

# 基于 INTERNET 的水稻基因数据库信息系统

鄂志国<sup>1</sup> 庄杰云<sup>2</sup> 曹永生<sup>3</sup> 钱 前<sup>2</sup> 王 磊<sup>1,\*</sup>

(<sup>1</sup>中国水稻研究所 科技信息中心, 浙江 杭州 310006; <sup>2</sup>中国水稻研究所 水稻生物学国家重点实验室/国家水稻改良中心, 浙江 杭州 310006;

<sup>3</sup>中国农业科学院 作物科学研究所 信息研究室, 北京 100081; \* 通讯联系人, E-mail: wang\_lei@hzrc.com)

## Information System of Rice Gene Database Based on INTERNET

E Zhi guo<sup>1</sup>, ZHUANG Jie yun<sup>2</sup>, CAO Yong sheng<sup>3</sup>, QIAN Qian<sup>2</sup>, WANG Lei<sup>1,\*</sup>

(<sup>1</sup>Information Service Center, China National Rice Research Institute, Hangzhou 310006, China; <sup>2</sup>State Key Laboratory of Rice Biology/

Chinese National Center for Rice Improvement, China National Rice Research Institute, Hangzhou 310006, China; <sup>3</sup>Information/Documenta-

tion Department, Crop Research Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China; \* Corresponding author, E

mail: wang\_lei@hzrc.com)

**Abstract:** Rice Gene Database is one of the key modules of the China National Crop Information Sharing Center, and also one of the sub-systems of Rice Database. It includes various rice gene data, i.e. gene sequences, gene products, gene functions, references, etc. Constructed with ASP.NET technology, this database is more secure, more quick running and easier to maintain. Users can easily access information in the database through internet interface <http://gene.ricedata.cn/>.

**Key words:** rice; gene; database; ASP.NET

**摘 要:** 水稻基因数据库系统是在建的国家作物科学数据共享中心的一个重要模块,也是国家水稻数据中心的一个重要子系统。功能上,该子系统提供了有关水稻基因各方面的数据信息,包括基因序列、基因产物、基因功能、参考文献等;系统设计上,基于 ASP.NET 的程序骨架使得系统在网络上运行时更快捷、更安全、更易维护。用户可通过因特网(<http://gene.ricedata.cn/>)对数据库进行检索操作。

**关键词:** 水稻; 基因; 数据库; ASP.NET

中图分类号: N37; Q943; S511

文献标识码: A

文章编号: 1001-7216(2006)06-0670-03

基因是染色体上具有一定座位的遗传单位,是 DNA 分子中一定长度的核苷酸序列。水稻的生长发育是基因在多种代谢和生理过程中时空表达的综合现象。分离潜在的各种有价值的基因,对水稻品种改良具有重要意义。

随着水稻全基因组精细序列图谱的完成,以功能基因组学研究为代表的后基因组时代已经到来。功能基因组学的主要任务是破译只用 A、T、G、C 等 4 个碱基组成的基因组 DNA 全序列,分离和克隆基因,阐明基因的结构和功能<sup>[1]</sup>。在继续利用自然变异的同时,国际水稻研究所、日本、韩国等国家和国际科研机构的科学家还致力于水稻突变体库的构建,国内的中国科学院植物生理与生态研究所、中国水稻研究所水稻生物学国家重点实验室等也相继建立了大规模的突变体库。相信随着这些突变体库的建立,将有大批的水稻基因被发现、认识和利用。

目前,美国的 Gramene 网站和日本的 Oryzabase 网站相继建立了水稻数据库,对已发现的水稻基因进行收集、整理和公布。截至 2006 年 6 月, Gramene 报道了 1 798 个水稻基因和 7 926 个水稻 QTL(数量性状座位)<sup>[2]</sup>, Oryzabase 公布了 2 364 个水稻突变体和相关基因<sup>[3]</sup>。这些水稻基因数据的集中发布,对国内外的水稻科研人员了解和利用当今水稻生物学研究成果提供了捷径。但是,这两个水稻基因库由于其自身因素的限制,给国内用户的访问带来不小的困难:首先,两个库的语言均为英文,这就给国内一些英语不是很熟练的育种专家带来相当的难度;其次,这两个基因库所在网站的服务器分别位于美国和日本,且在国内没有镜像站点,导致国内用户的访问速度有时非常慢甚至无法访问。另外,

