

RiceDB : 基于 Web 界面的水稻基因芯片注释整合数据库

何 飞 施清芸 陈 铭* 吴 平*

(浙江大学 生命科学学院 植物生理与生物化学国家重点实验室, 浙江 杭州 310058; * 通讯联系人, E-mail: mchen@zju.edu.cn)

RiceDB : a Web Based Integrated Database for Annotating Rice Microarray

HE Fei, SHI Qingyun, CHEN Ming*, WU Ping*

(State Key Laboratory of Plant Physiology & Biochemistry, College of Life Sciences, Hangzhou 310058, China; * Corresponding author, E-mail: mchen@zju.edu.cn)

Abstract : RiceDB, a web based integrated database to annotate rice microarray in various biological contexts was developed. It provides convenient query for comprehensive and auto updated annotation compared with other similar databases. It is composed of eight modules. RiceMap module archives the process of Affymetrix 57K probe sets mapping to different databases about rice, and aims to the genes represented by a microarray set by retrieving annotation information via identifier or accession number of every database; RiceGO module indicates the association between a microarray set and gene ontology categories consisting of biological process, molecular function and cellular component; RiceKO module is used to annotate a microarray set based on the KEGG biochemical and regulatory pathways; RiceDO module indicates the information of domain associated with a microarray set; RiceUP module is used to obtain promoter sequences for all genes represented by a microarray set; RiceMR module lists potential microRNA which regulated the genes represented by a microarray set; RiceCD is used to annotate the genes represented by a microarray set in the context of chromosome distribution; RiceGF is used to annotate the genes represented by a microarray set in the context of rice paralogous gene family distribution.

Key words : microarray annotation; *Oryza sativa*; molecular database

摘 要 : 开发了 1 个基于 Web 界面在生物学背景下注释水稻基因芯片的整合数据库 RiceDB, 它的注释信息比同类针对水稻 Affymetrix 57K 全基因组表达谱芯片的注释数据库更全面、更快捷且可以自动更新。它由 8 个功能模块组成, RiceMap 建立水稻 Affymetrix 57K 探针组与基因的对应关系, 从而提供来自各个水稻生物信息学数据库功能描述信息; RiceGO 整合 Gene Ontology 数据库从生物学途径/分子功能和细胞构建对 Affymetrix 57K 探针组进行注释; RiceKO 确定 Affymetrix 57K 探针组代表的基因可能参与的代谢与调控路径; RiceUP、RiceDO、RiceMR、RiceCD、RiceGF 分别确定 Affymetrix 57K 探针组代表的基因上游小于 1 kb 启动子序列、结构功能域组成、受哪些水稻 MicroRNA 调控、染色体定位、基因家族分类。

关键词 : 芯片注释; 水稻; 分子数据库

中图分类号: Q75; Q943; S511.01

文献标识码: A

文章编号: 1001-7216(2007)03-0235-07

水稻是最重要的粮食作物之一, 全世界 60% 的人口以水稻为主食。国际水稻基因组测序计划 (IRGSP) 和中国科学院基因组信息中心分别完成了水稻两个亚种粳稻 (*Oryza sativa* L. ssp. *japonica*) 和籼稻 (*Oryza sativa* L. ssp. *indica*) 全基因组工作框架图的测序^[1-3]; 水稻日本晴全长 cDNA 测序计划 (KOME) 产生大约 32 000 条全长 cDNA 序列^[4]; 大约 300 000 条水稻表达序列标签 (EST) 存储在美国国立生物技术信息中心数据库 (NCBI)。美国基因组研究所 (TIGR)、日本水稻全长 cDNA 数据库 (KOME)、谷类比较基因组数据库 (Gramene)、植物基因组数据库 (PlantGDB) 等已开始对水稻基因组进行注释, 并建立了相关水稻基因功能注释数据库。

随着模式植物水稻基因组序列、全长 cDNA 序列、EST 序列不断产生和功能注释不断更新完善, 表达谱基因芯片技术逐渐应用在水稻功能基因组研究中。目前应用最为广泛的表达谱基因芯片平台属

