基于 Solexa 高通量测序的黄曲条跳甲转录组学研究

贺华良#, 宾淑英#, 吴仲真, 林进添*
(仲恺农业工程学院外来有害生物预警与控制研究所, 广州 510225)

摘要: 黄曲条跳甲 *Phyllotreta striolata* (Fabricius) 是十字花科蔬菜的重要害虫。为深入了解其遗传信息, 本研究应用新一代高通量测序技术 Illumina's Solexa 平台对黄曲条跳甲成虫的转录组进行测序, 并结合 SOAPdenovo 拼接聚类等分析软件, 获取大量的 EST 和挖掘功能基因。本文最终获得了 4 924 条序列重叠群(contig), 其中包含 2 209 种与黑腹果蝇 *Drosophila melanogaster* 蛋白基因具直系同源的独立基因(unigene) 和 610 种黄曲条跳甲物种特有的 unigene。结合 Gene Ontology (GO) 数据库进行分析, 发现大部分的 unigene 具结合能力(binding capability) 和催化活性(catalytic activity); 上百种 unigene 可聚类于生物学过程分类中的配子发生、生殖腺发育和交配行为等重要功能。另外, 结合 KEGG Pathway 数据库分析发现, 共有 363 种 unigene 参与或涉及了 40 种代谢路径, 其中生物钟调控路径和植物次生代谢物路径等相关基因的发现, 有助于深入研究黄曲条跳甲行为发生的内在机理。Solexa 高通量测序技术作为昆虫功能基因组研究的重要手段, 为发掘黄曲条跳甲功能基因发挥了重要作用, 也为在分子水平上研发黄曲条跳甲的防治新策略提供了更翔实的基因信息。

关键词: 黄曲条跳甲; Solexa 测序; 序列重叠群; 独立基因; 转录组
中图分类号: Q966 **文献标识码:** A **文章编号:** 0454-6296(2012)01-0001-11

Transcriptome characteristics of *Phyllotreta striolata* (Fabricius) (Coleoptera: Chrysomelidae) analyzed by using Illumina's Solexa sequencing technology

HE Hua-Liang#, BIN Shu-Ying#, WU Zhong-Zhen, LIN Jin-Tian* (Institute for Management of Invasive Alien Species, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China)

Abstract: The striped flea beetle, *Phyllotreta striolata* (Fabricius), is an important pest damaging cruciferous vegetables. In order to investigate the profile of gene expression and elucidate the functional genes, we sequenced the transcriptome of the adult of *P. striolata* by Illumina's Solexa sequencing technology, and analyzed the data of expressed sequence tags (ESTs) by using SOAPdenovo system. A total of 4 924 contigs were obtained including 2 209 unigenes of orthologous genes relating to *Drosophila melanogaster* and 610 species-specific unigenes of *P. striolata* based on Gene Ontology and KEGG databases. We found that most of unigenes contain function domains with binding capacity and catalytic activity. More than 100 unigenes are involved in gamete generation, ovarian follicle cell development and mating behavior. Three hundred sixty-three unigenes may be involved in 40 different metabolic pathways based on KEGG database. The finding that 363 unigenes are involved in regulation pathway of biological rhythm and plant secondary metabolites will be useful to clarify the mechanism of behaviors of this insect such as oviposition rhythm, *etc*. Moreover, the sequence resources presented in this study provide useful information to develop new strategies to manage this pest.

Key words: *Phyllotreta striolata*; Illumina's Solexa sequencing technology; contig; unigene; transcriptome

黄曲条跳甲 *Phyllotreta striolata* (Fabricius) 俗称狗虱虫、菜蚤子、土跳蚤、黄跳蚤等, 隶属鞘翅目叶甲科昆虫。黄曲条跳甲是十字花科蔬菜的世界性害虫(Tahvaanainen, 1983), 广泛分布于我国南北菜区, 主要为害芥菜、菜心、萝卜、白菜、芥蓝、油菜等(高泽正等, 2000)。国内外对黄曲条跳甲在

基金项目: 国家自然科学基金青年科学基金项目(31101500)
作者简介: 贺华良, 男, 1977 年生, 湖南攸县人, 博士, 讲师, 从事农业害虫综合防治及分子生物学研究, E-mail: hhl_1234@126.com; 宾淑英, 女, 1964 年生, 广东封开人, 副教授, 从事农业害虫综合防治及推广研究, E-mail: binsuying@163.com
# 共同第一作者 Authors with equal contribution
* 通讯作者 Corresponding author, E-mail: linjtian@163.com
收稿日期 Received: 2011-07-26; 接受日期 Accepted: 2011-10-10

形态学、生态学及抗药性监测方面已有了广泛的研究(张茂新和梁广文,2000; Feng *et al.*,2000; 侯有明等,2003; 周先治和吴刚,2004; 傅建炜等,2006)。但是长期以来倚重化学防治的策略使得黄曲条跳甲对很多化学药剂都出现了不同程度的抗药性,而抗药性的动态发展又使得对其危害的控制越来越困难。目前,广东省部分地区黄曲条跳甲的危害甚至已超过小菜蛾的危害。黄曲条跳甲产生猖獗危害的另一个重要原因是因其幼虫是在土缝中孵化并危害寄主植物的根系(聂河兴,2007; 王玲等,2009),而常规化学防治法对幼虫却鞭长莫及。因此,目前需要积极探讨控制黄曲条跳甲的防治新策略。Zhao 等(2011)首次成功鉴定了黄曲条跳甲精氨酸激酶 PsAK 和特异性气味受体 PsOr1 基因的 cDNA 序列,并通过 RNA 干涉技术抑制了靶标基因的表达,结果表明上述两个基因功能的受损可导致黄曲条跳甲死亡、产卵选择性和取食选择性发生改变等现象,为在分子水平探讨基于黄曲条跳甲行为调控的防治新策略提供了新的参考。

