分子植物育种(网络版)，2011 年，第 9 卷，第 1017−1030 页
Fenzi Zhiwu Yuzhong (Online), 2011, Vol.9, No.1017−1030
http://mpb. 5th.sophiapublisher.com

水稻基因命名法系统

翻译：巩鹏涛
译者单位：东北林业大学盐碱地研究中心；海南省农作物分子育种重点实验室

分子植物育种, 2011 年, 第 9 卷, 第 3 篇　　doi: 10.5376/mpb.cn.2011.09.0003
收稿日期：2010 年 10 月 12 日
接受日期：2010 年 12 月 01 日
发表日期：2011 年 01 月 19 日

引用格式：
Susan R. McCouch & CGSNL (Committee on Gene Symbolization, Nomenclature and Linkage, Rice Genetics Cooperative), 2008, Gene Nomenclature System for Rice, Rice, 1(1):72-84 (doi:10.1007/s12284-008-9004-9)

Gene Nomenclature System for Rice

Susan R. McCouch [1], CGSNL (Committee on Gene Symbolization, Nomenclature and Linkage, Rice Genetics Cooperative) [2]
1. Department of Plant Breeding and Genetics, Cornell University, 162 Emerson Hall, Ithaca, NY 14853-1901, USA
2. International Rice Research Institute (IRRI), DAPO Box 7777, Metro Manila, Philippines
Corresponding author, srm4@cornell.edu;　Authors

Rice, 1(1):72-84, doi : 10.1007/s12284-008-9004-9
本文完整的电子版本可以从以下网址获得：http://www.springerlink.com/content/q047075330507t73/

研究背景

生物学界一直梦想有一套统一的基因命名系统。目前已经有多个基因命名法系统来描述：拟南芥(TAIR, 2005)、番茄(http://www.sgn.cornell.edu/documents/solanaceae-project/docs/tomato-standards.pdf)、玉米(http://www.maizegdb.org/maize_nomenclature.php)、苜蓿(VandenBosch and Frugoli, 2001)、酵母(http://www.yeastgenome.org/gene_guidelines.shtml)、小鼠(MGNC, 2005, http://www.informatics.jax.org/mgihome/nomen/)和人类(Wain et al., 2004)。一种跨越不同物种间通用遗传语言的使用将是科学界的一个巨大进步，将极大的便利于物种间的基因和遗传变异的结构、功能和进化比较。随着对基因和基因产物分子和生化特性的重视，建立一套能反映特定基因、基因模型或基因家族生化特点和在特定遗传背景下一个特定等位基因的表型结果的水稻基因命名法系统，显然十分重要。

目前水稻基因名称和基因符号规则是基于水稻遗传协作组织(CGSNL)中的基因符号、命名法和连锁委员会(Kinoshita, 1986)的建议制定的。绝大多数早期的基因命名和符号是对可见表型的描述，为该基因的存在提供最早的证据，这些基因名称和符号被水稻研究领域普遍使用。随着水稻全基因组测序的完成(IRGSP, 2005)，和新的鉴定、定义和描述基因的新方法不断涌现，概述一套描述基因的标准程序的命名法系统成为了迫切需要，描述是以生化特性和 DNA，RNA 和蛋白序列分析(Wu et al., 1991)，及以前规范基因和表型联系的规则(Kinoshita, 1986)为基础。

本文集中总结了水稻的基因命名法规则，尽一切可能来在水稻基因命名法系统和其他模式物种的之间协调一致。本文种描述了一套染色体命名和基于生物学功能、突变表型和序列鉴定座位、基因和等位基因的规则，并就如何处理不同来源的多种基因组组装注释种中的别名(同义名)、序列差异和座位提出了建议。这个命名法规则是以原有的水稻基因命名法系统(Kinoshita, 1986)为基础的，但新的命名法系统被扩展来吸收国际水稻基因组测序项目(International Rice GenomeSequencing Project, IRGSP)成员在两次水稻注释项目(RAP)会议(RAP−1, 2004 年 12 月, 日本, 筑波; RAP−2, 2005 年 12 月, 菲律宾, 马尼拉)总结的有关序列信息的建议。这些规则也已经获得了水稻遗传协作组织中 CGSNL 小组委员会的讨论通过(http://www.shigen.nig.ac.jp/rice/oryzabase/

rgn/office.jsp)。

尽管有记录以来水稻遗传研究已经有超过一个世纪的时间，但最近在籼稻(*Oryza sativa* ssp. *indica*)和粳稻(*Oryza sativa* ssp.*japonica*)大规模诱变实验和EST测序方面取得的进展，大大的增加了我们对基因网络、基因功能、等位基因和序列多态性的认识。因此，本次报告中概述的命名法主要是归纳出以生物功能为基础的基因和等位基因的命名规则，方便多个测序和注释项目中基因注释的交叉参考，这些项目包括：国际水稻基因组测序计划(IRGSP) (IRGSP, 2005)、水稻注释计划(RAP) (Ohyanagi et al., 2006)、美国基因组研究所(TIGR) (Yuan et al., 2005)、慕尼黑蛋白质序列信息中心(MIPS) (Karlowski et al., 2003)、美国国家生物技术信息中心(NCBI) (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=4530)、先正达(Syngenta) (Goff et al., 2002)和北京基因组研究所(BGI) (Zhao et al., 2004)，同时这将为来自不同种质资源测序(Ammiraju et al., 2006; McNally et al., 2006)中存在的基因差异体的注释提供内在连贯性。

1基因组组装和系统座位标识码(systematic_locus_ID)

一个简单的水稻物种也许可以支撑多重的遗传、物理、序列图谱、基因注释和基因组组装。目前水稻(*O.sativa*)基因组分别被粳稻栽培种 Nipponbare 基因组序列(IRGSP 测序)和籼稻栽培种 93−11 基因组序列(BGI 测序)所代表。Nipponbare 的序列已经被几个研究小组进行了注释，包括了 RAP (Itoh et al., 2007)、(Ohyanagi et al., 2006)、TIGR (Yuan et al., 2003)、NCBI-GenBank (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=4530)、MIPS (Karlowski et al., 2003)和先正达(Goff et al., 2002)，然而栽培种 93−11 的序列注释工作几乎全部来自 BGI (Zhao et al., 2004)。对 Nipponbare 来说，IRGSP 测序得到的原始序列数据被来 RAP 和 TIGR 各自独立组装和注释，因此水稻研究界目前管理着三个独立的基因组组装(两个来自栽培种 Nipponbare 和一个来自栽培种 93−11)。

这些组装代表着各自一套相互之间独立并有细微差别的对座位的注释，这些座位代表着沿假设分子锚定排列的基因模型/转录单元。一个座位定义为基因组上的一个位置，因为每个注释小组都独立的依据在假设分子上的位置，分配座位标识符(locus IDs)给所有的基因、转录本和蛋白。相同的基因可能因基因组、组装和注释软件的不同而赋予了不同的系统座位 ID (systematic_locus_ID)。每个注释小组使用的为核基因/转录本/蛋白、细胞器基因/转录本/蛋白和转座本分配的系统座位 IDs 规则，在 RAP 数据库(Ohyanagi et al., 2006)、TIGR Osa1 数据库(Yuan et al., 2005)和 BGI-RIS (Zhao et al., 2004)中有具体描述。来自(RAP)数据库的有关分配系统座位(IDs)的引证例子相关建议在本文的结尾部分有列举。