国内还没有一个数据库正式系统地介绍国内外科学家所定位或克隆的水稻基因数据。

正是基于此,在建的国家作物科学数据共享中心(<http://icgr.caas.net.cn/>)的 6 个科学主体数据库(作物遗传资源、作物育种、作物栽培、作物生理生化、作物分子生物学、作物生产)中的作物分子生物学中的水稻模块,以国家水稻数据中心为数据平台,建立水稻基因数据库信息系统,向用户提供已克隆或定位的水稻基因数据,包括基因序列、基因产物、基因功能、参考文献等内容,期望为水稻的分子辅助育种和生物技术研究提供有效的数据服务。

水稻基因数据库是国家水稻数据中心这一大型水稻综合数据共享平台的一个子系统。因为基于 ASP.NET/VB.NET 的程序将比传统网络程序运行得更快、更安全<sup>[4]</sup>,所以系统采用流行的 ASP.NET 技术设计,SQL server 2000 为底层数据库,借助 VB.NET 语言开发。

### 1 数据来源及分类

基因数据主要来自国内外已公开发表的研究结果,相当一部分来自 Gramene 网站或 Oryzabase 网站公布的研究结

收稿日期: 2006-03-27; 修改稿收到日期: 2006-07-27。

基金项目: 国家科技基础条件平台建设项目(2005DKA31800 01 01); 浙江省重大科技攻关项目(2004C12032); 浙江省自然科学基金资助项目(Y305083); 中国水稻研究所青年创新基金资助项目。

第一作者简介: 鄂志国(1977-),男,助理研究员。

表 1 水稻基因分类及举例

Table 1 .Classification and examples of rice genes .

基因分类 Gene classification	基因名称 Gene name
营养器官 Vegetative organ	
根 Root	<i>CRL1</i>
茎 Culm	<i>MOC1 , BC1</i>
叶 Leaf	<i>DL</i>
生殖器官 Reproductive organ	
穗 Panicle	<i>FON1 , OsG1</i>
小穗 Spikelet	<i>LHS1</i>
种子 Seed	<i>ALK</i>
异时性 Heterochrony	<i>mori1</i>
色素变异 Coloration	
叶绿素 Chlorophyll	<i>Ads</i>
花青素 Anthocyanin	<i>OsCHLH , IPL1</i>
其他 Others	<i>Bh1 , gfl</i>
育性 Fertility	
育性恢复 Fertility restoration	<i>Rf1</i>
雄性不育 Male sterility	<i>tms1 , tms2</i>
雌性不育 Female sterility	<i>esa 1 , esa 2 , esa 3</i>
杂种不育 Hybrid sterility	<i>SA1 , SA2</i>
其他 Other sterilities	<i>Cps</i>
株高 Plant height	
矮秆 Dwarf	<i>D1 , d2</i>
半矮秆 Semidwarf	<i>sd1 , sd11</i>
高秆 Giant	<i>ao</i>
产量 Yield and productivity	<i>GN1</i>
稻米品质 Grain quality	<i>amy6 , CR , br7</i>
生物和非生物胁迫反应 Tolerance and resistance	
抗病性 Disease resistance	<i>Xa21 , Pib</i>
抗虫性 Insect resistance	<i>Bph1 , Wph1</i>
抗逆性 Stress tolerance	<i>bel , bzs</i>
激素反应 Response to hormone	<i>eui1</i>

果,但同时注意收集我国科学家定位或克隆的水稻基因,例如,水稻生物学国家重点实验室合作克隆的4个水稻基因:与分蘖相关的基因 *MOC1*、糊化温度基因 *ALK*、脆茎基因 *BC1*、与穗粒数相关的基因 *Gn1a*<sup>[5-8]</sup>以及华中农业大学克隆的抗白叶枯病基因 *Xa26*<sup>[9]</sup>等。另外也注重收集录入水稻育种的直接目标基因,如抗病基因、抗虫基因、与产量和米质有关的基因等。

