棉花主要害虫绿盲蝽转录组动态变化特征及其与棉铃虫和棉蚜的比较分析

陈殿阳^{1,2}, 陈芳艳^{1,2}, 陈春雨^{1,2}, 陈晓亚^{1*}, 毛颖波^{1*}

1. 中国科学院上海生命科学研究院, 植物生理生态研究所, 分子植物科学卓越创新中心, 植物分子遗传国家重点实验室, 上海 200032;

2. 中国科学院大学, 北京 100049

* 联系人, E-mail: xychen@sibs.ac.cn; ybmao@sibs.ac.cn

收稿日期: 2017-05-10; 接受日期: 2017-06-01; 网络版发表日期: 2017-09-10

中国科学院战略性先导科技专项(批准号: XDB11030000)和农业部转基因生物新品种培育重大专项(批准号: 2014ZX08009001-009)资助

摘要 绿盲蝽(*Apolygus lucorum*)危害多种植物, 是主要的农业害虫之一. 本文对绿盲蝽3个不同发育期, 包括二龄若虫、四龄若虫和成虫的转录组测序和分析, 获得了98236条绿盲蝽单基因簇(unigenes), 平均长度1335 nt, 其中50640条获得了功能注释, 包括大量消化酶和细胞色素P450基因. 将其与棉铃虫(*Helicoverpa armigera*)和棉蚜(*Aphis gossypii*)的转录组做比较分析, 发现绿盲蝽消化酶基因的丰富性介于棉铃虫和棉蚜之间, 揭示了绿盲蝽作为细胞取食者的特征. 同时, 本研究分析了绿盲蝽从幼年到成年的过程中基因表达的动态变化. 高质量的转录组数据和基因表达谱为深入研究盲蝽类害虫的生理和发育、寻找防治新策略提供了宝贵的数据.

关键词 绿盲蝽, 转录组, 消化酶, 发育阶段

绿盲蝽(*Apolygus lucorum*)隶属半翅目, 盲蝽科, 是一种在亚洲、非洲、欧洲以及北美都有发生的世界性农业害虫^[1,2]. 20世纪, 棉铃虫(*Helicoverpa armigera*)曾是中国棉田的最主要害虫, 绿盲蝽在棉田中的数量较少, 造成的农业损失很小. 为了遏制棉铃虫的蔓延, 曾经不得不大量使用化学杀虫剂, 这不仅对棉铃虫起到了杀灭作用, 也控制了绿盲蝽在内的其他棉田次要害虫的发展^[3]. 表达苏云金芽孢杆菌(*Bacillus thuringiensis*)杀虫蛋白的转基因抗虫棉(Bt棉)对鳞翅目昆虫有较为专一的抗性, 但对包括绿盲蝽在内的刺吸式昆虫影响很小. 随着Bt棉在中国的广泛种植, 棉铃虫虫害得到了有效控制^[4], 田间化学杀虫剂的使用

量显著减少, 随之而来是绿盲蝽种群数量大幅增长. 绿盲蝽不仅严重危害棉花产量, 而且危害棉田周边的作物, 包括樱桃(*Cerasus pseudocerasus*)、葡萄(*Vitis vinifera*)、苹果(*Malus pumila*)、梨(*Pyrus* spp.)和茶树(*Camellia sinensis*)等^[2,5].

作为一种杂食性昆虫, 绿盲蝽可取食的植物超过150种^[3]. 同时, 绿盲蝽也会捕食其他昆虫作为营养补充, 如蚜虫和棉铃虫虫卵^[1,6]. 包括绿盲蝽在内的盲蝽科昆虫主要是刺吸式的昆虫, 绿盲蝽若虫和成虫均使用刺吸式口器取食植物组织. 杂食性盲蝽的食物构成可能会随着年龄的变化而发生改变: 通常来说, 低龄若虫倾向于植物性食物, 随着年龄增长, 动物性食物的

引用格式: 陈殿阳, 陈芳艳, 陈春雨, 等. 棉花主要害虫绿盲蝽转录组动态变化特征及其与棉铃虫和棉蚜的比较分析. 中国科学: 生命科学, 2017, 47: 936–948, doi: [10.1360/N052017-00096](https://doi.org/10.1360/N052017-00096)

英文版见: Chen D Y, Chen F Y, Chen C Y, et al. Transcriptome analysis of three cotton pests reveals features of gene expressions in the mesophyll feeder *Apolygus lucorum*. Sci China Life Sci, 2017, 60: 826–838, doi: [10.1007/s11427-017-9065-3](https://doi.org/10.1007/s11427-017-9065-3)

比例逐渐增加^[7]。蚜虫(*Aphidoidea*)和粉虱(*Aleyrodidae*)属于维管取食者,吸食植物韧皮部的汁液。与之不同的是,绿盲蝽是细胞取食者,取食植物组织的整个细胞,包括细胞核和细胞质^[8]。先前的研究表明,绿盲蝽取食时造成的机械损伤和分泌的唾液都是植物产生虫害表型的重要原因^[9,10]。

除了具有广泛的宿主范围,盲蝽类害虫迅速发展的另一个不容忽视的因素是对杀虫剂的抗性^[11-13]。但是,迄今对绿盲蝽的生理生化特性所知甚少。转录组测序技术已经用于多种盲蝽类昆虫^[14-17],包括最近发表的绿盲蝽三龄若虫和成虫的研究结果^[18]。植物介导的RNA干扰技术已经被证明是一种具有巨大发展潜力的防治植物害虫的有效手段^[19,20],对绿盲蝽同样有效^[21]。然而,绿盲蝽的转录组数据还需要进一步完善,特别是在序列平均长度、功能注释以及不同发育时期基因表达特征方面有待加强。本文选取了绿盲蝽3个不同发育期共9个样品,提取总RNA进行转录组测序,鉴定出98236条unigenes,其中15934条在成虫中的表达量明显高于二龄若虫,包括大量细胞色素P450单加氧酶和消化酶基因。本研究对绿盲蝽的研究与防治具有重要价值。

