

doi:10.3969/j.issn.1672-5565.2013.02.05

基于 Make2D - DB II 构建水稻 二维电泳 - 质谱联动数据库

刘海军, 朱克明, 杨艳华, 邱旭华, 陈克平*

(江苏大学生命科学研究院, 江苏 镇江 212013)

摘要: 基于 Make2D - DB II 软件包构建了一个水稻二维电泳 - 质谱联动数据库 Rice2DDB, 介绍了该数据库的构建方法和步骤, 为快速发展的水稻蛋白质组学研究提供了数据管理和交流的平台, 对其他物种的蛋白质组学研究也具有参考价值。

关键词: 水稻; 二维电泳; 数据库; Make2D - DB II

中图分类号: N37; Q943; S511 **文献标识码:** A **文章编号:** 1672-5565(2013)-02-105-04

Two-dimensional gel electrophoresis and mass spectrometry database for rice proteomics research constructed by make2D - DB II Package

LIU Hai-Jun, ZHU Ke-ming, YANG Yan-hua, QIU Xu-hua, CHEN Ke-ping*

(Institute of Life Sciences, Jiangsu University, Zhenjiang 212013, China)

Abstract: A two-dimensional gel electrophoresis and mass spectrometry database for rice proteomics research named Rice2DDB has been constructed by means of Make2D-DB II Package, which provides a data management and exchange platform for rapidly developing rice proteomics research. The database constructing methods and steps are introduced in this paper that will be useful references for proteomics research on other species.

Keywords: Rice; Two-dimensional Gel Electrophoresis; Database; Make2D-DB II

水稻是世界上最重要的粮食作物之一,也是单子叶植物遗传和分子生物学的模式生物。自2002年水稻全基因组测序结果发布以来,对水稻的研究逐步从功能基因组水平拓展到了蛋白质组水平^[1-2]。近年来,高通量的蛋白质组学技术,特别是二维电泳技术和质谱技术的广泛应用,加速了水稻基因和蛋白质功能的研究发展,在水稻各组织发育及亚细胞蛋白质表达谱、水稻抗病抗逆相关蛋白、转基因水稻相关蛋白等方面获得了大量的研究成果和相关数据^[3-6]。一是获得了大量的水稻特定组织蛋白的二维电泳图谱,二是获得了大量的水稻蛋白的表达特性数据和质谱鉴定数据,这些数据大多都是分散在众多的文献中,不利于科研人员查找分析和

系统研究,因此有必要采用一种专门的数据库系统来对这些数据进行收集和管理。

目前国内外关于水稻的专门数据库已经建立了数十种之多,如国际水稻基因组测序计划的 IRGSP、日本农业生物资源研究所(NIAS)的 RGP 和 Rice-GAAS、日本遗传研究所(NIG)的 Oryzabase、美国密歇根州立大学(MSU)的 RGAP、美国科学基金会(NSF)的 Gramene、德国慕尼黑蛋白质序列信息中心(MIPS)的 MOsDB、北京基因组研究所(BGI)的 RIS、中国农科院水稻研究所的 Ricedata 等^[7-15]。这些数据库主要都侧重于保存水稻的基因组、EST 和蛋白质序列及其注释等方面数据。与水稻蛋白质组学相关的数据库则只有日本 NIAS 建立的 RPD

收稿日期:2012-04-18;修回日期:2012-05-28.

基金项目:江苏大学高级人才科研启动基金(06JDG029);江苏省高校自然科学研究计划项目(06KJB180016);江苏省博士后科研资助计划项目(1101037C)和转基因生物新品种培育科技重大专项(2009ZX08012-018B)。

作者简介:刘海军,男,湖北荆门人,助理研究员,研究方向为生物信息学,E-mail: liuhjun@126.com.

* 通讯作者:陈克平,男,博士,研究员,研究方向为蛋白质组学,E-mail: kpchen@ujs.edu.cn.