注释的基因包括了蛋白质编码基因(open reading frames, ORFs/CDSs)、非编码RNA基因(ribosomal RNA (rRNA), 转移RNA (tRNA), 微RNA, 小干扰RNA (siRNA)，小核RNA (snoRNA)等)和假基因组。系统座位IDs的使用(将在本文后部详细描述)为基因标识符的分配提供了一个系统的方法，同时也为座位在已测序水稻基因组的位置提供了容易的识别。作为结果，座位ID可以被用来鉴定和在一个特定的基因组组装中追踪一个座位，在一个基因模型和功能注释基因之间建立关联。目前大多数的序列和注释基因是未知功能(实验确认)的，系统座位ID也为跟踪这些假基因的功能提供了一个有用的办法。如在表1总结的，基因可以根据计算机鉴别的序列与已知基因(推测的同源基因，直系同源基因或旁系同源基因)、蛋白或共有序列的特征(像某蛋白的功能结构域)的相似性来进行分类。当序列的相似性不足以保证基因名的分配时，对基因特征的记述信息就做出了关键的贡献。尽管系统座位标识符在一个基因组组装和注释数据集种提供独有的命名法，但不同注释小组使用的分配座位IDs的方法有细微的差别，加上基因组组装和基因指令表的差别(亦即籼稻和粳稻)，使得不同基因组组装版本之间最终的基因和座位交叉参考变得十分困难。因此，随着基因功能或表型的实验证实和描述，CGSNL就提供了一个统一的独立于不同基因组组装和注释版本之外的基因追踪系统。正如下边描述的，每一个在CGSNL注册的基因都可以通过基因的全称和一个基因符号获得独一无二的鉴别。

在CGSNL数据库登记基因将有助于多个注释系统，及等位基因和序列变异体之间的基因的交叉

表 1 CGSNL 建议的已测序基因分类规则

Table 1 Rules for Classifying Sequenced Genes as Suggested by the CGSNL

类型 Categories	分类法 Classification	标准手册 Standard protocol	描述 Description
类型Ⅰ Category Ⅰ	和已知功能水稻蛋白相同 Identical to rice protein with known function	和已知水稻蛋白的相同性>=98%,长度覆盖度=100% [blastx] Identity > = 98%, length coverage=100% to known rice protein [blastx]	赋予相同原始的基因名称 Receive the same, original gene name
类型Ⅱ Category Ⅱ	与已知蛋白相似 Similar to a known protein	与已知蛋白相似性>=50% [blastx] Identity > = 50% to a known protein. [blastx]	赋予“原始推测基因名称” Receive “original gene name, putative”
类型Ⅲ CategoryⅢ	InterPro蛋白结构域 InterPro domain-containing protein	不在分类Ⅰ或 II, 但包含 InterPro 结构域 Not in category Ⅰ or Ⅱ, but contains InterPro domain	赋予“InterPro 名蛋白结构域名称” Receive “InterPro name domaincontaining protein”
类型Ⅳ CategoryⅣ	保守假设蛋白 Conserved hypothetical protein	与假设蛋白的相同性>=50%, 长度覆盖度>=50% [blastx] Identity > = 50%, length coverage > = 50% to hypothetical protein [blast x]	赋予“保守假设蛋白” Receive “conserved hypothetical protein”
分类Ⅴ Category Ⅴ	假设蛋白 Hypothetical protein	不属于分类Ⅰ到Ⅳ If no t in category Ⅰ to Ⅳ	赋予假设蛋白 Receive “hypothetical protein”

注: 像 CGSNL 建议的, 上述了一个根据与已报告研究基因的序列相似性来对测序基因进行分类的系统; 在粳稻 Nipponbare 基因组中存在的预测或已知的基因根据序列分析可以分为五类(栏 1); 只有在有充实的实验证据证明一个基因在序列上和一个已知功能水稻基因相同, 基因才被分配给一个基因名和基因符号(分类Ⅰ); 如果证据支持不充分以分配给其一基因功能(分配为分类Ⅱ-Ⅴ), 这个基因名称字段将留空, 描述或定义字段(栏 2 和 4)将被用来记录有关这个基因的特征和阐释

Note: This describes a system for classifying sequenced genes into categories based on their sequence similarity to previously reported genes, as recommended by the CGSNL; The genes predicted and/or known to be present on the *O. sativa* ssp. *japonica* cv. Nipponbare, based on sequence analysis are classified into five categories (column 1). Genes are assigned a gene name and a gene symbol only if there is substantial experimental evidence confirming that a gene is identical in sequence to a previously characterized rice gene of known function (category I). If the evidence is considered insufficient to substantiate assigning a gene function (assigned categories II–V), the gene name field is left empty and the description/definition field (columns 2 and 4) is utilized to document what is known about the characteristics of the gene

参考。审订过基因名和符号的基因将和一个基因功能或表型相关联，在一切可能的情况下，在登记的时间研究人员被要求为来自RAP注释数据库的新基因标示一个系统座位标识符。在任何可能的情况下，为系统座位IDs在其他注释数据库(即TIGR, BGI等)建立联系。因此，在注册登记的时间，当一个系统座位标识符被提供给CGSNL的时间，组装和注释的版本也必须被储存以提供全面的细致的相关档案。如果可能，GenBank或DDBJ登录号也应该被提供。这将有助于确保在表2描述的交叉参考信息的正确和适用性。

水稻假分子之间的交叉索引需要人工认真的选择和整理。就旁系同源基因，特别是当串联排列时，基因模型结构中在多套相同基因组序列组装版本之间存在的细微差别都会带来巨大的挑战。如果研究人员对一个新基因的这些特定性状谙熟，自己将处于最有利的位置为提供这个基因准确信息方面，这将确保不同水稻基因组注释差别之间的渐进改进和更新。

2水稻中染色体名和基因符号命名规则

应这个命名法系统的需要，一个基因被定义为一段有已知或预测功能或表型的DNA序列。那些测序的基因如果没有实验证明的明确功能或表型，CGSNL将不会分配其一个基因名或基因符号(图1)。那些功能或表型已经被经典遗传学确认的基因，但如果没有和其关联的序列，将分配给一个基因名和基因符号，但也许不会分配给一个系统座位ID。

2.1染色体名称

以Khush和Kinoshita (Khush and Kinoshita, 1991)

表 2 *SD1* 基因的例子及其相关信息

Table 2 Example of the *SD1* Gene and Its Associations

物种 Species	*Oryza sativa*
基因符号 Gene symbol	*SD1*
基因名称 Gene name	*SEMIDWARF 1*
基因同义物 Gene synonym(s)	*dee-geo-woo-gen dwarf, d49, d47, green revolution gene, C20OX2, GA C20oxidase2, GA20 oxidase, Gibberellin-20 oxidase*
图谱位置 Map location	
序列图谱 Sequence maps	
RAP数据库(版本#4) RAPdb (build #4)	Os01g0883800 (*O. sativa* ssp. ssp. japonica cv. Nipponbare)
TIGR_osa1(版本#4) TIGR_osa1 (build #4)	LOC_Os01g66100 (*O. sativa* ssp. japonica cv. Nipponbare)
BGI_RIS	OsIBCD004089 (*O. sativa* ssp. indica cv. 93–11)
遗传图谱 Genetic maps	JRGP RFLP map: sd1, linkage group−1, 149.1–151 cM 水稻形态图谱 sd1, linkage group−1, 73 cM Rice morphological map: sd1, linkage group-1, 73 cM
引用 Citation	PMID: 12077303, 11961544, 11939564, etc.
GenBank 登录号 GenBank accession number	AB077025, AF465255, AF465256, AY114310, U50333
Uniprot 登录号 Uniprot accession number	Q8RVF5, Q8S492, Q0JH50, Q2Z294