1 材料与方法

1.1 昆虫饲养

绿盲蝽由中国农业科学院植物保护研究所吴孔明实验室提供,于2005年收集于河北廊坊(39.53°N, 116.70°E)。成虫和若虫在人工培养箱中饲养,温度24℃,相对湿度60%,长日照条件(16 h光照/8 h黑暗)。若虫使用鲜玉米(*Zea mays*)、成虫使用菜豆(*Phaseolus vulgaris*)和豇豆(*Vigna unguiculata*)喂养。对于测序用的绿盲蝽个体,收集同一天孵化的一龄若虫使用豇豆喂养,依照形态特征收集适龄虫体。二龄若虫体长约1.4 mm;五龄若虫体长3~4 mm,翅芽清晰可见;成虫体长5~6 mm,翅膀发育完全。

1.2 测序样品制备和RNA测序

取适量绿盲蝽(二龄若虫50只左右,五龄若虫15只,成虫10只)加入2 mL TRIzol(Invitrogen, 美国)依照产品说明进行RNA提取, DNase I 处理去除基因组DNA,使用Nanodrop 2000(NanoDrop products, 美

国)和Agilent 2100 Bioanalyzer(Agilent Technologies, 美国)检测RNA浓度和完整性。取10 μg总RNA使用带有Oligo (dT)的磁珠(Invitrogen)富集mRNA,再将mRNA打断为100~400 nt的片段,接着使用SuperScript II RT试剂盒(Invitrogen), random hexamer-primer为引物,进行第一链cDNA的合成。构建好的文库用Agilent 2100 Bioanalyzer和ABI StepOnePlus Real-Time PCR System质检合格后,由测序公司(壹基因, 杭州)使用Illumina HiSeq™ 4000测序平台(Illumina, 美国)进行双端测序。

1.3 测序结果组装

在正式组装前,去除原始数据低质量序列($Q < 20$, N含量>5%, 接头污染),得到的clean reads使用Trinity RNA-Seq-2.0.6^[22]进行从头组装,自定义参数(--min_contig_length 200--min_kmer_cov 2--min_glue 3--seqType fq),同一发育时期样品的测序结果组装为一套重叠群(contig),组装结果混合后使用TGICL^[23]去冗余得到unigene库,再用Langmead等人^[24]将测序得到的双端短序列比对到最终的绿盲蝽unigene库中。最终的绿盲蝽单基因簇(unigene)包含一致簇(consensus cluster)序列(由2~107条高度同源的unigene组成)和单元簇(singleton, 仅有一条unigene)。所有的原始数据上传至NCBI(<https://www.ncbi.nlm.nih.gov/>)(BioProject accession number: PRJNA347890)。

1.4 序列的功能注释和分析

使用BLASTx或BLASTn将unigene与以下数据库进行比对: GenBank NR(Non-Redundant Protein)和NT(Nucleotide), SwissProt(Swiss Protein), KEGG(Kyoto Encyclopedia of Genes and Genomes), COG, GO。选取P450单加氧酶基因的氨基酸序列进行系统进化分析,使用ClustalX^[25]进行多重比对,MEGA 5.1^[26]中使用最大简约法、默认设置、1000次重复进行系统进化树构建。潜在的SSR分子标记使用MISA^[27]进行鉴定。

1.5 差异表达基因分析

使用RSEM^[28]对FPKM(fragments per kilobase per million mapped reads)值进行差异表达计算,使用Benjamini多重检验校正^[29]计算错误发现率(false discovery rate, FDR),差异表达基因的筛选标准为差异倍数在2倍或以上,且FDR<0.001。

1.6 实时定量PCR检测(qRT-PCR)

使用NCBI的在线程序(www.ncbi.nlm.nih.gov/tools/primer-blast/)设计qRT-PCR引物(网络版附表1)。取1 μg总RNA使用反转录试剂盒(Transgene, 北京)合成cDNA, 反转录产物稀释10倍。使用SYBR green PCR Mix(Biotool, 美国)在Bio-Rad CFX Connect Real-Time PCR system (Bio-Rad, 美国)上进行qRT-PCR检测, 热循环步骤为: (i) 95℃ 5 min; (ii) 95℃ 15 s; (iii) 60℃ 30 s; 步骤(ii)和(iii)重复40个循环。选取5个在测序结果中表达量稳定的基因(ribosomal protein L26 and S15, transferrin, polyubiquitin, 26S proteasome)作为内参, 数据分析采用 $2^{-\Delta\Delta C_t}$ 法。

2 结果与分析

2.1 序列组装和功能基因注释

分别选取绿盲蝽二龄若虫、五龄若虫和成虫3个不同发育期虫体总RNA样本, 每个发育时期3个重复, 进行转录组测序。原始测序数据过滤后, 得到452503786条clean reads (约67.9 Gb), 每个发育期样本约含22 Gb数据(网络版附表2)。绿盲蝽3个发育期的样本分别组装, 得到6575, 80299和78219条contigs, 平均长度分别

为1015, 1113, 1197 nt, 数据合并后, 共得到98236条绿盲蝽unigene序列, 平均长度1335 nt, N_{50} 为2337 nt(图1A和B)。最终绿盲蝽总unigenes包含34481条contiguous序列和63755条singleton序列。