截至目前,在 GenBank 上注册的黄曲条跳甲的 cDNA 或 EST 序列仅仅只有 5 种,分别是上文提到的 PsAK 和 PsOr1,以及抗药性发生相关的 2 个非专一性酯酶基因和 1 个乙酰胆碱酯酶基因的 EST 序列。黄曲条跳甲已发展成为当前蔬菜产业的重要害虫,而其基因组及转录组研究的滞后使得对黄曲条跳甲的深入研究具有一定的困难。基因表达序列标签(expressed sequence tags, EST)技术被认为是一种研究转录组的有效方法,广泛应用于新基因发现、基因表达分析和蛋白质组学(Ewing *et al.*,1999)。新一代高通量测序技术 Illumina's Solexa 是对传统测序方法的一次革命性变革(Nagalakshmi *et al.*,2008; Rosenkranz *et al.*,2008)。Solexa 测序性价比最高,运行成本较低,高通量,高精确性,可以同时检测上亿个核苷酸片断。虽然该技术的序列读取长度较短,但其序列的拼接过程最终能达到高精准度。模式昆虫埃及伊蚊 *Aedes aegypti* 和冈比亚按蚊 *Anopheles gambiae* 的 EST 大规模测序中,采用的就是 Solexa's Illumina 技术(Gibbons *et al.*,2009)。最近,一种非模式生物蜗牛 *Radix balthica* 也采用该技术完成了转录组的测序任务(Feldmeyer *et al.*,2011)。高通量转录组测序可在短时间内获得的大量的 unigene 信息。目前,国际上已构建了 unigene 的数据库,即 UniGene。UniGene 是从属于 GenBank 的一部分,专门收集非冗余性的基因来源的 clusters 数据。每一个 UniGene cluster 包含代表单一基因的序列和相关的信息,可为科学研究快速提供有用信息,例如基因表达的组织类型和图谱定位信息(Schuler,1997; Pontius *et al.*,2003)。因此,应用 Solexa 高通量测序技术对农业害虫进行转录组研究,可大大降低测序所需时间和成本,使我们能够对部分重要的非模式生物或农业害虫启动高通量水平的基础研究及后续的应用研究。

本实验将 Solexa 高通量测序技术应用到黄曲条跳甲的转录组学研究中,并应用生物信息学方法对所得序列与模式昆虫黑腹果蝇 *Drosophila melanogaster* 和赤拟谷盗 *Tribolium castaneum* 等的基因组及转录组序列进行比对分析,从功能基因组水平上鉴定一批黄曲条跳甲的重要基因。本研究主要检测与黄曲条跳甲生殖发育和生殖行为等相关的关键基因;参与寄主蔬菜次生代谢或挥发性化合物分子代谢的基因;以及生物钟代谢路径相关的核心生物钟基因、调控因子等。上述关键基因的鉴定将为在分子水平开展黄曲条跳甲行为调控及防治新策略的研究奠定前期数据基础。

1 材料与方法

1.1 供试昆虫

黄曲条跳甲 *P. striolata* (Fabricius) 成虫来源于广州市市郊黄埔古港蔬菜基地,采集位置为芥菜 *Brassica juncea* 的心叶部位,采集时间为 2010 年 6 月 11 日的 11:00 – 15:00。虫源采集点即时温度为 27℃,相对湿度为 75%,光照度为 4 290 lx,收集成虫后立即用液氮冷冻,存于 – 80℃备用。

1.2 RNA 提取和 cDNA 文库构建

采用总 RNA 提取试剂盒(Qiagen)提取黄曲条跳甲成虫总 RNA。以 2 μg 总 RNA 为模板,cDNA PCR Library Kit (TaKaRa) 反转录合成双链 cDNA,并 PCR 扩增,扩增条件为:94℃ 1 min; 94℃ 30 s, 60℃ 30 s, 72℃ 3 min,进行 10 个循环。采用 PureLinkTM PCR Purification Kit (Invitrogen) 去除体系中小于 300 bp 的片段。通过多次的 PCR 扩增、纯化、浓缩,最终共收集到双链 cDNA 10 μg,浓度超过 1 μg/μL,送往华大基因公司 Solexa 高通量测序平台。

1.3 Solexa 文库构建和测序

应用新一代高通量测序平台 Illumina's Solexa Genome Analyzer Ⅱ对 cDNA 样品测序。5 μg 双链

cDNA 打断为 150 bp 左右的片段后，两端添加特异性衔接子 A 和 B，变性为单链连接到磁珠上，经 emPCR 富集后，置于 PicoTiterPlate 板上，上机测序。两端测序，每一个序列读取片段(read)的读长约 90 bp。

1.4　序列拼接、功能注释及分类

采用 GS-FLX Software 去除衔接子区域和低质量序列，屏蔽 cDNA 文库 PCR 引物，采用 SOAPdenovo 软件对每一个序列读取片段聚类进行拼接(Li *et al*.，2010)，形成序列不间断的 contig/unigene。后续采用序列比对的方法对所得序列注释，使用 Blastn 与 NCBI 的非冗余核酸序列数据库(non-redundant nucleotide database，nt)进行比较(E 值为 1e-10)，进一步使用 blastx 与 NCBI 的非冗余蛋白序列数据库(non-redundant protein database，nr)和黑腹果蝇的蛋白组数据库进行比较(E 值为 1e-5)，所有分析使用默认参数。unigene 的功能域注释及分类分析主要结合 Gene Ontology (GO) 数据库(http://amigo.geneontology.org/cgi-bin/amigo/blast.cgi?session_id=9036amigo1316253192)、SMART 数据库(http://smart.embl-heidelberg.de/)和 KEGG 数据库(http://www.genome.jp/kegg/pathway.html)。Gene Ontology (GO) 分类分析又可进一步按生物学过程(biological process)、分子功能(molecular function)和细胞组分(cellular component)三大亚类进行分类。对所有注释信息整理，重点搜索与黄曲条跳甲的生殖发育及生殖行为相关的关键基因及可能参与的调控因子等。

1.5　系统发育分析

从 NCBI 上下载昆虫纲中部分代表性昆虫 innexin2 的全长 cDNA 序列，使用 ClustalX 1.8 软件对这些序列进行比对，输出后缀为 * .phy 格式的文件。采用分子进化遗传分析软件 PHYLIP3.68 进行遗传距离分析，具体步骤如下: 把 * .phy 文件拷贝到 PHYLIP 目录下，更名为 infile; 用 Seqboot 分析，复制数为 1 000，运行后生成 1 000 套比对序列的文件，将此文件更名为 infile; 运行最大简约法程序 DNAPARS，生成两个文件 outfile 和 treefile; 利用多重树构建一致树，即打开 CONSENSE 软件，将刚才生成的 treefile 文件更名后输入，生成两个文件 outfile 和 treefile，完成进化树的生成。其中 treefile 用 TREEVIEW 打开，即可浏览一致树。遗传距离分析过程中没有设置外类群。

2　结果

2.1　Solexa 测序和序列拼接

采用 Solexa 高通量测序技术对处于跳跃活动盛期的黄曲条跳甲成虫的转录组进行测序，测序共获得 13 176 562 个序列读取片段，每一个序列读取片段的长度为 90 bp，即该次测序总的 cDNA 碱基读取量约为 1 185 Mb。采用 SOAPdenovo 软件聚类拼接，设置参数 kmer = 24，最终得有效的 contig 共 1 702 083 条，序列分析过程相关的统计数据见表 1。对于拼接序列的长度分布特征，鉴于只有 24 bp 的序列片段太多，本文中只统计 ≥50 bp 的 contig 的长度分布特征，具体结果见表 2。

