Gene Nomenclature for Protein-Coding Loci in Fish

JAMES B. SHAKLEE[1]

Washington Department of Fisheries
115 General Administration Building, Olympia, Washington 98504, USA

FRED W. ALLENDORF

Division of Biological Sciences
University of Montana, Missoula, Montana 59812, USA

DONALD C. MORIZOT

University of Texas M. D. Anderson Cancer Center
Science Park Research Division
Post Office Box 389, Smithville, Texas 78957, USA

GREGORY S. WHITT

Department of Ecology, Ethology, and Evolution
University of Illinois, Urbana, Illinois 61801, USA

Abstract. — The Fish Genetics Section of the American Fisheries Society established its Nomenclature Committee to develop and promote standardized genetic nomenclatures. Here, following public comments on previously published draft guidelines, we present the committee's revised version of a nomenclature for protein-coding loci in fish. This nomenclature closely parallels the one used for human genetics, but improves on it in several respects. The fish system (1) includes standardized abbreviations for commonly analyzed proteins, and provides formal symbols for gene loci encoding these proteins; (2) specifies typographic conventions for distinguishing between genes and proteins and for identifying alleles; (3) provides for multilocus isozyme systems, isoloci, regulatory loci, and pseudogenes; (4) allows important basic information (such as subcellular distributions of gene products, active substrate isomers, recent gene duplicates, and orthologous relationships among loci) to be specified in gene symbols via prefixes, suffixes, and multilocus designations; (5) provides three alternative systems of allele designation and distinguishes alleles having quantitative effects; and (6) includes guidelines for specifying genotypes, linkage relationships, and biochemical phenotypes. This nomenclature is intended to be comprehensive, flexible, unambiguous, and applicable to all fish species as well as to other organisms. The American Fisheries Society has adopted this system as an editorial standard for its publications.

The members of the Genetic Nomenclature Committee of the American Fisheries Society's Fish Genetics Section have worked for the past 2 years to develop a nomenclature for protein-coding loci that will meet the specific needs of fish geneticists, facilitate the understanding and use of genetics throughout ichthyology and fisheries biology, and promote communication among geneticists. In developing this nomenclature, we attempted to address and reconcile in one unified system the diverse requirements associated with inheritance studies; genetic characterization and analysis of population structure; investigations of genetic components of metabolism, early development, and ecological adaptation; and studies of molecular evolution, including gene structure, enzyme structure and function, and genome organization. Despite our own diverse backgrounds and biases, we found that this multiplicity of goals could be met by a single internally consistent yet flexible nomenclature similar to the one used by human geneticists.

Our proposed guidelines were published in this journal last year (Shaklee et al. 1989), and responses from the genetics community were invited. We sent the proposal to all 238 members of the Fish Genetics Section and to 35 other researchers who work with various vertebrate taxa, and asked for their comments. The guidelines were summarized and discussed during the Section's 1989 business meeting in Anchorage, Alaska. Twenty-five people sent us written critiques. These and numerous verbal communications have led

---

[1] The authors constitute the Fish Genetics Nomenclature Committee (J. B. Shaklee, Chair) of the Fish Genetics Section, American Fisheries Society.

us to clarify or modify certain points. Some correspondents objected to one or another feature of the nomenclature, favoring other practices, but we received no majority opposition to any of the system's principles.

The Nomenclature Committee's revised guidelines for protein-coding loci have been accepted by the American Fisheries Society as its editorial standard for genetics papers to which they can be applied (Kendall 1990). Those guidelines follow. In an appendix, we review the principal comments that we received during the review period, and explain our responses to them. Although we believe that this gene nomenclature is a substantial improvement over all other current alternatives, we emphasize that it should not be viewed as static or final. We expect it to evolve in response to both usage and new knowledge. We welcome and encourage future expansion and modification of the system so it can meet its goal of facilitating communication among geneticists. Comments may be sent to the Chair of the Nomenclature Committee (currently J. B. Shaklee) either directly or via the American Fisheries Society's editorial office.

Gene Nomenclature for Protein-Coding Loci in Fish

Fish biochemical genetics has been made unnecessarily difficult and confusing by the diversity of terminologies used to describe genetic variation at loci that encode enzymes and other proteins. The Fish Genetics Nomenclature Committee was established by the American Fisheries Society's Fish Genetics Section to devise a nomenclature for protein-coding loci that would (1) standardize the designations of fish loci, alleles, and proteins, (2) provide a basis for editorial policy within the American Fisheries Society, and (3) promote communication among fish geneticists, fishery scientists, and other biologists. We believe that the unified system given below meets these needs, yet it is flexible and can be modified and expanded as necessary. Though designed for fish, the system is applicable to other vertebrate taxa, thereby encouraging comparative genetic studies. This standardized nomenclature also will facilitate manuscript preparation, review, and editing both within and among journals.

The nomenclature is founded on two recognized standards. First, we have based the system on enzyme names recommended by the International Union of Biochemistry's Nomenclature Committee (IUBNC 1984) because we believe that, whenever possible, gene symbols should reflect the names of the proteins (enzymes) they encode.[2] Second, we have adopted with little or no change several conventions used in human genetics (Shows et al. 1987) because of their usefulness, clarity, and generality. However, the human gene nomenclature as a whole has several drawbacks: enzyme names often deviate from IUBNC recommendations, locus names often are based on patterns of tissue expression, gene names and symbols are lacking for several enzyme systems currently being studied in fishes, only a single cytosolic and a single mitochondrial form of each enzyme are presumed to exist, and there is no framework for dealing with recently duplicated gene loci. Nevertheless, human genetics holds a dominant position in the field, and fish geneticists will realize several advantages by adopting those aspects of the human gene nomenclature that can be applied to fishes. To the extent the two systems converge, it will be easier to do comparative studies among vertebrate groups, easier to communicate with nonfish geneticists, and easier to promote the use of fish in genetic experiments. We believe the proposed nomenclature will promote and facilitate genetic studies of fish.

Enzymes and Proteins

Enzyme and Protein Names

Enzyme names (Table 1) should be the names recommended by the IUBNC (1984)—with two exceptions.

Exception 1: cytosolic aminopeptidases.—We believe the IUBNC names for cytosolic aminopeptidases are inaccurate or misleading. These enzymes have been studied extensively in fishes (Frick 1983, 1984), and their distinctiveness suggests an ancient evolutionary origin of the loci that encode them. The fish gene nomenclature recognizes five well-characterized peptidases: dipeptidase (abbreviated PEPA), tripeptide aminopeptidase (PEPB), peptidase-C (PEPC), proline dipeptidase (PEPD), and peptidase-S (PEPS). (We accept peptidase-S as the name of one peptidase,

---

[2] An additional set of nomenclatural guidelines, for DNA sequences whose functions are unknown or do not involve protein coding, will be needed to accommodate research at the nucleic acid level.

TABLE 1.—Names, numbers, and abbreviations recommended for enzymes and other proteins commonly analyzed in fish genetics work. The names and numbers follow IUBNC (1984).