为了与棉铃虫转录组进行比较, 对一套来自棉铃虫脂肪体、血淋巴细胞以及二龄幼虫的RNA Hiseq数据^[30]进行了组装。去除冗余后, 获得了116046条棉铃虫unigene序列, 平均长度988 nt, N_{50} 为2224 nt, 包含40144条contiguous序列和75902条 singleton序列。在unigene数量和平均长度上, 本文的绿盲蝽转录组数据棉铃虫数据相近。

使用BLASTx或BLASTn将两种昆虫的转录组参考序列与常用的蛋白质或核酸数据库比对, 共有50640条(51.5%)绿盲蝽unigene序列以及48472条(41.8%)棉铃虫unigene序列可以获得至少一个数据库的功能注释(表1, 图1C)。显然, 尚有大量绿盲蝽和棉铃虫序列没有获得功能注释, 分别占总数的48.5%和58.2%。除了缺乏基因组数据, 另一个可能的原因是昆虫中丰富的可变剪辑^[31]。从E-value分布上看, 60.4%被注释的绿盲蝽unigenes与NR数据库中的蛋白质具有非常高的同源性($E\text{-value} < 1 \times 10^{-30}$), 其余序列E-value值介于 $1 \times 10^{-30} \sim 1 \times 10^{-5}$ (图2A)。进一步分析显示, 68.7%的序列

图1 绿盲蝽转录组拼装结果

A: 绿盲蝽不同发育期转录组组装结果统计; B: 绿盲蝽单基因簇(unigene)长度分布; C: 绿盲蝽转录组序列与NR, Swiss-prot, COG和KEGG数据库比对结果的韦恩图, 图中4个圈代表4个数据库, 重叠处的数字代表BLASTx比对后, 对应数据库之间共有的unigene数量

表 1 注释到NR/NT/SwissProt/KEGG/COG/GO数据库的基因数量^{a)}

	NR [*]	NT ^{**}	SwissProt ^{***}	KEGG ^{****}	COG ^{*****}	GO ^{*****}	All
绿盲蝽	47,629	18,816	33,711	30,964	17,566	10,098	50,640
棉铃虫	38,950	32,402	27,250	25,103	14,489	7,289	48,472

a) *: GenBank Non-Redundant Protein; **: GenBank Non-Redundant Nucleotide; ***: Swiss Protein; ****: Kyoto Encyclopedia of Genes and Genomes; *****: Clusters of Orthologous Groups; *****: Gene Ontology

图 2 绿盲蝽参考序列同源比对结果特征

通过BLASTx将绿盲蝽unigene比对到NR数据库后, 根据比对结果得到E-value分布(A), 基因相似性(B), 以及注释到的物种(C)

与NR数据库中的序列有超过40%的相似性(图2B). 被注释到最多条目的物种是温带臭虫(*Cimex lectularius*, 47629条中的32.8%), 其次是茶翅蝽(*Halyomorpha halys*, 11.3%), 这两种昆虫与绿盲蝽一样都属于半翅目(图2C).

2.2 GO和COG功能分析

GO(Gene Ontology)是一个国际标准化的基因功能分类体系, 在GO数据库的分类体系中, 将功能蛋白从3个大类进行描述: 分子功能(molecular function)、细胞组分(cellular component)和生物学过程(biological process). 通过Blast2GO分析得到10098条绿盲蝽unigene序列至少有一条GO注释结果, 共计51962条GO注释条目; 棉铃虫参考序列中, 有7289条unigene序列具有至少一条GO注释结果, 共计39041条GO注释条目. 绿盲蝽GO注释结果可以进一步分为54个类别, 其中“cellular process”(5102)和“metabolic process”(5024)在生物

学过程大类中最为丰富. 在细胞组分大类中, 3715个注释条目归属于“cell”, 3688条归属“cell part”; 在分子功能大类中, 分别有4642和4005个注释条目属于“binding”和“catalytic activity”, 使之成为该大类下最丰富的两个条目(图3A).

COG(Cluster of Orthologous Groups of proteins)是基于真核生物系统进化关系构建的蛋白质分类数据库^[32]. 通过分析, 获得了17566条绿盲蝽unigene序列的COG注释, 共计43732个COG注释条目, 这些条目可归为25类. 除了“general function prediction only”(6917条), 最大类群为“replication, recombination and repair”(4523条), 接着为“transcription”(3362条)和“translation, ribosomal structure and biogenesis”(3376条). 最小的类群为“nuclear structure”, 只有8个注释条目(图3B). 绿盲蝽和棉铃虫在COG注释的条目上总体接近, 但是两者之间仍有一些差异. 一些COG分类条目, 例如, “carbohy-

图3 GO和COG注释
A: 绿盲蝽参考序列GO注释结果分布; B: 绿盲蝽和棉铃虫COG注释结果分布

drate transport and metabolism”, “replication, recombination and repair”以及“transcription”, 相关基因在绿盲蝽中更为丰富; 一些与免疫相关的COG分类条目, 例如, “defense mechanisms”(270条), “intracellular trafficking, secretion, and vesicular transport”(1546条)在两种昆虫转录组中的丰富度基本相同。