提出的约定为基础，水稻的12条核染色体赋予了阿拉伯数字，连锁群与染色体建立了对应关系，并得到命名。为方便数据库建设的目的，每个染色体被赋予了01到12的一个两位数字符，但1到9的单位数字经常在出版物中出现。断臂和长臂分别标示为“S”和“L” (例如：1S，1L)，因此chr.1S和chr.1L或Chr.2S和Chr.2L的缩写是被接受的。尽管由于当时评估染色体大学和臂比技术的准确性不够，染色体和染色体臂命名的约定目前公认有不一致性，但本次并没有对存在的染色体命名法提供修订建议。环状的线粒体或质体和叶绿体染色体分别分配了英文符号“Mt”和“Pt”，而没有使用类似核染色体中的阿拉伯数字。这些不具有着丝粒的染色体将不被指定短臂或长臂。这些染色体可以缩写为chr.Pt或Chr.Mt。

2.2基因全称

一个基因全名包括了基因名和这个基因座的标识符编号。基因名全称所有字母大写斜体，在名称和座位编号之间有一个空格(即, *SHATTERING 1*)。这个基因名应该简要的描述与基因产物的生化功能或由于突变或基因的等位基因形式所呈现出的表型所关联的显著特征。基因座位编号由 1 到 3 位数的数字组成，用来区分一个特定基因座位上的基因与其他座位上基因的具有相似的功能和表型。座位标志符使用的数字表明一个特定基因或基因家族的被鉴定的次序，不应该和系统座位 ID 或其所在的染色体/连锁群混淆。默认的情况是，任何基因名如何没有一个座位标志符，就代表这个基因是第一被鉴定的此类基因，将分配给座位标志符“1”，例如，*PURPLE NODE* 就被标示为 *PURPLE NODE 1*。基因全称的写法是全部大写、斜体，这和先前的规则：基因全称第一个字母大写代表第一个等位基因显性，小写代表隐性；其他所有的其他字母小写、斜体。有关此的更多内容，请参考“显性/隐性关系”章节。

当一个表型定位到一个复杂座位，这个座位包

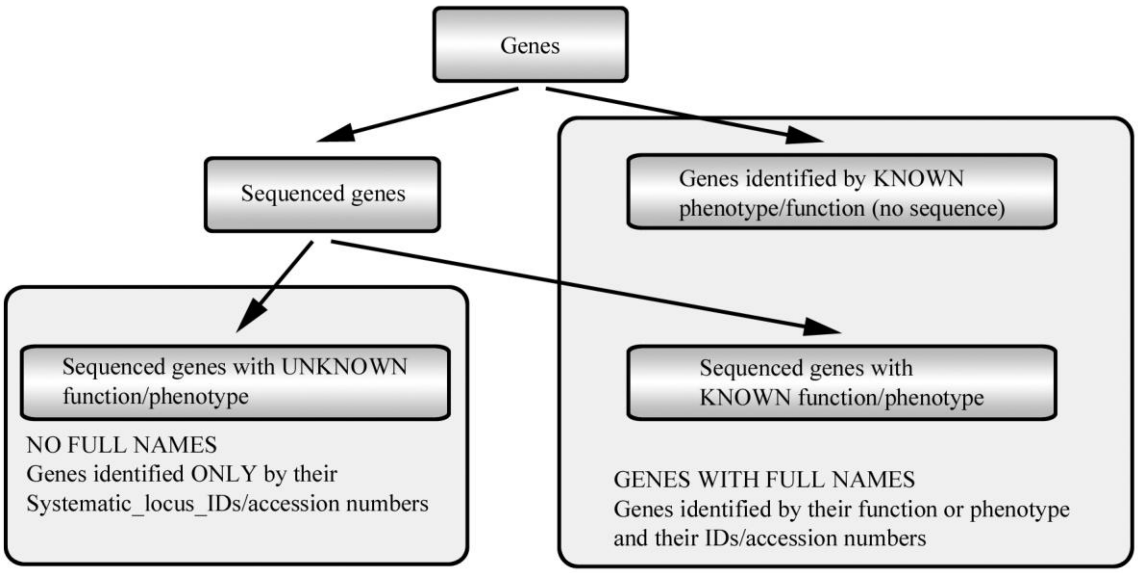

图1 基因赋予全称的图示"基因"是指水稻的全套基因

注: “测序基因”可能来自有一个完全测序的基因组或者其他核酸序列数据库集。“已知功能或表型的测序基因”这一亚类会赋予基因名。“已知基因功能和表型, 但没有序列”的基因, 参考没有序列信息的基因, 但将基于其功能和表型赋予基因名, 这些基因通常定位在遗传图谱或物理图谱上。“未知功能和表型的测序基因”(包括预测基因, 有全长cDNA序列的基因等), 因为其没有实验证据证明其功能, 不会被赋予基因名。然而, 不同的水稻基因组注释计划提供的系统座位IDs将是作为这些基因名称的位置标志符存在, 直到这些基因被评估归入已知功能测序基因类别, 此时这些基因将被分配给基因名和基因符号

Figure 1 Schematic representation of genes receiving full names

Note: “Genes” refers to the set of all genes in rice; “Sequenced genes” may be from a completely sequenced genome or other nucleotide sequence data sets. The subset of “sequenced genes of known function/phenotype” receives a gene name; Genes with “known function/phenotype, but without sequence” refers to genes that have no sequence information but do receive a gene name (based on their function/phenotype); often these are mapped on a genetic or physical map; “Sequenced genes of unknown function/phenotype” (this includes predicted genes, genes with full-length cDNA support, etc.) do not receive a name because they do not have experimental evidence supporting their function; However, various rice genome annotation projects provide systematic_locus_IDs that will serve as placeholders for names of these genes until they can be elevated into the category of sequenced genes of known function, at which time they will be assigned a gene name and a gene symbol

含一个串联的基因家族(例如, *XANTHOMONAS ORYZAE PV. ORYZAE RESISTANCE 21, XA21*,或者 *SUBMERGENCE 1, SUB1*)的情况下，这个串联排列基因家族中的每个基因都会被分配给一个独立的座位标识符(即 *SUB1, SUB2, SUB3* 等)。

如有一个基因是通过序列信息得到验证的，这个基因在后边被证明和一个基于表型验证的基因相同(如 Kinoshita(1986)列出的那些)，那么这个基因全称和规则的施用将基于表型，赋予其他名称作为同义名。如果不同的基因使用同样的名或同一个基因使用不同名称造成重复、冗余或混淆，第一个公开出版的基因名将被保留，CSGNL 将和出版发布这个基因的作者联系，赋予这个基因一个新的基因名和基因符号来用于以后对这个基因或座位的使用。质体和叶绿体基因组中鉴定的基因将按照 Uniprot 描述的那样分配名称和符号，线粒体基因组中的基因将按照推荐的规则分配名称和符号(http://ca.expasy.org/cgi-bin/lists?plastid.txt)。

基因的名称是基于实验验证的基因功能或表型效应来分配的。实验证据也许表明了一个分子功能、生物途径中的作用，与另一个基因的相互作用，或与这个基因有联系的表型(图1)。那些基于计算机验证的与同源物、直系同源物、旁系同源物、或一个保守区域如Interpro结构域(Mulder et al., 2005)的序列相似性的基因，只有在有充分的实验证据确认基因功能时才会分配给基因名。2004年11月在日本

筑波市举办的水稻注释计划(RAP−1)会议的参会人员同意，数据库的管理人员可以使用一个标准的“证据分类”系统来表明证据的类型或出版提供的有关核基因注释的实验证据。有关这些分类的描述可以在表1找到。如CGSNL规定:如果证据被认为不足以证明分配的基因功能，这个基因名字段将保留空白，描述/定义字段将被用来对这个基因特征的已知内容的进行描述(表1)。