Enzyme or protein name[a,b]	Enzyme number	Enzyme or protein abbreviation[b,c]
Aspartate aminotransferase [glutamic-oxaloacetic transaminase]	2.6.1.1	sAAT (GOT1)
	2.6.1.1	mAAT (GOT2)
Acid phosphatase	3.1.3.2	ACP
Adenosine deaminase	3.5.4.4	ADA
Alcohol dehydrogenase	1.1.1.1	ADH
Adenosine kinase	2.7.1.20	ADK
Aconitate hydratase [aconitase]	4.2.1.3	sAH (ACO1)
	4.2.1.3	mAH (ACO2)
Adenylate kinase	2.7.4.3	AK
Alanine aminotransferase [glutamic-pyruvate transaminase]	2.6.1.2	ALAT (GPT)
Alkaline phosphatase	3.1.3.1	ALP
alpha-Amylase	3.2.1.1	aAMY (AMY)
Adenine phosphoribosyltransferase	2.4.2.7	APRT
alpha-L-Arabinofuranosidase	3.2.1.55	aARAF
Carbonate dehydratase [carbonic anhydrase]	4.2.1.1	CAH (CA)
Catalase	1.11.1.6	CAT
Creatine kinase	2.7.3.2	CK
Dihydropteridine reductase	1.6.99.7[d]	DHPR
Enolase	4.2.1.11	ENO
Esterase	3.1.1.-	EST
Esterase-D	3.1.-.-	ESTD (ESD)
Fructose-bisphosphate aldolase	4.1.2.13	FBALD (ALDO)
Fructose-bisphosphatase	3.1.3.11	FBP
Fumarate hydratase [fumarase]	4.2.1.2	FH
alpha-L-Fucosidase	3.2.1.51	aFUC (FUCA)
alpha-Galactosidase	3.2.1.22	aGAL (GLA)
beta-Galactosidase	3.2.1.23	bGAL (GLB)
beta-N-Acetylgalactosaminidase	3.2.1.53	bGALA
Galactokinase	2.7.1.6	GALK
Glyceraldehyde-3-phosphate dehyrogenase	1.2.1.12	GAPDH (GAPD)
Guanine deaminase	3.5.4.3	GDA
Glucose dehydrogenase	1.1.1.47	GDH
Glutamate-ammonia ligase [glutamine synthetase]	6.3.1.2	GLAL (GLNS)
alpha-Glucosidase	3.2.1.20	aGLU (GAA)
beta-Glucosidase	3.2.1.21	bGLU (GBA)
N-Acetyl-beta-glucosaminidase [hexosaminidase]	3.2.1.30	bGLUA (HEX)
Glutamate dehydrogenase	1.4.1.-	GLUDH (GLUD)
Glycerate dehydrogenase	1.1.1.29	GLYDH
Glycerol-3-phosphate dehydrogenase	1.1.1.8	G3PDH (GPD)
Glucose-6-phosphate dehydrogenase	1.1.1.49	G6PDH (G6PD)
Glucose-6-phosphate isomerase	5.3.1.9	GPI
Glutathione reductase	1.6.4.2	GR (GSR)
Guanylate kinase	2.7.4.8	bGUK
beta-Glucuronidase	3.2.1.31	GUS (GUSB)
Hydroxyacylglutathione hydrolase	3.1.2.6	HAGH
Hemoglobin		HB
3-Hydroxybutyrate dehydrogenase	1.1.1.30	HBDH
Hexokinase	2.7.1.1	HK
Hypoxanthine phosphoribosyltransferase	2.4.2.8	HPRT
L-Iditol dehydrogenase [sorbitol dehydrogenase]	1.1.1.14	sIDDH (SORD)
Isocitrate dehydrogenase (NADP+)	1.1.1.42	mIDHP (IDH1)
	1.1.1.42	IDHP (IDH2)
L-Lactate dehydrogenase	1.1.1.27	LDH
Lactoylglutathione lyase [glyoxalase I]	4.4.1.5	aLGL (GLO1)
alpha-Mannosidase	3.2.1.24	sMAN
Malate dehydrogenase	1.1.1.37	sMDH (MDH1)
	1.1.1.37	mMDH (MDH2)
Malic enzyme (NAD+)	1.1.1.38 or 1.1.1.39	ME
Malic enzyme (NADP+)	1.1.1.40	sMEP (ME1)
	1.1.1.40	mMEP (ME2)
Mannose-6-phosphate isomerase	5.3.1.8	MPI
Nucleoside-phosphate kinase [uridine monophosphate kinase]	2.7.4.4	NPK (UMPK)
Nucleoside-triphosphate pyrophosphatase [inosine triphosphatase]	3.6.1.19	NTP (ITP)
Ornithine carbamoyltransferase	2.1.3.3	OCT (OTC)

TABLE 1.—Continued.

Enzyme or protein name[a,b]	Enzyme number	Enzyme or protein abbreviation[b,c]
Octanol dehydrogenase	1.1.1.73	ODH
Parvalbumin		PVALB
Dipeptidase	3.4.-.-	PEPA
Tripeptide aminopeptidase	3.4.-.-	PEPB
Peptidase-C	3.4.-.-	PEPC
Proline dipeptidase	3.4.13.9	PEPD
Peptidase-S	3.4.-.-	PEPS
6-Phosphofructokinase	2.7.1.11	PFK
Phosphoglycerate mutase	5.4.2.1[e]	PGAM
Phosphogluconate dehydrogenase	1.1.1.44	PGDH (PGD)
Phosphoglycerate kinase	2.7.2.3	PGK
Phosphoglucomutase	5.4.2.2[f]	PGM
Phosphoglycolate phosphatase	3.1.3.18	PGP
Pyruvate kinase	2.7.1.40	PK
Purine-nucleoside phosphorylase	2.4.2.1	PNP (NP)
Inorganic pyrophosphatase	3.6.1.1	PP
General (unidentified) protein		PROT
Superoxide dismutase	1.15.1.1	sSOD (SOD1)
	1.15.1.1	mSOD (SOD2)
Tyrosine aminotransferase	2.6.1.5	TAT
Transferrin		TF
Thymidine kinase	2.7.1.21	TK
Triose-phosphate isomerase	5.3.1.1	TPI
UDPglucose-hexose-1-phosphate uridylyltransferase [galactose-1-phosphate uridylyltransferase]	2.7.7.12	UGHUT (GALT)

[a] Bracketed synonyms for enzymes are names that have been used in some previous biochemical literature. These names are presented for information only. They are not the IUBNC's recommended names, and their use should be discontinued.

[b] In this table, the Roman versions of Greek letters are used, spelled out for names, shortened for abbreviations. Use of the Greek characters themselves in either names or abbreviations is acceptable.

[c] Abbreviations in parentheses are used in human genetics. They may be used as gene symbols (in the proper typographic format) when comparisons with other vertebrate groups (especially mammals) are made, provided that the equivalences between abbreviations for fish and humans are clearly presented in the text. When no acceptable human nomenclature equivalent is listed, the abbreviation (gene symbol) specified for fish is to be used in all cases.