随着已发表基因的相继录入和新发现基因的不断补充,水稻基因库的数据量将会迅速增加,如果没有一个好的分类方法,基因数据库会显得杂乱无章,不利于用户的查询。在咨询相关生物学和育种专家后,并参照 Kurata 等<sup>[10]</sup>的方法,将基因数据分成营养器官、生殖器官、异时性、色素变异、育性、株高、产量、稻米品质、生物和非生物胁迫反应以及激素反应10大类,20个亚类(表1)。其中,部分亚类还会再细分,如生殖器官类中的种子亚类会细分成胚、胚乳、种子形状、生理特性等。

## 2 系统结构设计

水稻基因数据库是一个关于水稻基因数据的综合信息系统,该系统建立的关系型数据库主要用来存储基因名称、

染色体位置、DNA 序列、基因产物、遗传图谱及参考文献等相关信息。为了既方便用户进行检索操作又利于管理人员执行输入校验工作,在设计上将系统分为管理系统和应用系统两个部分。管理系统为管理员提供一个可视化的界面,用来执行基因数据的增添、修改、删除等操作;应用系统面向终端用户,向访问者提供模糊或精确检索的功能。

基因库通过基因编号(ID)确定基因和编号的一一对应关系,也就是说,就像人的身份证一样,每条水稻基因数据都有一个唯一的编号,这样,基因库内部就不会发生混淆。同时,水稻基因库不是孤立的,它是整个水稻数据信息系统的有机组成部分,它通过库中特定关键字分别与品种库、参考文献表等关联。这样,对任一检索出的水稻基因,可以立即追溯到发现该基因的水稻供体的相关信息,如该供体的来源、抗性、品质等以及与该基因有关的文献报道等。

## 3 基本功能

目前水稻基因数据库的系统结构、数据的管理和检索等基本功能已经开发完成,基因库中已经积累了一定量的数据信息,已收录近200个水稻基因数据,其中已克隆基因34个,约占全球已报道水稻基因总数的10%。这些基因是农业育种中用得比较多的,如抗性基因、育性相关基因、矮秆基因等。基因数据库系统提供了数据库管理和使用的基本功能,如数据输入、校验、查询和结果显示等,同时,为方便用户的检索查询,系统对一些具体的问题作了专门的设计:

(1)从育种应用角度,对基因数据进行二次整理开发,将育种上用得较多的各种基因,如抗病基因、矮秆基因等,分门别类,建立独立页面,不需用户检索,结果一目了然而且全面,例如抗白叶枯病基因建立的列表([http://www.ricedata.cn/gene/gene\\_xa.htm](http://www.ricedata.cn/gene/gene_xa.htm))列出了截至2005年6月、经国际注册确认和期刊报道的水稻白叶枯病抗性基因30个;

(2)从检索效率角度,检索程序为用户提供了多种检索途径,如根据基因类别或所在染色体等,大大缩小了检索范围,提高了检索速度;

(3)从数据质量控制的角度,为了对基因数据进行有效的维护、更新和修改,数据库后台都有特定的校验程序实时监控,若因管理员疏忽而输入不符合设定的信息,管理程序将报错,从而有效保证数据质量。

## 4 展望

优质、多抗、高产作物新品种的选育和推广是实现我国粮食安全的重要途径。目前大多数育种工作仍然建立在表型选择和育种家的经验之上,影响了育种效率的进一步提高;另一方面,生物信息数据库积累的数据量极其庞大,由于缺乏必要的数据整合技术,可供育种工作者利用的信息往往有限<sup>[11-13]</sup>。为弥补这一缺憾,国家科技基础条件平台建设启动了作物科学数据的整合与共享项目,整合全国作物科学数据,建立国家作物科学数据共享中心,实现作物科学数据的共享和高效利用,为作物育种、生物技术和农业生产提供数据支撑。作为作物分子生物学主题数据库的重要子系统,水稻基因数据库信息系统是一个为用户提供水稻基因检索和发布的综合数据库系统,同时,水稻基因数据库信息系统也是我们正在建设的国家水稻数据中心的一个最新子