表 1　Solexa 高通量测序的序列拼接分析

Table 1　Sequence assembly after Illumina's Solexa sequencing

读取序列的拼接分析 Reads assembled	数量 Counting
读取片段总数 Total number of reads	13 176 562
读取片段的碱基数量总和(nt) Total nucleotide length of reads	1 185 890 580
重叠群总数 Total number of contigs	1 702 083
重叠群平均长度(nt) Mean contig length	43
重叠群长度总和(nt) Total contig length	72 508 369
序列骨架数量总和 Total number of scaffolds	50 332
序列骨架平均长度(nt) Mean scaffold length	302
序列骨架长度总和(nt) Total scaffold length	15 185 104

2.2　序列比对分析、注释及 unigene 的特征分析

对于长度 ≥500 bp 的 4 924 条 contig，结合现国际上已公布全基因组序列的 5 种模式昆虫的基因组数据库，进行 blastx 比对分析，见表 3。

首先对比对分析中 E value ≤1e-100 的保守基因数量进行了分析。表 3 显示，黄曲条跳甲与同为鞘翅目的赤拟谷盗之间的保守基因数量最多，达 788 种，占总数的 16.0%。另外，与表 3 中 5 种昆虫之间都保守的基因数量也有 120 种，占总数的 2.4%。本研究以高保守(E value = 0)的间隙连接蛋

表 2 contig 的长度分布特征分析

Table 2 Length distribution of contigs

	Contig≥50 bp	Contig≥100 bp	Contig≥200 bp	Contig≥500 bp
重叠群总数(个) Total number of contigs	245 572	60 239	25 480	4 924
重叠群平均长度(nt) Mean contig length	121	243	310	730
重叠群长度总和(nt) Total contig length	29 823 382	14 629 041	9 735 642	3 595 484

表 3 4 924 条 contig 与 5 种模式昆虫基因组数据的比对分析

Table 3 Alignment analysis of 4 924 contigs with sequences in genome database of 5 model insects

blastx 比对分析 Sequence alignment using blastx	同源基因数量 Total number of homologous genes (E value≤1e-100)	同源基因数量 Total number of homologous genes (E value≤1e-5)	非同源基因数量 Total number of non-homologous genes (E value>1e-5)			
与家蚕基因组信息比对分析 Hit to *Bombyx mori*	164	3 408	1 516			
与黑腹果蝇基因组信息比对分析 Hit to *Drosophila melanogaster*	345	3 699 (2 209**)	1 225	906 (631+275)		
与冈比亚按蚊基因组信息比对分析 Hit to *Anopheles gambiae*	375	3 709	1 215		631 (610+21)	
与西方蜜蜂基因组信息比对分析 Hit to *Apis mellifera*	426	3 827	1 097			610
与赤拟谷盗基因组信息比对分析 Hit to *Tribolium castaneum*	788	3 858	1 066	1 066 (631+435)		
与非冗余蛋白数据库比对分析 Hit to non-redundant protein database	120 (120*)	4 314	610	610	610	

* 与上述 5 种昆虫比对分析的 E value 都小于或等于 1e-100 的 unigene 总数 Total number of contig hits to above five insects with E value≤1e-100.

** 与黑腹果蝇基因直系同源的 unigene 的数量 Total number of contig orthlogs to genes of *D. melanogaster*.

白(innexin2)基因为代表，对 6 种昆虫的 innexin2 直系同源基因进行了系统进化分析(图 1)。图 1 显示，innexin2 的系统发生树与上述各种昆虫的进化地位基本相符。图中黄曲条跳甲的 innexin2 与赤拟谷盗的 innexin2 聚为一支；但鳞翅目昆虫的 innexin2 与双翅目蚊科昆虫 innexin2 的进化距离比双翅目昆虫蚊科与果蝇科 innexin2 之间的进化距离更小，这可能与 innexin2 在不同昆虫种类中的进化速率不同有关系。

再对比对分析中 E value≤1e-5 的同源基因数量进行了分析。表 3 显示，黄曲条跳甲与鳞翅目昆虫蜜蜂的同源基因数量最少，而与赤拟谷盗的同源基因的数量最多，基本反映了黄曲条跳甲与 5 种昆虫亲缘关系的远近。其中，在上述 4 924 条 contig 中，有 906 条 contig 未发现与家蚕、黑腹果蝇、冈比亚按蚊及蜜蜂的基因同源，但其中有 275 条可与赤拟谷盗的基因同源，推测该 275 条 contig 可能是鞘翅目特有的基因。同样，结合与 nr 数据库的比对分析结果，推测有 610 条 contig 可能是黄曲条跳甲特有的基因。由于缺乏功能注释参考，该 610 种黄曲条跳甲种特异性基因的功能还有待进一步解析。

进一步有针对性地结合黑腹果蝇的序列数据库和 contig 序列进行相互性的 blastx 比对分析，筛选得分最高的匹配对，进而获取与黑腹果蝇蛋白基因直系同源的 unigene。最终获得了 2 209 条 contig 或 2 209 种 unigene(表 3)，促进了该批次序列的后续注释及功能分析的高可信度。

2.3 与黑腹果蝇具直系同源的 2 209 种 unigene 的特征分析

统计结果表明，该批次的 2 209 种 unigene 涉及的转录组序列读取总长度为 1.78 Mbp，平均序列长度为 804 bp，其中≤1 000 bp 的 unigene 占主要部

分，达79.7%，1 001 ~ 1 500 bp 大小区间的 unigene 占 18.2%，而 > 1 500 bp 的 unigene 只占 2.1%。随机测序所获得的同一基因的 EST 的数目可以在一定程度上代表该基因在该组织中的表达丰度。统计分析上述每一个 unigene 包含的序列读取片段的数量总和分布趋势发现，所含序列读取片段数量分布在 11 ~ 50 次区间的 unigene 最多，占总数的 68.5%（图 2：A）。含序列读取片段的数量超过 150 次的 unigene 只有 5 个（0.2%），分别是与黑腹果蝇相应基因同源的含 WD 功能域基因家族成员（1 280 bp）、胱硫醚 β 合成酶基因（567 bp）、lodestar 基因（711 bp）、尿（核）苷磷酸化酶（uridine phosphorylase）基因（1 035 bp）、类异戊二烯生物合成酶（isoprenoid biosynthesis enzyme）基因（1 350 bp）。但是，由于每一个 unigene 拼接后的长度不同，因此，并不能完全由其所包含序列读取片段的数量精确推断 unigene 的表达趋势。本文采用了两种方法进行进一步的估计。

