评述

昆虫学研究进入基因组学时代

张岫^①, 谷少华^①, 李显春^{①②*}^① 中国农业科学院植物保护研究所, 植物病虫害生物学国家重点实验室, 北京 100193;^② Department of Entomology and BIO5 Institute, The University of Arizona, Tucson AZ 85721, USA

* 联系人, E-mail: lxc@email.arizona.edu

收稿日期: 2016-06-29; 接受日期: 2016-08-06; 网络版发表日期: 2016-09-19

国家自然科学基金(批准号: 31401737, 31321004, 31210103921)、北京市科技新星项目(批准号: Z1511000003150118)和北京市青年拔尖个人项目(批准号: 2015000021223ZK29)资助

摘要 自 2000 年果蝇(*Drosophila melanogaster*)基因组公布以来, 昆虫基因组学得到了快速发展并为昆虫学的研究思路和方法带来了革命性转变. 本文介绍了近年来 DNA 测序技术和生物信息学的发展及其对昆虫基因组学研究的促进作用, 综述了模式昆虫、卫生害虫、社会性昆虫和农业昆虫的基因组学研究现状, 并对昆虫基因组学研究的意义和发展趋势进行了总结和展望.

关键词 昆虫, 基因组学, 比较基因组学, DNA 测序, 生物信息学

生命科学是一门古老而又朝气蓬勃的学科, 古往今来人类探索生命本质的脚步从未停止过. 近现代以来, 生命科学研究不断迎来“大发现时代”, 取得了一系列里程碑式的重大突破^[1]. 19 世纪达尔文进化论的提出使人类逐步开始科学地认识生命起源和进化的真像, 细胞学说的问世使生命科学的研究进入微观层面. 20 世纪以来, 生物学经历了由表及里、由浅到深的发展过程, 遗传和基因成为了生命科学的主旋律, 孟德尔遗传定律的重新肯定和摩尔根基因连锁定律的发现催生了遗传学, 肺炎双球菌转化实验使研究者们明确 DNA 才是生命遗传延续的物质基础. 1953 年, DNA 双螺旋模型的提出成为生命科学史的里程碑事件, 标志着现代分子生物学的兴起, 开启了生命科学领域大跨步式前进的时代. 此后, 遗传密码的破译、蛋白质测序、DNA 测序、PCR、基因工

程等技术的革新成为生命科学快速发展的决定性条件, 为后来组学的兴起奠定了坚实基础. 1990 年, “人类基因组计划”(Human Genome Project, HGP)的启动, 开启了基因组学和生物信息技术在生命科学研究中的大规模应用. 随后十几年间, “国际人类基因组单倍体图计划”、“国际千人基因组计划”、“万种微生物基因组计划”、“千种动植物基因组计划”、“万种脊椎动物基因组计划”等项目的实施使基因组学扩展到生物学的各个领域. 与此同时, 测序、生物信息、分子克隆、物质谱、免疫组化等技术快速发展并相互交融, 促成了转录组学、蛋白质组学、RNA 组学、代谢组学、宏基因组学、糖组学、脂质组学、表型组学、互作组学、单细胞组学乃至系统生物学等学科的兴起. 以人类基因组计划为开端、在多种组学分支的带动下, 生命科学的研究进入了组学时代.

引用格式: 张岫, 谷少华, 李显春. 昆虫学研究进入基因组学时代. 中国科学: 生命科学, 2016, 46: 1162–1173

Zhang S, Gu S H, Li X C. Entomology moves into the genomics era. Sci Sin Vitae, 2016, 46: 1162–1173, doi: 10.1360/N052016-00222

1 昆虫学研究的革命——基因组学时代的来临

“组学”一词英文词根表示为“-ome”，指一类个体的系统集合。组学研究之所以受到广泛关注并取得快速发展，正是因为其系统性、高通量的特点，这为生命科学的研究思路与方法带来了巨大变革，生物学研究模式经历了由单基因到全基因组、由解释表象到阐述分子机制、由单一个体到系统规模、由低产出到通量化、由宽泛笼统到精细化个性化的转变。通过生命组学的研究，能在微观层面更全面地挖掘生物体内的基因、蛋白功能、调控网络等信息，宏观上则能更好地阐述系统进化以及生态学问题，组学时代下人类对生命本质的探索必将更加快速和深入。