(Rice Proteome Database) 一个^[16], 这与水稻蛋白质组学研究数据迅猛增长是不相称的。本文利用瑞士生物信息学研究所 (SIB) 提供的免费开源软件 Make2D-DB II 构建了一个标准化的水稻二维电泳-质谱联动数据库 Rice2DDB^[17], 与 RPD 相比, 它不仅能记录二维电泳图谱和蛋白点信息, 还能记录质谱峰图数据及更详细的注释信息, 并且安装方便易于维护, 对水稻蛋白质组学数据进行管理、交流和利用。

1 材料和方法

1.1 实验材料和数据

构建 Rice2DDB 数据库的数据来源于作者所在实验室所承担的农业部转基因专项课题实验研究数据, Rice2DDB 数据库也属于该课题任务的一部分, 用于建立水稻二维电泳-质谱联动检测技术体系。目前实验室已收集了 48 份我国不同地区、不同品种的水稻种子材料, 包括 21 份粳稻品种 (*O. sativa* L. subsp. japonica)、25 份籼稻品种 (*O. sativa* L. subsp. indica) 和 2 份广亲和品种 (Wide-compatibility varieties), 已经完成 12 份水稻绿苗、25 份黄化苗、24 份三叶期叶片、4 份种子的蛋白质组表达谱, 同时完成了 2 054 个蛋白点的质谱鉴定工作。表 1 列出了部分已完成表达谱鉴定的水稻品种。由于该课题仍在进行中, 还有大量二维电泳实验和蛋白质鉴定工作仍在继续, Rice2DDB 数据库的数据信息还将继续扩充。

表 1 部分已完成表达谱鉴定的水稻品种

Table 1 Partial rice varieties completed protein identification of expression profiling

品种名称	亚种	取材	鉴定蛋白数
日本晴	粳稻	绿苗/黄化苗	57/15
徐稻 3 号	粳稻	绿苗	25
武育粳 3 号	粳稻	绿苗/黄化苗	18/63
南粳 46	粳稻	黄化苗	27
武运粳 8 号	粳稻	黄化苗	31
中花 11	粳稻	黄化苗/三叶期	32/165
Balilla	粳稻	黄化苗/三叶期	20/43
9311	籼稻	绿苗/黄化苗	153/30
Kasalath	籼稻	黄化苗	219
NJ11	籼稻	黄化苗	41
协青早	籼稻	黄化苗/三叶期	34/58
培矮 64	籼稻	黄化苗/三叶期	23/29
明恢 63	籼稻	黄化苗/三叶期	4/19
珍汕 97B	籼稻	黄化苗/三叶期/种子	27/232/62
02428	广亲和	黄化苗	27
Dular	广亲和	三叶期	118

1.2 数据库构建平台和方法

Rice2DDB 数据库主要基于免费开源软件包 Make2D-DB II v2.50.1 来构建完成。Make2D-DB II 是瑞士生物信息学研究所开发的一款专业二维凝胶电泳数据库创建发布软件包, 利用它科研人员可以轻松建立自己的专业二维电泳数据库, 而不需要太多的数据库开发知识^[17-19]。Make2D-DB II 可以把处理好的电泳图及蛋白点注释文件等自动导入到数据库管理系统中, 并自动生成一个基于 Web 查询管理页面的数据库系统, 该系统不仅可以管理多个子数据库, 还可以通过网络与外界同平台的数据库如 SWISS-2DPAGE、WORLD-2DPAGE 等进行互联互通实现远程联合查询^[20]。Make2D-DB II 建立了一套标准的电泳图谱数据检索浏览方式, 但又允许个性化的 Web 界面设计, 并能够整合包容多样化的注释信息及文档, 可以说功能非常强大。目前国内外利用 Make2D-DB II 构建完成并成功发布的二维电泳数据库已有 8 个, 包括人、牛、老鼠、家蚕、线虫、酵母、细菌和拟南芥等物种^[21], 水稻方面则还没有, Rice2DDB 数据库最终将加入这一平台联盟。

利用 Make2D-DB II 构建 Rice2DDB 数据库的主要步骤如下:

(1) 首先需要根据 Make2D-DB II 的要求准备好相应的数据文件, 包括水稻二维电泳胶图原图及其缩略图、所有胶图的名称及相关信息列表文件 (existing. maps)、与每张胶图对应的已鉴定蛋白点编号、图上坐标及鉴定的详细信息列表文件等。