第一种方法是通过均一化计算模型分析 unigene 的相对表达丰度。计算公式为：相对表达丰度 = 拼接涉及的所有序列读取片段的数量总和 × Solexa 测序仪读取片段的长度参数即 90 bp/拼接后 unigene 的长度。对 2 209 种 unigene 的整体分析结果显示，表达丰度在 2.0 ~ 4.9 之间的 unigene 占大多数，接近 60%（图 2：B）。表达丰度超过 20 的 unigene 只有 6 个（0.3%），其序列长度范围为 524 ~ 600 bp，分别与黑腹果蝇的功能未知蛋白基因（hypothetical protein）、adiponectin receptor、dribble、线粒体核糖体大亚基蛋白（39S ribosomal protein L3，mitochondrial）、肌浆/内质网型钙离子 ATP 酶通道蛋白（calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type）、胱硫醚 β 合酶基因具有同源性。上述统计数据说明高丰度表达基因比较少，大多数呈中低丰度表达。

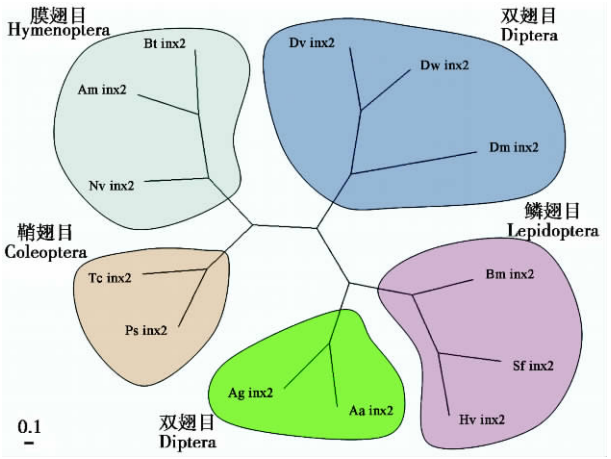

图 1　多种昆虫中 innexin2（inx2）基因的系统发生树分析

Fig. 1　Phylogenetic relationship of innexin2 (inx2) in different insects

使用 PHYLIP 软件中的最大简约性法（DNAPARS）构建系统发生树。图中的比例尺（0.1）表示 10% 的差异。Phylogenetic tree generated by method of Maximum Parsimony (DNAPARS) in PHILIP. The scale bar (0.1) indicates a 10% difference. 所分析种类的 inx2 相关序列的来源及 GenBank 登录号 GenBank accession numbers related to the species analyzed: Ps_inx2: 黄曲条跳甲 *Phyllotrata striolata*; Tc: 赤拟谷盗 *Tribolium castaneum* (XP968805); Ag: 冈比亚按蚊 *Anopheles gambiae* (XM321635); Aa: 埃及伊蚊 *Aedes aegypti* (XM001649705); Dm: 黑腹果蝇 *Drosophila melanogaster* (NM132147); Dw: 果蝇近缘种 *D. willistoni* (XM002071107); Dv: 果蝇近缘种 *D. virilis* (XM002057877); Am: 西方蜜蜂 *Aphis mellifera* (XM003251623); Bt: 熊蜂 *Bombus terrestris* (XM003397631); Nv: 丽蝇蛹集金小蜂 *Nasonia vitripennis* (XM001603984); Bm: 家蚕 *Bombyx mori* (NM001043738); Sf: 草地贪夜蛾 *Spodoptera frugiperda* (AY196138); Hv: 烟芽夜蛾 *Heliothis virescens* (AY633755).

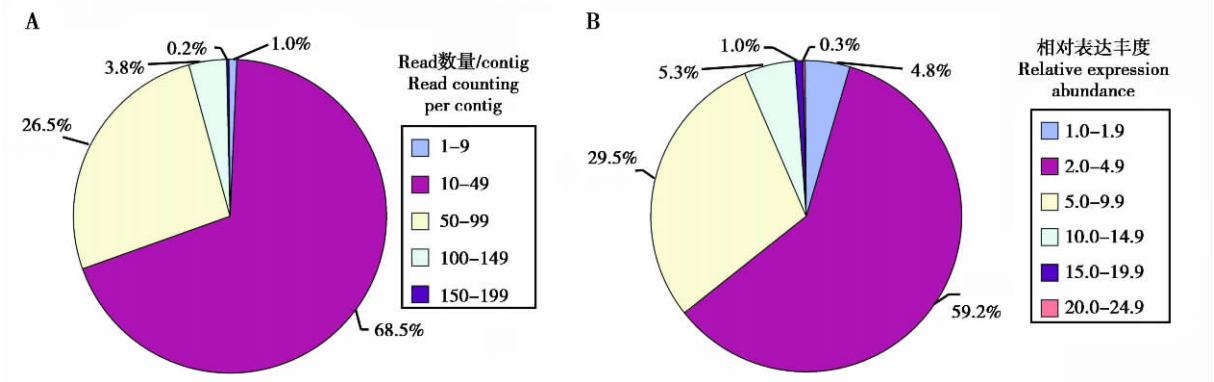

图 2　Unigene 包含的序列读取片段数量分析及其表达丰度分析

Fig. 2　Read counting and expression level analysis of unigenes

A: unigene 包含序列读取片段数量的分布 Distribution of reads counting of unigenes; B: unigene 相对表达丰度的分布 Distribution of relative expresssion abundance of unigenes.

第二种方法是分析 unigene 的拼接过程中重叠区的序列读取片段的数量(图 3)。通过分析不同重叠区域内序列读取片段的数量总和，选择最高值。如图 3(A)中重叠区涉及序列读取片段总和最高数量为 5。因为本文中对 cDNA 片段的测序是两端测序法，因此 unigene 的相对表达丰度约为拼接序列读取片段数量总和的 1/2，即图 3(A)中 unigene 基因相对表达丰度可初步计数为 2.5。同理，如图 3(B，C，D)中该 unigene 基因相对表达丰度可初步计数分别为 7，7.5 和 10.5。与利用第一种计算模式得出 4 个 contig 计算出来的相对表达丰度分别为 1.8，6.5，7.0 和 8.8 的数值相比，总体趋势相对偏大，但其可信度更高。