昆虫种类繁多，数量庞大，分布广泛，不仅与人类社会的生存发展密切相关，而且在地球生态系统中占有重要地位，因此昆虫学的研究始终是生命科学的重要部分。为了深入探索和认知昆虫世界，更好地控制有害昆虫给人类带来的危害，开发有益昆虫的经济效益和生态效益，昆虫学的研究将更加侧重从分子水平和进化角度阐述昆虫丰富多彩的生长发育模式、生物学特性、环境适应性和表观多样性及可塑性，而组学必能在其中发挥重要作用。2000 年，研究人员破译了黑腹果蝇(*Drosophila melanogaster*)的基因组，开创昆虫基因组学研究之先河，组学技术开始广泛运用到昆虫学研究之中。时至今日，已完成或正在实施的昆虫基因组测序项目已达二百多例，基因组学成为昆虫学研究的重要内容和手段之一，标志着昆虫学研究已进入基因组学时代。

2 昆虫基因组学时代来临的技术支撑

昆虫基因组学研究的流程通常是测序-组装-注释-结构基因组学分析-功能基因组学、比较基因组学分析-多组学联合分析(图 1)。组学研究通量化、系统化的特点迫切需要测序、生物信息等技术的支撑。因此，近年来与基因组学相关的 DNA 测序技术和生物信息学得到快速发展，为昆虫基因组学的兴起奠定了坚实的技术基础。

20 世纪 90 年代基因组学兴起时，真核生物基因组测序大都依赖遗传图谱和 BAC, YAC 文库，费时费力且限制了没有遗传学研究基础的非模式生物基因组的研究。全基因组鸟枪法(whole genome shotgun, WGS)策略的提出打破了这一限制，但其可行性却存在较大争议。在此背景下，黑腹果蝇作为昆虫基因组学研究的先驱，其基因组测序对 WGS 策略进行了成功尝试^[2]，为基因组学领域带来了里程碑式的策略革新，极大地加速了“人类基因组计划”等项目的进程。因此昆虫基因组学的研究不仅促进了昆虫学的发展，更革命性地改变了基因组测序的策略，推动了整个基因组学领域的前进。

自 20 世纪 70 年代以来 DNA 测序技术飞速发展，不断向自动化、通量化、低成本化和长读长的方向改进(图 2)，使得研究者们能够在更短时间内完成更高质量的基因组测序。第一代测序以 Sanger 双脱氧链终止原理和荧光标记为基础，具有极高的准确性和相对较长的读长，在基因组学发展初期承担了酿酒酵母(*Saccharomyces cerevisiae*)、秀丽线虫(*Caenorhabditis elegans*)以及人类等诸多基因组的测序任

图 1 基因组研究的一般流程

务^[3], 然而一代测序也有着速度慢、通量低等缺点, 无法满足基因组学快速发展的需要. 2000 年前后, 以高通量为显著特点的第二代测序技术发展起来并实现商业化, 成为近十几年来基因组和转录组研究中的主要测序技术. 二代测序均基于微阵列并行分析技术, 根据具体原理的不同, 其测序平台主要有 3 个: Roche 454 的 GS、Illumina 公司的 Solexa 和 ABI 公司的 SOLiD. 其中 Roche 454 的读长最长, 但通量和成本不占优势; Illumina Solexa 是目前应用最广的二代测序平台, 虽然读长不长, 但通量高、成本低; ABI SOLiD 的读长和通量一般, 但准确性最高^[3]. 由于昆虫基因组通常具有较高的杂合度和较多的重复序列, 因此昆虫基因组 *de novo* 测序一般需要 Roche 454 乃至一代测序与 Illumina 平台结合的测序方案, 以提高后续序列组装的准确性, 而昆虫基因组重测序以及转录组测序中主要应用 Illumina Solexa 或 ABI SOLiD 平台. 二代测序尽管在通量和成本上具有优势, 但仍然存在读长较短、依赖序列扩增、前期建库过程复杂等问题^[4], 显然限制了基因组学研究的发展. 由此, 第三代测序技术应运而生并在近几年快速发展. 三代测序以单分子直接测序和长读长为显著特征, 更适于无参基因组的 *de novo* 测序. 目前最成熟的三代测序平台当属基于单分子实时(single molecule real time, SMRT)测序技术的 PacBio RS, 已被应用到许多

