



基因组时代的昆虫学研究*

叶恭银** 方 琦

(农业部农业昆虫学重点实验室 浙江大学昆虫科学研究所 杭州 310029)

摘 要 昆虫种类繁多,它与生态系统中的生物多样性,以及人类的日常生活和生产密切相关。自2000年黑腹果蝇 *Drosophila melanogaster* 全基因组测序完成以来,至今已先后开展了88种昆虫全基因组测序工作,这标志着昆虫学研究进入了基因组时代。本文综述了近年来昆虫基因组测序进展,以及基于基因组的昆虫学研究方法及应用等两方面的研究成果。同时,着重介绍了昆虫全基因组测序进程,昆虫基因组在个体生物学、多物种间及种群,及系统生物学研究中的应用等方面的内容。最后,还探讨了基因组时代昆虫学研究所面临的挑战。

关键词 昆虫, 基因组测序, 功能基因组学, 比较基因组学

Entomological research in the genomic age

YE Gong-Yin** FANG Qi

(Key Laboratory of Agricultural Entomology of Ministry of Agriculture, Institute of Insect Sciences,
Zhejiang University, Hangzhou 310029, China)

Abstract The many different species of insects on earth comprise a significant proportion of global biodiversity and play important roles in the ecosystem, as well as impacting on human beings. Since the sequencing of the *Drosophila melanogaster* genome was completed in 2000, the genomes of 88 different kinds of insects have now been sequenced. This indicates that entomological research has entered the genomic age. In this review, we summarize recent progress in insect genomic sequencing and review research methods and applications of genome-based entomology. We highlight the process of sequencing insect genomes, as well as the application of sequencing data to individual organism, species, population and systems biology. Finally, we discuss the challenge of the entomological research in the genomic age.

Key words insect, genome sequencing, functional genomics, comparative genomics

1 引言

昆虫种类繁多,估计有数百万种,约占所有动物种类总数的3/4,它对地球生物的多样性组成占有举足轻重的地位。此外,其与人类的日常生活和生产亦有着非常密切的关系。如家蚕、蜜蜂等经济资源类益虫能够为人类提供丰富的生产资料与生活资源。反之,世界上每年因害虫危害造成农作物和牲畜的损失总价约合260亿美元左右。在我国,20世纪90年代棉铃虫的大爆发对棉花生产造成了毁灭性打击,巨大的损失仍让我们记忆

犹新;在非洲,冈比亚按蚊所传播的疟疾肆虐,每年引起数百万人感染并引起大量死亡,但至今仍缺乏有效治疗的手段(赵小凡,2005)。因此,昆虫学研究历来备受世界各国政府及相关研究机构重视,为此投入了大量的人力、物力和财力。自1989年人类基因组计划启动以来,已有多种生物体的基因组全序列被先后完成测定,生物学研究也正式步入了飞速发展的基因组时代。期间,昆虫基因组学的研究也得到了长足的进步与发展。基因组学的深入发展对于学者们进一步了解生物的遗传、进化、系统发育及具体的生物学过程等方面的

* 资助项目:国家杰出青年基金项目(31025021)、浙江省自然科学基金重点项目(Z3090191)。

**通讯作者, E-mail: chu@zju.edu.cn

收稿日期:2011-10-25,接受日期:2011-11-08

信息及本质起到了重要的推动作用,并对从分子水平上改造或改良其固有的遗传背景或性状提供了依据与保障。针对昆虫基因组的深入研究不仅为传统昆虫学科的发展提供了崭新的机遇,而且对深入了解昆虫多样性及其生物学特性与本质具有重大意义(薛建等,2009)。

2 昆虫全基因组测序进展

2.1 昆虫基因组大小预测

在测定昆虫基因组序列之前,首先必须对待测物种的基因组大小进行估计,以便合理地设计测序计划与方案,估算测序所需花费的人力、时间及经济成本。因此,昆虫基因组大小的预测是昆虫基因组测序工作的重要前期准备环节之一。目前,预测昆虫基因组大小的方法较多,除全基因组测序法外,主要还有流式细胞仪测定法、Feulgen 光密度法、Feulgen 图像分析光密度法、甲基绿光密度法、静态细胞荧光测定法、紫外线显微镜法等,其中应用最为广泛的是流式细胞仪测定法(薛建等,2009)。流式细胞技术是一种利用激光束散射与折射原理来对液体中的细胞或其他微小颗粒状物体进行快速分析与分选的技术(Gregory *et al.*, 2003)。利用该技术预测昆虫基因组大小的结果较为准确,可通过提取昆虫细胞中的细胞核,采用染料对其中的 DNA 进行染色,而后对细胞核进行分析,即可测出细胞核中的 DNA 及异染色质含量等信息(Gregory *et al.*, 2003; Nardon *et al.*, 2003)。随着流式细胞仪及相关试剂已大规模商品化生产与应用,流式细胞技术已成为当今预测昆虫基因组大小的最为常用、有效和便捷的方法。在预测昆虫基因组大小时,通常采用黑腹果蝇 *Drosophila melanogaster* 细胞的 DNA 作为测定的参照物,并选取待测昆虫的脑、血细胞或精子等作为测定的组织样品(薛建等,2009)。迄今,已有 700 多种不同种类昆虫的基因组大小通过各种方法得到预测(<http://www.genomesize.com/>),这为昆虫基因组测序工作的顺利开展提供了前提和保障。