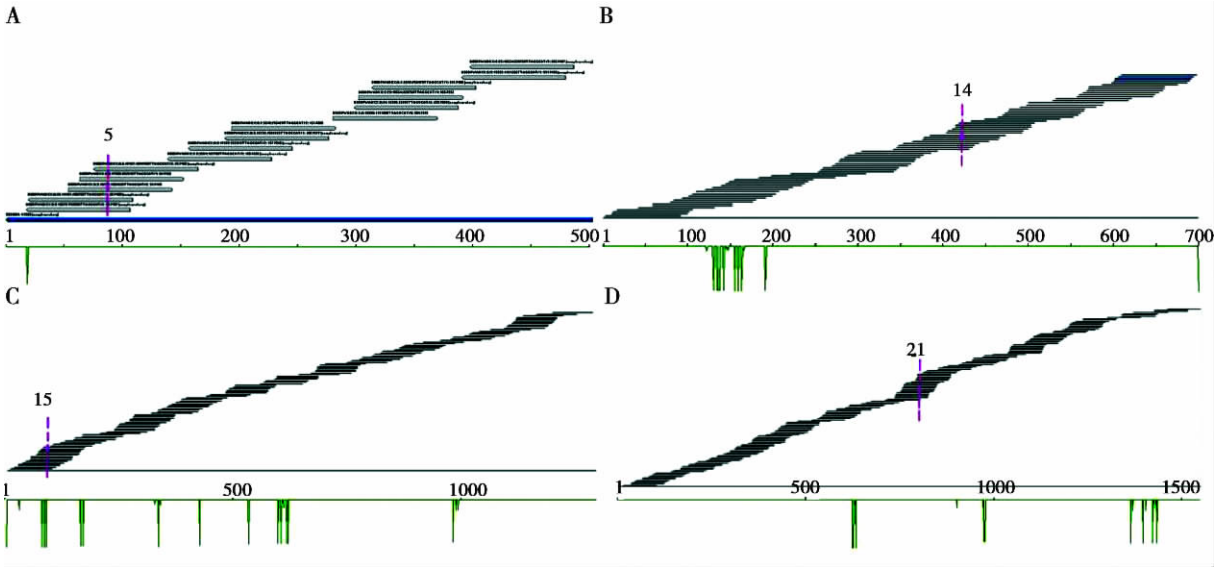

图 3 contig 中序列读取片段的拼接位置及表达量分析

Fig. 3 Mapping and abundance calculation of reads in assembled contig

结合 Vector NTI Contig Express 软件分析 contig 中序列读取片段的拼接位置及表达量 Mapping of reads in the corresponding assembled contig using software of Vector NTI Contig Express and calculating the relative abundance of unigene. A，B，C 和 D 分别表示含有 10，50，100 和 150 对(SOAPdenovo 算法)序列读取片段的 contig 代表序列的拼接图；4 条 contig 的长度分别为 502，691，1 280 和 1 539 bp。A，B，C and D demonstrate the assembled contig candidates with different length (502，691，1 280 and 1 539 bp，respectively) covering different number of reads (10，50，100 and 150，respectively). 图中红色虚竖线代表该重叠区域涉及的序列读取片段数量最多，可作为该 unigene 被测到转录本的数量；图中绿色实竖线表示为多个序列读取片段在同一重叠位点显示为不同的碱基，竖线的长度代表该位置碱基相异的程度，越长代表碱基相异的频次越高。图中 contig 所含序列读取片段的数量与前期 SOAPdenovo 算法提供的数量不完全相同，如图 1(A)中只定位到 7.5 对序列读取片段，这是因为此处序列读取片段位置分析所用的 Vector NTI Contig Express 软件与前期华大基因公司测序拼接所用的 SOAPdenovo 的算法不完全相同，造成小部分序列读取片段在 Vector NTI Contig Express 分析中未能成功定位。The number on the top of red dashed line demonstrates the relative abundance of unigene in this sequencing research; green lines demonstrate different nucleotide in the same assembly site from different reads. Numbers of reads mapped using software of Vector NTI Contig Express in these four figures are different from the statistics numbers by method of SOAPdenovo，due to a different algorithm design between these two models.

2.4 与黑腹果蝇具直系同源的 2 209 种 unigene 的功能注释分析

结合 GO 数据库对黄曲条跳甲的 2 209 种 unigene 进行功能注释分类分析，共有 2 099 种 unigene 获得了功能分类。GO 数据库又可分为 3 个亚类，即基因所涉及的分子功能(molecular function)、生物学过程(biological process)和细胞组分(cellular component)。本文首先按 unigene 可能参与的分子功能进行分类分析，发现在 2 209 种 unigene 中，被赋予功能的基因累计达到 1 516 条，其分子功能类别共涉及 128 种。以分子功能分类中的一级子目录所涉及的 unigene 数量进行统计分析发现，该批次 unigene 以具结合能力(binding capability，GO: 0005488)和具有酶活性(catalytic activity，GO: 0003824)为主。由于同一种基因可能具有多个功能域或多种活性(一因多效)，因此本研究进一步以分子功能分类中的二级子目录进行统计分析(图 4)。结果显示，具有核苷酸结合能力

(GO: 0000166) 的 unigene 的累计数量最多，具有离子结合能力(GO: 0043167) 和具有水解酶活性(GO: 0016787) 的 unigene 的累计数量也比较多，分别居第 2 位和第 3 位。

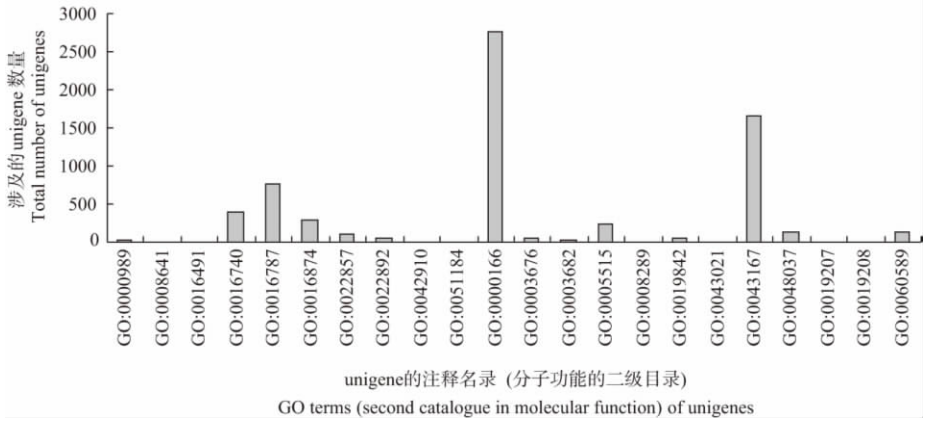

图 4　unigene 可能参与分子功能的聚类统计分析

Fig. 4　Statistical analysis of molecular function of unigenes

GO: 0000989 转录因子结合转录因子活性 Transcription factor binding transcription factor activity; GO: 0008641 小分子蛋白激活酶活性 Small protein activating enzyme activity; GO: 0016491 氧化还原酶活性 Oxidoreductase activity; GO: 0016740 转移酶活性 Transferase activity; GO: 0016787 水解酶活性 Lydrolase activity; GO: 0016874 连接酶活性 Ligase activity; GO: 0022857 跨膜转运活性 Transmembrane transporter activity; GO: 0022892 底物特特异性转运活性 Substrate-specific transporter activity; GO: 0042910 异源物转运活性 Xenobiotic transporter activity; GO: 0051184 辅助因子转运活性 Cofactor transporter activity; GO: 0000166 核苷酸结合能力 Nucleotide binding capability; GO: 0003676 核酸结合能力 Nucleic acid binding capability; GO: 0003682 染色质结合能力 Chromatin binding capability; GO: 0005515 蛋白结合能力 Protein binding capability; GO: 0008289 脂类分子结合能力 Lipid binding capability; GO: 0019842 维生素结合能力 Vitamin binding capability; GO: 0043021 核蛋白结合能力 Ribonucleoprotein binding capability; GO: 0043167 离子结合能力 Ion binding capability; GO: 0048037 辅助因子结合能力 Cofactor binding capability; GO: 0019207 激酶调控活性 Kinase regulator activity; GO: 0019208 磷酸酶调控活性 Phosphatase regulator activity; GO: 0060589 核苷三磷酸酯酶调控活性 Nucleoside-triphosphatase regulator activity.