2.3基因符号

基因符号是基因全称的缩写，用斜体表示。一个基因符号包括两个部分，即基因的分类符号包括2到5个字母，和对应的座位标识符包括1到3位字符。基因符号衍生于先前讨论的基因名全称，因此和基因名全称一样使用相同的座位标识符。基因符号的两部分应该写到一起，中间不留空格、连字符或其他符号(例: *SH1*, *GLH2*)。基因分类号和座位标识符一起组成了基因符号，必须对这个座位和基因组来说是唯一的。基因符号的分配原则就是容易和一个基因名全称对应并辨识。在任何地方，如何存在的标识符如果不能完全符合这个规则，就应该被保留，例如：*C* (*CHROMOGEN FOR ANTHOCYANIN*)，*A* (*ANTHOCYANIN ACTIVATOR*)和 *WX* (*GLUTINOUS ENDOSPERM*)。对任何没有座位标识符的基因符号来说，会被默认为座位标识符为“1”，例如，*GLUTINOUS ENDOSPERM* (*WX*) 应该指定为 *GLUTINOUS ENDOSPERM 1*(*WX1*)。所有具有相似特性的新基因将被 CGSNL 根据发现的顺序分配给一个新的座位标识符。CSGNL 将保证先前鉴证的基因符号和新鉴定登记的基因分配到一个唯一的基因符号，避免名称和符号的混淆。

使用后缀“(t)”和“*”来表明一个假设性的座位标识(当一个新的描述基因和一个先前已知基因的等位基因关系不是很清楚(Kinoshita, 1986))被暂时使用，在假定其为新的座位的情况下，新基因将被分配给一个新的座位标识符。如果这个新基因在以后被证明和原来已知座位是等位的，两个相关记录将被合并，最初的基因符号将被按照程序规则采纳。其他符号将会作为同义词引用。以前分配的基因符号将不会删除，这可以避免相同符号重新使用导致的混乱。分配一个符号给一个基因的时候应像上边描述的那样，保持和基因名全称的一致性。

作者在其文章种涉及到已知功能的水稻基因的时间，一定要引用核准的基因名全称和符号，如果有可能要引用基因组注释中心之一的系统座位ID和GenBank登录号。当完整的信息不存在的情况下，除非有额外的实验证据的提供，否则系统座位ID或基因符号将不会被使用。只有通过CGSNL的审查，基因名才可以被分配使用。

3物种名在基因名和符号中的使用

出版物中在基因名和基因符号前使用物种特异性前缀如“Os” (*O.sativa*)也许是有用的，但这并不在官方已经命名规则中，因为对已经和物种信息关系的递交/注册基因来讲显得有些冗余。而且，也会导致基因名*Oryza sativa-X*的扩散。基因和物种之间的关系会在所有的基因组数据库和学历数据库中清楚的保留。然而，作者可以在出版物中附加上物种特殊性的前缀，以此来避免在任何时候参考一个基因时所带来的物种名重复。在任何情况下，物种符号都不应该变成被采用的基因符号或基因名全称的一部分。特别需要指出的是，符号“Os”在系统座位ID中是允许使用的，例如，Os05g0000530、LOC_Os03g01590和OsIBCD000082这些分别为RAP (http://rapdb.lab.nig.ac.jp/index.html)、TIGR (http://www.tigr.org/tdb/e2k1/osa1/tigr_gene_nomenclature.shtml)和BGI-RIS (http://rise.genomics.org.cn/rice/index2.jsp)数据库所采用。

3.1等位基因变异体

同一个基因的不同的等位基因是通过添加数字后缀来区分的，数字和基因全称或基因符号之间有破折号或连字符分开，例如，SHATTERING 1−1(SH1−1)，PGI1−1和PGI1−2。在历史上，有几种情况下曾用过字母(t)或星号(*)而不是数字来指明一个等位基因，因为这些字母符号在描述等位基因变异体时广泛使用并被水稻遗传研究学界接受，这些字符将在出版物中作为例外保留，并将在数据库中作为同义名标明。

3.2显性/隐性关系

历史上，基因全称是全部小写斜体，如果文献中第一个等位基因是显性的就以一个大写字母开头，如是隐性就以小写字母开头。由近来在大规模

基因组测序努力为基础在基因鉴定取得的进展，那些仅仅在单倍体细胞(即花粉或卵)中表达基因，一个等位基因的显性或隐性或一个座位的变异体，也许就是未知的或无关紧要的。然而，在调查基因功能的时候，一个等位基因的显/隐性对遗传研究来说依然是十分重要的。因此，为出版的目的，建议每一个特定的种质资源被描述的地方，隐性等位基因都使用小写字母，显性等位基因第一个字母大写，其他全部小写，这两种情况下所有字母都是斜体(如先前的约定; 表3)。但正式的(全称)基因名称所有字母需大写，特定等位基因的显隐性表型将作为这个等位基因的属性，而不是在数据库中基因名称的一部分被记录。

表 3 在文献中基因名全称和符号使用的例子

Table 3 Example of a Gene Full Name and Symbol for Use in Publications

类型 Type	基因全称 Gene Full name	基因符号 Gene Symbol
座位/基因 Locus/gene	*NARROW LEAF 1*	*NAL1*
隐性等位基因 Recessive allele	*narrow leaf 1-1*	*nal1-1*
显性等位基因 Dominant allele	*Narrow leaf 1-2*	*Nal1-2*
序列变异体 1 Sequence variant 1	*NARROW LEAF 1-s1*	*NAL1-s1*
序列变异体 2 Sequence variant 2	*NARROW LEAF 1-s2*	*NAL1-s2*

注: 基因名全称和符号将斜体大写, 显性等位基因第一个字母大写斜体, 后边全部小写并斜体; 隐性等位基因全部小写并全部斜体

Note: The gene full name and symbol will be written in italics and all caps; Dominant alleles begin with an upper case first letter followed by lower case letters and recessive alleles are indicated with all lower case letters and all in italics

3.3序列变异体

考虑到一个基因就是一段具有已知或预测功能或表型的DNA片段，一旦一个基因通过一个系统座位ID被命名并定位到一个序列图谱上，它就可以被一群定位在同一个遗传座位的等位基因和序列变异体表示。在不同植物材料中仅仅通过序列鉴定的分子序列变异体将被赋予一个名称、符号和登录标识码，有关这个序列变异体的相关信息将交叉参考形成有关这个种质资源(包括对应的种质登记ID)的特定信息，从这些种质中DNA/RNA材料被提取。然而，序列变异体将不会被CGSNL认为是“等位基因”，除非它们具有分子功能或有描述的表型，并且经过等位性试验的证明。那些特定功能未知的“序列变异体”将通过添加给一个等位基因名后缀“-sX”来区分“等位基因”，“s”表示“已测序”，“X”是鉴别一个特定序列变异体的编码。一个分子变异体的名称和符号获得一个对应基因的名称和符号，这和一个等位基因的约定相似，除了其带后缀的描述，并且由于不能分配给这些测序变异体等位表现而全部大写(表3)。

如果一个测序的变异体以后被证据并赋予一个新的特定表型或功能，它将被分配给一个新的等位基因标识符，或如果一个测序变异体被证明和一个先前命名的等位基因对应一个已知基因相同，它将被基于程序规则分配给一个存在的等位基因标识符，原来其他标识符将作为同义名保留。一个推荐的基因座位、全称和等位基因命名的例子在表3。序列变异体鉴定相关的种质名和其登记号信息将不会在正式的名称和符号中记录。这部分信息将单独记录在数据库中，以方便遗传学界的交叉参考。提交序列变异体的作者有责任来搜寻这个新的序列形式是否和以前报告的序列变异体或等位基因相同。在文献中，作者可以选择等位基因名、序列变异体后缀和种质资源串联起来，避免对读者来说过度的重复。

3.4蛋白质名和符号

一个特定基因编码的蛋白，其名称在该基因名称是基于表型或分子功能的情况下(参考“基因全称”章节)，应该和这个基因的全称保持一致，但蛋白的名称将全部大写并且不需要斜体。如果在后期阶段，一个基因和它对应的蛋白产物被证明有一个生化分子功能，如是一个酶或一个大分子复合物的结构组成部分(亚基)，这个蛋白应该按照IUPAC酶学委员会建议的酶命名法或IUBMB的大分子命名法分配一个同义名(Committees Biochemical Nomenclature, 2006)。对于一个特定的蛋白来说可能会有多个功能分配(即, 基于表型试验, 生化分析或分子功能)，因此一个蛋白名会有几个异名(和基因名全称相似)。蛋