The prefixes m and s denote subcellular mitochondrial and cytosolic localizations, respectively.

[d] Formerly 1.6.99.10.

[e] Formerly 2.7.5.3.

[f] Formerly 2.7.5.1.

although it is out of alphabetical sequence with the others, because it is widely used in the literature, it is used in the human gene nomenclature, and because Frick [1984] demonstrated that its locus is orthologous between fish and mammals.)

Exception 2: malic enzymes. — The IUBNC assigned the name "malate dehydrogenase" to four enzymes. We preserve this name for the familiar dimeric enzyme 1.1.1.37. The other three enzymes—1.1.1.38, 1.1.1.39, and 1.1.1.40—apparently are tetramers activated by magnesium and manganese ions. We believe these should have a distinguishing name, and recommend "malic enzyme," a name in widespread use in the literature, for them (Table 1).

If there is substantial uncertainty about the identity of an enzyme, a tentative phenomenological name may be used. Such a name should relate the unknown enzyme, according to its catalytic activity, to one of known characteristics by the suffix "-like": xanthine oxidase-like, for example. Phenomenological names should be used only as a last resort when reasonable efforts have failed to identify an enzyme.

Enzymes known to occur in two or more subcellular compartments (e.g., cytosolic, mitochondrial, lysosomal) should be so identified to maximize information and minimize ambiguity: mitochondrial malate dehydrogenase, cytosolic superoxide dismutase, and so on.

Nonenzymatic proteins should be referred to by their generally accepted names: hemoglobin, transferrin, etc. Unknown proteins (which may or may not be enzymes), such as those visualized by nonspecific protein stains, are best named generically: protein-1, protein-2, etc.

When a protein name contains a Greek letter, either the Greek symbol or the spelled-out Roman

transliteration may be used. We have used the latter form throughout this report and recommend it.

Enzyme Numbers

The first use of an enzyme name in a publication should be accompanied by the four-part IUBNC number (Table 1) whenever the enzyme's identity is certain (e.g., mannose-6-phosphate isomerase 5.3.1.8). Partial ambiguity should be indicated by one or more hyphens in the numerical code (e.g., esterase 3.1.1.-). If the enzyme identity is completely unknown, the number is omitted altogether.

Enzyme and Protein Abbreviations

Enzyme abbreviations are derived from IUBNC names; our recommendations for these and for nonenzymatic proteins reflect common usage among fish geneticists (Table 1). Enzyme abbreviations should be typed UPPERCASE without an underline (nonitalic).

We advocate the following standard endings for classes of enzymes:

aminotransferases	. . . AT (e.g., AAT)
deaminases	. . . DA (e.g., ADA)
dehydrogenases	. . . DH (e.g., LDH)
hydratases, dehydratases	. . . H (e.g., AH)
isomerases	. . . I (e.g., GPI)
kinases	. . . K (e.g., CK)
phosphatases	. . . P (e.g., ACP)
reductases	. . . R (e.g., GR)
transferases	. . . T (e.g., HPRT)

The goals of abbreviation are clarity and brevity. Clarity takes precedence when the two conflict. Hyphens should not be used within an enzyme or protein abbreviation (e.g., G3PDH not G-3-PDH, PNP not P-NP, etc.). Numbers, Greek letters, and stereochemical isomer symbols should not be included in an enzyme abbreviation unless they are necessary to avoid ambiguity (e.g., PGDH not 6-PGDH, GPI not G6PI, LDH not L-LDH; but G3PDH and G6PDH not GPD or GPDH). Abbreviations should be defined the first time they are used in each report.

Additional specifications for enzymes are designated by prefixes and suffixes attached to the abbreviations.

Greek prefix. —When a Greek letter is part of an enzyme's name, it normally is preserved—directly or with a Roman surrogate—as a prefix to the abbreviation. The Roman surrogates are these lowercase letters:

a = alpha;
b = beta;
g = gamma;
d = delta;
e = epsilon.

Thus, beta-galactosidase is symbolized βGAL or bGAL.

Subcellular prefix. —If an enzyme encoded by multiple gene loci is known to occur in two or more subcellular compartments, the localization of each form should be indicated by one of the following single-letter, lowercase prefixes:

l = lysosomal;
m = mitochondrial;
p = peroxisomal;
s = cytosolic (="supernatant" or "soluble").

There are currently six well-characterized enzyme systems in fish that are known to occur in mitochondrial and cytosolic forms (AAT, AH, IDHP, MDH, MEP, and SOD). When subcellular localization is known, this should be specified by the appropriate prefix (e.g., sSOD and mSOD). When subcellular localization is not known, abbreviations without such prefixes should be used to communicate this uncertainty (e.g., SOD-1 and SOD-2). However, investigators are encouraged to make every reasonable effort to establish the subcellular localization of such gene products prior to submission of manuscripts involving them. The subcellular prefix precedes a Greek prefix if both apply.

NAD(P) suffix. —Some vertebrate enzymes, or groups of similar enzymes, have forms that use NAD$^+$ coenzymes *and* other forms (similar enzymes) that use NADP$^+$ coenzymes. These must be distinguished. The NAD$^+$-dependent forms are the more common and need no additional designation. The NADP$^+$-dependent forms take an uppercase "P" suffix closed up to the stem (e.g., IDH 1.1.1.41 versus IDHP 1.1.1.42).

Multilocus isozyme designation. —When two or more forms of the same enzyme are known, these isozymes (and their subunits) should be identified with an Arabic number or uppercase letter that follows the enzyme abbreviation and is separated from it by a hyphen: IDHP-1 and IDHP-2; LDH-A and LDH-B. Letters are used *only* when orthology with isozymes in other taxa is well established. Two terms, orthology and paralogy, are commonly used to describe homologous molecular rela-

tionships. Orthology reflects the descent of species (e.g., the gene loci encoding the LDH-A subunits of all fish species are orthologous genes). Paralogy reflects the descent of genes (e.g., the gene loci encoding the LDH-A, LDH-B, and LDH-C subunits of striped bass are paralogous genes). Paralogous gene loci coexist in the same species, orthologous gene loci do not.

Enzyme subunits should not be given letter designations representing the tissue source of the protein. Rather, numbers (or letter designations where orthology is well established) should be assigned consecutively beginning with 1 (or A). The correspondence between isozyme and tissue source should be given in the explanatory text of each report.

Two different conventions are commonly used to assign numbers to multiple isozymes (and to the multiple loci that encode them). In both cases, numbers are assigned sequentially in relation to the electrophoretic mobilities of the homomeric isozymes. In one case, numbering begins with 1 for the isozyme closest to the anode and proceeds toward the cathode; in the other, numbering starts with the isozyme closest to the cathode.