本研究还对 2 209 种 unigene 所涉及的生物学过程(biological process) 和细胞组分(cellular component) 进行了统计分析。结果表明，有 1 393 种 unigene 涉及 471 种注释名录。表 4 显示了所涉及 unigene 在数量上排前 10 位的注释名录，主要有基因转录调控、代谢、分子转运和雌配子发生等生物学过程。另外，有 944 种 unigene 涉及了 121 种细胞组分的合成和构建。上述 3 种功能分类结果显示了黄曲条跳甲成虫跳跃盛期的基因表达谱的总体情况。

2.5　生殖发育的相关基因分析

表 4 中的信息提示与有性生殖和胚子发生生物学过程相关的 unigene 数量较多。为了深入研究，本文进一步详细整理了上述 471 种注释名录中与黄曲条跳甲生殖行为和生殖发育相关的 unigene。表 5 中显示了部分 unigene 可能参与卵泡细胞迁移、性别分化、生殖腺发育和交配行为等。通过对与表 5 中的“交配行为”相关的 unigene 的序列逐一分析，发现这 9 个基因分别是与黑腹果蝇或赤拟谷盗的快速交配基因(quick-to-court)、发动蛋白(dynamin)、多巴脱羧酶(dopa decarboxylase)、肾上腺皮质铁氧还蛋白(NADPH: adrenodoxin oxidoreductase)、核糖体蛋白 S7(ribosomal protein S7)、学习缺陷突变体(dunce)、睾丸特异性的富含亮氨酸的蛋白(testis specific leucine rich repeat protein)、甘露糖转移酶(beta-1,4-mannosyltransferase egh) 和葡萄糖脱氢酶(glucose dehydrogenase) 等基因具有同源性。对于 quick-to-court 基因的功能，国外研究已表明，果蝇 quick-to-court 表达量的提高会导致雄性对雄性(male-male) 求偶和一个迅速发生的雄性对雌性(male-female) 求爱的表示(Gaines *et al.*, 2000)。因此，上述功能分类信息的分析及整理为今后黄曲条跳甲的生殖发育及生殖行为调控的研究提供了宝贵的序列信息。

2.6　代谢路径分析

结合黑腹果蝇的 KEGG Pathway 数据库，对上述 2 209 种 unigene 可能参与或涉及的代谢路径进行分析。本研究发现共 363 种 unigene，涉及到 40

表 4 biological process 分类中涉及 unigene 最多的前 10 种注释名录

Table 4 Terms in biological process with Top 10 highest number of related unigenes

排序 Rating	GO 注释名录(生物学过程) Terms of biological process	GO 分类号 Accession no. of GO term	unigene 数量 unigene counting
1	转录调控 Regulation of transcription	GO: 0045449	167
2	RNA 分子代谢调控 Regulation of RNA metabolic process	GO: 0051252	141
3	多细胞生物的生殖 Multicellular organism reproduction	GO: 0032504	141
4	有性生殖 Sexual reproduction	GO: 0019953	140
5	配子发生 Gamete generation	GO: 0007276	136
6	囊泡介导转运 Vesicle-mediated transport	GO: 0016192	131
7	蛋白定位 Protein localization	GO: 0008104	125
8	细胞骨架构成 Cytoskeleton organization	GO: 0007010	122
9	转录 Transcription	GO: 0006350	119
10	蛋白水解 Proteolysis	GO: 0006508	118

表 5 biological process 分类中与生殖发育相关独立基因的注释名录

Table 5 Unigenes categorized into GO subcategories of biological process involved in reproductive biology

排序 Rating	GO 注释名录(生物学过程) Terms of biological process	GO 分类号 Accession no. of GO term	unigene 数量 unigene counting
1	卵泡细胞发育 Ovarian follicle cell development	GO: 0030707	49
2	卵泡细胞迁移 Ovarian follicle cell migration	GO: 0007297	22
3	性别分化 Sex differentiation	GO: 0007548	12
4	生殖腺发育 Gonad development	GO: 0008406	10
5	初级性征发育 Development of primary sexual characteristics	GO: 0045137	10
6	交配行为 Mating behavior	GO: 0007617	9

种代谢路径。含相关 unigene 数量较多的代谢路径主要有: 泛素代谢路径(ubiquitin mediated proteolysis)、氨酰 tRNA 合成路径(aminoacyl-tRNA biosynthesis)、柠檬烯和蒎烯代谢路径(limonene and pinene degradation)等。其中柠檬烯和蒎烯代谢路径中 unigene 主要有: 转谷氨酰胺酶(transglutaminase)、溶血磷脂酰基转移酶(lysophospholipid acyltransferases)、反式异戊烯转移酶(*trans*-prenyltransferase)、γ-谷氨酰转移酶(gamma glutamyl transpeptidase)、短链脱氢酶

(short-chain dehydrogenase) 和触角特异性细胞色素 P450(antennae-rich cytochrome P450) 等，上述 6 种 unigene 都可在 nr 数据库中找到已注释的同源基因。昆虫的柠檬烯和蒎烯代谢路径涉及对植物次生代谢物的趋性或防御体系。因此，上述 unigene 的发现对于阐明黄曲条跳甲对植物次生代谢分子或挥发性气味分子的识别和信息反馈机理研究具有非常重要的意义。

本研究还成功鉴定到与生物钟代谢路径相关的 5 种生物钟基因：Ps_clk，Ps_cyc，Ps_tim，Ps_per 和 Ps_vri 基因。参考黑腹果蝇的生物钟代谢路径分析，初步推断 5 个基因在生物钟代谢路径中可能的位置见图 5 中的星号标记，真实的路径逻辑关系还需进一步的实验验证。上述核心生物钟基因的鉴定，对于深入研究黄曲条跳甲昼夜节律行为及相关行为的分子、细胞和进化的基础提供了宝贵的数据信息。

