水稻害虫转录组数据库

吴昊^①, 李飞^②, 华红霞^③, 王睿^①, 高亮^①, 侯吉祥^①, 陈浩^①, 周菲^①, 林拥军^{①*}

① 华中农业大学生命科学技术学院, 作物遗传改良国家重点实验室, 国家植物基因研究中心(武汉), 武汉 430070;

② 南京农业大学植物保护学院昆虫学系, 南京 210095;

③ 华中农业大学植物科学技术学院, 武汉 430070

* 联系人, E-mail: yongjunlin@mail.hzau.edu.cn

收稿日期: 2014-05-10; 接受日期: 2014-06-15

国家科技重大专项转基因生物新品种培育专项(批准号: 2011ZX08001-001)和国家高技术研究发展计划(批准号: 2012AA10A304)资助项目
doi: 10.1360/052014-109

摘要 对于亚洲的一些国家, 水稻是最重要的粮食作物, 然而其产量却遭受数百种害虫的危害. 尽管目前已存在许多控制虫害的策略, 但是损失依然非常严重. RNA 干涉被认为在农业害虫防控领域具有良好的应用前景, 但是遗传信息的匮乏严重限制了该技术的应用. 水稻害虫转录组数据库是一个综合性的数据库, 其数据来源于 8 种最重要的水稻害虫的新一代测序结果. 本数据库致力于为昆虫功能基因组学研究、昆虫治理及其他领域提供一个有条理的遗传信息平台, 以满足上述研究领域的需要. 另外, 本数据库通过进化树的构建揭示出这些昆虫的进化关系. 所有的信息公布于免费网站: <http://rptdb.hzau.edu.cn>.

关键词

转录组数据库
水稻害虫
新一代测序
进化树

水稻(*Oryza sativa* L.)是世界上种植面积最广的作物. 2010 年, 水稻全球种植面积超过 1.53×10^8 hm^2 , 其中近 0.3×10^8 hm^2 在中国(<http://faostat.fao.org/site/567/default.aspx>). 作为十几亿中国人的主粮, 水稻受超过 200 种害虫的侵害, 其中最主要的种类隶属于鳞翅目(Lepidoptera)和半翅目(Hemiptera). 鳞翅目中危害最大的昆虫为三化螟(*Tryporyza incertulas* (Walker))、二化螟(*Chilo suppressalis* (Walker))和稻纵卷叶螟(*Cnaphalocrocis medinalis* (Guenée)); 而半翅目主要有褐飞虱(*Nilaparvata lugens* (Stål))、灰飞虱(*Laodelphax striatellus* (Fallen))和白背飞虱(*Sogatella furcifera* (Horvath))^[1]. 另外, 近十年, 在中国的部分区域, 分属于鳞翅目和半翅目的大螟(*Sesamia inferens* (Walker))和黑尾叶蝉(*Nephotettix cineticeps* (Uhler))逐渐发展为危害水稻产量的先锋种. 尽管已

采取多种策略来控制虫害, 但收效有限, 损失依然巨大, 仅二化螟每年就可造成 115 亿人民币的损失^[2,3]. 另外, 为抑制虫害, 杀虫剂的滥用又导致诸多环境问题. 通过转基因技术培育出的 Bt 抗虫水稻被认为是最有效且最安全的抗虫手段, 然而潜在的对 Bt 蛋白的抗性风险将限制该策略的应用, 所以需要提前开发出抗虫的新技术^[4]. RNA 干涉(RNA interference, RNAi)作为一种能够有效抑制内源 mRNA 转录水平的工具已被用于一些害虫治理的实验中, 并被认为是具有巨大的应用潜力^[5-7]. 然而, 对于一些重要的水稻害虫, 遗传资源的匮乏严重阻碍了该技术的应用, 为此, 利用新一代测序技术收集了这些必需的遗传信息以便更深入地研究, 完成了水稻最重要的 8 个害虫的转录组测序. 通过对这些非重复序列(Unigene)的注释和整理, 发布了一个综合且有序的水稻害虫

引用格式: 吴昊, 李飞, 华红霞, 等. 水稻害虫转录组数据库. 中国科学: 生命科学, 2014, 44: 832-836

Wu H, Li F, Hua H X, et al. Rice pest transcriptome database. SCIENTIA SINICA Vitae, 2014, 44: 832-836, doi: 10.1360/052014-109

转录组数据库(Rice Pest Transcriptome Database, RPTDB, <http://rptdb.hzau.edu.cn>).

1 材料与方法

1.1 数据来源

鳞翅目害虫样品均采集于其幼虫阶段, 而半翅目害虫样品均采集于其若虫阶段. 利用 Invitrogen 公司的 TRIzol 试剂盒(Cat. No. 15596-026, 美国)提取昆虫的总 RNA 用于测序.