Number designations for loci and isozymes should be considered as transitory, because the eventual aim is to elucidate evolutionary relationships and assign letter designations to loci. Further, any number sequence based upon electrophoretic mobilities is vulnerable to changes in electrophoretic conditions and discoveries of additional isozymes. Therefore, we recommend (1) that numbers be assigned to isozymes and loci by either of the two schemes outlined above, but that existing practices (if established) be followed for the taxa under study—for example, numbering from cathode to anode for salmonids, from anode to cathode for centrarchids; (2) when more than one numbering system has been used for taxa, that authors provide a correlation between their system and those used in other studies; and (3) that authors make every effort to indicate presumed evolutionary relationships among loci across taxa and studies by assigning numbers that are consistent with previous assignments.

Characters that may indicate such relationships among proteins and loci include tissue specificities or ontogenetic patterns of expression, amino acid or nucleotide sequences, catalytic or physicochemical properties, and antigenic cross-reactivities. The absence of a multilocus isozyme suffix implies that an enzyme is only known to be coded by a single locus in the species under study.

So far, orthology has been established convincingly for only four multilocus systems:

- creatine kinase: CK-A, CK-B, CK-C, and CK-D (Scholl and Eppenberger 1972; Fisher and Whitt 1978);
- glucose-6-phosphate isomerase: GPI-A and GPI-B (Avise and Kitto 1973; Dando 1974);
- L-lactate dehydrogenase: LDH-A, LDH-B, and LDH-C (Shaklee et al. 1973; Markert et al. 1975); and
- cytosolic malate dehydrogenase: sMDH-A and sMDH-B (Kitto and Lewis 1967; Bailey et al. 1970; Fisher et al. 1980).

Number designations should be used for all other multilocus isozyme systems until orthologies are clearly established.

When multilocus enzyme systems have different isozymes occurring in two or more subcellular compartments, the isozymes should be numbered separately for each compartment: mIDHP-1, mIDHP-2, sIDHP-1, sIDHP-2.

"Like" suffix.—When an enzyme has not been clearly identified and bears a "-like" designation (e.g., xanthine dehydrogenase-like), a closed-up "l" (lowercase letter el) suffix is appended to the putative enzyme's abbreviation (XDHl). Authors should specify whether the "l" designation results from characterizations done at the phenotypic (enzyme) or genotypic (DNA) level, because this suffix can arise from both approaches.

Multimer subscripts.—Biochemists customarily denote the subunit composition of multimeric proteins, when this is known, by numeric subscripts. This convention is preserved in the present nomenclature. For example, $LDH-A_4$ is the homotetramer of LDH that has four A-type subunits; $LDH-A_2B_2$ is the heterotetramer with two A and two B subunits.

The nomenclature of allelic isozymes (allozymes) follows the conventions described below for alleles.

Genes

Because genes and their alleles are named in our system by the proteins they code, they normally are represented only by symbols. It is presumed that the respective protein names will have been spelled out in each report. "Gene" and "locus" (the location of a gene on a chromosome) are interchangeable terms as far as this nomenclature is concerned.

Gene Symbols

Genes have the same alphanumeric symbols as the abbreviations of the proteins they code. Gene symbols are distinguished from protein abbreviations by two typographic conventions: they are underlined in typescript and italicized in typeset publications, and they should include an asterisk. We believe that the use of both italics and an asterisk is desirable to provide unambiguous distinction between gene loci (and genotypes) and enzymes and other proteins (and phenotypes). Although the use of asterisks as specified in the human gene nomenclature is not universally practiced at present, we strongly encourage adoption of this convention as a recommended but not mandatory component of gene symbols. The asterisk follows the locus designation (which mimics the usual protein abbreviation) and precedes the allele designation if there is one. For example, *MPI** is the symbol for the gene locus that encodes the enzyme mannose-6-phosphate isomerase MPI; *MPI*1* signifies the *1 allele of the MPI locus.

Locus Symbols

Locus symbols take the same prefixes and suffixes for Greek letters, subcellular localization of product, NADP+ coenzymes, and multilocus expression (of isozymes) as proteins; for example, *βGAL** or *bGAL**, *mAAT**, *IDHP-1**, *LDH-A**. Several additional symbols are described below. Although these are most relevant to gene symbols, they also would be part of respective protein designations.

Multilocus designation.—When two or more loci encode subunits of the same enzyme, the locus terms should have the same trailing Arabic numbers or uppercase letters that identify the subunits (*see* "Multilocus isozyme designation," above). The multilocus designation should be separated from the locus stem (and attached suffix, if present) by a hyphen. Multiple loci encoding isozymes of the same type of enzyme that occur in two or more subcellular compartments should be treated as separate systems when multilocus designations are assigned (e.g., *mAAT-1**, *mAAT-2**, *sAAT-1**, *sAAT-2**).

Recently duplicated loci, such as many of the loci in ancestrally polyploid fishes, present a special nomenclatural problem common to many fish species and several other lower vertebrates. These should be dealt with in either of two ways depending on the amount of information available for a system.

(1) When little or nothing is known about the evolutionary relationships of the loci involved, multiple loci encoding the same enzyme should be distinguished by Arabic numerals (e.g., *EST-1**, *EST-2**, *EST-3**, *ADA-1**, *ADA-2**). Even when there is reason to believe that multiple loci within a taxon (or restricted taxonomic group) may be recent duplicates (paralogous loci), they should be designated with sequential numbers (e.g., *TPI-1** and *TPI-2**, *TPI-3** and *TPI-4** in salmonids) until the multilocus system is thoroughly characterized for diverse fish species and, if necessary, orthologies among the ancient duplicates are formally established and letter designations are assigned to these loci. We strongly encourage thorough comparative studies to characterize such multilocus systems so important evolutionary information can be encoded in the gene nomenclature.

(2) When orthology among ancient duplicates has been clearly established and letter designations have been assigned (e.g., *LDH-A**, *LDH-B**, *LDH-C**), parologies among recent duplicates should be coded by multilocus designations consisting of letters followed by Arabic numbers (e.g., *LDH-A1** and *LDH-A2**, *LDH-B1** and *LDH-B2**).

Isoloci.—Isoloci are recently duplicated loci that have not yet diverged to the point at which their most common allelic products are electrophoretically distinguishable. Isoloci are known in many ancestrally tetraploid species (e.g., salmonids). Isoloci should be identified by a comma between the two numbers designating the loci involved in the isolocus pair (e.g., *sAAT-1,2**; *sMDH-A1,2**; *sMDH-B1,2**; *GPI-B1,2**).

Regulatory locus suffix.—Many loci affect the tissue-specific, ontogenetic, physiological, or molecular expression of other loci; in a broad sense, these all may be called regulatory loci or genes. Regulatory loci should be denoted with lowercase italic "r" suffixes attached to the symbol of the locus they regulate (e.g., *GPIr** and *LDH-Ar** are loci that regulate the expressions of the *GPI** and *LDH-A** loci). When more than one regulatory locus is known to affect a structural gene locus, they are assigned Arabic numbers in order of their discovery (*PGM-1r1**, *PGM-1r2**, *PGM-1r3**, etc.).