于假阳性率最低、重复性最好、质控标准最高、最省时和性价比最高的 Affymetrix (<http://www.affymetrix.com/>)^[5]。它生产的水稻全基因表达谱芯片是针对粳稻和籼稻来自 Unigene、GenBank 和 TIGR (Osa1) 版本 2 总共 51 279 个表达序列来设计探针组 (probe set) 的。每个探针组由 11 ~ 20 对 25 个核苷酸组成的完全匹配 (perfect match) 和 1 个碱基错配 (mis match) 的探针序列组成。

芯片数据的产生对于 Affymetrix 商业化生产的芯片平台来说技术已经成熟且标准化。芯片数据分析过程涉及数据预处理、确定候选基因和生物学意义挖掘, 后者统称为芯片数据的生物学分

收稿日期: 2006-10-08; 修改稿收到日期: 2006-11-13。

基金项目: 国家重点基础研究发展计划资助项目 (2005CB120900); 国家自然科学基金资助项目 (30500106); 浙江省教育厅专项资助项目 (20051038)。

第一作者简介: 何 飞 (1982-), 男, 硕士研究生。

析^[6]。我们针对水稻 Affymetrix 57K 芯片数据生物学分析过程遇到的困难进行简单介绍。

首先,确定水稻 Affymetrix 57K 每个探针组对应于哪个水稻基因,然后搜索有关水稻的生物信息学数据库获取该基因的功能描述、基因本体论功能分类 (gene ontology)、可能参与的代谢与调控路径、染色体定位、结构功能域组成等注释信息。虽然 Affymetrix 开发的芯片注释数据库 NetAffx 提供水稻 Affymetrix 57K 探针组的注释信息,但是信息不全面且更新不及时。水稻基因组测序完成后,利用生物学实验研究和计算机程序预测方法产生的海量基因注释信息被分散在不同的水稻生物信息学数据库中。这些数据库由于对水稻基因的注释方法不同而产生互补和冲突的结果,迫切需要通过数据库整合手段为研究人员查询提供方便。

其次, Affymetrix 57K 全基因表达谱芯片实验通常产生一组或几组差异表达探针组。研究人员面临的挑战就是差异表达探针组的生物学分析。假设差异表达探针组数量是 1~30 个,很容易通过单个查询各个水稻生物信息学数据库找到该探针组代表基因的注释信息,但是数量增加到 100~1000 个时,手工访问数据库的工作庞大而烦琐。为了解决这个难题必须通过计算机程序语言和数据库技术自动解析和存储这些数据库的基因注释信息,并相应根据基因本体论功能分类、生化代谢与调控路径、基因家族、结构功能域、染色体定位对芯片差异表达探针组进行归类分析。

1 现有注释系统

为了解决以上存在的问题,目前几个基于 Web 服务的数据库和工具已经实现了自动化注释水稻 Affymetrix 57K 全基因表达谱芯片探针组,它们可以对实验产生的差异表达探针组进行功能注释和归类分析。

1.1 NetAffx

Affymetrix 的 NetAffx 数据库提供属于探针组序列号的每个探针序列和它在靶标序列位置的查询^[7],同时提供来自 GenBank、Unigene、SwissProt、TIGR (Osa1) 版本 2 等公共数据库对探针组序列号代表基因的注释信息。通过 NetAffx 可以批量查询注释信息,但信息不全面、更新不及时,大部分探针组只对应于 GenBank 基因标识和简单的功能描述,只有小部分探针组有对应的 TIGR (Osa1) 版本 2 的基因座位号和 Swissprot 基因标

识。

1.2 Rice Multiple Microarray Search

水稻多种芯片探针注释查询系统 (Rice Multiple Microarray Search) 是针对美国国家科学基金会水稻寡核苷酸 20K、45K 芯片探针注释而建立的,它也提供水稻 Affymetrix 57K 表达谱芯片探针组序列号与 TIGR (Osa1)、KOME、the TIGR Rice Gene Index Release 15 这些数据库的水稻基因标识的对应关系,但没有直接提供注释信息。

1.3 TIGR/PlantGDB Genome Browser

TIGR/PlantGDB 基因组序列浏览工具整合了 TIGR (Osa1) 版本 4 日本晴水稻全基因组组装序列和水稻 Affymetrix 57K 芯片探针组序列,该工具的特点是芯片探针序列通过比对定位到水稻基因组上的特定位置,通过该位置所处的水稻基因来建立芯片探针组序列号与水稻基因的对应关系,但也不能直接通过探针序列号获取注释信息。