2.3 消化酶基因

广食性昆虫一个显著的特点就能够消化不同类

型的植物成分。作为棉田中的3种重要害虫, 穿刺后细胞取食的盲蝽、咀嚼式取食的棉铃虫和维管取食的棉蚜(*Aphis gossypii*)有着不同的口器和取食方式。对3个棉蚜的Illumina RNA Hiseq测序数据^[33,34]进行组装、注释, 并与绿盲蝽和棉铃虫一起, 分析比较消化酶基因(表2)。通常, 在植物组织如叶片中, 蛋白质和脂类成分含量较低。绿盲蝽和棉铃虫的unigene库中包含大量编码蛋白酶的序列, 绿盲蝽转录组中鉴定到10类胰蛋白酶(trypsin-like protease)基因, 棉铃虫转录组则包

表2 绿盲蝽、棉铃虫、棉蚜转录组中消化酶基因的种类和数量

	绿盲蝽	棉铃虫	棉蚜
类胰蛋白酶	10	25	0
羧肽酶	161	186	64
氨肽酶	129	125	59
磷脂酶	181	222	62
脂肪酶	308	490	112
α -淀粉酶	2	13	0
果糖酶	1	14	0
β -葡萄糖苷酶	21	61	16
内切葡聚糖酶	2	0	1

含多达25条,但在棉蚜中没有发现这一类基因的表达。本课题组在绿盲蝽中鉴定出161条编码羧肽酶(carboxypeptidase)和129条氨肽酶(aminopeptidase)的基因,棉铃虫羧肽酶和氨肽酶基因数量与绿盲蝽接近,分别为186条和125条;棉蚜中鉴定出的羧肽酶和氨肽酶数量较少,只有64条和59条。这些数据表明,细胞取食者绿盲蝽的蛋白消化酶基因的丰富性比咀嚼式昆虫棉铃虫略低。除了蛋白质,昆虫还需要充分利用宿主植物中有限的脂类,其中植物类固醇是昆虫合成胆固醇必需的原料^[35]。脂肪酶(lipase)和磷脂酶(phospholipase)是昆虫消化脂类成分的关键酶,在鳞翅目幼虫的中肠、脂肪体和唾液腺中已经有许多具有功能的脂肪酶和磷脂酶被报道^[36]。本研究组在棉铃虫转录组中鉴定出490条编码脂肪酶的序列,222条编码磷脂酶的序列;在绿盲蝽中,相关基因数目比棉铃虫大约低1/4,分别为308条和181条;在棉蚜中,脂肪酶和磷脂酶基因的数量进一步减少至112条和62条。和蛋白酶基因一样,编码脂肪酶和磷脂酶的基因数在这3种不同取食方式的昆虫中存在差异。与低含量的蛋白质和脂类相比,植物组织中各种不同种类的碳水化合物的含量十分丰富。研究发现,在双翅目(dipteran)昆虫中, α -淀粉酶(alpha-amylase)基因种类丰富并且形成一个大家族^[37]。本课题组在棉铃虫转录组中发现了13条 α -淀粉酶的基因序列,在绿盲蝽中只鉴定出2条。本研究组还鉴定出绿盲蝽2条果糖苷酶(fructosidase)基因序列,而在棉铃虫中则多达14条。然而,在棉蚜转录组数据中未能发现果糖苷酶和 α -淀粉酶基因。有趣的是,在绿盲蝽转录组中发现了两条unigenes序列,所编码的

蛋白与内切葡聚糖酶(endoglucanase)高度相似;内切葡聚糖酶可能帮助绿盲蝽利用植物组织中的纤维素成分。在绿盲蝽转录组数据中还鉴定出数个核糖核酸酶(ribonuclease, RNases)基因,其中一条unigene编码分泌性核糖核酸酶。除了消化植物组织中的RNA,分泌性核糖核酸酶也可以在受伤的植物组织中诱发对病原菌的防御反应^[38]。

2.4 细胞色素P450基因

细胞色素P450单加氧酶在脊椎动物、昆虫、真菌、植物、细菌、古细菌和病毒中广泛存在,组成较大的蛋白质家族,它们或催化内源代谢物的生物合成,或参与外源化合物的代谢^[39]。全基因组测序结果表明,P450基因在不同的昆虫中都具有丰富的多样性:东亚飞蝗(*Locusta migratoria*)基因组中有94个P450基因^[40],地中海实蝇(*Ceratitis capitata*)有88个^[41],赤拟谷盗(*Tribolium castaneum*)有134个^[42],家蚕(*Bombyx mori*)有84个^[43]。在绿盲蝽转录组数据库中,有197条unigene序列预测为编码P450,棉铃虫则为194条。由于可变剪辑的存在,这两种昆虫的P450基因的实际数目可能小于根据转录组推测的数量。对115条大于300 aa的绿盲蝽P450蛋白序列进行系统分析(图4,网络版附表3),可以看到绿盲蝽P450酶主要分为5类,分别属于CYP2, CYP3, CYP4和CYP6家族,其中CYP3家族有12个成员,这一家族的昆虫P450通常参与外源化合物代谢进而形成对杀虫剂的抗性(抗药性),其基因表达可以被苯巴比妥、杀虫剂和天然产物诱导^[44]。棉铃虫P450基因HaCYP6AE14可以被棉酚诱导并参与棉酚的解毒代谢^[19,45],发现了绿盲蝽7个CYP6家族基因,它们可能在棉酚解毒代谢中发挥作用,从而增强绿盲蝽对棉花的适应性。

2.5 不同发育阶段的基因表达变化

为了进一步了解基因表达的动态变化,对绿盲蝽3个不同发育期的转录组做了动态分析。根据FPKM值和FDR值来判断差异表达基因,并根据KEGG对差异表达基因进行不同途径的富集分析。比较绿盲蝽五龄若虫与成虫的转录组,发现1265条在成虫中表达量上升的unigene序列,同时1343条unigenes的表达量在五龄若虫中更高。在二龄若虫-成虫转录组之间的差异表达基因更多,共有15934条unigenes在成虫中的表达