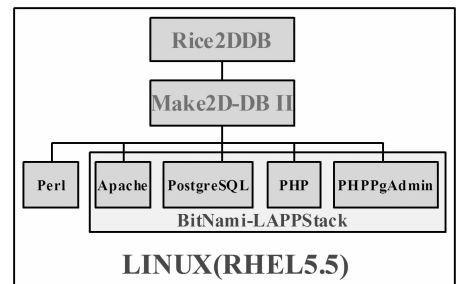


图 1 Rice2DDB 数据库构建平台

Fig. 1 The software platform to run Rice2DDB database

(2) 其次需要配置好 Make2D-DB II 运行所需的系统环境。Make2D-DB II 必须在类 Linux 操作系统上运行, 基于 PostgreSQL 数据库管理系统, 需要 Apache Web 服务和 Perl 脚本支持, 即所谓的 LAPP 平台 (Linux + Apache + PostgreSQL + Perl)。Rice2DDB 实际采用了 RHEL5.5 的操作系统平台及 BitNami 提供的 LAPP 集成安装包 bitnami-lappstack-1.2-2 (已集成 Apache 2.2.14 和 PostgreSQL 8.4.4)^[22], 其中 RHEL5.5 系统已自带 Perl 语言解释器, 如图 1 所示。

这些支持软件都是免费和开源的,可以从网上下载得到,这为构建 Rice2DDB 数据库提供了便利。

(3)之后要编辑 Make2D-DB II/config 文件夹下的 basic_include.pl、include.cfg、2d_include.pl 三个文件(后两个文件可由交互命令自动生成),分别对系统全局参数、数据库主要参数、Web 界面参数进行配置,三个文件均为纯文本文件,可用任何纯文本编辑器打开进行修改。

(4)最后则依次运行 Make2D-DB II 的三个 Perl 脚本命令 Perl make2db.pl -m [config|check|transform]来完成数据库的自动创建和数据导入。config 是对 include.cfg 和 2d_include.pl 两个配置文件进行设置或确认,check 是对第一步准备的数据文件进行格式检查,transform 是创建新数据库,把数据导入到数据库中并生成 Web 服务器。

关于初始数据文件格式、LAPP 平台的安装调试、Make2D-DB II 配置文件中参数的详细设置等,仍需要参考 Make2D-DB II 软件包的使用手册^[19]。

2 Rice2DDB 系统

构建完成的 Rice2DDB 数据库是一个基于 Web 访问,功能完备、界面友好的水稻蛋白质表达谱数据库,对实验产生的二维电泳胶图、已鉴定蛋白质点等数据能够进行多种方式的搜索查询和浏览(如图 2)。主页面左侧为导航栏,分为蛋白质检索、图谱浏览和子数据库及外部网络数据库选择三个主要的功能模块。蛋白质检索可按蛋白质登录号(如 Genbank 登录号)、关键词或基因名称、数据或文献提交作者、蛋白点编号、蛋白质鉴定方法、等电点或分子量范围及组合查询等七种方式来检索。图谱浏览则以缩略图汇总的形式提供了查询每张胶图的编号、实验方案、物种组织名、蛋白点数及其详细列表等信息的一个访问入口,分为实验信息、蛋白质列表和图片浏览三种方式。图谱浏览通过各种链接使得科研人员对数据库中的所有电泳图谱和蛋白质点信息能够一目了然。数据库选择模块可以让科研人员选择检索或图谱浏览的子数据库,检索时还可选择外部关联数据库(如同时检索 SWISS-2DPAGE、WORLD-2DPAGE 等数据库),或仅检索外部数据库排除本数据库(Exclude local DBs)。主页面右侧为首页和检索内容页,首页内容可进行个性化定制,还可对数据库后台及私有数据进行管理(可限制部分数据对外公开及被检索到)。