2.2 昆虫全基因组测序进程

在昆虫中,最早完成全基因组测序工作的是黑腹果蝇,这是昆虫学乃至生物学研究领域的里程碑式进展。黑腹果蝇也成为了继酿酒酵母

Saccharomyces cerevisiae 和秀丽隐杆线虫 *Caenorhabditis elegans* 之后,第 3 个完成全基因组测序工作的真核生物(Goffeau *et al.*, 1996; *C. elegans* Sequencing Consortium, 1998)。选取黑腹果蝇作为第 1 个用于全基因组测序的昆虫是具有一定理由的:首先,它是生物学及基础医学研究领域中最重要模式生物材料之一;其次,它具有易饲养、生活周期短、繁殖能力强等生物学特性;最后,其基因组较小(仅 4 对染色体),遗传学研究背景清晰,可利用的遗传标记与品系较为丰富。继黑腹果蝇完成全基因组测序之后,冈比亚按蚊 *Anopheles gambiae*、家蚕 *Bombyx mori*、意大利蜜蜂 *Apis mellifera*、埃及伊蚊 *Aedes aegypti*、赤拟谷盗 *Tribolium castaneum*、3 种金小蜂科寄生蜂(*Nasonia giraulti*、*N. longicornis* 及 *N. vitripennis*),以及豌豆蚜 *Acyrtosiphon pisum* 等也相继完成了全基因组测序工作。目前已在完结或正在实施的昆虫基因组测序计划共有 88 个(<http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi>),其中果蝇 *D. virilis* (TSC# 15010 – 1051.88) 及 *D. simulans* (Mosaic) 这 2 个测序计划已完结;另有 51 个昆虫全基因组已初步完成测序工作,目前正在对测序结果进行拼接组装;此外,还有 35 个昆虫的测序工作正在进行中。而与之对应的数据在 2009 年时分别为 1、24 和 28 个(薛建等,2009),这说明在 2009 年至 2011 年期间,昆虫全基因组测序工作得到了飞速发展,这与昆虫基因组学研究在全世界范围内日益备受重视及测序技术的迅猛发展密切相关。在这些全基因组测序计划中,果蝇属的有 31 个,占到总数的 35%;冈比亚按蚊则有 19 个品系在进行全基因组测序,占总数的 22%。从上述数据不难看出,相关研究机构在选择测序昆虫的种类时,具有一定的指向性,即选择重要的模式昆虫进行全基因组测序。此外,值得欣喜的是,重要的农业害虫如豌豆蚜、小菜蛾 *Plutella xylostella* 以及烟粉虱 *Bemisia tabaci* 等也已成功进入了全基因组测序昆虫的大家庭。

在上述这些已进行全基因组测序的昆虫中,已有多数昆虫(如黑腹果蝇、冈比亚按蚊、家蚕、意大利蜜蜂、埃及伊蚊、赤拟谷盗、金小蜂、豌豆蚜等)的基因组计划已完成测序和拼接组装阶段,并发表了相关论文。论文数据表明,黑腹果蝇基因组的大小为 180 Mb,约共编码 13 600 个基因

(Adams *et al.*, 2000)。果蝇细胞核内共含 3 对常染色体和 1 对性染色体(分别标记为 X 和 Y),据粗略估计,其中 2/3 为常染色质,1/3 为异染色质。其中,含 120 Mb 基因组的常染色质位于 2 个较大的 Y 染色体(2 和 3 号)、X 染色体和 4 号 Y 染色体上。异染色质集中于所有染色体的调集区和着丝点附近,其主要包括简单重复序列(如微卫星 DNA)、中度重复元件(如转座子及核糖体 DNA),以及一些单拷贝 DNA。由于重复元件在克隆、绘图及拼接上存在较大困难,因此直到 2007 年异染色质才被测序成功(Hoskins *et al.*, 2007; Smith *et al.*, 2007)。冈比亚按蚊的基因组大小为 278 Mb,其大小比黑腹果蝇约大 100 Mb,该基因组约编码 14 000 个基因。对其免疫相关基因注释的完成及大规模表达序列标签(expressed sequence tag, EST)的发掘,为进一步了解吸血性昆虫的生理适应机制及疟疾的发病机理提供了理论依据(Holt *et al.*, 2003)。家蚕基因组大小为 428.7 Mb,约编码 18 510 个基因。全基因组测序结果表明,家蚕基因组的实际大小要小于 Cot 分析法预测的家蚕单倍体基因组大小,家蚕基因组中所包含的基因数则远多于黑腹果蝇所拥有的基因数(Mita *et al.*, 2004; Xia *et al.*, 2004)。意大利蜜蜂基因组的大小为 220 Mb,约编码 10 157 个基因。蜜蜂基因组中含有近 67% 的 A/T 碱基,该数量明显高于黑腹果蝇(58%)和冈比亚按蚊(56%)。意大利蜜蜂基因组内富含 CpG,同时与黑腹果蝇相比,其中转座子缺失较为严重(Honeybee Genome Sequencing Consortium, 2006)。埃及伊蚊基因组的大小达 1 376 Mb,约编码 167 899 个基因。其基因组大小为同一科的冈比亚按蚊基因组的 5 倍,这与埃及伊蚊基因组中含近 50% 的转座元件有关(Nene *et al.*, 2007)。赤拟谷盗的基因组大小为 204 Mb,约编码 16 404 个基因。赤拟谷盗基因与其他昆虫基因具有正向同源性,研究人员认为它的发育模式比果蝇更具代表性,这可能与其基因的含量与功能相关(*Tribolium* Genome Sequencing Consortium, 2008)。丽蝇蛹集金小蜂基因组大小为 239.8 Mb,约编码 10 734 个基因。与同为膜翅目的意大利蜜蜂不同,丽蝇蛹集金小蜂富含转座元件,其多样性比此前认为转座元件最为丰富的家蚕还多 30%。此外,丽蝇蛹集金小蜂有一个与脊椎动物相似的“DNA 甲基化工具包”(The