图4 绿盲蝽P450单加氧酶系统树

长度大于300 aa的P450蛋白序列被用来构建系统进化树, 5个主要聚类用不同颜色的阴影区分; 棉铃虫P450蛋白HaCYP6AE14也包含在系统树中, 该序列与其他7个绿盲蝽P450一起位于图中粉色区域

水平高于二龄若虫, 2698条unigenes在成虫中则明显降低。将二龄若虫与五龄若虫的转录组相比, 差异表达基因的数量也很多, 共有10351条unigenes在五龄若虫中表达量高于二龄若虫, 1237条unigenes在五龄若虫中表达量降低(图5A)。上述结果表明, 成年或近成年期的绿盲蝽有着更为丰富和活跃的基因表达。本研究组鉴定到356条和366条unigene序列在成虫中的表达量分别显著升高或降低, 但在二龄若虫-五龄若虫之间表达量保持稳定(图5B), 这类表达模式的基因可能包含了与绿盲蝽生殖和性成熟相关的基因, 其中有37条仅存在于成虫转录组样本中。

KEGG pathway富集结果表明, 在二龄若虫-成虫差异表达基因中, *D*-精氨酸和*D*-鸟氨酸代谢(*D*-arginine and *D*-ornithine metabolism)有着最高的富集程度, 绿盲蝽转录组中22条属于该途径的unigenes中, 16条的表达有显著性差异, 其次是mRNA监控通路(mRNA surveillance pathway), 唾液分泌(salivary secretion)排名

第4, 而RNA转运(RNA transport)有着最多的基因数量(图5C)。差异表达基因KEGG pathway富集结果表明, 绿盲蝽从幼年期到成年期, 一些与RNA、蛋白代谢与分泌相关的途径发生了很大变化。P450基因的表达模式也随着发育进展而变化, 在鉴定出的197个P450基因中, 有70个在若虫到成虫的发育过程中表达量发生显著变化, 其中16个表达量在成虫中下降, 54个在成虫中升高(网络版附表3, 网络版附图1)。有5条P450序列只在成虫中有较高的表达量(FPKM值比二龄若虫和五龄若虫的FPKM值高出10倍以上), 这些P450可能与绿盲蝽的生殖发育相关。挑选20个二龄若虫-成虫差异表达的P450基因做实时定量荧光PCR分析, 结果表明这些unigene转录本的相对表达量与转录组测序数据基本相符(图6A和B)。

在绿盲蝽性相关(sex-biased)和繁殖相关(reproduction-related)的基因中, 有66条气味结合蛋白(odorant-binding proteins)、38条羧酸酯酶(carboxylesterase)、34

图5 绿盲蝽不同发育期的差异表达基因

A: 差异表达基因的数目。黄色柱子代表上调的unigene数目, 蓝色柱子代表下调的unigene数目; B: 仅在绿盲蝽成虫中上调与下调的基因。绿盲蝽unigene根据FPKM值筛选后进行分析, FPKM需要在至少一个样本中高于5, 并且与二龄若虫、五龄若虫相比, 在成虫中表达(差异倍数 ≥ 2 或 ≤ 0.5 , $FDR \leq 0.001$)。Y轴代表基因表达量高低(取 $\log_{10}$ FPKM), 以五龄若虫的表达量为参考; C: 二龄若虫和成虫差异基因富集程度前20的生物学途径统计。Y轴代表富集程度, 点的大小和颜色分别代表基因的数量和Q-value

条表皮蛋白(cuticular protein)和4条变形-2蛋白(transformer-2 protein)的基因序列(网络版附表4)。进一步分析发现, 有13个(19.7%)气味结合蛋白基因和11个(32.4%)表皮蛋白基因在五龄若虫和成虫之间差异表达, 比五龄若虫-成虫之间差异表达基因的平均出现占比(2.6%)要高出许多。只有1个羧酸酯酶基因但没有transformer-2基因在五龄若虫-成虫的转变过程中表达量发生显著变化。气味结合蛋白和表皮蛋白的差异表达基因提示, 在绿盲蝽从若虫到成虫的转变过程中, 嗅觉和表皮结构发生了较大变化, 并可能与生殖发育密切相关。

消化酶基因在绿盲蝽不同发育时期的表达模式同样发生变化。以类胰蛋白酶为例, 实时定量荧光PCR(qRT-PCR)结果表明不同类胰蛋白酶基因在3个发育期有不同的表达量(图6C)。在4个mRNA水平最高的类胰蛋白酶基因中, 有2个在不同发育时期表达量保持稳定; 其余2个则显示出相反的表达模式: 一个仅在

成虫中高表达, 另一个在成虫中的表达量大幅下降, 最终这两个类胰蛋白酶基因的转录本在成虫中的表达水平几近相同(图6C, 网络版附图2), 说明绿盲蝽蛋白酶基因的表达模式具有发育时期特异性。综上, 转录组分析结果表明, 作为一种渐变态昆虫, 绿盲蝽在从低龄若虫向高龄若虫以及成虫转变的过程中基因表达谱发生了巨大变化, 而五龄若虫和成虫之间转录组相对较为相似, 印证了它们之间较为相似的形态学特征和取食习性。当然, 五龄若虫和成虫之间的差异表达基因, 可能包含了绿盲蝽繁殖和变态过程中重要的靶基因。

2.6 SSR鉴定

简单重复序列(simple sequence repeat, SSR)作为一种分子标记, 已经在模式昆虫例如果蝇(*Drosophila melanogaster*)^[46]和家蚕(*Bombyx mori* L.)^[47]中报道。转录组测序为研究者们提供了一个快速、灵敏、经济