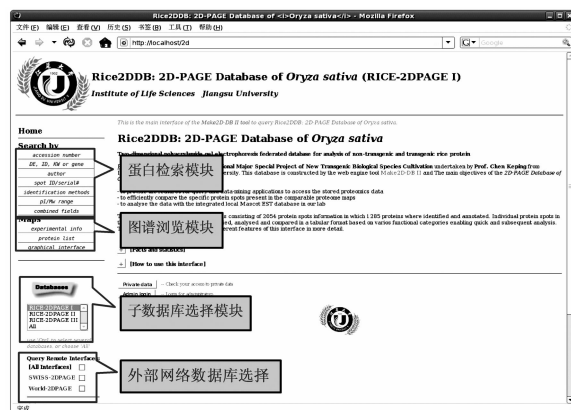


图 2 Rice2DDB 数据库主页介绍

Fig. 2 Rice2DDB database homepage introduction

Rice2DDB 数据库包含了 3 个子数据库,分别为绿苗、黄化苗和三叶期叶片的表达谱数据。构建数据库的初始数据主要包含了二维电泳表达谱扫描原图(JPG 格式)、原图对应的 100×80 像素左右的缩略图(JPG 格式)、每张图对应的已鉴定蛋白点信息文件和一个图列表文件(existing.maps)。其中每个蛋白点信息文件与其对应的图文件同名,为电子表格形式(CSV)的纯文本文件(.txt 后缀),每个文件中记录了对应图上已鉴定蛋白点的编号(SPOT)、坐标(X/Y)、等电点(PI)、分子量(MW)、登录号(AC)、鉴定信息(MS IDENT-URI)等一系列信息。其中在鉴定信息中加入了 Mascot PMF 比对结果的链接,可以直接点击该链接联网 Mascot 数据库查看详细的质谱比对结果,因而能与质谱数据形成联动。图列表文件也是一个 CSV 形式的纯文本文件,汇总记录了每个子数据库中所有胶图的文件名、全长名、胶图像素大小、等电点范围、分子量范围、taxid 分类号、品种名(strain)、取样组织、跑胶实验方案链接(protocols)、电泳图处理软件等一系列信息。

3 讨论和总结

Make2D-DB II 是一个功能强大的二维电泳数据库构建和发布工具,它在很大程度上缩小了开发建立专业二维电泳数据库所需的知识和成本,但在实际操作中仍存在一些值得注意的问题。如需要开发人员具备一定的 LAPP 平台配置知识并熟悉 Linux 系统下的命令行操作方式,需要注意 Make2D-DB II 与 PostgreSQL 及 Perl 版本的兼容性问题,需要注意二维电泳初始数据文件的格式问题等,否则一旦某个环节准备不充分,很容易造成数据库安装失败。

构建完成的 Rice2DDB 数据库具有友好的操作界面和丰富强大的检索查询功能,特别是能包含详细的蛋白质鉴定信息如质谱峰图数据和 Mascot 比对结果等,可以说实现了蛋白质组学研究数据的一站式管理和服务。由于目前 Rice2DDB 数据库只包含了部分实验数据,还有大量图谱注释和蛋白质鉴定工作未完成,因此 Rice2DDB 数据库还没有上网对外发布,但该数据库的构建方法对广大蛋白质组学科研人员仍具有重要的参考价值,将来完成发布也必将促进水稻蛋白质组学的研究和交流。

参考文献 (References)