Nasonia Genome Working Group, 2010)。豌豆蚜基因组的的大小约为 464 Mb,据估计约编码 35 000 个基因。其中,约有 13 000 个在基因组上存在重复,其功能包括染色质修饰、microRNA 合成,以及糖转运等。基因重复现象对揭示蚜虫所具有的独特生物学特性的分子基础起了关键作用(International Aphid Genomics Consortium, 2010)。

3 基于基因组的昆虫学研究方法及应用

随着越来越多的昆虫进行或完成全基因组测序,现阶段昆虫学研究已经步入基因组时代。与传统昆虫学研究方法相比,基于基因组的昆虫学研究方法已被赋予新的时代特征,同时它也将面临新的挑战。全基因组测序仅仅是一种实验手段,它的完成并不意味着针对测序物种的科学研究的完结。相反,我们应该思考如何利用测序获取的昆虫基因组数据去解决个体生物学问题,如何利用基因组数据去解决多物种或种群学问题,以及解决系统生物学问题(Severson and Behura, 2012, 校对)。

3.1 在个体生物学研究中的应用

正是由于大量昆虫的全基因组测序工作的开展,才造就了如今高通量大规模的转录模式研究,包括全转录组芯片及基于第 2 代高通量短片段读取测序技术的转录组学研究。其中,已完成注释的基因集合为全转录组芯片的探针设计提供了序列信息,全基因组数据也为高通量转录组测序提供了匹配的参考序列。此前,转录组学一直被认为是一种探索性的研究方法,现今却已逐步转变成成为具备高信息通量特点的实用的实验手段。此外,转录组学研究也被认为是基础生物学研究的引领者,尤其对黑腹果蝇、冈比亚按蚊以及家蚕等模式生物而言。例如,采用 Affymetrix 芯片研究人员揭示了冈比亚按蚊肠道及唾液腺生理学的分子基础(Neira Oviedo *et al.*, 2008, 2009),以及胚胎发育早期的耐旱分子机制(Goltsev *et al.*, 2009)。然而随着第 2 代测序技术的飞速发展,及其测序成本的下降,该技术大有取代全转录组芯片技术之势(Marioni *et al.*, 2008; Simon *et al.*, 2009)。与芯片相比,第 2 代测序技术在应用方面更具优势,它既能够被用于差异表达研究,又可以被用来做单核苷酸多态性(single nucleotide

polymorphism, SNP), 以及插入/删除突变分析; 此外, 还可以用于 RNA 可变剪切, 以及基因组中未被注释的新基因的研究 (Wilhelm and Landry, 2009)。例如, 利用 Illumina RNA 测序平台研究人员就埃及伊蚊与冈比亚按蚊的比较功能基因组学与进化基因组学进行了深入研究 (Gibbons *et al.*, 2009)。另外, 随着昆虫全基因组测序计划的不断开展, 昆虫蛋白质组学也已逐渐成为反映和利用基因组信息的手段。随着蛋白质质谱技术的发展 (目前已能对全蛋白质组的混合肽段直接进行鉴定), 蛋白质组学已成为研究昆虫学的有力武器之一 (Mallick and Kuster, 2010)。例如, 结合丽蝇蛹集金小蜂的基因组数据, 研究人员已从该蜂毒液蛋白中鉴定出了 79 种潜在的组分 (de Graaf *et al.*, 2010)。正是得益于全基因组测序计划, 该蜂成为目前毒液蛋白组分被研究的最为透彻的寄生蜂。

3.2 在多物种间及种群研究中的应用

物种生物学最有价值的信息往往来自于该物种不同个体间基因组的系统性比较分析。随着现在进行全基因组测序的昆虫种类不断增多, 不同物种间的比较基因组学研究的重要性更为明显 (Kaufman *et al.*, 2002)。例如埃及伊蚊和冈比亚按蚊的比较基因组学分析可深入揭示该 2 种蚊子在其基因与基因组结构上的进化关系 (Waterhouse *et al.*, 2008)。同样, 针对同一种群中的不同个体进行基于全基因组数据的大规模 SNPs 分析亦能为研究该种群的分化或动态演化提供相关数据与支持。就目前而言, 比较基因组学的主要内容主要包括: 宏观及微观共线性 (macrosyneny and microsyteny); 调控元件, 以及特异性调控元件与基因表达模式之间的关系; 密码子使用的偏好性及其与基因表达水平间的关系, 以及 tRNA 基因的相对丰度; 小 RNA 的组成、及其表达模式及其靶标基因; 线粒体基因组及其核基因插入效率等方面的研究 (Severson and Behura, 2012)。此外, 除了比较基因组学以外, 随着基因组在多物种间或种群研究中的不断应用还衍生出了另一门新兴交叉学科, 即进化基因组学 (phylogenomics), 基于进化学观点的基因组学研究必将成为生命科学研究领域的新亮点。多种昆虫全基因组测序完成将为从进化的角度比较研究昆虫的发育、分化、免疫和

生态适应性等重要基础科学问题提供巨大帮助。随着昆虫基因组序列测定的完成, 大量的 DNA 分子标记得以有效应用于昆虫分子系统学与进化方面的研究 (Munstermann and Conn, 1997; Krzywinski *et al.*, 2001)。此外, 基于全基因组的 ESTs 已被证实可有效应用于转录模式研究, 以揭示功能遗传学及功能基因组学的基础问题, 而且研究人员认为 ESTs 极有可能成为可靠的分子标记系统, 为昆虫生态基因组学及进化基因组学研究奠定基础 (Behura, 2006)。