生命组学的研究之中. 例如, 2014 年斯坦福大学研究人员利用该平台检测了人的类淋巴母细胞转录组, 鉴定了 20 个基因的等位基因型以及 12000 个新的可变剪接^[5]; 2015 年 VanBuren 等人^[6]运用 SMRT 技术以 72x 的深度完成了一种耐旱草类 *Oropetium thomaeum* 的全基因组测序, 得到近乎完整的基因组框架图, 其 contig N₅₀ 达 2.4 Mb, 组装出包括端粒、着丝粒以及串联重复等在内的基因组元件, 序列覆盖度达 99% 以上; Greer 等人^[7]利用 SMRT 技术首次在秀丽线虫基因组中检测出大量甲基化位点, 并明确其特殊的甲基化方式——腺嘌呤甲基化, 同时鉴定出腺嘌呤甲基化酶和去甲基化酶, 揭示了线虫中腺嘌呤甲基化在表观遗传方面的可能性. 此外, 其他一些三代测序技术, 如纳米孔单分子技术、非光学显微镜成像测序技术等, 亦受到广泛关注并在快速试验和完善之中^[8-10]. 然而, 由于技术和成本等因素, 三代测序的应用范围仍然有限, 商业化程度不高, 尤其是在昆虫基因组测序中尚未见应用. 可以预见, 随着测序技术的革新以及组学发展的需求, 三代测序在昆虫基因组学领域的运用推广将会很快实现.

随着测序通量的提高和组学研究的深入, 生物数据大量积累, 促使组装、注释等生物信息技术得到飞速发展. 目前针对基因组测序, 根据不同的算法原理, 已开发出了高度自动化、流程化的组装和注释工

图 2 DNA 测序技术的发展

具. 基因组序列组装的算法主要有 3 种: OLC (overlap-layout-consensus)法、extension-based 法和 De Bruijn(Eulerian)图法^[11]. 其中 OLC 算法适于组装长片段, 如 Sanger 或 Roche454 测序得到的 read; extension-based 算法适于短测序 read 的拼接, 但对重复序列和 SNP 较敏感; De Bruijn 图是二代测序数据组装中运用最多的算法, 但对服务器 RAM 要求较高^[12,13]. 基于这些算法开发出来的组装工具多达数十种, 它们各有特点, 用于不同情况的序列拼接. 基因组注释一般基于软件预测和同源比对两种方式. Yandell 和 Ence^[14]将基因组注释分为两步: 第一步是“计算阶段”(computational phase), 即运用注释工具, 结合已有的转录组、ESTs 数据对目标基因组进行比对、识别和预测, 第二步是“注释阶段”(annotation phase), 即运用软件整合和管理注释结果, 并进行质量控制. 随着基因组研究成果的积累以及转录组和分子生物学的发展, 可用于注释的资源愈加丰富, 注释工具亦呈现多样化的发展, 有助于更加深入地挖掘基因组中的信息, 同时后基因组时代的到来也促使注释分析更加侧重于蛋白功能、进化关系、调控网络等方面. 通常的, 基因组注释过程中需要选择多种工具, 将多个来源的注释信息整合, 从而保证注释结果的全面和可信度. 在昆虫基因组, 尤其是越来越多的非模式昆虫基因组学的研究中, 通过与已破译基因组的近缘种的共线性比对或将大量转录组信息比对到待注释基因组上, 能为其基因注释及后续分析提供帮助. 生物信息学的发展促进了昆虫基因组学研究的发展, 越来越丰富的生物信息工具和数据资源使低成本通量化地获得并处理生物数据、开展大规模组学分析成为可能, 为昆虫基因组学时代的到来提供了有力支撑.