图6 绿盲蝽3个发育期20个P450基因和10个类胰蛋白酶基因的相对表达水平的qRT-PCR分析

A: 成虫中表达量下降的基因; B: 成虫中表达量上升的基因; C: 类胰蛋白酶。横坐标编号为1: 二龄若虫; 2: 五龄若虫; 3: 成虫。二龄若虫和五龄若虫mRNA水平以成虫的表达量为参考值, 相对mRNA表达量的 x 和SD在柱形图上表示, 生物学重复数 $n=3$

和高通量的研究非模式生物SSR的方法^[48]. 通过对绿盲蝽转录组参考序列的搜索, 鉴定出6868个绿盲蝽SSR, 包含了双核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸重复序列. 大部分的SSR少于18个重复单元, 5~8个重复单元较为普遍. 三核苷酸重复占总数的58.22%, 是最丰富的SSR类型; 第二丰富的是双核苷酸重复, 占总数的36.35%; 紧接着分别是四核苷酸(2.46%)、五核苷酸(1.95%)和六核苷酸(1.03%)重复(图7).

3 讨论

转录组测序已经成为一种重要的生物学研究手段, 通过转录组测序, 可以获得大量的序列信息. 在本研究获得了绿盲蝽3个不同发育时期样本的转录组信息, 拼装得到了98236条绿盲蝽unigenes作为绿盲蝽参考序列, 其中有超过50%的序列与常用的蛋白质或核酸数据库有较高的同源比对结果, 这个结果略优于公共数据库中棉铃虫转录组. 相比已经报道的盲蝽类昆虫转录组数据, 如绿盲蝽^[18]和中黑盲蝽(*Adelphocoris suturalis*)^[15], 本文的测序结果在序列数量、平均长度、 N_{50} 等参数上有着更高的质量. 昆虫基因组测序近几年有了很大发展^[49], 转录组可以用于基因组测序结果的优化^[50], 为将来绿盲蝽基因组研究提供支撑. 此外, 转录组结果不仅提供了大量昆虫生理生化方面的信息, 也帮助孕育防治绿盲蝽的新思路和新方法,

例如, 利用miRNA^[20]和dsRNA^[19]对昆虫靶基因进行调控, 以及最近发展起来的基因编辑技术^[51]. 报道显示, 跨界的miRNA参与植物-哺乳动物^[52]、真菌-植物的相互作用^[53], 棉花-昆虫之间潜在的miRNA互作有望通过昆虫转录组与棉花miRNA数据库^[54]共同分析获得线索. 绿盲蝽转录组序列总体上与温带臭虫(*Cimex lectularius*)高度相似, 与同属于半翅目的昆虫——蚜虫和褐飞虱(*Nilaparvata lugens*)相比, 盲蝽类昆虫与温带臭虫在基因序列相似性上更为接近, 这与盲蝽和臭虫同属于半翅目下的臭虫次目的分类系统是一致的.

尽管缺少绿盲蝽的全基因组序列, 本文报道的转录组数据提供了绿盲蝽3个不同发育期的基因表达谱, 从中可以发现很多具有重要功能的基因, 这大大丰富了对这种农业害虫的了解. 二龄若虫与五龄若虫、成虫相比, 差异表达的基因大多数在发育早期表达量较低. 本研究组鉴定了16个在成虫期下调以及54个在成虫期上调的P450单加氧酶基因. 许多与性以及繁殖相关的基因, 在褐飞虱^[55]和烟粉虱(*Bemisia tabaci*)^[56]中报道的羧酸酯酶、表皮蛋白以及气味结合蛋白等基因在本研究组的转录组数据中都有发现. 五龄若虫和成虫相比, 差异表达基因的数量相对要少, 说明五龄若虫、成虫在转录组上较低龄若虫相比更为接近, 这与绿盲蝽3个发育时期的虫体在外部形态、生活习性上的差异成都颇为一致——五龄若虫与成虫在外部形态和习性上更为相似. 虽然五龄若虫-成虫转变期

图7 绿盲蝽转录组SSR统计

黑色条柱代表具有对应重复单元的SSR数量; 彩色条柱和饼图代表双核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸重复序列的数量

间变化的基因数量较少,但是在这一期间表达量发生显著变化的基因往往与性器官发育和生殖相关,可作为绿盲蝽防控的潜在靶基因。本文转录组数据也为将来鉴定绿盲蝽嗅觉系统、寻找化学防治靶点、研究绿盲蝽发育、性别决定机制提供了重要的信息。

维管取食的刺吸式昆虫如蚜虫和烟粉虱,缺乏鳞翅目昆虫那样的复杂的消化酶系统,这可能是由于植物维管束韧皮部细胞富含游离氨基酸、单糖和寡糖,取食维管汁液的昆虫可以更加容易地消化食物中的营养成分。和这些维管取食的刺吸式昆虫不同,绿盲蝽属于细胞取食的刺吸式昆虫,主要取食叶肉以及其他组织的整个细胞,并且在取食前有一系列体外消化的过程,这样的取食方式要求绿盲蝽拥有相对复杂的