Pseudogene and cross-hybridizing sequence suffix.—Sequences of DNA that are not transcribed (to messenger RNA and thence to proteins), but that are homologous to transcribed loci, generally are termed pseudogenes. When pseudogenes are

identified and well characterized, they should be named in relation to the transcribed loci they resemble, and their gene symbols take a lowercase "*p*" suffix: *LDH-Ap** would be the documented pseudogene of *LDH-A**. If evidence for homology is unavailable, the appropriate suffix is lowercase letter "*l*" for "-like": *LDH-Al** denotes a locus with sequence similarity to *LDH-A**. As noted previously, authors should specify whether the letter "*l*" implies similarity at the protein or DNA level, because this notation can be ambiguous otherwise.

Allele Symbols

Three alternative systems of allele designation are acceptable. Alleles may be identified by Arabic numbers (allele **1*, allele **2*, etc.), lowercase letters (allele **a*, allele **b*, etc.), or relative electrophoretic mobilities (allele **100*, allele **50*, allele **175*, etc.).

Symbols for alleles are always italicized (underlined in manuscript) and should be preceded by an asterisk (e.g., the *MDH-B*1* allele; the *ADA-1*a* allele; the *EST-2*150*, *EST-2*75*, *EST-2*60*, and *EST-2*43* alleles). Once the gene and allele symbols have been identified in the text, the italicized allele symbols (including the asterisks) can be used without the gene symbols provided such usage is clear and unambiguous (e.g., "four variant alleles of *EST-2** were observed: **1*, **2*, **4* and **6*" [or **100*, **75*, **150*, and **43*]).

Of the three acceptable systems for allele designation, the committee recommends the use of sequential number codes. Number (and letter) codes are somewhat independent of observed relative electrophoretic mobilities, and usually are unaffected by minor variations in procedure (among laboratories) or major differences in technique (buffer choice, etc.) that affect protein mobilities. In addition, number (and letter) codes can be used to identify all types of allelic variation—whether or not they result in detectable changes in electrophoretic mobility. Number codes allow more alleles to be designated than letter codes, which are limited by the size of the alphabet. Furthermore, use of Arabic number codes is consistent with one of the two acceptable conventions for allele designation in the human gene nomenclature.

Nevertheless, authors have the option to use any of the three acceptable systems of allele designation. As for isozymes and loci, we encourage authors to adopt a system of allele designation that is as consistent as possible with that used in previous studies of the same loci and taxa—except when the previous system is considered confusing or inadequate—and to indicate the correspondence of alleles between studies when more than one nomenclature has been used commonly for the taxa under study.

Sequential number or letter systems.—Designation of allelic products with either sequential Arabic number codes (starting with "1") or lowercase letters (starting with "a") can be especially useful in taxonomic surveys when several taxa exhibit different allelic forms and brevity of data presentation increases clarity. Number or letter codes are also useful for labeling electrophoretically cryptic variants (see below). When alleles are designated by number or letter, authors should relate the codes to relative mobilities under a set of specified electrophoretic conditions (either in a table in the methodology section or in text). Such numbers or letters are part of the genotype and should be italicized (e.g., the *LDH-A*1*, **2*, and **3* or *LDH-A*a*, **b*, and **c* alleles; the *EST-1*1* or *EST-1*a* allele).

When numeric codes or lowercase letters are used to designate specific allelic forms of a protein (enzyme), they should be presented as superscripts immediately following the isozyme designation in nonitalic type without a preceding asterisk, because they are intended to describe the biochemical phenotype, not the genotype (e.g., for the LDH-A^1 subunit, use LDH-A$^1_4$ to designate the LDH-A homotetramer coded by the **1* allele; the CK-A^b subunit). Note that this system of superscript designations is to be used only for allelic products (proteins) not alleles (genes).

Relative mobility system.—When alleles are identified according to the relative electrophoretic mobility system, all alleles at a locus are designated by the electrophoretic mobilities of the homomeric gene products they encode relative to the electrophoretic mobility of a specified standard (based upon careful, side-by-side comparisons under defined electrophoretic conditions). Acceptable standards include (a) the most common allelic product at the locus in a specified genetic stock or hatchery strain, (b) a designated protein from the same or another species, and (c) a specific dye or other marker molecule. In practice, a relative mobility is calculated by dividing the electrophoretic mobility of the homomeric allele product being described by the electrophoretic mobility of the standard and multiplying the quotient by 100. A minus sign is assigned to any allele product exhibiting cathodal mobility. Inherent in

the use of relative mobilities is the need to document the specific electrophoretic conditions used in the separation and identification of the allelic products (electrophoresis buffers used, including the concentration and pH of each; presence or absence of coenzymes or other factors that affect relative electrophoretic mobilities; etc.).

When relative mobilities are used as allelic designations, they are part of the genotype and should be italicized (e.g., the *LDH-A*100, *125*, and **75* alleles; the *EST-1*75* allele). When relative mobilities are used to designate specific forms of a protein (enzyme) they should be presented in nonitalic type without the preceding asterisk, because they are intended to describe the biochemical phenotype, not the genotype. Such designations can be in either regular or superscript type (e.g., the LDH-A 150 subunit; the LDH-A$_4$ 150 isozyme; the GPI-A^{125} subunit; the HAGH^{75} subunit). As noted above, superscript designations should be used only for allelic products (proteins) not alleles (genes).

In cases of cryptic electrophoretic variation (i.e., when two or more allelic products co-migrate in one buffer system but exhibit different mobilities on a second buffer), allele nomenclature is complicated, especially when the relative mobility nomenclature is used. In cases of cryptic alleles, it is necessary to provide detailed descriptions of the electrophoretic behavior of the different allelic products in all appropriate buffers so that the characteristics and identities of the different forms are clearly documented. We suggest presentation of a table (or a description in text) that outlines the relative electrophoretic mobilities of all such cryptic alleles. Identification of alleles by Arabic number or lowercase letter codes rather than by relative mobilities may be preferable in such cases as a means of reducing confusion.

Alleles at regulatory loci should be identified with Arabic number or lowercase letter codes.

Quantitative variants.—Alleles encoding quantitative variants should be designated by one of the three following codes: "QO" (uppercase letter "O", not zero) for alleles associated with no detectable quantity (activity) of the product ("null" alleles); "QE" for alleles associated with elevated amounts of the protein product or the enzyme's reaction product (depending upon the method of detection); and "QL" for those alleles associated with lowered amounts of the protein or reaction product (e.g., the *PGM-2*100, *150, *QO*, and **25* alleles). When the relative mobility of an allele encoding a quantitative variant can be determined (e.g., by extrapolation from the mobility of a het-

eropolymer containing the product of a null allele but expressing enzymatic activity), the designation of the quantitative variant allele should include both the relative mobility (or Arabic number or lowercase alphabetic code) of the allele product and the quantitative variant symbol as a suffix to the basic allele code (e.g., *GPI-B*24QO, sIDHP-3*2QE, sMDH-2*aQL*).