3 昆虫基因组学研究现状

2000 年以来, 果蝇、蚊类、蜜蜂(*Apis cerana* Fabricius)、家蚕(*Bombyx mori*)、豌豆蚜(*Acyrtosiphon pisum* Harris)、帝王蝶(*Danaus plexippus*)、小菜蛾(*Plutella xylostella*)、东亚飞蝗(*Locusta migratoria*)、褐飞虱(*Nilaparvata lugens*)等一大批重要昆虫的基因组相继得到破译, 尤其是近 5 年内昆虫基因组测序呈现出爆发式增长的态势. 2012 年前后, 国际昆虫学界发起了宏伟的“i5k 计划”(5000 insect genome

project)^[15]和“1KITE 计划”(1K Insect Transcriptome Evolution)(www.1kite.org), 旨在通过约 5000 种昆虫基因组和 1000 种昆虫转录组的研究, 从组学角度系统揭示昆虫环境适应性和遗传进化机制. 截至 2016 年 2 月, NCBI-genome 数据库中(<http://www.ncbi.nlm.nih.gov/genome/browse/>)已登陆了涉及 15 个目、约 200 个物种的昆虫基因组测序项目, 此外还有部分昆虫物种的基因组研究成果已发表但序列尚未公布, 如大白蚁(*Macrotermes natalensis*)^[16]、竹节虫(*Timema cristinae*)^[17]等. 以下对具有代表性的模式昆虫、卫生害虫、社会性昆虫、农业害虫以及部分具有特殊习性的昆虫的基因组学研究现状进行介绍.

3.1 模式昆虫基因组学

黑腹果蝇作为重要的模式生物, 是首个完成全基因组测序的昆虫物种, 也是继酿酒酵母^[18]和秀丽线虫^[19]之后第 3 个完成基因组测序的真核生物. 该基因组测序首次尝试了全基因组鸟枪法的策略, 使测序进程提前了 1 年. 黑腹果蝇的常染色质区约 120 Mb, 预测得到了 13601 个基因. 通过功能注释, 鉴定了与 DNA 复制、染色体行为、转录和基因调控等相关的蛋白家族, 并与酵母、线虫及脊椎动物进行了比较分析; 糖通透酶(sugar permease)、线粒体载体蛋白和 ATP 结合转运子等跨膜蛋白家族在编码基因中占很大比例, 表明膜蛋白在生命活动中的重要作用; 细胞色素 P450 超家族共鉴定出 90 个基因, 分属 25 个家族, 其中 CYP4 和 CYP6 家族的成员最多, 占 P450 基因总数的 1/2; 在生物合成代谢通路中, 鉴定出 31 种三酰甘油脂肪酶(triacylglycerol lipases)和 32 种二磷酸尿苷糖基转移酶(uridine diphosphate glycosyltransferase, UGT), 分别参与脂类分解和甾醇糖苷合成^[2].

2005 年 Richards 等人^[20]完成第 2 种果蝇 *D. pseudoobscura* 的测序. 2007 年, *Drosophila* Genome Consortium^[21]发布了另外 10 种果蝇(*D. sechellia*, *D. simulans*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. persimilis*, *D. willistoni*, *D. mojavensis*, *D. virilis* 和 *D. grimshawi*)的测序. 这些果蝇生存环境各异, 相应的形态、行为学特性也有很大差别^[22], 比较基因组研究表明, 其在基因组大小、基因数量、转座子分布及密码子使用等方面表现出高度保守的特性, 但在一些蛋白编码基因、非编码 RNA、顺式调节区存在非

中性变异,许多涉及环境互作和生殖的基因进化较快,这可能是不同种类果蝇在生态学、行为学上差异的分子基础。Hahn 等人^[23]借助 12 种果蝇的基因组做了基因家族进化分析,结果表明,尽管基因总数相近,但种间同一基因家族内的扩张或收缩却普遍存在,蛋白编码基因的分化甚至比 microRNA 更普遍,表面上果蝇基因数量的保守掩盖了种间基因快速进化的本质,而这种基因频繁的增减翻覆可能是造成种间形态、生理、代谢上差别的重要原因。Stark 等人^[24]进行了 12 种果蝇基因组中功能元件、RNA 编码基因、miRNA 基因以及转录调控 motif 的鉴定,进一步校正和丰富了果蝇基因组注释结果。