1.4 BarleyBase

BarleyBase 整合了 NetAffx 数据库中水稻 Affymetrix 57K 芯片探针序列、靶标序列和对应的水稻基因序列^[8];通过植物同源基因关系建立水稻 Affymetrix 57K 芯片的探针序列号与拟南芥 ATH1 22K 芯片、大麦 Barley 22K 芯片的探针序列号的对应关系,从而实现了跨物种的芯片探针组注释;此外,它对水稻 Affymetrix 57K 芯片实验产生的差异表达探针组进行基因本体论功能归类分析;BarleyBase 还提供了拟南芥 ATH1 22K 芯片探针组在代谢与调控路径、基因家族、结构功能域的注释信息,但是没有针对水稻做这方面的工作。

以上几个涉及水稻 Affymetrix 57K 芯片探针组的注释数据库都有各自的特点和缺陷,为了弥补以上这些注释数据库建设的不足,高效、准确、全面地提供水稻 Affymetrix 57K 芯片探针组的注释信息,我们开发了基于 Web 服务的在生物学背景下注释水稻基因芯片的数据库。

2 RiceDB 注释系统

RiceDB 是 1 个针对水稻在生物学背景下注释基因芯片的 MySQL 数据库管理系统 (<http://www.genomics.zju.edu.cn/ricedb>)。它是由 NetAffx、TIGR (Osa1) 版本 4、KOME、KEGG、Gene Ontology、TAIR、Interpro、miRBase 这些有关水稻的生物信息学数据库整合而成的综合性注释数据库。RiceDB 可通过 Perl 开发的程序代码自动更

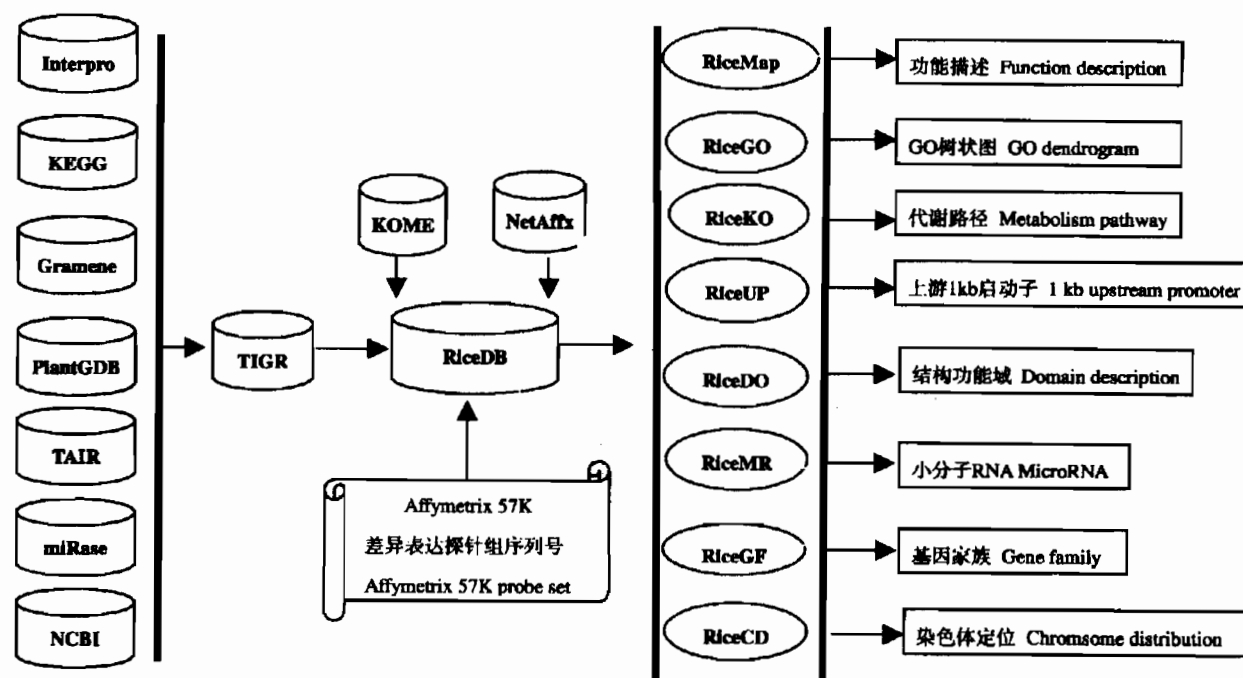


图 1 RiceDB 系统结构示意图

Fig. 1. The system architecture of RiceDB.