白符号应该一直和采纳的基因符号保持一致，除了蛋白符号全部大写不用斜体，紧接一个空格和数字的座位标识符。例如，GLUTINOUS ENDOSPERM 1 (WX1)基因编码谷粒表面淀粉合成酶(EC:2.4.1.11)。蛋白名是GLUT-INOUS ENDOSPERM 1和符号“WX1”。蛋白名‘WAXY’、‘WAXY 1’和GRANULE-BOUND STARCH SYNTHASE (GBSS)将被当作同义名记录。如果一个名称不能基于表型、已知生化或其他实验证据支持其他功能而得到分配，一个系统座位标识符(上述)和一个和表1描述一致的名称必须用来描述这个基因，直到其功能得到确认。

3.5转录后修饰

当存在转录后修饰时，如蛋白质剪接导致两个或更多具有不同活性或功能的蛋白质分子的形成，剪接蛋白分子将获得和它们分子功能或关联表型一致的蛋白质名称和符号，初始分子的名称和符号作为同义名。

3.6假基因

分子技术鉴定了和结构基因序列十分相似的相同序列，但不被转录，这些序列称为假基因。为了表明这些假基因和功能基因的相关性，假基因将被使用和结构/功能基因相同的基因符号，名称斜体表示，后边加上字母“.P”(符号“.”和大写字母“P”)。这将取代通常使用的希腊符号“psi”作为假基因的标记；一个例子是RPS14.P取代RPS14. psi作为假核糖体蛋白S14的符号。同样的命名法建议用于线粒体和质体(叶绿体)基因组的假基因，例子就是ACTB.P1 (ACTIN BETA PSEUDOGENE 1)、ACTB.P2 (ACTIN BETA PSEUDOGENE 2)等。假基因可能在不同的染色体上或者和功能基因紧密靠近，据此得到它们的名称并存在变化的编号。从命名法的角度来看，一个假基因就是一个没有功能的基因(http:// pseudogene.org/main.html.)。如果一个假基因以后被证明可以转录并调控另外一个基因的表达或转录的mRNA具有一定功能，这个基因必须重新划归另一个基因类别如fnRNA或potogene (Brosius and Gould, 1992)。

3.7未定位基因

由于一个物种中固有的遗传变异性，来自某一种质的一个基因有可能在水稻两个已测序全基因组中，由于插入/缺失多态性和基因家族的扩增/收缩而不能定位。同样，在分离群体中通过表型鉴定出的一个基因也可能不存在于两个双亲的基因组中。在这种情况下，即使没有定位信息，基因名称和符号也应该分配给这些等位基因变异体。当给一个未定位座位分配基因名称时，必须有有效的实验证据支持这个基因和其功能的存在。如有一个确认相似未定位测序基因第二审存在的话，最好的相互匹配的方法应该是更加严格的确认是否事实上它和先前鉴定的基因是同样的。在二审表现型确认基因存在的情况下，等位性或互补试验作为必需证据被考虑进去。如有任何这些证据的缺失，这个基因就将被分配一个新的基因名称和符号。同时，唯一标识符将被分配给CGSNL登记的这个基因，如果有Genbank登录号存在的话，将作为占位符。

3.8数量性状遗传座位(QTL)

QTLs座位基因的位置标识符对基因组功能特征有贡献。一个QTL代表一个统计学上的可测表型，概括来说就是一个数量遗传性状。通过分离群体的连锁遗传作图鉴定的QTLs，每一个QTL被定义为至少通过两个紧密连锁的定位遗传标记限定的染色体上特定区间。

水稻QTL命名法规则(McCouch et al., 1997)指出，每一个QTL名称都要以小写斜体字母“q”来表明是其是一个QTL，紧跟着就是2到5个字母的标准“性状名称”(如, SW标示粒宽)，一个数字表明其所在的水稻染色体(1−12)，一个“.”和一个区别于其他在同一个染色体上的单个QTLs的唯一标识符(例如, qSW5.1)当QTLs进入一个基因组数据库例如Gramene (Jaiswal et al., 2006)，它们将被分配给一个来自与性状主体的标准性状术语(TO; 例如, 籽粒宽度, 登录号#TO: 0000140) (Jaiswal et al., 2002)来方便查询，也许可能分配一个新的唯一标识符来避免研究中间的混乱。在任何情况下，这个数据库赋值将作为这个QTL记录中的一个同义名，原始的文献中的QTL名称将为检索的目的而得到保留。

当这些负责具体表型变异的QTL就是这些基因，并且第一次基于对应的QTL而被鉴别出来时，基因名称全称可以反映QTL的标示(除了除掉前缀‘q’外, 还要使用斜体(例如, SW5))；然而，如果

对应QTL的基因和一个先前命名和描述的基因想符合，流程规则的使用和原始基因的名称必须被保留。但是，建议基因和QTLs之间的关系要在和基因名关联的同义名列表中注明。

4系统座位ID分配：一个RAP数据库例子

4.1核基因的系统座位ID

系统座位标识符将依次分配给水稻(*O. sativa* ssp. *japonica*, cv. Nipponbare)假分子(水稻测序基因组组装染色体重叠群)基于自动基因预测程序、直系同源联配、和/或ESTs和全长cDNAs联配鉴定出的基因，遵循酵母(S.cereviseae) (http://www.yeastgenome.org/gene_guidelines.shtml) 和拟南芥(A.thaliana) (TAIR, 2005)。系统标识符被分配给蛋白编码基因(ORFs)、RNA编码基因(snoRNA, snRNA, rRNA, tRNAs, and microRNAs)、和假基因。一个核基因座位ID将包括：(a)一个大写字母“O”和小写字母“s”来标明水稻物种O .sativa；(b)两个位数的数字表明特定的水稻染色体(01, 02, 03, …12)；(c)一个字母“g”来标明这个座位ID是一个基因；(d)一个7位数数字(假设每个染色体上有不少于10 000个基因)标明基因在染色体上的顺序，按照从端粒的短臂(北端)到端粒的长臂(南端)的升序排列。标明基因顺序的数字是独立于染色体链的极性的(+/-或Watson/Crick)，并且起始的时间就分配了100的增量，为新基因的发现扩增留下空间。例如，染色体5上的第三个和第四个基因被标示为Os05g0000300和Os05g0000400。

在测序过程中或有新的实验证据表明一个新基因在两个基因注释的基因间被辨别出来，这个新基因将被使用后边第十位数字位置，分配一个两个先前已的注释基因之间的一个数字。例如，在基因Os05g0000300和Os05g0000400发现的基因可以分配给Os05g0000350，而且留下扩增的空间。尽管这个策略有很明显的优势，但在一些情况下在一个特定的染色体片段中基因的顺序并没有遵循基于基因发现优先顺序的升序/降序规则；然而，这些缺陷并没有遮盖这套系统整个的价值。系统座位IDs将分配给所有的基因，包括那些已知的通过一个器官基因组(质体和/或线粒体)的一部分插入核基因组，这些基因常常被证明是没有功能的或是假基因。

对于那些水稻基因组序列不完整的区间，例如端粒和着丝粒区域间的间隙或更小的内部间隙，一个座位ID空间保留是合适的。在端粒和着丝粒区间，一个座位ID空间可以接纳每一个间隙的1 000个基因，每个基因约2 kb空间间隙。

需要注意的，水稻栽培种、亚种或种，而非水稻粳稻栽培种Nipponbare中基因组中鉴定的座位必须征询CGSNL的命名。数据库的监护者和个体的研究人员必须在CGSNL注册并且审核通过时，才可以分配名称和符号。