3.3 在系统生物学研究中的应用

系统生物学是一门新兴的交叉学科, 它的研究内容集中于生物系统内复杂的相互作用, 它主要用于解析在特殊的系统或机制内, 上述相互作用如何能够触发以形成特定的生物学功能与行为。预测基因网络是基于全基因组的系统生物学的中心任务。例如, 利用系统生物学的研究人员已就登革热病毒、埃及伊蚊及人之间所写成的基因网络进行了研究 (Doolittle and Gomez, 2011)。该研究结果表明, 有近 4 000 对相互作用存在于登革热病毒与人之间, 而仅有 176 对相互作用存在于病毒和埃及伊蚊之间。同时, 有关埃及伊蚊用于应答登革热感染的蛋白网络也已被解析, 发现共有 4 214 个蛋白参与应答, 共形成 10 209 对相互作用, 其中 3 500 个蛋白间发生了无标度网络互连 (Guo *et al.*, 2010)。此外, 有关埃及伊蚊糖代谢, 抗登革热病毒时所诱导产生的 Toll 以及 JAK-STAT 等免疫信号转导途径的系统生物学网络已被解析 (Xi *et al.*, 2008; Souza-Neto *et al.*, 2009; Vital *et al.*, 2010)。由此可见, 全基因组测序工作是昆虫开展系统生物学研究的基石, 而系统生物学对基因组学研究范畴提供了很好的延伸。

4 基因组时代昆虫学研究所面临的挑战

4.1 未来发展趋势

在过去的 10 年中, 复杂多细胞生物体的全基因组测序迎来其发展的黄金时期, 而且全基因组测序技术也日臻成熟, 基于全基因组散弹法 (whole genome shotgun) 的测序方法也在昆虫全基因组测序过程中得以应用, 且十分有效 (所有昆虫均采用散弹法, <http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi>)。而在此期间昆虫全基因组测

序工作亦进展十分迅速,昆虫学研究也已步入基因组时代且发展势头良好,有关该方面的研究已经有大量的文献报道。在 2011 年 2 月,综合 *Science* 期刊上为了纪念人类基因组公布 10 周年所发表的所有短文与评论,我们能够得出一个结论:大多数研究人员认为,人类基因组计划对于人类健康的重要性已毋庸置疑。因此,我们认为与人类基因组计划相比,不同种类昆虫的全基因组测序工作的重要性也必将得到昆虫学研究人员的充分认可。此外,随着第 2 代测序技术的不断发展与进步,高通量、高覆盖率的短片段读取测序法必将取代传统的测序方法,随之而来的就是测序速度的加快与测序成本的下降。据推测,在为期不远的将来,测定 1 个人全基因组序列所需要的费用将会下调至 1 000 美金。当然,由于商品化程度上存在巨大差异,1 个昆虫全基因组测序费用未必能降低至人的预期水平,但预计其降幅较大。近期,10 位来自名叫 i5k Ad Hoc Launch Group 组织的科学家在 *Science* 上发表公开信,声明将在未来 5 年内对 5 000 种昆虫或相关节肢动物开展全基因组测序工作,该项目亦被称作“i5k 计划”或“昆虫学的曼哈顿计划”。相对于传统的昆虫全基因组测序计划,i5k 计划具有变革性,因为它主要针对的测序目标是:1) 世界性与农业、食品安全、医药以及能源生产密切相关的昆虫种类;2) 生物学研究领域的重要模式生物;3) 在全世界生态系统中具有很高丰度的昆虫种类;3) 位于昆虫系统发育树上每一进化分支的代表性昆虫种类。如上所述,未来 5~10 年期间,昆虫学将迈入基因组时代的全盛期。当大量昆虫全基因组测序工作完成时,昆虫学研究将步入后基因组时代,即功能基因组时代(<http://arthropodgenomes.org/wiki/i5K>)。

所谓功能基因组学,即在完成昆虫全基因组测序的基础上,针对其上所有编码基因及非编码序列进行鉴定与研究,主要研究内容包括各编码基因的转录、表达及生物学功能,以及非编码序列的功能。针对功能基因组学研究,通常采用 2 种策略,即基因对基因(gene-by-gene)以及基因组范围(genome-wide)筛选法。通常采用的手段主要有:昆虫转基因、昆虫突变体构建、RNA 干扰等方法,其中前 2 项主要适用于黑腹果蝇、埃及伊蚊及家蚕等模式昆虫,而 RNA 干扰技术则既适用于模式昆虫又适用于非模式昆虫(Boutros and

Ahringer, 2008)。此外,基于全基因组的转录组芯片, RNA 深度测序、蛋白质芯片、蛋白质二维电泳以及蛋白质质谱等技术亦对昆虫功能基因组学研究起到一定的辅助作用。昆虫功能基因组学研究与传统的昆虫学研究不同,因此,我们需要将传统的生态学、行为学、生理学及分子生物学的相关研究逐步向更为全面的系统生物学过渡;逐渐将个体生物学、进化学向多物种间或种群内不同个体间的比较基因组学与进化基因组学过渡。