新,8 个用 Perl 语言开发的核心模块 (RiceMap、RiceGO、RiceKO、RiceUP、RiceDO、RiceMR、RiceCD、RiceGF)调用 RiceDB 数据库资源满足用户的需要(图 1)。

2.1 RiceMap

RiceMap 执行的功能是获取水稻 Affymetrix 57K 芯片探针组对应的基因和注释信息。用户可以根据需要选择不同的数据库,RiceMap 结果以网页的形式返回指定数据库的基因标识和注释信息(表 1)。

RiceMap 实现方法是 Affymetrix 57K 芯片探针组对应的靶标序列和 TIGR(Osa1)或 KOME 的全部水稻 cDNA 序列进行 BLASTn 联配,以 0.000 001 作为 E 值初步筛选,通过是否符合以下两个标准来确定两者之间的对应关系:1) 以靶标序列为基准,100% 匹配且匹配的序列长度大于等于 50 bp(base pair) 或 98% 匹配且匹配的序列长度等于

靶标序列;2) 如果第一个标准无法找到与靶标序列匹配的 cDNA 序列,适当降低标准为 98% 匹配,并且匹配序列长度大于等于 50 bp。

考虑到 TIGR (Osa1) 预测的水稻基因基本包含 Gramene 数据库收录的水稻基因,所以对来自这两个数据库中的水稻蛋白质序列进行 BLASTp 联配,用 Perl 语言编写的脚本提取 BLAST 结果中最优匹配的两条序列,实现 TIGR (Osa1) 基因标识与 Gramene 基因标识相互对应;考虑到水稻基因的功能注释往往大部分来自拟南芥的同源基因的功能注释,所以对 TIGR (Osa1) 水稻蛋白质序列与拟南芥信息资源中心 (TAIR) 拟南芥蛋白质序列进行 BLASTp 联配,用 Perl 语言编写的脚本提取 BLAST 结果中最优匹配的两条序列,实现 TIGR (Osa1) 基因标识与 TAIR 基因标识相互对应;此外, Gene Ontology 标识、Interpro 数据库的结构功能域标识和基因家族分类可以从 TIGR (Osa1)、

表 1 水稻生物信息学数据库的基因标识以及网址

Table 1. IDs reference into rice bioinformatics databases and websites.

标识 ID	数据库 Database	网址 Website
Affymetrix 探针组序列号 Affymetrix probe set	NetAffx	http://www.affymetrix.com/analysis/index.affx
TIGR 基因座位号 TIGR locus identifier	TIGR (Osa1)	http://rice.tigr.org
KOME 基因标识 KOME accession number	KOME	http://cdna01.dna.affrc.go.jp/cDNA/
PlantGDB 基因座位号 PlantGDB locus identifier	PlantGDB	http://www.plantgdb.org
Interpro 结构功能域标识 Interpro domain identifier	Interpro	http://www.ebi.ac.uk/interpro/
TAIR 基因座位号 TAIR locus identifier	TAIR	http://www.arabidopsis.org/
Gene Ontology 标识 Gene Ontology identifier	Gramene	http://www.gramene.org/

Gramene、KOME 数据库中直接获取,并且通过超链接直接访问相应的数据库,拟南芥的代谢与调控路径注释信息可以直接来自 TAIR 的 Aracyc (<http://www.arabidopsis.org/tools/aracyc/>) 和 KEGG 代谢与调控路径数据库。

2.2 RiceGO

基因本体论 (Gene Ontology, GO) 是 20 世纪 90 年代被开发应用于基因组功能注释和基因之间功能关联性分析的工具,每个 Gene Ontology 具有一个数字标识和简化的本体论术语。本体论术语被归类成三个方面:生物途径、分子功能和细胞构建。对水稻 Affymetrix 57K 芯片探针组进行基因本体论功能分类注释可以推断一组差异表达或者表达模式类似的基因可能参与哪些生物途径,或属于哪些细胞构建,或行使哪些分子功能。

