of a specific gene on the genetic map can be found using probes representing coding or non-coding sequences. The gene can be considered

a haplotype. The gene name should be used for the locus symbol on genetic maps whether the probe represents a coding sequence or not. However, the anonymous nature of the probe should be clearly retained in publications and databases, and its anonymous locus name should be used in fine structure mapping.

NAMING CHROMOSOMES AND LINKAGE GROUPS

Autosomes will be numbered in descending order by size. The sex chromosomes will not be numbered but called Z and W. Very few linkage groups are now assigned to chromosomes. Therefore, the classical linkage groups should be designated in Roman numerals as assigned by Somes (1988). The linkage groups assigned in the Compton and East Lansing reference populations are not associated in many cases and will be called C01-nn and E01-nn until chromosomal assignments can be achieved. It may be necessary, before all linkage groups are assigned to chromosomes, to develop a distinct system of naming common linkage groups between the East Lansing and Compton maps that have not been assigned to chromosomes.

Microchromosomes, defined as autosomes smaller than chromosome 8, will be temporarily defined by the first single-copy gene that is assigned to them by fluorescent in situ hybridization, and given a number greater than eight and less than 39 roughly consistent with its relative size. This arbitrary definition of microchromosome is based on those that do not have internationally accepted banding patterns (see below). Any gene linked to that locus, by physical or genetic means, will be considered to be on that microchromosome. Endogenous viral and other repetitive genes will not be used to define a microchromosome.

CHROMOSOMAL AND PHYSICAL MAPPING NOMENCLATURE

A standard banding nomenclature was discussed at the North American Colloquium on Domestic Animal Cytogenetics and Gene Mapping held in Guelph, ON, 13–16 July 1993. Standard banding nomenclature for the Z, W and the eight largest autosomes was agreed upon (Ladjali, Tixier-Boichard, Bitgood, and Ponce de Leon, *International standardization of the chicken karyotype*, in preparation). Such standardization is necessary for the integration of physical and genetic maps. Genes that are assigned to a unique location in the genome can be named as outlined above even though Mendelian segregation has not yet been detected. As physical mapping progresses, nomenclature for expanded DNA fragments or contigs will need to be addressed.

REFERENCES

Andresen E, Broad T, Di Stasio L, Dolling CHS, Hill D, Huston K, Larsen B, Lauvergne JJ, Leveziel H, Mahler X, Pillar P, Rae AL, Renieri C, Tucker EM (1991) Guidelines for gene nomenclature in ruminants 1991. *Genet Sel Evol* 23, 461-466

Bitgood JJ, Somes Jr RG (1990) Linkage relationships and gene mapping. In: *Poultry Breeding and Genetics* (RD Crawford, ed), Elsevier, Amsterdam, 469-495

Bitgood JJ, Somes Jr RG (1993) Gene map of the chicken (*Gallus gallus* or *G domesticus*). In: *Genetic Maps*, 6th ed (SJ O'Brien, ed), Cold Spring Harbor Laboratory Press, Plainview, NY, 4, 333-342

COGNOSAG ad hoc committee (1995) Revised guidelines for gene nomenclature in ruminants 1993. *Genet Sel Evol* 27, 89-93

McAlpine PJ (1993) The 1992 catalog of mapped genes and report of the nomenclature committee. *Genome Priority Reports 1*, 11-42

Shows TB, McAlpine PJ, Boucheix C, Collins FS, Conneally PM, Frezal J, Gershowitz H, Goodfellow PN, Hall JG, Issitt P, Jones CA, Knowles BB, Lewis M, McKusick VA, Meisler M, Morton NE, Rubenstein P, Schanfield MS, Schmickel RD, Skolnick MH, Spence MA, Sutherland GR, Traver M, Van Cong N, Willard HF (1987) Guidelines for human gene nomenclature: an international system for human gene nomenclature (ISGN, 1987). *Cytogenet Cell Genet* 46, 11-28

Somes RG Jr (1980) Alphabetical list of the genes of domestic fowl. *J Hered* 71, 168-174

Somes RG Jr (1988) International Registry of Poultry Genetic Stocks. *Bulletin 476*, Storrs Agricultural Experiment Station