4.2 存在的问题与对策

综合国内外现有的有关昆虫结构基因组学及功能基因组学的研究进展,发现基因组时代的昆虫学研究进展虽然迅速,但也存在以下 4 个主要问题与不足:第一,全基因组测序用的昆虫样本准备较为困难:由于全基因组测序需要纯合的基因组 DNA,因此需对供试昆虫(单倍体昆虫除外)种群进行多代自交继代饲养(一般要求 10~20 代),以确保待测样本高度纯合;而自交饲养多代以后,供试昆虫种群的生活力与繁殖力显著下降,最终可能导致断种;因此,自交继代饲养费时耗力又极具风险性。第二,测序方法选择错误:由于人力、费用及时间等问题,多数研究人员会选用第 2 代高通量短片段读取的方法直接对昆虫的基因组 DNA 进行测序;第 2 代测序方法所产生的片段较短,后续拼接组装难度较大,极易造成测序缺口;此外,昆虫体内时常含有共生微生物且较难去除,此亦会造成待测样品纯合度下降,影响第 2 代测序、拼接与组装的效果。第三,多重视结构基因组草图绘制,忽视遗传标记筛选及遗传图谱构建,从而缺乏绘制精细物理图谱的有效手段。第四,除果蝇、家蚕、丽蝇蛹集金小蜂等少数模式昆虫之外,多数非模式昆虫缺少相应的突变体库及遗传群体,采用正向遗传学手段对功能基因进行定位较为困难;至今昆虫仍缺乏有效的转座系统, RNA 干扰效率不高且效果很不稳定,采用反向遗传学的方法开展功能基因组研究较为困难。

针对基因组时代的昆虫学研究所存在的 4 大问题,特提出以下对策:第一,建立并优化供试昆虫多代自交继代饲养体系,自交 8 代以后,每一代均冻存部分供试昆虫样品,以防断种造成样品损失。第二,建议采用传统建立人工染色体文库与第 2 代测序技术相结合的方法,这样虽然造成测

序费用上升,但它可有效减轻因供试昆虫基因组 DNA 纯合度不够而对测序、拼接及组装的不利影响。第三,建议在开展昆虫全基因组测序工作之前,先初步完成对供试昆虫遗传分子标记筛选及遗传图谱构建。第四,根据已完成全基因组测序工作的昆虫的自身特性,选择合适的功能基因组学研究手段,有序的对其基因组上的各个编码基因及非编码序列开展研究。如果根据上述对策妥善解决所存在之问题,即可深信基因组时代的昆虫学必将迎来更大的发展与突破,并取得更辉煌的新成就。

参考文献 (References)

- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YHC, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Miklos GLG, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies A, de Pablos B, Delcher A, Deng ZM, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferreira S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong FC, Gorrell JH, Gu ZP, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston DA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke ZX, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai ZW, Lasko P, Lei YD, Levitsky AA, Li JY, Li ZY, Liang Y, Lin XY, Liu XJ, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacle JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RDC, Scheeler F, Shen H, Shue BC, Siden-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirkas R, Tector C, Turner R, Venter E, Wang AHH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, Woodage T, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang GG, Zhao Q, Zheng LS, Zheng XQH, Zhong FN, Zhong WY, Zhou XJ, Zhu SP, Zhu XH, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC, 2000. The genome sequence of *Drosophila melanogaster*. *Science*, 287 (5461):2185—2195.
- Behura SK, 2006. Molecular marker systems in insects: current trends and future avenues. *Mol. Ecol.*, 15 (11): 3087—3113.
- Boutros M, Ahringer J, 2008. The art and design of genetic screens; RNA interference. *Nat. Rev. Genet.*, 9 (7): 554—566.
- C. elegans* Sequencing Consortium, 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*, 282(5396):2012—2018.
- de Graaf D, Aerts M, Brunain M, Desjardins C, Jacobs FJ, Werren JH, Devreese B, 2010. Insights into the venom composition of the ectoparasitoid wasp *Nasonia vitripennis* from bioinformatics and proteomic studies. *Insect Mol. Biol.*, 19(Suppl. 1):11—26.
- Doolittle JM, Gomez SM, 2011. Mapping protein interactions between Dengue virus and its human and insect hosts. *PLoS Negl. Trop. Dis.*, 5(2):e954.
- 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.
- Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG, 1996. Life with 6000 genes. *Science*, 274(5287):546, 563—567.
- Goltsev Y, Rezende GL, Vranizan K, Lanzaro G, Valle D, Levine M, 2009. Developmental and evolutionary basis for drought tolerance of the *Anopheles gambiae* embryo. *Dev. Biol.*, 330(2):462—470.
- Gregory TR, Nedved O, Adamowicz SJ, 2003. C-value estimates for 31 species of ladybird beetles (Coleoptera: Coccinellidae). *Hereditas*, 139(2):121—127.
- Guo X, Xu Y, Bian G, Pike AD, Xie Y, Xi Z, 2010. Response of the mosquito protein interaction network to dengue infection. *BMC Genomics*, 11:380.

- Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Loftus B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft CL, Abril JF, Anthouard V, Arensburger P, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Boscus D, Barnstead M, Cai S, Center A, Chaturverdi K, Christophides GK, Chrystal MA, Clamp M, Cravchik A, Curwen V, Dana A, Delcher A, Dew I, Evans CA, Flanagan M, Grundschober-Freimoser A, Friedli L, Gu Z, Guan P, Guigo R, Hillenmeyer ME, Hladun SL, Hogan JR, Hong YS, Hoover J, Jaillon O, Ke Z, Kodira C, Kokoza E, Koutsos A, Letunic I, Levitsky A, Liang Y, Lin JJ, Lobo NF, Lopez JR, Malek JA, McIntosh TC, Meister S, Miller J, Mobarry C, Mongin E, Murphy SD, O'Brochta DA, Pfannkoch C, Qi R, Regier MA, Remington K, Shao H, Sharakhova MV, Sitter CD, Shetty J, Smith TJ, Strong R, Sun J, Thomasova D, Ton LQ, Topalis P, Tu Z, Unger MF, Walenz B, Wang A, Wang J, Wang M, Wang X, Woodford KJ, Wortman JR, Wu M, Yao A, Zdobnov EM, Zhang H, Zhao Q, Zhao S, Zhu SC, Zhimulev I, Coluzzi M, della Torre A, Roth CW, Louis C, Kalush F, Mural RJ, Myers EW, Adams MD, Smith HO, Broder S, Gardner MJ, Fraser CM, Birney E, Bork P, Brey PT, Venter JC, Weissenbach J, Kafatos FC, Collins FH, Hoffman SL, 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science*, 298(5591): 129—149.
- Honeybee Genome Sequencing Consortium, 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature*, 443(7114):931—949.
- Hoskins RA, Carlson JW, Kennedy C, Acevedo D, Evans-Holm M, Frise E, Wan KH, Park S, Mendez-Lago M, Rossi F, Villasante A, Dimitri P, Karpen GH, Celniker SE, 2007. Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science*, 316 (5831): 1625—1628.
- International Aphid Genomics Consortium, 2010. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.*, 8(2):e1000313.
- Kaufman TC, Severson DW, Robinson GE, 2002. The *Anopheles* genome and comparative insect genomics. *Science*, 298(5591):97—98.
- Krzywinski J, Wilkerson RC, Besansky NJ, 2001. Evolution of mitochondrial and ribosomal gene sequences in anophelinae (Diptera: Culicidae): implications for phylogeny reconstruction. *Mol. Phylogenet. Evol.*, 18(3): 479—487.
- Mallick P, Kuster B, 2010. Proteomics: a pragmatic perspective. *Nat. Biotechnol.*, 28(7):695—709.
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y, 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.*, 18(9):1509—1517.
- Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, Kanamori H, Namiki N, Kitagawa M, Yamashita H, Yasukochi Y, Kadono-Okuda K, Yamamoto K, Ajimura M, Ravikumar G, Shimomura M, Nagamura Y, Shin-I T, Abe H, Shimada T, Morishita S, Sasaki T, 2004. The genome sequence of silkworm, *Bombyx mori*. *DNA Res.*, 11(1):27—35.
- Munstermann LE, Conn JE, 1997. Systematics of mosquito disease vectors (Diptera, Culicidae): impact of molecular biology and cladistic analysis. *Annu. Rev. Entomol.*, 42: 351—369.
- Nardon C, Weiss M, Vieira C, Biémont C, 2003. Variation of the genome size estimate with environmental conditions in *Drosophila melanogaster*. *Cytometry A*, 55(1):43—49.
- Neira Oviedo M, Ribeiro JM, Heyland A, VanEkeris L, Moroz T, Linser PJ, 2009. The salivary transcriptome of *Anopheles gambiae* (Diptera: Culicidae) larvae: a microarray-based analysis. *Insect Biochem. Mol. Biol.*, 39 (5/6):382—394.
- Neira Oviedo M, VanEkeris, Corena-Mcleod MD, Linser PJ, 2008. A microarray-based analysis of transcriptional compartmentalization in the alimentary canal of *Anopheles gambiae* (Diptera: Culicidae) larvae. *Insect Mol. Biol.*, 17 (1):61—72.
- Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, Loftus B, Xi Z, Megy K, Grabherr M, Ren Q, Zdobnov EM, Lobo NF, Campbell KS, Brown SE, Bonaldo MF, Zhu J, Sinkins SP, Hogenkamp DG, Amedeo P, Arensburger P, Atkinson PW, Bidwell S, Biedler J, Birney E, Bruggner RV, Costas J, Coy MR, Crabtree J, Crawford M, Debruyne B, Decaprio D, Eiglmeier K, Eisenstadt E, El-Dorri H, Gelbart WM, Gomes SL, Hammond M, Hannick LI, Hogan JR, Holmes MH, Jaffe D, Johnston JS, Kennedy RC, Koo H, Kravitz S, Kriventseva EV, Kulp D, Labutti K, Lee E, Li S, Lovin DD, Mao C, Mauceli E, Menck CF, Miller JR, Montgomery P, Mori A, Nascimento AL, Naveira HF, Nusbaum C, O'leary S, Orvis J, Perteira M, Quesneville H, Reidenbach KR, Rogers YH, Roth CW, Schneider JR, Schatz M, Shumway M, Stanke M, Stinson EO, Tubio JM, Vanzee JP, Verjovski-Almeida S,