目前,利用基因本体论功能分类原理开发的工具有基于 Web 服务的应用程序 FatiGO^[9]、Goduster^[10]、NetAffx GO Mining Tool^[11]、DAVID^[12] 和基于 JAVA 的应用程序 GoMiner^[13]。这些工具大部分运用统计学分析来确定一组或几组差异表达基因在哪个 Gene Ontology 标识分布最密集,从而揭示生物学意义。

RiceGO 是一个针对水稻 Affymetrix 57K 芯片实验产生的差异表达探针组基因本体论功能分类注释工具。Gramene、KOME 和 TIGR (Osa1) 各自存储水稻基因本体论功能分类注释,它们对水稻同一个基因的注释存在互补性。先前我们利用 RiceMap 把所有水稻基因统一成为国际上普遍采用的 TIGR (Osa1) 的水稻基因标识,进而通过统一的水稻基因标识来提取这些数据库的基因本体论功能分类注释信息。

RiceGO 的实现方法是用 Perl 语言编写的脚本从 Gramene、KOME、TIGR (Osa1) 提取每个基因对应的基因本体论功能分类注释信息。系统读取用户提交的芯片实验产生的差异表达探针组序列号,自动运行 RiceMap 获取与之对应的 TIGR (Osa1) 的基因标识。RiceGO 提取 Gramene、KOME、TIGR (Osa1) 数据库基因本体论功能分类注释信息以网页的形式返回给用户,RiceGO 产生的网页中每个 Gene Ontology 标识提供超链接到 AmiGO 浏览器 (<http://www.godatabase.org/cgi-bin/amigo/go.cgi>)。RiceGO 也可以读取用户提交的 Gene Ontology 标识返回属于这个 Gene Ontology 标识的所有探针组序列号。

2.3 RiceKO

KEGG 是目前应用最为广泛的多物种代谢与调控路径数据库。Pathway Processor^[14]、PathMAPA^[15]、ArrayXPath^[16] 都是基于 KEGG 对芯片探针注释的工具。KOBAS^[17] 是以 KO (KEGG Ortholog, KO) 为基础的自动注释系统,KO 是用来整合基因组注释和代谢调控路径注释的桥梁。如果水稻所有基因赋予 KO 标识的话,就可以在代谢与调控路径的水平上来分析它们的功能。跟 KOBAS 类似,RiceKO 执行的功能是对水稻 Affymetrix 57K 芯片探针组进行 KEGG 代谢与调控路径的注释。

RiceKO 的实现方法是用 Perl 语言编写的复杂的正则表达式对来自 KEGG 数据库版本 32.0 中的基因数据文件和 KO 数据文件进行解析。结果包括 5 540 个 KO,179 个代谢调控路径,748 177 个基因核酸序列和氨基酸序列,232 637 个基因与 KO 的对应关系。这些信息用 SQL 数据库语言装进 MySQL 数据库中。系统读取用户提交的水稻 Affymetrix 57K 实验产生的差异表达探针组序列号,自动运行 RiceMap 获取与之对应的 TIGR (Osa1) 的基因标识,RiceKO 通过基因标识获取该基因的蛋白质序列,利用 BLAST 程序与 KEGG 数据库的所有基因蛋白质序列进行相似性匹配,最终获得相应的 KO 标识。提交的基因被赋予 KO 标识需满足以下条件:1) 必须含有已知 KO 标识的命中序列;2) 提交序列与命中序列匹配相似度 E 值小于 0.000 01;3) 存在不超过 5 条 E 值更低的命中序列。最后,RiceKO 用 Perl 语言编写的代码可以解析每个 KO 标识对应的代谢与调控路径标识,从而完成对水稻 Affymetrix 57K 芯片探针组进行代谢与调控路径的注释。

2.4 RiceUP

RiceUP 执行获取水稻 Affymetrix 57K 芯片探针组对应的基因上游小于 1 kb 启动子序列。系统读取用户提交的水稻 Affymetrix 57K 芯片实验产生的差异表达探针组序列和指定的序列长度,RiceUP 返回每个探针组对应的水稻基因上游启动子序列。同时提供超链接到 PLACE 数据库做后续分析。PLACE 数据库是一个专门从已报道文献中收录在植物中发现的具有顺式调控作用的 DNA 序列元件数据库,同时也为每个序列提供了一些简短的描述和相关的报道文献^[18] (<http://www.dna.affrc.go.jp/PLACE/>)。它所开发的一个作用元件