1.2 序列拼接

SOAPdenovo 软件用于对测序原始数据的组装进而获得 Unigene.

1.3 注释

将获得的序列比对 4 个主要数据库(NR, Swiss-Prot, KEGG 和 COG)以获取注释信息, 阈值均设定为 E 值 ≤ 0.00001 , 覆盖度 $\geq 50\%$, 一致性 $\geq 30\%$. 利用这些数据和 CDS(coding sequences)预测软件 ESTScan, 获得部分蛋白序列. 随后又通过 HMMER 预测功能结构域, E 值阈值设置为广泛接受的 0.01 ^[8,9].

1.4 直系同源(ortholog)和旁系同源(paralog)预测

使用 RBH(reciprocal best hit)法^[10]预测不同物种间的直系同源. 通过对两个物种的 all-against-all BLASTP, 定义覆盖度 $\geq 50\%$, 一致性 $\geq 50\%$, 且 E 值 $\leq 10^{-5}$ 的最佳匹配序列对为直系同源. 同时, BSR(BLAST score ratio)法^[11]被用于预测旁系同源, 阈值设定为覆盖度 $\geq 50\%$, 一致性 $\geq 50\%$, $BSR \geq 0.4$, 且 E 值 $\leq 10^{-5}$.

1.5 进化树构建

如上文所述, 在两个目昆虫间用 RBH 法预测了可能的直系同源. 半翅目害虫中的黑尾叶蝉由于预测出的直系同源序列数量有限未被用于进化树构建, 同时, 三化螟被作为外群. 提取所有的保守编码区, 作为 PAML 软件包中 mcmctree^[12,13]的输入数据以构建进化树. 为了预测进化关系和物种分化时间, 使用了果蝇(*Drosophila melanogaster*)的中性替换率(the neutral substitution rate of fruit fly (*Drosophila*

melanogaster) (11.1×10^{-9})^[14].

相似的工作也在鳞翅目昆虫中进行, 同样因为缺乏直系同源序列, 大螟未被用于进化树构建. 该进化树构建过程中, 褐飞虱被设置为外群.

1.6 数据库实现

采用 MySQL 作为数据库的管理系统, 系统地管理数据并构建出一个运行于 linux 服务器上的数据库. HTML 和 PHP 被用作网络页面的输出.

2 结果

2.1 数据分析结果

通过 SOAPdenovo 软件对原始数据的组装, 每个昆虫获得对应的 Unigene 集合, 随后通过对多个数据库的比对及若干软件的使用, 预测获得各个昆虫的蛋白序列集合. 不同昆虫的集合间相互比对, 直系同源与旁系同源也被预测得到. 进化树的构建揭示出分别隶属于鳞翅目和半翅目两组昆虫的进化关系(图 1 和表 1).

2.2 数据浏览

RPTDB 对每一个昆虫都有一个友好的用户入口. 用户点击害虫的名字或者图片以浏览对应昆虫的主页. 每个昆虫的所有 Unigene 采用 COG(cluster of orthologous groups of proteins)功能分类体系, 被列入

图 1 两个目昆虫的进化树

A: 半翅目; B: 鳞翅目. 黑尾叶蝉和大螟由于直系同源的序列数量有限未被用于建树. 果蝇的中性替换率被用于预测物种分化时间

图2 数据库内容概观

A: RPTDB 中的 8 个物种; B: 白背飞虱的所有 Unigene 根据 COG 所属类别分类; C: 其中一个类别的所有序列列表; D: 其中一个 Unigene 的注释页面

表 1 RPTDB 中 8 个转录组的基本信息

学名	中文名	目: 科	Unigenes ^{a)}	蛋白 ^{b)}
<i>Chilo suppressalis</i>	二化螟	鳞翅目: 螟蛾科	116262	41967
<i>Tryporyza incertulas</i>	三化螟	鳞翅目: 螟蛾科	99709	38336
<i>Cnaphalocrocis medinalis</i>	稻纵卷叶螟	鳞翅目: 螟蛾科	124232	60876
<i>Sesamia inferens</i>	大螟	鳞翅目: 夜蛾科	49390	4414
<i>Nilaparvata lugens</i>	褐飞虱	半翅目: 飞虱科	76406	33093
<i>Laodelphax striatellus</i>	灰飞虱	半翅目: 飞虱科	69612	32355
<i>Sogatella furcifera</i>	白背飞虱	半翅目: 飞虱科	92724	41287
<i>Nephotettix cineticeps</i>	黑尾叶蝉	半翅目: 叶蝉科	81225	9655

a) RPTDB 中该物种 Unigene 总数; b) RPTDB 中该物种蛋白总数

A~Z 的分类中, 未归类的列入 other, 而每个分类有单独的界面, 用户单击该界面中的 Unigene ID 即可浏览该序列的细节(图 2).