Phenotypes and Genotypes

Genotypes should be underlined in manuscripts and printed in italics in publications. In addition, we strongly recommend the use of an asterisk between the locus and the first allele in a genotype. Subsequent alleles in the genotype need not be preceded by asterisks. Note this convention differs from the human gene nomenclature recommendations, which specify that each allele in the genotype be preceded by an asterisk, but we believe the resulting simplification and shortening of genotypes (especially for isoloci) without loss of information justifies this change. Multiple alleles in a genotype should be separated by forward slashes: *LDH-A*107/100, EST-1*100/55, sMDH-2*a/ b, sAAT-1,2*130/100/100/80*, etc.

Phenotypic designations should have the same characters as genotypic symbols; however, phenotypes should be presented in nonitalic type, and the asterisk separating locus and allele characters in genotypes is omitted in phenotype designations. As with genotypes, a forward slash between the allelic forms should be used in phenotypic designations: MPI 1/2; PEPA a/c; LDH-A 107/100; PGM-2 220/100.

Authors should note that gene symbols are italicized (or underlined) and should end with an asterisk (e.g., *LDH-A*, EST-1*, mMDH-2*, MPI**), whereas enzyme and protein abbreviations are not italicized and do not carry asterisks (e.g., LDH-A, EST-1, mMDH-2, MPI). Care should be taken to use the proper forms throughout a manuscript to distinguish between gene loci and their protein or polypeptide products.

When reference is made to multimeric proteins, it is customary to designate the precise subunit composition, when known, by using accepted subunit designations with the appropriate numeric subscripts (e.g., LDH-A$_4$ for the homotetramer containing four A-type subunits, LDH-A$_2$B$_2$ for the heterotetramer containing two A-type and two B-type subunits, etc.).

Linkage and Phase

The human gene nomenclature contains detailed instructions (with examples) for designating

TABLE 2.—Summary of recommended nomenclatural conventions for protein-coding loci in fish.

Feature	Example	Comment
Proteins		
Enzyme name and number	Aspartate aminotransferase, 2.6.1.1; Glucose-6-phosphate isomerase, 5.3.1.9; L-lactate dehydrogenase, 1.1.1.27	Recommendations of the International Union of Biochemistry Nomenclature Committee (IUBNC) should be followed
Enzyme abbreviations	AAT; GPI; G3PDH; G6PDH; LDH	Derived from IUBNC-recommended names. Internal hyphens and substrate isomer letters and numbers should be deleted if this will not cause ambiguity
Molecular designations of multimeric enzymes	$MDH-A_2$; $MDH-B_2$; MDH-AB $LDH-A_4$; $LDH-A_2B_2$	Subscripts identify subunit composition
Protein abbreviations and gene symbols		
Proteins	IDHP; MPI; sMDH-B	Uppercase, nonitalic (with lowercase prefix as appropriate)
Gene locus	*IDHP**; *MPI**; *sMDH-B**	Same as protein abbreviations except loci are italicized (underlined in manuscripts) and preferably end with an asterisk
Prefixes and suffixes (illustrated for gene loci only)		
Isomer prefix	*αAMY** or *aAMY**; *βGAL** or *bGAL**	Either the lowercase Greek letter or its Roman abbreviation should be used (a = alpha, b = beta, etc.)
Subcellular prefix		
Mitochondrial (m)	*mMDH-1**; *mSOD**	Used only when two or more localizations
Cytosolic (s)	*sMDH-B**; *sSOD**	are known for an enzyme
NAD(P) suffix		
NAD⁺-dependent enzyme	*IDH**	No additional symbol
NADP⁺-dependent enzyme	*IDHP**	Suffix is a capital "*P*"
Single encoding locus	*ADH**; *MPI**	No additional symbol
Multiple encoding loci		Multilocus designation is hyphenated
Unknown relationships	*EST-1**; *EST-2**	Suffix is an italic Arabic number
Known orthologies	*LDH-A**; *LDH-B**	Suffix is an italic capital letter
Recent gene duplicates		
Unknown evolutionary relationships	*sIDHP-1**, *sIDHP-2** *mAH-1**, *mAH-2**, *mAH-3**, *mAH-4**	Assign sequential italic Arabic numbers
Known evolutionary relationships with established orthologies	*CK-A1**, *CK-A2**; *CK-C1**, *CK-C2**	Add italic Arabic number after letter
Isolocus designation	*sIDHP-1,2**; *GPI-B1,2**	Component loci are separated by a comma
Pseudogene suffix	*LDH-Ap**	Suffix is a lowercase "*p*"
"Like" suffix	*XDHl**	Suffix is a lowercase "*l*" (el)
Regulatory locus suffix	*PGM-1r**; *GPIr**	Suffix is a lowercase "*r*"
Alleles		
Three alternative notations		
Number codes	**1*; **2*; **3*	Alleles are italicized or underlined and preferably begin with an asterisk. Number
Letter codes	**a*; **b*; **c*	codes are Arabic. Letter codes are lower-
Relative mobilities	**150*; **100*; **75*	case. Mobility codes are percentages of the mobility of a specified standard
Quantitative allele suffix	**aQE*, **75QL*; **QO*	*Q* denotes "quantitative"; *E*, "elevated" activity; *L*, "lowered" activity; *O* (capital oh), a null allele
Genotypes (locus plus alleles)		
Individual loci	*GPI-A*1/2*; *G3PDH*175/50*	Always italicized or underlined. Locus pref-
Isoloci	*sIDHP-1,2*a/a/a/b* *GPI-B1, 2*200/200/100/75*	erably is separated from alleles by an asterisk; otherwise, a space is left. Alleles are separated from each other by a slash

linkage relationships among gene loci (Shows et al. 1987, pages 14–15). These guidelines for specifying linkage and phase are good, and we adopt them without change into the fish gene nomenclature.

Alphabetization

Protein, enzyme, and isozyme abbreviations and gene symbols should be alphabetized on the basis of the uppercase letters in the symbol. Lowercase letters in the symbol (e.g., prefixes designating subcellular localization of the protein product of a locus, substrate isomer, etc.) should be ignored for this purpose.

Summary and Future Developments

The major features of the fish gene nomenclature are summarized in Table 2. As stated in the Introduction, the conventions presented in this report have been adopted as editorial standards by the American Fisheries Society, and authors will be expected to conform to them in future Society publications. Questions regarding nomenclatural details and proposals for changes in the nomenclature should be addressed to the chairman of the Nomenclature Committee or brought to the attention of the editor at the time of manuscript submission. The Society's Fish Genetics Section, acting through its Nomenclature Committee, is prepared to maintain, expand, and update this nomenclature to accommodate and incorporate advances in knowledge and experience, and in technology and techniques. Revisions and updates to the fish gene nomenclature will be published from time to time in this journal.

We encourage all fish geneticists to adopt and use this gene nomenclature whenever possible, both informally and in formal scientific and technical publications. To speak a common language, we all will have to change our habits to a greater or lesser extent. A standard gene nomenclature for fish that is close to the one used for humans will, we believe, enhance communication among all geneticists.