搜索工具可以为用户提交的序列确定是否存在顺式调控作用元件。

2.5 RiceDO

RiceDO 执行获取水稻 Affymetrix 57K 芯片探针组对应基因的蛋白质结构功能域标识。同时对每个结构功能域标识提供超链接到 Pfam (<http://www.sanger.ac.uk/Software/Pfam/>) 和 Interpro 数据库获取详细的功能描述和相关文献^[19]。Pfam 是一个利用多序列比对程序和隐马尔可夫模型对所有物种的蛋白质序列进行基因家族归类 and 提取存储结构功能域信息的数据库。

2.6 RiceMR

考虑到水稻 Affymetrix 57K 芯片实验产生的芯片表达数据中具有相似表达模式的一组基因可能受同一个或同一类 MicroRNA 调控, RiceMR 执行获取水稻 Affymetrix 57K 芯片探针组对应的基因可能受哪些 MicroRNA 调控。我们利用 miRU 对来自 miRBase 所有水稻的 MicroRNA 序列预测靶标序列。miRU 是一个植物 MicroRNA 靶标序列预测服务器^[20]。

2.7 RiceCD

近来有研究报道,真核生物具有相似表达模式

的一组基因与它们在基因组所处的位置是有关联的^[6, 21]。为此, RiceCD 执行获取水稻 Affymetrix 57K 芯片探针组对应基因的染色体基因组位置信息。

2.8 RiceGF

RiceGF 执行获取水稻 Affymetrix 57K 芯片探针组对应基因属于哪个基因家族的信息。我们的数据来自 TIGR(Osa1)的根据是否存在共同的结构功能域的基因家族分类信息。

3 系统实现

RiceDB 存储在安装了 Linux 操作系统、Apache Web 服务器、MySQL 数据库、ActivePerl 5.8.7 接口程序的 PC 服务器中。系统根据用户输入信息用 Perl 跟 CGI 动态生成网页技术通过 Apache HTTP 服务器运送网页给客户端(图 2)。BioPerl 用来解析所有有关水稻生物信息学数据库的注释信息 (http://bioperl.org/wiki/Main_Page)。

4 讨论

我们已经开发了一个基于 Web 界面在生物学



图 2 RiceDB 系统主页面

Fig. 2. The mainpage of RiceDB system.

表 2 RiceDB 与其他注释数据库的比较

Table 2 . Comparison of RiceDB with other annotation databases .

注释内容 Annotation content	NetAffx	Rice Multiple Microarray Search	TIGR/PlantGDB Genome Browser	BarleyBase	RiceDB
NCBI 注释 NCBI annotation	+	+	+	+	+
TIGR 注释 TIGR annotation	-	+	+	+	+
KOME 注释 KOME annotation	-	+	+	+	+
GO 注释 GO annotation	+	-	-	+	+
结构功能域 Domain description	+	-	-	-	+
代谢与调控路径 Pathway	-	-	-	-	+
上游启动子序列 Upstream promoter	-	-	-	-	+
染色体定位 Chromosome distribution	+	-	+	+	+
小分子 RNA 调控 MicroRNA regulatory	-	-	-	-	+

“+”代表具备的功能，“-”代表不具备的功能。

“+”represents function contained,“-”represents function not contained.

背景下注释水稻基因芯片的整合数据库。任何一个完整的基因芯片注释数据库需要它能够提供全面、及时更新的信息,方便的数据查询功能。RiceDB 的发展仍然在初期阶段,它为继续整合有关水稻生物信息学数据库和开发水稻 Affymetrix 57K 全基因表达谱芯片的信息查询、处理、管理以及可视化奠定了基础。RiceDB 可以对水稻 Affymetrix 57K 芯片实验产生的差异表达探针组分别在功能描述、基因本体论功能分类、代谢与调控路径、结构功能域、染色体定位、基因家族等水平上提供注释信息。与现有针对水稻 Affymetrix 57K 芯片探针组注释数据库比较发现 RiceDB 注释的深度和广度明显优于其他数据库,但是也有不足之处需要改进(表 2)。

RiceDB 正在不断更新版本,未来有以下几个方面需要改进:

1)整合其他物种的生物信息学数据库(如拟南芥、玉米、大麦等),提供这些物种的 Affymetrix 芯片探针组的注释信息,实现跨物种的 Affymetrix 芯片探针组序列号查询。