平台。除该子库外,国家水稻数据中心网站(<http://www.ricedata.cn/>)已经开发了水稻优异种质库、水稻突变体库、水稻品种库。这些子库都是独立的子系统,拥有各自的用户检索界面和查询结果,同时,这些子库内部又是相互关联的一个整体,用户从一个查询结果能很快进入另一个子系统寻找关联信息而不需要重新检索。我们希望水稻基因数据库信息系统的开发和广泛应用能为育种家进一步提高水稻育种效率提供数据支撑,也能为分子生物学家了解基因与相关农艺性状的关系提供渠道,真正成为生物数据和育种需求之间的一座桥梁,所以我们将 在现有工作的基础上,继续收录未录入和新发表的基因,丰富水稻基因数据库,而且将建立独立的分子标记数据库,并整合从不同群体构建的基因定位信息,同时进一步完善水稻数据中心其他子库的建设工作,并进一步优化各个子库之间的联系,真正做到各个子库之间的无缝连接。

水稻基因数据库的设计已经完成,目前正在进行数据录入工作,其中可能会存在或大或小的程序 BUG,欢迎用户反馈使用情况并提出宝贵意见。最新的信息可以通过访问 <http://gene.ricedata.cn/> 获取,如有疑问,可以直接在站内留言。

谢辞:在水稻基因数据库的系统设计以及数据的收集、整理过程中,得到了中国水稻研究所国家水稻改良中心和水稻生物学国家重点实验室多位专家的帮助,在此表示衷心的感谢。

#### 参考文献:

[1] 钱前,程式华.水稻遗传学和功能基因组学.北京:科学出版社,2006.

出版社,2006.

[2] <http://www.gramene.org/>.

[3] <http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp>.

[4] 赵斌,张滨义,董清波.ASP.NET从入门到精通.北京:人民邮电出版社,2002.

[5] Li X Y, Qian Q, Fu Z M, et al. Control of tillering in rice. *Nature*, 2003, 422(10): 618-621.

[6] 高振宇,曾大力,崔霞,等.水稻稻米糊化温度控制基因 *ALK* 的图位克隆及其序列分析.中国科学: C 辑, 2003, 33(6): 481-487.

[7] Li Y H, Qian Q, Zhou Y H, et al. *BRITTLE CULM1*, which encodes a COBRA like protein, affects the mechanical properties of rice plants. *Plant Cell*, 2003, 15: 2020-2031.

[8] Motoyuki A, Hitoshi S, Lin S Y, et al. Cytokinin oxidase regulates rice grain production. *Science*, 2005, 309(5735): 741-745.

[9] Sun X L, Cao Y L, Yang Z F, et al. *Xa26*, a gene conferring resistance to *Xanthomonas oryzae* pv. *oryzae* in rice, encodes an LRR receptor. *Plant J*, 2004, 37(4): 517-527.

[10] Kurata N, Miyoshi K, Nonomura K I, et al. Rice mutants and genes related to organ development, morphogenesis and physiological traits. *Plant Cell Physiol*, 2005, 46: 48-62.

[11] 万建民.作物分子设计育种.作物学报, 2006, 32(3): 455-462.

[12] 王磊,鄂志国,余汉勇,等.国际水稻信息系统及其应用.中国水稻科学, 2005, 19(2): 193-194.

[13] 李建,王磊.试论水稻专业性信息支撑体系的构建.农业图书情报学刊, 2005, 17(3): 31-33.

## 欢迎订阅 2007 年《农业质量标准》

主管:中华人民共和国农业部 主办:中国农业科学院

协办:农业部农产品质量安全中心 承办:中国农业科学院农业质量标准与检测技术研究所

主要栏目:本刊特稿、本刊专访、专家点评、专题论坛、政策法规、农产品质量安全、农业标准化、无公害食品行动、标准制定与实施、质量认证与管理、质量监督与检验、检验检测体系建设、农业标准公告、研究与探讨、质检中心之窗、名企名品、市场信息与动态、海外博览、编读园地、广告信息等。

读者对象:与农产品质量安全和农业质量标准有关的各级农业行政管理、科研教学、检验检测、技术推广、生产企业等部门的有关人员。该刊为双月刊,逢双月 10 日出版。大 16 开本,彩色四封,56 页。全国各地邮政局(所)均可订阅,也可直接到编辑部办理订阅手续。邮发代号 82-223,每册定价 6.80 元,全年共 40.80 元。

通讯地址:北京市中关村南大街 12 号中国农业科学院农业质量标准与检测技术研究所;邮政编码:100081;联系电话/传真:(010)62138026;E-mail: aqs@caas.net.cn。

欢迎各界朋友赐教、赐稿、订阅和刊登广告。