Acknowledgments

References

Allendorf, F. W., and F. M. Utter. 1979. Population genetics. Pages 407–454 *in* W. S. Hoar and D. J. Randall, editors. Fish physiology, volume 8. Academic Press, New York.

Avise, J. C., and G. B. Kitto. 1973. Phosphoglucose isomerase gene duplication in the bony fishes: an evolutionary history. Biochemical Genetics 8:113–132.

Bailey, G. S., A. C. Wilson, J. E. Halver, and C. L. Johnson. 1970. Multiple forms of supernatant malate dehydrogenase in salmonid fishes. Journal of Biological Chemistry 245:5927–5940.

Buth, D. G. 1983. Duplicate isozyme loci in fishes: origins, distribution, phyletic consequences, and locus nomenclature. Isozymes: Current Topics in Biological and Medical Research 10:381–400.

Dando, P. R. 1974. Distribution of multiple glucose-phosphate isomerases in teleostean fishes. Comparative Biochemistry and Physiology B, Comparative Biochemistry 47:663–679.

Fisher, S. E., J. B. Shaklee, S. D. Ferris, and G. S. Whitt. 1980. Evolution of five multilocus isozyme systems in the chordates. Genetica 52/53:73–85.

Fisher, S. E., and G. S. Whitt. 1978. Evolution of isozyme loci and their differential tissue expressions: creatine kinase as a model system. Journal of Molecular Evolution 12:25–55.

Frick, L. 1983. An electrophoretic investigation of the cytosolic di- and tripeptidases of fish: molecular weights, substrate specificities, and tissue and phylogenetic distributions. Biochemical Genetics 21:309–322.

Frick, L. 1984. Isolation and characterization of five cytosolic di- and tripeptidases from the skipjack tuna (*Katsuwonus pelamis*). Comparative Biochemistry and Physiology B, Comparative Biochemistry 77:533–540.

IUBNC (International Union of Biochemistry, Nomenclature Committee). 1984. Enzyme nomenclature. Academic Press, Orlando, Florida.

Kendall, R. L. 1990. Towards a standardized genetic nomenclature for fish. Transactions of the American Fisheries Society 119:1.

Kitto, G. B., and R. G. Lewis. 1967. Purification and properties of tuna supernatant and mitochondrial malate dehydrogenases. Biochimica et Biophysica Acta 139:1–15.

Markert, C. L., J. B. Shaklee, and G. S. Whitt. 1975. Evolution of a gene. Science (Washington, D.C.) 189: 102–114.

Scholl, A., and H. M. Eppenberger. 1972. Patterns of isoenzymes of creatine kinase in teleostean fish. Comparative Biochemistry and Physiology B, Comparative Biochemistry 42:221–226.

Shaklee, J. B., F. W. Allendorf, D. C. Morizot, and G. S. Whitt. 1989. Genetic nomenclature for protein-coding loci in fish: proposed guidelines. Transactions of the American Fisheries Society 118:218–227.

Shaklee, J. B., K. L. Kepes, and G. S. Whitt. 1973. Specialized lactate dehydrogenase isozymes: the molecular and genetic basis for the unique eye and liver LDHs of teleost fishes. Journal of Experimental Zoology 185:217–240.

Shows, T. B., and 24 coauthors. 1987. Guidelines for human gene nomenclature. An international system for human gene nomenclature (ISGN, 1987). Cytogenetics and Cell Genetics 46:11–28.

Appendix: Reviewer Comments and Committee Responses

Twenty-five scientists wrote to the Nomenclature Committee about the published draft guidelines (Shaklee et al. 1989). They all broadly supported the establishment of a standardized nomenclature for protein-coding loci in fish, but several of them differed with the committee's proposal in particular respects. Several suggestions led to useful refinements in the nomenclature and have been incorporated in the system. Others also had merit but we felt we could not adopt them without compromising the system's universality. Below, we briefly discuss the topics that earned most of the reviewers' comments, and summarize the committee's position on each.

Relationship to human gene nomenclature. — Eight reviewers specifically mentioned that they favored a nomenclature for fish that was similar to, and consistent with, the present one for human genes. A ninth reviewer, although not explicitly supporting conformance with the human system, emphasized the desirability of standardizing gene nomenclature across vertebrates (and eventually among all organisms). No correspondent opposed bringing the fish gene nomenclature as close as possible to the human one.

The committee feels strongly that the eventual adoption and use of a single, unified system of gene nomenclature will benefit not only research in fish genetics, but all of the biological sciences. We chose the human gene nomenclature (Shows et al. 1987) as our model because the field of human genetics is preeminent, because greater funding and more staff are dedicated to this taxonomic group than any other, because human genetics research is likely to exert an increasingly dominant effect on all of genetics because of the Human Genome Project, and because human geneticists have led the way in developing a standardized gene nomenclature. Therefore, we deviated from the human system only when it did not adequately accommodate specific situations encountered with fish or when substantial improvements in it seemed possible. In this way, we hoped to maintain maximal consistency with the human gene nomenclature yet meet all identified needs of fish genetics.

Locus designations. — Five reviewers disliked use of the asterisk (*) to identify loci. They felt that it is an unnecessary symbol, that it complicates the reading of text, or that it will be editorially unacceptable in certain journals because asterisks have another function there. In contrast, two reviewers specifically encouraged use of the asterisk because they felt it provides an unambiguous distinction between loci and proteins and also serves to identify alleles clearly. Five respondents opposed the use of italicized, all-uppercase letters in gene symbols and recommended that the first letter be uppercase and subsequent letters lowercase. They felt that such mixed-case designations for loci are already common in the fish genetics literature and that such designations are easier to read in text than all-uppercase gene symbols. Two reviewers suggested that typographic symbols such as hyphens, periods, or commas in locus designations might cause difficulties with certain computer software (internationally) or lead to ambiguity or confusion when text is typeset in justified columns.

The committee believes that the human gene nomenclature's use of all uppercase italic letters (and numbers as necessary) for loci is as informative and convenient as mixed upper- and lowercase letters (the most common alternative notation). We feel that conformance with this fundamental and highly visible aspect of the human system is important.

The strongly recommended (but not mandatory) use of asterisks as a "flag" in the fish gene

nomenclature to designate loci and to separate, yet connect, loci and alleles is entirely consistent with the human system:

> 3. The allele characters are separated from the locus characters by a new symbol—the asterisk—which serves to combine gene and allele symbols. There should be no spaces between gene, asterisk, and allele, and the entire symbol should be underlined or italicized.
>
> 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.
>
> ... It may be convenient in manuscripts, computer printouts, and in printed text to designate a gene symbol by following it with an asterisk (e.g., *PGM1**). [Shows et al. 1987, pages 13–14.]

We believe that adoption of this simple and distinctive convention will be highly beneficial, although the combined use of italics and asterisks to designate loci (as also recommended in the human system) is somewhat redundant. The redundancy should eliminate any possible confusion between gene loci and the proteins they encode; such confusion often arises from publications on biochemical genetics.