4.2细胞器基因的系统座位ID

主要的线粒体和叶绿体染色体是环状(也称为master circles)，没有臂。细胞器染色体上的基因座位IDs将使用符号‘Mt’来代表线粒体，‘Pt’代表质体(叶绿体)，而不是像核基因那样使用数字代表染色体。这些字母将紧跟着字母“g”标明这个座位对应一个基因，接着紧跟一个7位数的数字(假设每个染色体有少于10 000个基因)标明一个细胞器染色体中基因的顺序，不考虑链的极性，按照完整测序分子的第一个碱基到线性化分子(如测序作者向任何参考序列数据库, 即NCBI-GenBank, DDBJ或EMBL)的最后一个碱基。例如，OsPtg0000100标明水稻质体基因组上第一个基因。在GenBank登录号中寻找水稻栽培种Nipponbare的质体基因组，这可以参考基因PSBA (82−1, 143 bp)，如GenBank登录号NC_001320。

除了在master circles上鉴定座位的系统外，那些在质体上与线粒体线性和环状(也称为亚基因组环)发现的基因，将使用一个小写字母a-z (依据提交到GenBank的顺序)紧跟细胞器符号Mt或Pt。例如，OsMtag0000200表明2 135 bp长的线粒体质体B1上的基因2 (GenBank登录号NC_001751)。质体上的基因顺序编号将开始于全组装测序质体或亚基因组环中第一个碱基序列，按照提交到GenBank、DDBJ或EMBL的顺序。

4.3转录物ID

一个基因的每个已知或预测的转录本形式将被分配给一个系统标识符，和座位标识符不同的是代表基因的字母‘g’将被代表转录本的‘t’代替座位后缀，同时紧跟2位数的染色体标识符。这种命名约定将可以保证基因座位ID和其他转录本ID的一致性。例如，转录本Os05t0000300是座位

Os05g0000300的转录本，代表染色体5上的基因3。有时初期的转录本要经历选择性剪切。为了清楚的分辨转录本的不同剪切体，两位数的后缀将被加入到基因的系统转录ID中，通过一个破折号，按照发现的顺序分开，例如，-01，-02，-03，....-99。默认的第一个转录本(或唯一鉴定的转录本)的转录ID常常有编码“-01”座位转录ID的后缀。例如，座位Os05g0000300的转录ID，已知没有其他的剪接变异体，将是Os05t0000300-01。如果以后有报告说明这个座位有选择性剪切，如有三个选择性剪切体，如果三个中有任何一个和原始的转录本匹配，其就将作为原始的转录本ID保留，而另外其他两个ID就是Os05t0000300-02和Os05t0000300-03。剪切变异体分配的编号序列将基于鉴定过程、GenBank提交或可能cDNA的大小。任何附加的不同剪切体将依次编码。

4.4蛋白ID

一个基因序列或转录本通过实验或计算所推演出的所有多肽将被分配给一个和转录本标识符一样的系统标识符，除了标示转录本的字母‘t’被标示蛋白质的‘p’代替，这也保证了和基因座位ID和其转录本ID的一致性。例如，蛋白Os05p0000300-01翻译于转录本Os05t0000300-01，其又转录于代表了染色体5的基因3的座位Os05g0000300。为了避免单一座位不同转录选择性剪切体带来的蛋白质的混淆，蛋白质ID必须反应和其推衍出的对应的转录本，除了字母‘t’。

4.5未锚定在测序克隆上的基因

未在BAC/PAC克隆上锚定而鉴别出的基因依然使用命名法系统，据此基因随后标识测序中心分配的数字作为BAC/PAC克隆名称的后缀(例如，F23H14.13)。上边简述的系统座位ID命名法系统中，在这个区间的序列可以完全组装和完整时，将取代克隆的名称。在一些情况下，早期克隆为基础的座位标识符必须成为基因同义名或不同的ID。

5座位的增加、删除、编辑、合并和拆解

5.1编辑一个座位

一个特定座位的标识符、基因全称和基因符号在通常情况下其使用要具有一定的一致性。在基因的功能没有任何大的变动，尤其是在没有任何的变化导致座位的起始位点没有变化的情况下，一致性是可以被保持的。例如，命名法的一致性在基因编码的是一个ORF和注释修订变化仅在内含子-外显子的边界处的情况下是可能的，序列相同，要求外显子或内含子的增删，或功能或相关表型的改变或修饰的情况下，保持座位分配的一致性。与此相似，在ORF定义的注释、基因的全称、符号和GenBank/DDBJ/EMBL定义的记录更新变化，应该放映分子结构或功能的变化，但在以上的所有情况下，座位ID保持不变。

5.2删除一个座位

计算方面鉴定的基因当被实验证据证明是假阳性的时间，就必须删除这个座位。这种情况下，所有的记录和对应的标识符应该被标示OBSOLETE并保留，永远不能从数据库存储中删除。标示OBSOLETE确保相同的标识符不再一次被一个新的座位使用，因此避免了混乱情况的出现，如有需要，一个废除的基因还可以提供参考。

5.3拆分一个座位

这种情况一般是一个座位标识符实际上代表多于一个基因时(例如，两个基因被自动预测方法错误的鉴别为两个基因)，最接近座位开始位置的座位将保留原始的座位标识符、基因名和基因符号，远离座位开始位点的基因将被作为一个新的鉴定座位，并按照上述的方法赋予一个新的座位标识符、基因名和基因符号。在适用的情况下，基因名和基因符号的修订应该符合其新的功能。

5.4座位的合并

在有实验证据(如全称cDNA序列)证明两个先前鉴定的基因实际上是一个基因或是同一个座位的一部分的情况下，这两个座位必须合并为一个。新的座位必须保留最靠近新的合并座位起始位置的座位标识符、基因名和基因符号。对第二个基因来说，座位标识符变成了次级座位ID (和第一个关联)，第二个基因的名称和符号也借此成为第一个的同义名。

5.5转座因子座位ID

分配给包含一个转座因子(TE)座位的IDs和基因座位的相似，不同是基因座位ID中的‘g’为‘te’

取代，例如，Os05te0000300。既然目前大多数的TE注释是基于电子预测和计算分析，在RAP1的会议上决定这个系统在晚期实施。也建议TE的命名系统在实施前征询专家的建议。然而，如果一个TE被证明包含一个功能基因，将被分配给一个前边描述的基因座位的标识符。

6基因名称和符号的登记注册

为支持登记注册过程，一个基于网络的基因登记和命名法网站被建立了http://shigen.lab.nig.ac.jp/rice/oryzabase_submission/gene_nomenclature。CGSNL中的小组委员会将依据基因是否通过序列和表型鉴定，来负责处理登录注册请求。水稻人员被鼓励使用这个网站注册基因和感兴趣的等位基因。CGSNL将给予功能特征基因优先权，也会为了处理一个新请求，而要求提供实验证据。核准的基因名和符号在核准后立刻释放出来。尽管这套命名法系统将会对来自栽培水稻(O. sativa)的基因进行分类，CGSNL也将会尽一切努力来管理非栽培水稻的其他水稻的基因命名法，水稻界并鼓励使用相同的基因注册网站注册登记来自非栽培水稻的基因。

6.1登记注册过程

下列类型信息应该在注册登记一个新基因的时间提交：

(1)有关这个基因特征的描述信息，包括但不限制其分子功能，在生物过程中的作用，在亚细胞中的定位，在特定植物组织和生长期的表达和其表型效应；

(2)遗传性和等位性数据；

(3)来源种质(属，种，原种/品种/登录ID/种质库)。如果来自一个杂交登录号，提供其亲本的种质信息；

(4)染色体和图谱定位；

(5)序列数据和基因模型(内含子/外显子结构，启动子等)；

(6)GenBank登录号和/或至少来一个水稻基因组注释计划的座位ID (如果有的话)