- Werner D, White O, Wyder S, Zeng Q, Zhao Q, Zhao Y, Hill CA, Raikhel AS, Soares MB, Knudson DL, Lee NH, Galagan J, Salzberg SL, Paulsen IT, Dimopoulos G, Collins FH, Birren B, Fraser-Liggett CM, Severson DW, 2007. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science*, 316(5832):1718—1723.
- Severson DW, Behura SK, 2012. Mosquito genomics: progress and challenges. *Annu. Rev. Entomol.*, 57:143—166.
- Simon SA, Zhai J, Nandety RS, McCormick KP, Zeng J, Mejia D, Meyers BC, 2009. Short-read sequencing technologies for transcriptional analyses. *Annu. Rev. Plant Biol.*, 60:305—333.
- Smith CD, Shu S, Mungall CJ, Karpen GH, 2007. The release 5.1 annotation of *Drosophila melanogaster* heterochromatin. *Science*, 316(5831):1586—1591.
- Souza-Neto JA, Sim S, Dimopoulos G, 2009. An evolutionary conserved function of the JAK-STAT pathway in anti-dengue defense. *PNAS*, 106(42):17841—17846.
- The *Nasonia* Genome Working Group, 2010. Functional and evolutionary insights from the genomes of three *Nasonia* species. *Science*, 327(5963):343—348.
- Tribolium* Genome Sequencing Consortium, 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature*, 452(7190):949—955.
- Vital W, Rezende GL, Abreu L, Moraes J, Lemos FJ, Vaz Ida S Jr, Logullo C, 2010. Germ band retraction as a landmark in glucose metabolism during *Aedes aegypti* embryogenesis. *BMC Dev. Biol.*, 10:25.
- Waterhouse RM, Wyder S, Zdobnov EM, 2008. The *Aedes aegypti* genome: a comparative perspective. *Insect Mol. Biol.*, 17(1):1—8.
- Wilhelm BT, Landry JR, 2009. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods*, 48(3):249—257.
- Xi Z, Ramirez JL, Dimopoulos G, 2008. The *Aedes aegypti* toll pathway controls dengue virus infection. *PLoS Pathog.*, 4(7):e1000098.
- Xia Q, Zhou Z, Lu C, Cheng D, Dai F, Li B, Zhao P, Zha X, Cheng T, Chai C, Pan G, Xu J, Liu C, Lin Y, Qian J, Hou Y, Wu Z, Li G, Pan M, Li C, Shen Y, Lan X, Yuan L, Li T, Xu H, Yang G, Wan Y, Zhu Y, Yu M, Shen W, Wu D, Xiang Z, Yu J, Wang J, Li R, Shi J, Li H, Li G, Su J, Wang X, Li G, Zhang Z, Wu Q, Li J, Zhang Q, Wei N, Xu J, Sun H, Dong L, Liu D, Zhao S, Zhao X, Meng Q, Lan F, Huang X, Li Y, Fang L, Li C, Li D, Sun Y, Zhang Z, Yang Z, Huang Y, Xi Y, Qi Q, He D, Huang H, Zhang X, Wang Z, Li W, Cao Y, Yu Y, Yu H, Li J, Ye J, Chen H, Zhou Y, Liu B, Wang J, Ye J, Ji H, Li S, Ni P, Zhang J, Zhang Y, Zheng H, Mao B, Wang W, Ye C, Li S, Wang J, Wong GK, Yang H, Biology Analysis Group, 2004. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science*, 306(5703):1937—1940.
- 薛建, 程家安, 张传溪, 2009. 昆虫基因组及其大小. 昆虫学报, 52(8):901—906.
- 赵小凡, 2005. 基因组时代的昆虫学, 国际学术动态, (3):26—27.

蚜虫基因组及功能基因组研究进展*

张柳平^{1,2**} 卢利霞^{1,3**} 刘石娟² 康乐¹ 崔峰^{1***}

(1. 农业虫害鼠害综合治理研究国家重点实验室 中国科学院动物研究所 北京 100101;

2. 曲阜师范大学生命科学学院 曲阜 273165; 3. 东北林业大学生命科学学院 黑龙江 150040)

摘要 蚜虫作为刺吸式昆虫和植物病毒的传播者,已经成为严重威胁农业生产发展的重要害虫之一。近几年随着分子生物学的发展,尤其是基因组测序技术的进步,蚜虫基因组学和功能基因组学取得了重大突破,使我们对蚜虫特殊的生物学特征有了深层次的认识。本文就蚜虫与内共生菌关系、表型可塑性、发育和生殖、系统进化、解毒酶基因家族以及唾液腺方面在基因组和功能基因组水平上的研究进展进行了综述。

关键词 共生菌, 表型可塑性, 发育, 解毒酶, 唾液腺

Advances in research on aphid genomics and functional genomics

ZHANG Liu-Ping^{1,2**} LU Li-Xia^{1,3**} LIU Shi-Juan² KANG Le¹ CUI Feng^{1***}

(1. State Key Laboratory of Integrated Management of Pest Insects & Rodents, Institute of Zoology,

Chinese Academy of Sciences, Beijing 100101, China;

2. College of Life Sciences, Qufu Normal University, Qufu 273165, China;

3. College of Life Sciences, Northeast Forestry University, Heilongjiang 150040, China)

Abstract Aphids are one of most important agricultural pests which not only damage plants by feeding on their fluids but also transmit viruses. Advances in molecular biology, especially genome sequencing technology, have resulted in a better understanding of aphid genomics and functional genomics. This in turn has allowed us to gain a better understanding of the unique biological characteristics of aphids. Here we review progress in research on the relationship between aphids and symbionts, phenotypic plasticity, development, reproduction, phylogenetics, detoxification enzymes and aphid salivary glands.

Key words symbiont, phenotypic plasticity, development, detoxification enzyme, salivary gland

蚜虫类隶属于昆虫纲(Insecta)同翅目(Homoptera)胸喙亚目(Stemorrhyncha),包括2个总科,球蚜总科(Adelgoidea)和蚜总科(Aphidoidea)。蚜虫类是同翅目昆虫中一个较大的类群,世界已知蚜虫种类4700余种(von Dohlen *et al.*, 2006),中国蚜虫类资源丰富,已知1000余种(Qiao and Zhang, 2004)。蚜虫刺吸植物的韧皮部,吸食植物汁液,具有复杂的生活周期、生殖方式(孤雌生殖和有性生殖交替)和表型可塑性,与细菌有严格的共生关系,与寄主植物有严格的特化关系。另外蚜虫还可以传播多种植物病毒从而

对农业生产造成重大损失。蚜虫的这些特性,使其成为众多昆虫学家和进化生物学家研究的对象,以往多集中在蚜虫生物学、生态学和分类学方面的研究。近年来随着分子生物学技术的发展,尤其是基因测序技术的突飞猛进,致使测序成本大幅下降,蚜虫的基因组及功能基因组学方面的研究与日俱增,取得了实质性的进展,本文就当前蚜虫基因组方面取得的研究成果进行综述。