消化酶系统,并且不同于咀嚼式昆虫和维管刺吸式昆虫。从3种棉花害虫的转录组数据比较结果看,棉铃虫拥有最为发达的蛋白酶系统,包括大量不同种类的胰蛋白酶、羧肽酶和氨肽酶;相反,这些消化酶基因的种类在棉蚜的转录组中要少很多;绿盲蝽消化酶的复杂性则介于棉铃虫和棉蚜之间,这与这3种昆虫的口器特征以及取食方式的差异是相统一的,随着食物成分多样性的增加,消化酶的多样性也随之增加。虽然用于比较的转录组数据来自于不同的实验,这可能会对比较结果产生一定影响,但是考虑到用于比较的绿盲蝽和棉铃虫转录组数据质量有着很高的相似度,本文结果能够说明刺吸式昆虫消化酶体系的复杂度低于咀嚼式昆虫。

参考文献

- 1 Lu Y H, Qiu F, Feng H Q, et al. Species composition and seasonal abundance of pestiferous plant bugs (Hemiptera: Miridae) on Bt Cotton in China. *Crop Protect*, 2008, 27: 465–472
- 2 Watanabe K. Characteristic of Damages of *Lygocoris (Apolygus) lucorum* (Meyer-Dür) (Heteroptera Miridae) and *Halyomorpha halys* (Stål) (Heteroptera: Pentatomidae) on Cherry. Annual Report of the Society of Agricultural Chemicals of North Japan, 1996. 143–144
- 3 Lu Y, Jiao Z, Wu K. Early season host plants of *Apolygus lucorum* (Heteroptera: Miridae) in Northern China. *Jnl Econ Entom*, 2012, 105: 1603–1611
- 4 Wu K M, Lu Y H, Feng H Q, et al. Suppression of cotton bollworm in multiple crops in China in areas with Bt toxin-containing cotton. *Science*, 2008, 21: 1676–1678
- 5 Lu Y, Wu K, Jiang Y, et al. Mirid bug outbreaks in multiple crops correlated with wide-scale adoption of Bt cotton in China. *Science*, 2010, 328: 1151–1154
- 6 Yuan W, Li W, Li Y, et al. Combination of plant and insect eggs as food sources facilitates ovarian development in an omnivorous bug *Apolygus lucorum* (Hemiptera: Miridae). *Jnl Econ Entom*, 2013, 106: 1200–1208
- 7 Eubanks M D, Denno R F. The ecological consequences of variation in plants and prey for an omnivorous insect. *Ecology*, 1999, 80: 1253–1266
- 8 Rodriguez-Saona C, Crafts-Brandner S J, Williams III L, et al. *Lygus hesperus* feeding and salivary gland extracts induce volatile emissions in plants. *J Chem Ecol*, 2002, 28: 1733–1747
- 9 Kimmins F M, Tjallingii W F. Ultrastructure of sieve element penetration by aphid stylets during electrical recording. *Entomol Expts Appl*, 1985, 39: 135–141
- 10 Zhang L, Lu Y, Liang G. A method for field assessment of plant injury elicited by the salivary proteins of *Apolygus lucorum*. *Entomol Exp Appl*, 2013, 149: 292–297
- 11 Zhen C, Gao X. A point mutation (L1015F) of the voltage-sensitive sodium channel gene associated with lambda-cyhalothrin resistance in *Apolygus lucorum* (Meyer-Dür) population from the transgenic Bt cotton field of China. *Pesticide Biochem Physiol*, 2016, 127: 82–89
- 12 Zhen C, Miao L, Liang P, et al. Survey of organophosphate resistance and an Ala216Ser substitution of acetylcholinesterase-1 gene associated with chlorpyrifos resistance in *Apolygus lucorum* (Meyer-Dür) collected from the transgenic Bt cotton fields in China. *Pesticide Biochem Physiol*, 2016, 132: 29–37
- 13 Fleming D E, Krishnan N, Catchot A L, et al. Susceptibility to insecticides and activities of glutathione *S*-transferase and esterase in populations of *Lygus lineolaris* (Hemiptera: Miridae) in Mississippi. *Pest Manag Sci*, 2016, 72: 1595–1603
- 14 Hull J J, Geib S M, Fabrick J A, et al. Sequencing and *de novo* assembly of the western tarnished plant bug (*Lygus hesperus*) transcriptome. *PLoS ONE*, 2013, 8: e55105
- 15 Tian C, Tek Tay W, Feng H, et al. Characterization of *Adelphocoris suturalis* (Hemiptera: Miridae) transcriptome from different developmental