(7)蛋白质/基因家族的关系；

(8)支持材料包括突变表型的照片，RNA和/或蛋白表达数据，酶活性检测，序列匹配等。

提交的注册条目将通过电子递交的方式通过基因注册网站OryzaBase数据库(http://shigen.lab.nig.ac.jp/rice/oryzabase_submission/gene_nomenclature)送到CGSNL的召集人。在检查提交的信息后，决定一个基因是否是新的，并考虑命名规则，随后召集人通知提交者来验证新的基因全称和符号。审核通过后，注册基因将被分配给一个合适的基因名全称和基因符号。这必须在注释数据库和文献中报明。基因登记数据库将提高一个在线可下载的有关注册基因的列表，如果有的话，包括了批准的基因名、符号、同义名、注释数据库中定位的系统座位IDs和相关联的GenBank登录号(表2)。召集人必须与其他数据库和RGC成员沟通，以保证新的基因名和符号可以包括在OryzaBase (http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp)、Gramene (http://www.gramene.org)、IRIS (http://www.iris.irri. org)、RAP (http://rapdb.lab.nig.ac.jp)、TIGR (http://www.tigr.org/tdb/e2k1/osa1)和其他相关数据库和网站的基因/等位基因名单中出现。一个描述所有新批准基因/等位基因的研究备忘录在在Rice Genetics Newsletter (http://www.shigen.nig.ac.jp/rice/oryzabase/rgn/newsletter.jsp)中每半年出版一次。

6.2增补修订

对这些规则增补的修订建议可以通过OryzaBase网站http://shigen.lab.nig.ac.jp/rice/oryzabase_submission/gene_nomenclature的在线“建议”向CGSNL提交。增补修订将在期刊RICE、Rice Genetics Newsletter中和OryzaBase、Gramene、IRIS数据库和rice-e-net电子邮箱列表(http://chanko.lab.nig.ac.jp/list-touroku/rice-e-net-touroku.html)公布。用户可以通过向genenomenclature@chanko.lab.nig.ac.jp发送邮件取得联系。

7讨论

实验证明功能的基因(已知功能的基因)和仅仅依据序列分析预测的基因(基因模型)是不同的，为此水稻界在两种不同基因结构/功能分析方法之间通过一个灵活的严谨的系统建起了桥梁。为水稻中的每一个基因提供功能性描述是长期目标，最终每个基因模型(座位ID)应该和一个基因名关联。然而，随着测序成本的迅速降低和公共领域水稻测序基因组数目的快速增加，我们对水稻中基因库的理解

就不仅局限到单个的粳稻(*O. sativa* ssp. Japonica)和籼稻(*O. sativa* ssp. *Indica*)基因组序列。因此，水稻基因命名法系统采用了在已知功能基因和计算的基因模型之间建立一对多的联系的协议，多种类型的证据来支持每一个水稻(Oryza)基因的功能描述。

一个基因可以编码一个蛋白产物(CDS)或可以编码众多非编码RNA分子的一种，包括snoRNA、snRNA、tRNA、rRNA、microRNA、SiRNA或fnRNA(功能RNA)等。如果在将来有新的基因类别鉴定出来，我们将增补修订我们的分类系统。

在基因的命名中，英语是首选的，基因符号应该由拉丁字母和阿拉伯数字组成。基因的名称应该简要的描述表型和/或表达出基因产物功能的一些意思(如果知道的话)。所有新基因应该通过CGSNL注册和批准，以避免混淆和重复。水稻界一个基因第一次文献中出现的名称以优先权，但应该认识到的是名称的改变应该放映新的认识。在这个时间，我们不赞成采用一个严格或限制性的基因命名法系统，我们同意采用一个同义名系统，以此使测序的基因标识符和基于实验确认了生化功能或表型差异的名称之间建立对应关系。这种方法使得随着新技术的发展和新认识的积累，水稻基因命名法系统的持续发展更新成为了可能。

致谢

We kindly acknowledge the following researchers, Pankaj Jaiswal, Junjian Ni, and Immanuel Yap from the Gramene database (http://www.gramene.org) and the Department of Plant Breeding and Genetics at Cornell University, Ithaca, NY, USA; Toshiro Kinoshita from the Kita 6 Jo, Nishi 18 Chome, Sapporo 060-0006, Japan; David Mackill and Richard Bruskiewich from the International Rice Research Institute, DAPO 7777, Metro Manila, Philippines; C. Robin Buell from Department of Plant Biology, Michigan State University, East Lansing, MI 48824-1312, USA; Masahiro Yano, Takeshi Itoh, and Takuji Sasaki from the Department of Molecular Genetics, National Institute of Agrobiological Resources, Tsukuba, Ibaraki 305-8602, Japan; and Qifa Zhang from the National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, 430070, People's Republic of China for the help in preparing this manuscript and to numerous other experts for their help and useful suggestions on improving the rice gene nomenclature. We also thank all the members of the Rice Genetics Cooperative (RGC: http://www.shigen.nig.ac.jp/rice/oryzabase/rgn/office.jsp) for their support. Financial support was provided by NSF Grant DBI 0703908 (Cold Spring Harbor Subcontract 22930113 to Cornell University).

参考文献

Ammiraju J.S.S., Luo M., Goicoechea J.L., Wang W., Kudrna D., Mueller C., Talag J., Kim H., Sisneros N.B., Blackmon B., Fang E., Tomkins J.B., Brar D., MacKill D., McCouch S., Kurata N., Lambert G., Galbraith D.W., Arumugana- than K., Rao K., Walling J.G., Gill N., Yu Y., SanMiguel P., Soderlund C., Jackson S., and Wing R.A., 2006, The *Oryza* bacterial artificial chromosome library resource: Construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*, Genome Res., 16(1): 140-147

Brosius J., and Gould S.J., 1992, On "genomenclature": A comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA", Proc. Natl. Acad. Sci., USA, 89(22): 10706-10710

Cheng Z., Buell C.R., Wing R.A., Gu M., and Jiang J., 2001, Toward a cytological characterization of the rice genome, Genome Res., 11(12): 2133-2141

Committees Biochemical Nomenclature, 2006, IUPAC-IUBMB joint commission on biochemical nomenclature (JCBN). Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB)

Goff S.A., Ricke D., Lan T.H., Presting G., Wang R., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma H., Hadley D., Hutchison D., Martin C., Katagiri F., Lange B.M., Moughamer T., Xia Y., Budworth P., Zhong J.P., Miguel T., Paszkowski U., Zhang S.P., Colbert M., Sun W.L., Chen L.L., Cooper B., Park S., Wood T.C., Mao L., Quail P., Wing R., Dean R., Yu Y., Zharkikh A., Shen R., Sahasrabudhe S., Thomas A., Cannings R., Gutin A., Pruss D., Reid J., Tavtigian S., Mitchell J., Eldredge G., Scholl T., Miller R.M., Bhatnagar S., Adey N., Rubano T., Tusneem N., Robinson R., Feldhaus J., Macalma T., Oliphant A., Briggs S., 2002, A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*), Science, 296(5565): 92-100

IRGSP, 2005, The map-based sequence of the rice genome, Nature, 436: 793-800

Itoh T., Tanaka T., Barrero R.A., Yamasaki C., Fujii Y., Hilton P.B., Antonio B.A., Aono H., Apweiler R., Bruskiewich R., Bureau T., Burr F., Costa de Oliveira A., Fuks G., Habara