1 蚜虫基因组

随着人类基因组计划的开展,昆虫基因组学

* 资助项目:本研究由中科院重要方向项目(KSCX2-EW-N-5)、公益性行业(农业)科研专项(201103022)。

** 共同第一作者 E-mail: liuping2006_cool@163.com; lulixia0@126.com

*** 通讯作者, E-mail: cuif@ioz.ac.cn

收稿日期:2011-10-06,接受日期:2011-10-25

得到了长足的发展,目前至少 12 种昆虫的基因组已被测序或正在测序,最多的是双翅目昆虫。同翅目昆虫只有豌豆蚜 *Acyrtosiphon pisum* 的基因组完成了全测序,大豆蚜 *Aphis glycines* 基因组部分序列已公布,桃蚜 *Myzus persicae* 的功能基因组研究的较多。

1.1 豌豆蚜 *Acyrtosiphon pisum* 基因组

国际蚜虫基因组联盟 (IAGC, The International Aphid Genomics Consortium) 利用鸟枪测序法成功对豌豆蚜全基因组序列进行测定,该研究成果发表于 2010 年 2 月 23 日美国《*PLoS Biology*》杂志上。豌豆蚜基因组大小约为 464 Mb,综合各种基因预测方法估计有 35 000 个基因,在昆虫中数量最多,其中约有 13 000 个基因在基因组上发生重复,其功能包括染色质修饰、miRNA 合成和糖转运等。大量重复基因,如单性生殖基因、信号传输复制基因和对病毒媒介起重要作用的膜输送相关基因,对揭开蚜虫独特生物学特性的分子基础起到了重要作用。一些具有进化保守性的基因在基因组上发生丢失,如涉及 IMD 免疫通路、硒蛋白的利用、嘌呤补救途径和鸟氨酸循环的基因 (The International Aphid Genomics Consortium, 2010),与免疫相关的基因发生丢失对蚜虫识别微生物及信号传导有一定的危险性 (Gerardo *et al.*, 2010)。

1.2 桃蚜 *Myzus persicae* 功能基因组

Ramsey 等 (2007) 从 16 个已测序的桃蚜 cDNA 文库中建立了 26 669 个基因表达序列标签 (ESTs),建立的桃蚜 cDNA 文库包括有性生殖和无性生殖整个蚜虫虫体的文库以及肠、头部和唾液腺等组织的文库。通过对各 cDNA 文库进行比较,发现有些基因的表达具有组织特异性,有些基因的表达是取食烟草诱导的。此外,还鉴定了 2 423 个桃蚜特有的新基因。通过比较 3 个蚜虫家系的 cDNA 确定了单核苷酸多态性 (SNPs),可作为遗传标记,在某些情况下可以代表蛋白质的功能差异,特别是一个高表达肠道蛋白酶不保守氨基酸的替换可能对于桃蚜取食不同寄主植物的适应性有重要意义。

1.3 大豆蚜 *Aphis glycines* 基因组

Bai 等 (2010) 用 Roche-454 和 Illumina GA-II 2 种方法测定了长度为 2.78×10^8 bp 的大豆蚜

基因组,产生了 19 293 个转录组和 56 688 个基因组序列。从这些数据中,鉴定了 635 个 SNPs 和 1 382 个微卫星标记,而且开发了可能用于鉴定大豆蚜不同生物型的分子标记。此外,还发现了专性内共生菌 *Buchnera aphidicola* 和兼性内共生菌 *Hamiltonella defensa* 的基因序列。

2 蚜虫与共生菌

豌豆蚜与共生菌 *B. aphidicola* APS 是最好的研究昆虫专性共生关系的材料,基因组数据显示,必需氨基酸的生物合成在豌豆蚜和共生菌 *B. aphidicola* APS 之间共享 (Wilson *et al.*, 2010)。氨基酸氮元素的补给是将胺态氮经谷氨酸循环同化为谷氨酸盐,谷氨酸盐循环是一个整合蚜菌共生体氨基酸代谢补充氮的重要资源 (Hansen and Moran, 2011)。豌豆蚜缺乏精氨酸合成能力,其生长所需精氨酸是通过共生菌 *B. aphidicola* APS 产生的,而豌豆蚜具有编码氨基酸合成反应的基因,包括苏氨酸脱水酶和支链氨基酸转氨酶,而这些酶在 *B. aphidicola* APS 中不能编码。豌豆蚜缺乏编码核苷酸磷酸化酶和腺苷脱氢酶的嘌呤循环基因,其共生菌 *B. aphidicola* APS 中的嘌呤代谢基因可以加以补充,而 *B. aphidicola* APS 可通过蚜虫产生的鸟苷酸来满足自身核苷酸的需求 (Ramsey *et al.*, 2010a)。虽然 *Serratia symbiotica* 对寄主的代谢和耐热性有重大影响, Burke 和 Moran (2011) 用 *S. symbiotica* 感染豌豆蚜后,发现其基因表达的变化幅度很小。因此, *S. symbiotica* 对宿主的代谢影响有可能是共生体本身新陈代谢的结果,或宿主基因表达的转录后修饰。

3 蚜虫表型可塑性

Legaei 等 (2010) 结合豌豆蚜基因组高通量测序技术和生物信息分析的方法,确定了豌豆蚜 149 个 miRNA,包括 55 个保守 miRNA 和 94 个新型 miRNA,并研究了 miRNAs 对豌豆蚜不同形态的调节,分析了整个生殖模式转换中 miRNA 的表达,发现 miRNA 是调控蚜虫表型可塑性的重要基础。

Brisson 等 (2010) 研究了蚜虫不同翅形系统表型可塑性的分子机制,蚜虫基因组中存在果蝇属主要的翅发育基因如 *apterous* 和 *decapentaplegic* 的同源基因。Ghanim 等 (2006) 利用 cDNA 芯片,比较了桃蚜有翅成蚜和无翅成蚜的基因表达,发