- [1] 郭灿, 谢锦云. 水稻蛋白质组学研究[J]. 生命科学研究, 2003, 7(3):198-202.
- [2] 魏敏, 李阳生, 傅彬英. 水稻蛋白质组学研究进展[J]. 生物技术通报, 2005, 6:1-6.
- [3] 张晓婷, 郑奇君, 高飞, 李瑞, 赵全志. 蛋白质组学在水稻生长发育研究中的应用研究进展[J]. 河南农业科学, 2011, 40(1):1-5.
- [4] 程彦伟, 李建友, 姜爱良, 齐耀程, 朱倩, 赵昕, 张伟. 蛋白质组学在植物科学研究中的进展[J]. 江苏农业科学, 2010, 1:17-20.
- [5] 任丽萍, 范海延, 宋铁锋, 王珊珊, 刘晶瑜. 基于双向电泳技术的植物差异蛋白质组学研究进展[J]. 中国农学通报, 2011, 27(05):53-57.
- [6] 李雪梅, 李莹, 马莲菊, 刘旭, 陈亮. 逆境胁迫下水稻蛋白质组学研究进展[J]. 沈阳师范大学学报(自然科学版), 2009, 27(3):257-263.
- [7] International Rice Genome Sequencing Project. The map-based sequence of the rice genome[J]. Nature, 436(7052): 793-800.
- [8] Sasaki T. The rice genome project in Japan[J]. Proc. Natl. Acad. Sci. USA, 1998, 95(5):2027-8.
- [9] Sakata K, Nagamura Y, Numa H, Antonio BA, Nagasaki H, Itonuma A, Watanabe W, Shimizu Y, Horiuchi I, Matsumoto T, Sasaki T, Higo K. RiceGAAS: an automated annotation system and database for rice genome sequence[J]. Nucleic Acids Res., 2002, 30(1):98-102.
- [10] Kurata N., Yamazaki Y. Oryzabase: An Integrated Biological and Genome Information Database for Rice[J]. Plant Physiology, 2006, 140(1):12-17.
- [11] Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J, Buell CR. The TIGR Rice Genome Annotation Resource: improvements and new features[J]. Nucleic Acids Research, 2007, 35 (Database issue):D883-D887.
- [12] Liang C, Jaiswal P, Hebbard C, Avraham S, Buckler ES, Casstevens T, Hurwitz B, McCouch S, Ni J, Pujar A, Ravenscroft D, Ren L, Spooner W, Teclé I, Thomason J, Tung CW, Wei X, Yap I, Youens-Clark K, Ware D, Stein L. Gramene: a growing plant comparative genomics resource[J]. Nucl. Acids Res., 2008, 36 (Database issue): D947-D953.
- [13] Karlowski WM, Schoof H, Janakiraman V, Stuempflen V, Mayer KF. MOsDB: an integrated information resource for rice genomics[J]. Nucl. Acids Res., 2003, 31(1): 190-192.
- [14] Zhao W, Wang J, He X, Huang X, Jiao Y, Dai M, Wei S, Fu J, Chen Y, Ren X, Zhang Y, Ni P, Zhang J, Li S, Wang J, Wong GK, Zhao H, Yu J, Yang H, Wang J. BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics[J]. Nucl. Acids Res., 2004, 32 (Database issue): D377-D382.
- [15] 鄂志国, 庄杰云, 曹永生, 钱前, 王磊. 基于 Internet 的水稻基因数据库信息系统[J]. 中国水稻科学, 2006, 20(6): 670-672.
- [16] Komatsu S, Kojima K, Suzuki K, Ozaki K, Higo K. Rice Proteome Database based on two-dimensional polyacrylamide gel electrophoresis: its status in 2003[J]. Nucleic Acids Res, 2004, 32(Database issue): D388-92.
- [17] Mostaguir K, Hoogland C, Binz PA, Appel RD. The Make 2D - DB II package: Conversion of federated two-dimensional gel electrophoresis databases into a relational format and interconnection of distributed databases[J]. Proteomics, 2003, 3(8):1441-4.
- [18] Hoogland C, Baujard V, Sanchez JC, Hochstrasser DF, Appel RD. Make2ddb: a simple package to set up a two-dimensional electrophoresis database for the World Wide Web[J]. Electrophoresis, 1997, 18(15):2755-8.
- [19] Hoogland C, Mostaguir K, Sanchez JC, Hochstrasser DF, Appel RD. SWISS-2DPAGE, ten years later[J]. Proteomics. 2004, 4(8):2352-6.
- [20] Hoogl C, Mostaguir K, Appel RD. Creating 2DE databases for the World Wide Web[J]. Methods Mol Biol., 2009, 519:533-9.
- [21] SIB ExpASy server. World-2DPAGE Portal [EB/OL]. <http://world-2dpage.expasy.org/portal>, 2011-12-23.
- [22] BitNami. org. LAPPStack [EB/OL]. <http://bitnami.org/stack/lappstack>, 2011-11-24.