- stages. *Sci Rep*, 2015, 5: 11042
- 16 Showmaker K C, Bednářová A, Gresham C, et al. Insight into the salivary gland transcriptome of *Lygus lineolaris* (Palisot de Beauvois). *PLoS ONE*, 2016, 11: e0147197
 - 17 Yuan H B, Ding Y X, Gu S H, et al. Molecular characterization and expression profiling of odorant-binding proteins in *Apolygus lucorum*. *PLoS ONE*, 2015, 10: e0140562
 - 18 Cao D P, Liu Y, Wei J J, et al. A *de novo* transcriptomic analysis to reveal functional genes in *Apolygus lucorum*. *Insect Sci*, 2016, 23: 2–14
 - 19 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
 - 20 Guo H, Song X, Wang G, et al. Plant-generated artificial small RNAs mediated aphid resistance. *PLoS ONE*, 2014, 9: e97410
 - 21 Wu X M, Yang C Q, Mao Y B, et al. Targeting insect mitochondrial complex I for plant protection. *Plant Biotechnol J*, 2016, 14: 1925–1935
 - 22 Haas B J, Papanicolaou A, Yassour M, et al. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*, 2013, 8: 1494–1512
 - 23 Pertea G, Huang X, Liang F, et al. TIGR gene indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*, 2003, 19: 651–652
 - 24 Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 2009, 10: R25
 - 25 Thompson J, Gibson T J, Plewniak F, et al. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*, 1997, 25: 4876–4882
 - 26 Tamura K, Peterson D, Peterson N, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*, 2011, 28: 2731–2739
 - 27 Sharma P C, Grover A, Kahl G. Mining microsatellites in eukaryotic genomes. *Trends Biotech*, 2007, 25: 490–498
 - 28 Li B, Dewey C N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 2011, 12: 323
 - 29 Benjamini Y, Drai D, Elmer G, et al. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res*, 2001, 125: 279–284
 - 30 Xiong G H, Xing L S, Lin Z, et al. High throughput profiling of the cotton bollworm *Helicoverpa armigera* immunotranscriptome during the fungal and bacterial infections. *BMC Genomics*, 2015, 16: 321
 - 31 Chen Z X, Sturgill D, Qu J, et al. Comparative validation of the *D. melanogaster* modENCODE transcriptome annotation. *Genome Res*, 2014, 24: 1209–1223
 - 32 Tatusov R L, Fedorova N D, Jackson J D, et al. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, 2003, 4: 41
 - 33 Pan Y, Peng T, Gao X, et al. Transcriptomic comparison of thiamethoxam-resistance adaptation in resistant and susceptible strains of *Aphis gossypii* Glover. *Comp Biochem Physiol Part D Genomics Proteomics*, 2015, 13: 10–15
 - 34 Pan Y, Yang C, Gao X, et al. Spirotetramat resistance adaption analysis of *Aphis gossypii* Glover by transcriptomic survey. *Pesticide Biochem Physiol*, 2015, 124: 73–80
 - 35 Canavoso L E, Jouni Z E, Karnas K J, et al. Fat metabolism in insects. *Annu Rev Nutr*, 2001, 21: 23–46
 - 36 Dennis E A. Diversity of group types, regulation, and function of phospholipase A2. *J Biol Chem*, 1994, 269: 13057–13060
 - 37 Maczkowiak F, Da Lage J L. Origin and evolution of the Amyrel gene in the α -amylase multigene family of Diptera. *Genetica*, 2006, 128: 145–158
 - 38 Musser R O, Hum-Musser S M, Slaten-Bickford S E, et al. Evidence that ribonuclease activity present in beetle regurgitant is found to stimulate virus resistance in plants. *J Chem Ecol*, 2002, 28: 1691–1696
 - 39 Liu N, Ming L, Gong Y, et al. Cytochrome P450s—Their expression, regulation, and role in insecticide resistance. *Pestic Biochem Physiol*, 2015, 120: 77–81
 - 40 Wang X, Fang X, Yang P, et al. The locust genome provides insight into swarm formation and long-distance flight. *Nat Commun*, 2014, 5: 2957
 - 41 Papanicolaou A, Schetelig M F, Arensburger P, et al. The whole genome sequence of the Mediterranean fruit fly, *Ceratitidis capitata* (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species. *Genome Biol*, 2016, 17: 192
 - 42 Richards S, Gibbs R A, Weinstock G M, et al. The genome of the model beetle and pest *Tribolium castaneum*. *Nature*, 2008, 452: 949–955
 - 43 Xia Q, Guo Y, Zhang Z, et al. Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*). *Science*, 2009, 326: 433–436
 - 44 Feyereisen R. Evolution of insect P450. *Biochem Soc Trans*, 2006, 34: 1252–1255
 - 45 Tao X Y, Xue X Y, Huang Y P, et al. Gossypol-enhanced P450 gene pool contributes to cotton bollworm tolerance to a pyrethroid insecticide.

- [Mol Ecol](#), 2012, 21: 4371–4385
- 46 Stolle E, Kidner J H, Moritz R F A. Patterns of evolutionary conservation of microsatellites (SSRs) suggest a faster rate of genome evolution in Hymenoptera Than in Diptera. [Genome Biol Evol](#), 2013, 5: 151–162
- 47 Miao X X, Xub S J, Li M H, et al. Simple sequence repeat-based consensus linkage map of *Bombyx mori*. [Proc Natl Acad Sci USA](#), 2005, 102: 16303–16308
- 48 Finseth F R, Harrison R G. A comparison of next generation sequencing technologies for transcriptome assembly and utility for RNA-Seq in a non-model bird. [PLoS ONE](#), 2014, 9: e108550
- 49 Benoit J B, Adelman Z N, Reinhardt K, et al. Unique features of a global human ectoparasite identified through sequencing of the bed bug genome. [Nat Commun](#), 2016, 7: 10165
- 50 Chen G, Shi T, Shi L. Characterizing and annotating the genome using RNA-seq data. [Sci China Life Sci](#), 2017, 60: 116–125
- 51 Ran F A, Hsu P D, Lin C Y, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. [Cell](#), 2013, 154: 1380–1389
- 52 Zhang L, Hou D, Chen X, et al. Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. [Cell Res](#), 2012, 22: 107–126
- 53 Weiberg A, Wang M, Lin F M, et al. Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. [Science](#), 2013, 342: 118
- 54 Wang Y M, Ding Y, Yu D W, et al. High-throughput sequencing-based genome-wide identification of microRNAs expressed in developing cotton seeds. [Sci China Life Sci](#), 2015, 58: 778–786
- 55 Zhai Y, Zhang J, Sun Z, et al. Proteomic and transcriptomic analyses of fecundity in the brown planthopper *Nilaparvata lugens* (Stål). [J Proteome Res](#), 2013, 12: 5199–5212
- 56 Xie W, Guo L, Jiao X, et al. Transcriptomic dissection of sexual differences in *Bemisia tabaci*, an invasive agricultural pest worldwide. [Sci Rep](#), 2014, 4: 4088