家蚕是昆虫分子生物学和系统进化研究中常用的模式生物,同时也是重要的产丝昆虫和生物反应器。2004 年中国^[25]和日本^[26]的研究人员几乎同时报道了家蚕基因组测序成果,这是中国破译的首个昆虫基因组,对农业昆虫基因组学的研究具有重要参考价值,也使我国在这一领域走在了世界前列。2008 年 International Silkworm Genome Consortium^[27]在 fosmid, BAC 文库测序的基础上,将两个数据库融合,得到 432 Mb 大小的家蚕基因组精细图,并将 87% 的 scaffold 定位到了 28 条染色体上。家蚕基因组中重复序列相对果蝇、按蚊较多,约占 43.6%。注释得到 14623 个基因,其中超过 3000 个为家蚕特有。基因组中特有的 tRNA 和丝胶基因簇的存在,可能与家蚕的产丝功能相关;同时发现了一种新的、起源于细菌的蔗糖酶可能与专食桑叶习性和生物碱解毒代谢有关。2009 年, Xia 等人^[28]对 40 种野生品系家蚕进行了重测序,并与之前的家蚕基因组比对,构建了家蚕单碱基分辨率的遗传变异图谱,据此鉴定出 1600 万个单核苷酸多态性(single nucleotide polymorphisms, SNP),以及诸多插入缺失(InDel)或结构变异(structural variation)。该研究发现驯化品系与野生品系在遗传上有很大不同,但同时二者也保留了较高水平的遗传可变性,这表明家蚕的驯化是一个涉及大量个体的短期事件;同时在丝腺、中肠和精巢中筛出 354 个在家蚕驯化过程中可能起重要作用的基因。另外,对 41 种家蚕或野蚕线粒体基因组的比较和系统进化分析表明,现有的家蚕品系是由中国野蚕品系进化而来的,家蚕祖先的线粒体在进化中没有发生大规模重组事件,且二者种群数量都未经历过膨胀或缩减,但在分化中,驯化品系的线粒体基因 *cytochrome b* 明

显受到正向选择作用^[29]。这些研究更加清晰准确地展示了家蚕的进化历史,解决了关于驯化家蚕起源的争论。

赤拟谷盗(*Tribolium castaneum*)是鞘翅目中重要的储粮害虫,能在极端缺水的环境中生存,危害仓储食品而造成严重的经济损失,同时也是一种通用的模式生物,在遗传学、免疫学等研究领域占有重要地位。2008 年, Richards 等人^[30]报道了赤拟谷盗基因组,得到约 160 Mb 的基因组图谱和 16404 个预测基因,并通过高密度遗传图谱将 90% 的序列定位到 10 个连锁群上。化学感受、细胞色素 P450、利尿激素-抗利尿激素化学受体等基因家族的膨胀体现了赤拟谷盗对仓储环境的适应。此外,赤拟谷盗基因组内包含 3 个与 SID-1 多跨膜蛋白类似的编码基因,该类蛋白在秀丽线虫中与 dsRNA 吸收有关,因此赤拟谷盗体内可能具备系统化的 RNAi 体系。赤拟谷盗是首个完成基因组测序的鞘翅目昆虫,该基因组有助于提高仓储害虫的防治水平,而且作为比果蝇更接近于高等动物的模式生物,赤拟谷盗基因组为高等动物的研究提供了良好的遗传系统。

3.2 卫生害虫基因组学

卫生害虫的基因组学研究主要涉及双翅目蚊虫类以及虱目的人体虱,它们携带多种人类疾病的致病病原,在长期协同进化中形成了良好的共生关系,且生存能力和入侵扩散能力强。基因组学的研究有助于人类深入了解宿主昆虫生物学特性以及病原物与宿主相互适应的机制,为控制疾病传播奠定基础。