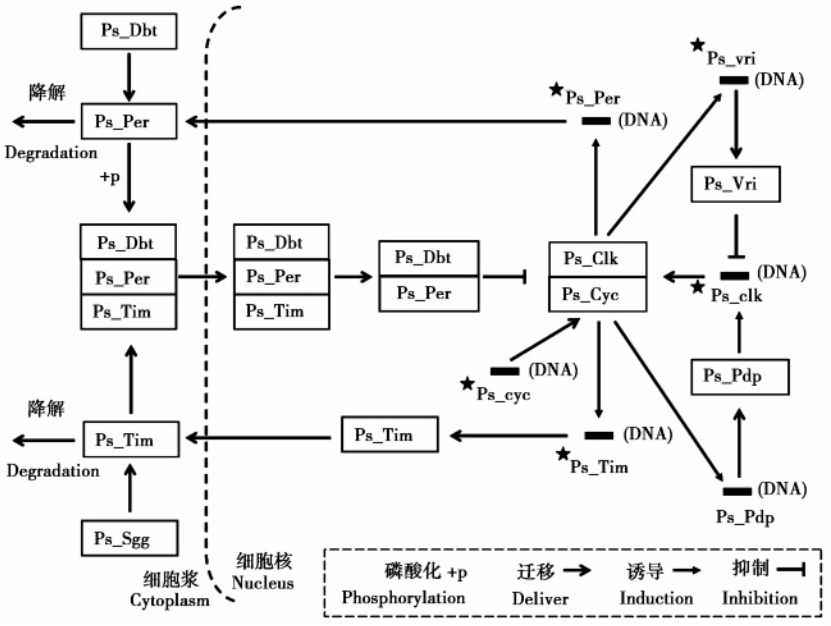

图 5　鉴定的 5 个生物钟基因在生物钟路径模型中可能的位置

Fig. 5 Five core biological clock genes showed in circadian rhythm model

生物钟路径图仿黑腹果蝇的模型路径(KEGG database，Circadian rhythm-Fly，Kanehisa Laboratories)；Dbt 为黑腹果蝇 double-time(dbt) 基因的蛋白产物；Sgg 是黑腹果蝇 glycogen synthase kinase 基因的蛋白产物；Pdp 是黑腹果蝇 hepatic leukemia factor 基因的蛋白产物。这 3 个基因的同源基因尚未在黄曲条跳甲转录组测序中被鉴定到。Circadian rhythm model of *Phyllotrata striolata* refers to the circadian rhythm of *Drosophila melanogaster* (dme04711，KEGG database)；Dbt is the protein product of double-time (dbt) of *D. melanogaster*; Sgg is the protein product of glycogen synthase kinase of *D. melanogaster*; Pdp is the protein product of hepatic leukemia factor of *D. melanogaster*. Gene of *P. striolata* orthology to above three genes was not identified in currently sequencing research.

3　讨论

本研究首次在国内外采用高通量测序技术对黄曲条跳甲成虫的转录组进行测序和功能分析，并重点挖掘与其行为及生殖发育相关的基因。本研究共获得 2 209 种具较高注释可信度的 unigene，表明在对黄曲条跳甲基因组及遗传背景几乎不清楚的情况下，高通量测序技术是批量发现黄曲条跳甲功能基因的有效手段。与传统测序相比，Solexa 高通量测序的长度完全可以满足序列数据分析的要求，且 Solexa 测序还具有速度快、通量高、成本低的优点。

本研究目前只优先分析了长度≥500 bp 的 2 209种 unigene。然而，本次转录组测序数据中还有 20 556 条 contig，其长度范围在 200 ~ 500 bp 之间。按同样分析方法进行初步筛选后，可获得 uingene 为 4 451 种。该批次的 4 451 种 unigene 的序列与上述的 2 209 种 unigene 的序列之间没有重叠关系，但可能存在为同一种全长 cDNA 序列上的两个或多个尚未相连拼接的序列区段。因此，本次转录组分析可获得的长度≥200 bp 的 unigene 种类数量会超过 2 209 种，数量介于 2 209 ~ 6 660 之间。由于黄曲条跳甲基因组水平研究的滞后，其基因组长度及基因数量都还未知。与黄曲条跳甲同属鞘翅

目的赤拟谷盗在2008年已成为第一种被基因组测序的农业害虫，基因组长度大约200 Mbp，具有基因(或蛋白)数量大约有16 000个(Tribolium Genome Sequencing Consortium，2008)。如果初步参考赤拟谷盗的基因数量，将黄曲条跳甲的基因数量假定为16 000个，则本次转录组测序分析获得unigene(≥200 bp)的数量可能占黄曲条跳甲总基因数目的13.8%～41.6%，为后期在基因组水平研究黄曲条跳甲提供了丰富的基因信息。

昆虫的行为节律是一种由生物钟控制的内源性节律。昆虫的交尾节律受内源性生物钟控制的现象已在部分昆虫中得到证实。国内学者对大猿叶虫卵孵化的时辰节律研究发现，卵的孵化主要发生在黎明和黄昏，且以黎明时孵化率最高(徐强和张庆，2007)。2009年，美国研究人员发现，果蝇的生物钟可以使它们在一天的某一时段，而非其他时段，对杀虫剂敏感得多。即在一天的时间里，与果蝇抵御能力最弱时段相比，在它最强时，要用3倍的杀虫剂剂量才会与前者有相同的致死效果(Hooven *et al.*，2009)。对于本研究检测的重要蔬菜害虫——黄曲条跳甲，其成虫的产卵习性以晴天为多，一天中以午后为多(与其午后活动盛期的时间区域基本相符)，也表现出一定的节律。本研究通过同源性搜索，获得了核心生物钟基因5个，同时也获得了与黄曲条跳甲生殖发育及生殖行为相关unigene共100多种。因此，诸如关键基因的鉴定及其功能研究将有助于从行为调控的角度创新发展害虫的防治策略。

本研究结合果蝇的KEGG Pathway数据库分析发现共363种unigene涉及40种代谢路径。涉及unigene数量相当较多的代谢路径主要有：泛素代谢路径(ubiquitin mediated proteolysis)、氨酰tRNA合成路径(aminoacyl-tRNA biosynthesis)、柠檬烯和蒎烯代谢路径(limonene and pinene degradation)等。其中柠檬烯和蒎烯代谢路径主要涉及昆虫对植物次生代谢物的趋性或防御体系。2010年，德国、台湾及瑞士科学家合作成功发现黄曲条跳甲雄虫可以分泌一种聚集信息素[(+)-(6R,7S)-himachala-9,11-diene]，而且这种性外激素可以通过α-雪松烯[α-himachalene(1R,7S)]为前体进行化学合成(Beran *et al.*，2010)。美国学者Bartelt等(2011)又报道了黄曲条跳甲雄成虫特有的倍半萜类的性外激素。本研究中柠檬烯和蒎烯代谢路径及其他代谢路径关键基因的鉴定，对于阐明黄曲条跳甲对外源或内源性信息素的代谢和信息反馈的研究具有非常重要的意义。

因此，黄曲条跳甲转录组学的研究，对于黄曲条跳甲的行为、代谢路径及生殖发育相关关键酶基因的发掘为克隆基因、研究基因功能提供了基础数据，为黄曲条跳甲行为调控的研究奠定了基础，同时为应用生物技术方法研发防治新策略提供了可行性。

致谢 本研究转录组测序工作及部分数据分析得到广州迈平生物科技公司的大力支持。

参考文献(References)

Bartelt RJ, Zilkowski BW, Cossé AA, Schnupf U, Vermillion K, Momany FA, 2011. Male-specific sesquiterpenes from *Phyllotreta* flea beetles. *J. Nat. Prod.*, 74(4): 585－595.