2.3 数据检索

RPTDB 提供 2 种检索方式: 快速检索和高级检索. 快速检索位于首页的右上方, 输入 Unigene 的 ID 即可获得对应的序列细节; 而高级搜索则为用户提供了多项参数. 关键词或其他数据的 ID 都可以用于提取数据, 同时还可以限制检索的物种. 另外, 也提供对数据库中所有序列的 BLAST 功能.

3 讨论: 物种分化时间

采用了果蝇的中性替换率用于估算物种分化时间. 事实上, 目前还没有更合适的昆虫的中性替换率以提高本研究中分化时间估算的准确性, 本研究中的 8 种昆虫由于拥有相对更长的生命周期, 其实际中性替换率可能会略小于果蝇, 因此实际的物种分化时间可能略久于展示的结果. 另外, 分别在半翅目和鳞翅目的建树过程中, 由于黑尾叶蝉和大螟与同组的其他昆虫的直系同源序列非常有限, 因此未将这两个物种用于建树.

参考文献

- Li L Y. Integrated rice pest control in the Guangdong province of China. *BioControl*, 1982, 27: 81–88
- Sheng C F, Wang H T, Sheng S Y, et al. Pest status and loss assessment of crop damage caused by the rice borers, *Chilo suppressalis* and *Tryporyza incertulas* in China. *Entomol Knowl*, 2003, 40: 289–294
- Sheng C F, Wang H T, Gao L D, et al. The occurrence status, damage cost estimate and control strategies of stem borers in China. *Plant Protect*, 2003, 29: 37–39
- Chen M, Shelton A M, Ye G Y. Insect-resistant genetically modified rice in China: from research to commercialization. *Annu Rev Entomol*, 2011, 56: 81–101
- Mao Y B, Cai W J, Wang J W, et al. Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. *Nat Biotechnol*, 2007, 25: 1307–1313
- Baum J A, Bogaert T, Clinton W, et al. Control of coleopteran insect pests through RNA interference. *Nat Biotechnol*, 2007, 25: 1322–1326
- Huvenne H, Smagghe G. Mechanisms of dsRNA uptake in insects and potential of RNAi for pest control: a review. *J Insect Physiol*, 2010, 56: 227–235
- Finn R D, Mistry J, Schuster-Böckler B, et al. Pfam: clans, web tools and services. *Nucleic Acids Res*, 2006, 34: D247–D251
- Finn R D, Mistry J, Tate J, et al. The Pfam protein families database. *Nucleic Acids Res*, 2010, 38: D211–D222
- Moreno-Hagelsieb G, Latimer K. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics*, 2008, 24: 319–324
- Rasko D A, Myers G S, Ravel J. Visualization of comparative genomic analyses by BLAST score ratio. *BMC Bioinformatics*, 2005, 6: 2
- Adachi J, Hasegawa M. MOLPHY Version 2.3: programs for molecular phylogenetics based on maximum likelihood. *Computer Sci Monogr*, 1996, 28: 1–150
- Adachi J, Hasegawa M. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J Mol Evol*, 1996, 42: 459–468
- Zhang X, Feng Y, Cheng H, et al. Relative evolutionary rates of NBS-encoding genes revealed by soybean segmental duplication. *Mol Genet Genomics*, 2011, 285: 79–90

Rice Pest Transcriptome Database

WU Hao¹, LI Fei², HUA HongXia³, WANG Rui¹, GAO Liang¹, Hou JiXiang¹, CHEN Hao¹,
ZHOU Fei¹ & LIN YongJun¹

1 National Key Laboratory of Crop Genetic Improvement and National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, China;

2 Department of Entomology, College of Plant Protection, Nanjing Agricultural University, Nanjing 210095, China;

3 College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

Rice is the most important crop in some Asian countries, but its yield is threatened by hundreds of kinds of pests. Although various strategies are used to control pests, the losses are still tremendous. RNAi (RNA interference) is believed to have a great potential in pest control, but the lack of genetic information vastly limits the application of this technology. The Rice Pest Transcriptome DataBase (RPTDB, <http://rptdb.hzau.edu.cn/>) is a comprehensive transcriptome database with data derived from next generation sequencing for eight important rice pests. RPTDB aims at providing well-organized information for researchers who require genetic materials for functional research on insect, pest management or others, and thus meets the needs in these research fields. Furthermore, two phylogenetic trees are constructed to reveal the evolutionary relationships of two pest orders. All of the information is freely available at <http://rptdb.hzau.edu.cn>.

transcriptome database, rice pests, next generation sequencing, phylogenetic tree

doi: 10.1360/052014-109