冈比亚按蚊(*Anopheles gambiae*)是疟疾的主要传播媒介,与疟原虫具有频繁的互作和专性适应,同时也在进化上影响着病原与动物宿主^[31,32]。作为重要的卫生害虫,冈比亚按蚊是第一个完成解析的蚊虫类基因组,并经多次完善得到染色体级别的精细图。冈比亚按蚊基因组共 278 Mb,预测得到约 14000 个基因,与免疫相关的 36 种凝集素和 24 种含钙黏蛋白结构域的蛋白得到鉴定;与果蝇相比,丝氨酸蛋白酶以及先天免疫相关基因家族的膨胀可能与按蚊的取食行为以及病原-按蚊-动物宿主的频繁互作有关;涉及视觉系统、细胞黏附及糖降解的基因家族膨胀,可能是按蚊频繁搜寻食物行为的分子基础;而转座子、解毒酶系、编码唾液腺成分以及合成分解代谢酶系的基因的过量表达可能与其较强抗药性相关^[33-35]。冈

比亚按蚊有 M 和 S 两个分子亚型, 二者在幼虫行为选择和成虫繁殖行为上出现分化^[36]. 两个种群间的基因交流极其有限, 因此很难凭此解释亚型形成的复杂过程. Lawniczak 等人^[37]通过两个亚型的比较基因组学研究, 观察到两个种群间存在大量异质性的基因组分化并发现了种群内的选择性清除(selective sweep)的证据. 通过选择性清除, 某些优势基因可在种群中相对较快地扩散, 这可能是原始种群产生分化、最终导致两个亚型形成的分子机制. 随后, 研究人员先后报道了达式按蚊(*Anopheles darlingi*)^[38]、中华按蚊(*Anopheles sinensis*)^[39]以及源自非洲、亚洲、欧洲和拉丁美洲的 16 种按蚊的基因组^[40], 研究比较了按蚊属物种在基因组层面上的保守与分化, 通过对比与生殖、免疫应答、杀虫剂抗药性和化学感受机制相关的基因, 揭示了不同按蚊在生存环境、病原-介体-寄主互作以及传毒能力方面的差异分子机制.

埃及伊蚊(*Aedes aegypti*)是登革热和黄热病的主要传播媒介, 为探究其高效携带、传播病原菌的机制, 提高对此类传染病的防治水平, 2007 年 Nene 等人^[41]完成了埃及伊蚊基因组的测序, 得到 1.38 Gb 的基因组和 15419 个预测基因, 其基因组内转座子含量达到 47%, 这是导致伊蚊基因组膨胀的主要原因, 但转座子对伊蚊与按蚊的基因组共线性的影响较小. 埃及伊蚊与冈比亚按蚊在基因密度上有所差别, 气味结合蛋白、细胞色素 P450 及表皮蛋白等基因的拷贝数增加. 基因组大小及基因数目的差异有助于解释二者不同的生物学特性, 例如, 对不同血液偏好性的不同、在选择宿主时的行为差异、传播特定病原体时个体能力的差别等.

致倦库蚊(*Culex quinquefasciatus*)是西尼罗病毒、圣路易斯脑炎病毒以及丝虫病线虫的媒介. 2010 年致倦库蚊基因组测序完成^[42], 结合冈比亚按蚊和埃及伊蚊基因组数据, 为蚊虫比较基因组学的研究提供了平台. 库蚊基因组中预测出 18883 个蛋白编码基因, 与冈比亚按蚊和埃及伊蚊相比, 气味受体、味觉受体、免疫蛋白、解毒相关的谷胱甘肽转移酶、细胞色素 P450 等基因家族显著膨胀, 这可能是致倦库蚊地理分布广泛的重要原因.

体虱(*Pediculus humanus humanus*)是重要的卫生害虫, 能传播斑疹伤寒、战壕热等疾病. 2010 年 Kirkness 等人^[43]等完成了对体虱及其体内 *Riesia pediculicola* 菌(体虱消化道的关键共生菌, 为体虱提

供必需的维生素)的基因组测序. 通过全基因组鸟枪法测序, 得到 108 Mb 的基因组, 预测得到 10773 个蛋白编码基因和 57 个 microRNA. 与其他昆虫相比, 感受与响应环境变化的基因家族显著收缩, 因此体虱在基因组大小和基因数量上均呈现收缩趋势. *Riesia* 菌基因组包括一条线状 DNA 和一个环状质粒, 仅有不到 600 个编码基因, 其中质粒上有独特的维生素 B5(泛酸)合成基因, 对于其寄主体虱非常重要, 因为体虱的食物中缺乏泛酸, 必须通过 *Riesia* 菌获得. 体虱、*Riesia* 菌及体虱携带的致病菌的基因组均比各自的近缘种的要小, 在长期进化中, 三者之间可能通过紧密的互利共生关系, 使自身在基因丢失的情况下能正常生存下去, 实现协同进化.