2)用 RiceKO 和 RiceGO 整合 R 语言统计学分析工具,实现水稻 Affymetrix 57K 芯片实验产生的差异表达探针组在基因本体论功能分类和代谢与调控路径上的显著分布分析。

3)注释信息可视化技术还有待加强。比如,RiceGO 实现 Gene Ontology 标识的图形化显示,以更直观的方式提供给用户。

面对这些挑战,我们将把这个基于 Web 界面的在生物学背景下注释水稻基因芯片的整合数据库做得更加完善。

参考文献:

[1] sasaki T, Burr B. International rice genome sequencing pro-

ject: The effort to completely sequence the rice genome. *Curr Opin Plant Biol*, 2000, 3: 138-141.

[2] Yu J, Hu S, Wang J, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, 2002, 296: 79-92.

[3] International Rice Genome Sequencing Project. The map based sequence of the rice genome. *Nature*, 2005, 436: 793-800.

[4] Kikuchi S, Satoh K, Nagata T, et al. Collection, mapping and annotation of over 28,000 cDNA clones from japonica rice. *Science*, 2003, 301: 376-379.

[5] Hennig L, Grussem W, Grossniklaus U, et al. Transcriptional programs of early stages of plant reproduction. *Plant Physiol*, 2004, 132: 640-652.

[6] Clarke J D, Zhu T. Microarray analysis of the transcriptome as a stepping stone towards understanding biological systems: Practical considerations and perspectives. *Plant J*, 2006, 45: 630-650.

[7] Liu G, Loraine A E, Shigeta R, et al. NetAffx: Affymetrix probesets and annotations. *Nucl Acids Res*, 2003, 31: 82-86.

[8] Shen L, Gong J, Caldo R A, et al. BarleyBase: An expression profiling database for plant genomics. *Nucl Acids Res*, 2005, 33: D614-D618.

[9] Al-Shahrour F, Diaz Uriarte R, Dopazo J, et al. FatiGO: A web tool for finding significant associations of gene ontology terms with groups of genes. *Bioinformatics*, 2004, 20: 578-580.

[10] Wrobel G, Chalmel F, Primig M, et al. Goduster integrates statistical analysis and functional interpretation of microarray expression data. *Bioinformatics*, 2005, 21: 3575-3577.

[11] Cheng J, Sun S, Tracy A. NetAffx gene ontology mining tool: A visual approach for microarray data analysis. *Bioinformatics*, 2004, 20: 1462-1463.

[12] Dennis G Jr, Sherman B T, Hosack D A, et al. DAVID: Database for annotation, visualization and integrated discovery. *Genome Biol*, 2003, 4: 3.

[13] Zeeberg B R, Feng W M, Wang G, et al. GoMiner: Resource for biological interpretation of genomic and proteomic data. *Genome Biol*, 2003, 4: 28.

- [14] Grosu P , Townsend J P , Hartl D L , et al . Pathway processor : A tool for integrating whole genome expression results into metabolic networks . *Genome Res* ,2002 , 12 : 1121-1126 .
- [15] Pan D , Sun N , Cheung K H , et al . PathMAPA : A tool for displaying gene expression and performing statistical tests on metabolic pathways at multiple levels for *Arabidopsis* . *BMC Bioinformatics* ,2003 , 4 : 56 .
- [16] Chung H J , Kim M , Park C H , et al . ArrayXPath : Mapping and visualization microarray gene expression data with integrated biological pathway resources using Scalable Vector Graphics . *Nucl Acids Res* ,2004 , 32 , 460-464 .
- [17] Mao X , Cai T , Olyarchuk J G , et al . Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary . *Bioinformatics* ,2005 , 21 : 3787-3793 .
- [18] Higo K , Ugawa Y , Iwamoto M , et al . Plant cis acting regulatory DNA elements (PLACE) database . *Nucl Acids Res* , 1999 , 27 : 297-300 .
- [19] Bateman A , Coin L , Durbin R , et al . The Pfam protein families database . *Nucl Acids Res* ,2004 , 32 : 138-141 .
- [20] Zhang Y J . miRU : An automated plant miRNA target prediction server . *Nucl Acids Res* ,2005 , 33 : 701-704 .
- [21] Spellman P T , Rubin G M . Evidence for large domains of similarly expressed genes in the *Drosophila* . *Genome J Biol* , 2002 , 1 : 5 .