Prefixes and suffixes. —Four reviewers recommended against the use of lowercase prefixes in gene symbols to identify the subcellular localization of products or substrate isomers. They felt that such prefixes will complicate alphabetization of loci in tables and lists, will interfere with the efficient use of computerized databases, or will make gene symbols too long and complicated to be useful. Two of these reviewers suggested that this information be omitted from the gene symbol itself and be presented in tabular form in each publication. One reviewer suggested that the prefixes should instead be suffixes to avoid the alphabetization problem and to make the gene symbols for fish look more like those used in the human gene nomenclature. One reviewer supported the use of prefixes in gene symbols because of their high information content.

The fish gene nomenclature diverges from the human system with respect to prefixes. The human gene nomenclature specifically prohibits use of lowercase letters and prefixes in gene symbols, and requires conversion of prefixes to uppercase

Roman suffixes. Further, loci that encode cytosolic and mitochondrial protein forms are generally distinguished in the human system by Arabic suffixes: 1 for the cytosolic form, 2 for the mitochondrial form (e.g., *ACO1** and *ACO2**).

We see several problems with these human conventions. First, many multilocus systems in fish consist of two or more genes that encode cytosolic isozymes (e.g., CK, GPI, G3PDH, LDH, TPI) or of multiple genes that encode both cytosolic and mitochondrial isozymes (e.g., AAT, AH, IDHP, MDH). In such cases, use of a 1 to identify the single gene encoding the one cytosolic form and of a 2 to identify the single gene encoding the one mitochondrial form is clearly inadequate. Furthermore, the profound metabolic and evolutionary differences between cytosolic and mitochondrial forms of most of these enzymes demand easily recognizable and distinctive symbols for these different gene loci. The mnemonic, single-letter prefixes (s = cytosolic [supernatant]; m = mitochondrial; etc.) of the fish gene nomenclature serve this purpose. In combination with multilocus designations, these prefixes offer a powerful and flexible means to clearly identify multiple loci that encode enzymes in the same subcellular compartment; the system easily accommodates multiple cytosolic or mitochondrial forms without loss of information or confusion. Prefixes already have been used in the fish genetics literature to identify subcellular localization, so the practice is not a new departure for gene nomenclature.

We also believe that lowercase prefixes should be used to distinguish between gene loci that encode enzymes that act on different isomeric forms of substrates (e.g., *aGAL** and *bGAL** for the alpha galactosidase and beta galactosidase loci). Commonly used prefixes for substrate isomers (a, b, g, d, e) do not presently overlap with those for subcellular localization (l, m, p, s), so there should be no ambiguity about the meaning of prefixes even when two are required.

It is a general rule in organic chemistry that names are alphabetized before prefixes are inserted; the prefixes only affect the order of chemicals with the same stem name. If the same principle is applied to gene and protein symbols, prefixes should cause no problem with alphabetical listings of these codes. We recommend adoption of this rule.

In the human gene nomenclature, uppercase suffixes are added to gene symbols to distinguish pseudogenes (P), less-well characterized "like" genes (L), and regulatory genes (R). We have retained these mnemonic letters in the fish gene no-

menclature, but in lowercase versions. As with prefixes, lowercase suffixes can be incorporated into a gene symbol without confusing or complicating its meaning. For example, in the fish gene nomenclature, *ALP** and (*ALPp**) are clearly the structural gene encoding alkaline phosphatase and an alkaline phosphatase pseudogene, *ADH** and *ADH1** are the structural gene for alcohol dehydrogenase and a presumably related alcohol dehydrogenase-like gene, and *GR** and *GRr** are the structural gene encoding glutathione reductase and a glutathione reductase regulatory gene. Lowercase suffixes do not affect the basic uppercase structure of the locus symbols, which therefore remain easily recognizable. Symbols such as *ACPp** (an alkaline phosphatase pseudogene), *MPIl** (a mannose-6-phosphate isomerase-like gene), and *GPIr** (a glucose-6-phosphate isomerase regulatory gene) are more easily recognized and understood than their counterparts in the human system (*ACPP**, *MPIL**, and *GPIR**).

Designation of multiple loci. —Four reviewers commented on the system for designating multiple loci with numbers when orthologies have not been clearly established, and with letters when orthologies are known. One reviewer strongly endorsed this system. A second stated that studies to establish orthologies and paralogies are very important and should be strongly encouraged. Two reviewers expressed concern that these guidelines might "encourage sloppy science" if they lead researchers to guess at locus identifications rather than indicate uncertainty. The committee encourages researchers to indicate uncertainty by use of numbers when letters might otherwise be appropriate, by omission of the subcellular prefix when localization is ambiguous, or by use of the "-like" designation when the true identity of the protein is unknown. Two reviewers indicated a need for formal nomenclatural recommendations regarding the assignment of numbers to multiple loci to achieve some consistency among publications. The fundamental question here is whether to begin numbering at the anodal or cathodal end of the gel (or at the origin).

The human gene nomenclature specifies that multiple related loci ". . . are designated by Arabic numerals placed immediately after the gene symbol, without any space between the letters and numbers used" (Shows et al. 1987). We feel that such juxtaposition of letters and numbers detracts from the recognition of locus symbols. For this reason, the fish gene nomenclature specifies that such "multilocus designations" (numbers or let-

ters) be separated from, but connected to, the locus stem by a hyphen (a commonly used convention in genetics). The fish nomenclature also specifies the use of single capital letters, not numbers, when evolutionary relationships within multiple-locus systems are known. This endows gene symbols with considerably more information and is consistent with much of the biochemical genetics literature.

We strongly encourage studies to elucidate orthologies and parologies within multilocus systems. Publication of such studies, with recommendations for appropriate modifications of existing gene designations, will contribute to the growth and improvement of the fish gene nomenclature.

The use of lowercase prefixes and suffixes and hyphenated multilocus designations in gene symbols has several advantages. These conventions facilitate the unambiguous identification of gene loci while preserving the clear meaning of each locus stem, they convey important information efficiently, and they allow easy alphabetization for grouping of related loci in lists and tables. The committee believes these strengths justify deviations from the human system for these aspects of gene nomenclature.

Allele designations. —Two reviewers objected that the designations for quantitative alleles (QE, QL, and QO) are too cumbersome; one reviewer specifically supported them. Two reviewers favored using superscripts to designate alleles.

We adopted the human conventions for alleles that affect the quantitative expression of a gene product because we feel this information is valuable and we could not devise a better way to provide it. We oppose superscripts as allele designations, although these occur in some literature, because they are typographically cumbersome, prohibited in the human system, and no more readable or informative than the alternative formats recommended.

Greek letters. —One reviewer strongly supported the use of Greek letters in gene symbols, another strongly opposed it.

The fish gene nomenclature allows the use of either Greek letters or their Roman surrogates in enzyme abbreviations and gene symbols. However, we encourage use of Roman letters because this conforms more closely with the human nomenclature standard and should facilitate computerized keyword searches of literature data bases.