3.3 社会性昆虫的比较基因组学

社会性昆虫具有高度社会化的组织, 其种群中具有严密的等级分化、劳动分工以及哺育异己后代的行为, 这种特性被称为真社会性(eusociality). 昆虫真社会性的关键特征是社会群体中某一等级的个体高度分化而丧失了其他等级个体的行为能力^[44]. 昆虫真社会性的起源与进化是昆虫学及进化生物学关注的焦点之一^[45], 对社会性昆虫的比较基因组学研究有助于理解昆虫社会行为的演化模式.

西方蜜蜂(*Apis mellifera*)是第一个完成全基因组测序的社会性昆虫. 与果蝇比, 其基因组中大多数细胞信号通路的基因高度保守, 但体内控制生理节奏、衰老、RNA 干扰及 DNA 甲基化的基因相差较大, 反而与人类等脊椎动物的同类基因更为接近. 此外解毒酶、表皮蛋白相关家族的收缩能够解释蜜蜂对杀虫剂的敏感性^[46].

真社会性昆虫主要包括等翅目的白蚁和膜翅目中的蚂蚁、胡蜂(*Paper wasp*)和蜜蜂. 白蚁属等翅目, 与膜翅目社会性昆虫分属不同的进化分支, 但却进化出与蜜蜂、蚂蚁等类似的社会习性, 因此是研究社会性进化的重要材料. 白蚁另一个显著特性在于其种群中存在雄性蚁王, 终生保持可育, 能多次交配, 而膜翅目种群中没有长久可育的雄性个体, 雄蜂(雄蚁)一般只能单次交配^[47]. 2014 年首个白蚁——湿木白蚁(*Zootermopsis nevadensis*)的基因组测序完成^[48]. 与膜翅目社会性昆虫的比较基因组学分析表明, 湿木白蚁基因组中与雄性生殖相关的基因家族显著扩张, 且这些基因在蚁王中是上调表达的, 这可能是蚁

王长期存在和多次交配的原因之一; 化学受体家族成员数量的差异可能与湿木白蚁隐蔽的生活环境有关; 而湿木白蚁与决定个体发育方向的基因家族的数量和表达水平与膜翅目社会性昆虫类似。白蚁基因组的破译为理解白蚁独特的生殖生物学和社会性昆虫的分子进化提供了关键信息, 也有助于人们寻找控制白蚁的新策略。2015 年, Korb 等人^[49]对两种社会性进化复杂度不同的白蚁做了比较基因组学研究。湿木白蚁 *Z. nevadensis* 较原始, 社会等级简单, 所有工蚁均具备全能性, 能发育成各种等级的个体, 而大白蚁在进化上更高级, 具有更为复杂的社会结构, 其工蚁形态高度分化, 且不能转变成具有繁殖能力的蚁后。对通讯、免疫防御、交配以及共生相关基因家族的分析表明, 尽管两种白蚁社会性分化程度差异明显, 然而在多数家族中二者的基因组成没有太大差异, 只是在基因结构和转座子上有所不同, 进而预测转座子、表观调控以及雄性生殖等方面的差异, 可能与社会习性的分化更加相关。

蚂蚁是严格的社会性昆虫。2010 年, Bonasio 等人^[50]对两种蚂蚁 *Camponotus floridanus* 和 *Harpegnathos saltator* 进行了比较基因组学和比较转录组学分析, 发现蚁后、雄蚁、工蚁和兵蚁个体间的基因组成相同, 但它们会因为表观遗传学的变化而发育成为分工等级不同的个体。在 *C. floridanus* 的群体中, 只有蚁后才会产下受精卵; 当该蚁后死亡时, 该蚂蚁的群体也随之死亡, 蚁群中蚁后与其他工蚁、兵蚁之间有着明显的生理和行为差异, 不会发生社会等级的转换; 相反, 在 *H. saltator* 的群体中, 蚁后与工蚁的差别相对较小, 而且在蚁后死亡时, 某个工蚁会转变成蚁后。这种社会等级分化程度的差异可能与 RNA 介导的信号转导和甲基化等基因调节的变化有关。这一研究不仅阐述了两只蚂蚁的不同等级间表观差异以及社会性分化程度的分子机制, 还建立了一个研究衰老和行为的表观遗传学的实验模型。