Beran F, Mewis I, Srinivasan R, Svoboda J, Vial C, Mosimann H, Boland W, Büttner C, Ulrichs C, Hansson BS, Reinecke A, 2011. Male *Phyllotreta striolata* (F.) produce an aggregation pheromone: identification of male-specific compounds and interaction with host plant volatiles. *J. Chem. Ecol.*, 37(1): 85－97.

Ewing RM, Ben Kahla A, Poirot O, Lopez F, Audic S, Claverie JM, 1999. Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. *Genome Res.*, 9: 950－959.

Feldmeyer B, Wheat CW, Krezdorn N, Rotter B, Pfenninger M, 2011. Short read Illumina data for the *de novo* assembly of a non-model snail species transcriptome (*Radix balthica*, Basommatophora, Pulmonata), and a comparison of assembler performance. *BMC Genomics*, 12: 317

Feng HT, Huang YJ, Hsu JC, 2000. Insecticide susceptibility of cabbage flea beetle *Phyllotreta striolata* (Fabricius) in Taiwan. *Plant Protection Bulletin* (*Taipei*), 42(1): 67－72.

Fu JW, Li JY, Qiu LM, Lin ZY, You MS, 2006. The regional diversity of susceptibility of striped flea beetle (SFB), *Phyllotreta striolata* (Fabricius), to insecticides in Fujian Province. *Journal of Fujian Agriculture and Forestry University* (*Natural Science Edition*), 35(3): 235－238. [傅建炜，李建宇，邱良妙，林泽燕，尤民生，2006. 福建省黄曲条跳甲药剂敏感性的地区差异. 福建农林大学学报(自然科学版)，35(3): 235－238]

Gaines P, Tompkins L, Woodard CT, Carlson JR, 2000. *quick-to-court*, a *Drosophila* mutant with elevated levels of sexual behavior, is defective in a predicted coiled-coil protein. *Genetics*, 154(4): 1627－1637.

Gao ZZ, Wu WJ, Cui ZX, 2000. Studies on the host range of *Phyllotreta striolata* (Fabricius). *Ecologic Science*, 19(2): 70－72. [高泽正，吴伟坚，崔志新，2000. 关于黄曲条跳甲的寄主范围. 生态科学，19(2): 70－72]

Gibbons JG, Janson EM, Hittinger CT, Johnston M, Abbot P, Rokas A, 2009. Benchmarking next-generation transcriptome sequencing for functional and evolutionary genomics. *Mol. Biol. Evol.*, 26(12):

2731 – 2744.

Hooven LA, Sherman KA, Butcher S, Giebultowicz JM, 2009. Does the clock make the poison? Circadian variation in response to pesticides. *PLoS ONE*, 4(7): e6469.

Hou YM, Pang XF, Liang GW, You MS, 2003. Evaluation of azadirachtin against striped flea beetle, *Phyllotreta striolata* (F.). *Chinese Journal of Applied Ecology*, 14(6): 959 – 962. [侯有明, 庞雄飞, 梁广文, 尤民生, 2003. 印楝素乳油对黄曲条跳甲种群控制作用评价. 应用生态学报, 14(6): 959 – 962]

Li RQ, Zhu HM, Ruan J, Qian WB, Fang XD, Shi ZB, Li YR, Li ST, Shan G, Kristiansen K, Li SG, Yang HM, Wang J, Wang J, 2010. *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res.*, 20(2): 265 – 272.

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M, 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 320(5881): 1344 – 1349.

Nie HX, 2007. The cause of serious damage of *Phyllotreta stiolata* and its control measures. *Hunan Agricultural Sciences*, (5): 122 – 124. [聂河兴, 2007. 黄曲条跳甲危害严重原因与防治对策. 湖南农业科学, (5): 122 – 124]

Pontius JU, Wagner L, Schuler GD, 2003. UniGene: a unified view of the transcriptome. In: The NCBI Handbook. National Center for Biotechnology Information, Bethesda, MD.

Rosenkranz R, Borodina T, Lehrach H, Himmelbauer H, 2008. Characterizing the mouse ES cell transcriptome with Illumina sequencing. *Genomics*, 92(4): 187 – 194.

Schuler GD, 1997. Pieces of the puzzle: expressed sequence tags and the catalog of human genes. *J. Mol. Med.*, 75: 694 – 698.

Tahvaanainen J, 1983. The relationship between leaf beetle and their cruciferous host plants: the role of plant and habitat characteristics. *Oikos*, 40(3): 433 – 437.

Tribolium Genome Sequencing Consortium, 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature*, 452(7190): 949 – 955.

Wang L, Zhang EQ, Yu ZP, 2009. Occurrence cause and integrated control of striped flea beetle in Jingzhou city in recent years. *Hubei Plant Protection*, 5(115): 21 – 22. [王玲, 张恩桥, 余周苹, 2009. 荆州市黄曲条跳甲偏重发生原因及综合防治技术. 湖北植保, 5(115): 21 – 22]

Xu Q, Zhang Q, 2007. Hatching rhythm of eggs in the cabbage beetle *Colaphellus bowringi*. *Jiangxi Plant Protection*, 30(3): 99 – 100. [徐强, 张庆, 2007. 大猿叶虫卵孵化的时辰节律研究. 江西植保, 30(3): 99 – 100]

Zhang MX, Liang GW, 2000. The influence of host plants on the experimental population of striped flea beetle [*Phyllotreta striolata* (F.)]. *Journal of South China Agricultural University*, 21(3): 25 – 28. [张茂新, 梁广文, 2000. 寄主植物对黄曲条跳甲实验种群增长的影响. 华南农业大学学报, 21(3): 25 – 28]

Zhao YY, Liu F, Yang G, You MS, 2011. PsOr1, a potential target for RNA interference-based pest management. *Insect Mol. Biol.*, 20(1): 97 – 104.

Zhou XZ, Wu G, 2004. Temporal and spatial dynamics of resistance to some commercial insecticides in *Phyllotreta striolata* (Fabricius) (Coleoptera: Chrysomelidae) in Fuzhou, China. *Journal of Fujian Agriculture and Forestry University* (*Natural Science Edition*), 33(2): 158 – 161. [周先治, 吴刚, 2004. 福州地区黄曲条跳甲的抗性监测. 福建农林大学学报(自然科学版), 33(2): 158 – 161]

(责任编辑：赵利辉)