T., Haberer G., Han B., Harada E., Hiraki A.T., Hirochika H., Hoen D., Hokari H., Hosokawa S., Hsing Y.I., Ikawa H., Ikeo K., Imanishi T., Ito Y., Jaiswal P., Kanno M., Kawahara Y., Kawamura T., Kawashima H., Khurana J.P., Kikuchi S., Komatsu S., Koyanagi K.O., Kubooka H., Lieberherr D., Lin Y.C., Lonsdale D., Matsumoto T., Matsuya A., McCombie W.R., Messing J., Miyao A., Mulder N., Nagamura Y., Nam J., Namiki N., Numa H., Nurimoto S., O'Donovan C., Ohyanagi H., Okido T., Oota S., Osato N., Palmer L.E., Quetier F., Raghuvanshi S., Saichi N., Sakai H., Sakai Y., Sakata K., Sakurai T., Sato F., Sato Y., Schoof H., Seki M., Shibata M., Shimizu Y., Shinozaki K., Shinso Y., Singh N.K., Smith-White B., Takeda J., Tanino M., Tatusova T., Thongjuea S., Todokoro F., Tsugane M., Tyagi A.K., Vanavichit A., Wang A., Wing R.A., Yamaguchi K., Yamamoto M., Yamamoto N., Yu Y., Zhang H., Zhao Q., Higo K., Burr B., Gojobori T., and Sasaki T., 2007, Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*, Genome Res., 17: 175-183

Jaiswal P., Ni J., Yap I., Ware D., Spooner W., Youens-Clark K., Ren L., Liang C., Zhao W., Ratnapu K., Faga B., Canaran P., Fogleman M., Hebbard C., Avraham S., Schmidt S., Casstevens T.M., Buckler E.S., Stein L., and McCouch S., 2006, Gramene: A bird's eye view of cereal genomes, Nucleic Acids Res., 34: D717-D723

Jaiswal P., Ware D., Ni J., Chang K., Zhao W., Schmidt S., Pan X., Clark K., Teytelman L., Cartinhour S., Stein L., and McCouch S., 2002, Gramene: Development and integration of trait and gene ontologies for rice, Compar. Funct. Genom., 3: 132-136

Karlowski W.M., Schoof H., Janakiraman V., Stuempflen V., and Mayer K.F.X., 2003, MOsDB: An integrated information resource for rice genomics, Nucleic Acids Res., 31: 190-192

Kinoshita T., 1986, Report of the committee on gene symbolization, nomenclature and linkage groups, Rice Genet. Newslett., 3: 4-8

McCouch S.R., Cho Y.G., Yano M., Paul E., Blinstrub M., Morishima H., and Kinoshita T., 1997, Report on QTL nomenclature, Rice Genet. Newslett., 14: 11-13

McNally K.L., Bruskiewich R., Mackill D., Buell C.R., Leach J.E., and Leung H., 2006, Sequencing multiple and diverse rice varieties. Connecting whole-genome variation with phenotypes, Plant Physiol., 141: 26-31

Mulder N.J., Apweiler R., Attwood T.K., Bairoch A., Bateman A., Binns D., Bradley P., Bork P., Bucher P., Cerutti L., Copley R., Courcelle E., Das U., Durbin R., Fleischmann W., Gough J., Haft D., Harte N., Hulo N., Kahn D., Kanapin A., Krestyaninova M., Lonsdale D., Lopez R., Letunic I., Madera M., Maslen J., McDowall J., Mitchell A., Nikolskaya A.N., Orchard S., Pagni M., Ponting C.P., Quevillon E., Selengut J., Sigrist C.J., Silventoinen V., Studholme D.J., Vaughan R., and Wu C.H., 2005, InterPro, progress and status in 2005, Nucleic Acids Res., 33: D201-D205

Ohyanagi H., Tanaka T., Sakai H., Shigemoto Y., Yamaguchi K., Habara T., Fujii Y., Antonio B.A., Nagamura Y., Imanishi T., Ikeo K., Itoh T., Gojobori T., and Sasaki T., 2006, The rice annotation project database (RAP-DB): Hub for *Oryza sativa* ssp. *japonica* genome information, Nucleic Acids Res., 34: D741-D744

Price C.A., Reardon E.M., and Lonsdale D.M., 1996, A guide to naming sequenced plant genes, Plant Mol. Biol., 30: 225-227

TAIR, 2005, Arabidopsis Nomenclature, http://www.arabidopsis.org/info/guidelines.jsp, 10 April 2008

VandenBosch K.A., and Frugoli J., 2001, Guidelines for genetic nomenclature and community governance for the model legume *Medicago truncatula*, Mol. Plant-Microb. Interact., 14: 1364-1367

Wain H.M., Lush M.J., Ducluzeau F., Khodiyar V.K., and Povey S., 2004, Genew: The human gene nomenclature database, 2004 updates, Nucleic Acids Res., 32: D255-D257

Wu R., Hirai A., Mundy J., Nelson R., and Rodriguez R., 1991, Guidelines for nomenclature of cloned genes or DNA fragments in rice, Rice Genet. Newslett., 8: 51-53

Yu J., Hu S., Wang J., Wong G.K-S., Li S.G., Liu B., Deng Y.J., Dai L., Zhou Y., Zhang X.Q., Cao M.L., Liu J., Sun J.D., Tang J.B., Chen Y.J., Huang X.B., Lin W., Ye C., Tong W., Cong L.J., Geng J.N., Han Y.J., Li L., Li W., Hu G.Q., Huang X.X., Li W.J., Li J., Liu Z.W., Li L., Liu J.P., Qi Q.H., Liu J.S., Li L., Li T., Wang X.G., Lu H., Wu T.T., Zhu M., Ni P.X., Han H., Dong W., Ren X.Y., Feng X.L., Cui P., Li X.R., Wang H., Xu X., Zhai W.X., Xu Z., Zhang J.S., He S.J., Zhang J.G., Xu J.C., Zhang K.L., Zheng X.W., Dong J.H., Zeng W.Y., Tao L., Ye J., Tan J., Ren X., Chen X.W., He J., Liu D.F., Tian W., Tian C.G., Xia H.A., Bao Q.Y., Li G., Gao H., Cao T., Wang J., Zhao W.M., Li P., Chen W., Wang X.D., Zhan Y.g, Hu J.F., Wang J., Liu S., Yang J., Zhang G.Y., Xiong Y.Q., Li Z.J., Mao L., Zhou C.S., Zhu Z., Chen R.S., Hao B.L., Zheng W.M., Chen S.Y., Guo W., Li G.J., Liu S.Q., Tao M., Wang J., Zhu L.H.,

Yuan L.P., and Yang H.M., 2002, A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*), Science, 296: 79-92

Yuan Q., Ouyang S., Liu J., Suh B., Cheung F., Sultana R., Lee D., Quackenbush J., and Buell C.R., 2003, The TIGR rice genome annotation resource: Annotating the rice genome and creating resources for plant biologists, Nucleic Acids Res., 31: 229-233

Yuan Q., Ouyang S., Wang A., Zhu W., Maiti R., Lin H.N., Hamilton J., Haas B., Sultana R., Cheung F., Wortman J., and Buell C.R., 2005, The institute for genomic research osa1 rice genome annotation database, Plant Physiol., 138: 18-26

Zhao W.M., Wang J., He X.M., Huang XB., Jiao Y.Z., Dai M.T., Wei S.L., Fu J., Chen Y., Ren X.Y., Zhang Y., Ni P.X., Zhang J.G., Li S.G., Wang J., Wong G.K-S., Zhao H.Y., Yu J., Yang H.M., and Wang J., 2004, BGI-RIS: An integrated information resource and comparative analysis workbench for rice genomics, Nucleic Acids Res., 32: D377-D382

Khush G.S., and Kinoshita T., 1991, Rice karyotype, marker genes, and linkage groups, In: Khush G.S., Toenniessen G.H., (eds.), Rice biotechnology, Wallingford, Oxon, UK and Manila, Philippines: CAB International and IRRI, pp. 83-108