2011 年, 另一种具有独特习性的蚂蚁——切叶蚁 (*Acromyrmex echinator*) 的基因组测序完成^[51]。切叶蚁的特殊之处在于, 它们能从树木和其他植物上切下叶子, 将叶片用来种植真菌, 并且用长出的真菌为整个蚁群提供食物^[52]。切叶蚁基因组中一组肽酶基因膨胀, 而精氨酸合成与解毒代谢基因有丢失现象, 这与切叶蚁和真菌的特殊共生关系有关。此外,

研究还发现切叶蚁基因组中含有与其他蚂蚁相似的神经肽基因, 且在几种蚂蚁基因组中十分保守, 这些神经肽很可能与蚂蚁社会等级的决定有关。据此推测所有蚂蚁的神经肽内分泌系统可能都含有一个非常类似的构造, 且可能在蚂蚁祖先分化之前就已进化形成。

随着测序技术的发展和测序物种的增加, 大量社会性昆虫基因组的横向比较逐渐成为社会性进化研究的主要手段之一。2013 年, Simola 等人^[53]对 8 种社会性昆虫 (7 种蚂蚁和西方蜜蜂) 以及 22 种独居性昆虫进行了比较基因组学研究, 寻找真社会性昆虫基因组中是否存在特有结构。结果表明, 社会性昆虫基因的表达调控和种系特异性基因与独居性昆虫明显不同, 基因调控的改变是真社会性起源的关键因素, 而种系特异性基因的组成差异则与社会性的种系特异性更加相关。2014 年, Roux 等人^[54]为鉴定蚂蚁适应社会性生活而特有的分子进化模式, 对 7 个已公布的蚂蚁基因组以及部分蜜蜂和果蝇的基因组做了比较, 寻找其中被正向选择的蛋白编码基因, 得出的结果颠覆了之前对于昆虫社会性演化的认知: (i) 与果蝇类似, 蚂蚁、蜜蜂的免疫相关基因在进化中也受到正向选择, 而不像之前认为的, 社会性昆虫由于存在清洁卫生的行为而降低了对免疫基因的选择压; (ii) 膜翅目昆虫中的嗅觉和神经发生基因在其社会性进化之前就开始受正向选择而富集, 即社会习性的产生和进化并非是这些基因家族扩张的起因; (iii) 通过比较发现了蚂蚁特有的基因组进化特征, 最典型的是线粒体功能相关基因受到反复的正向选择。2015 年, Kapheim 等人^[55]比较了 10 种代表不同社会化程度的蜜蜂的基因组, 总结出在从独居到社会性群居转变的过程中, 基因组特征的演化特点: (i) 随着社会性分化程度的提高, 由于选择压较小, 一些重要基因经历了中性突变的进化; (ii) 没有统一的社会性进化途径, 不同蜜蜂的社会性似乎是独立演化而来的; (iii) 尽管不同物种中社会性的进化是相互独立的, 但这些不同的进化过程最终“殊途同归”, 导致了社会性昆虫相似的遗传特性, 如基因调控复杂化、转座现象减少等。随着越来越多的社会性昆虫基因组测序完成, 基因组学、比较基因组学以及表观遗传学的研究将为深入理解昆虫真社会性的演化提供更多线索。

3.4 农业昆虫基因组学

模式物种基因组图谱的公布和后基因组时代的发展,为许多重要农业害虫的基因组学研究提供了可能和便利.近年来,越来越多的害虫基因组得到解析,有助于深入理解害虫发生的分子基础、遗传机制和进化过程,为害虫持续控制提供新思路 and 新技术^[56].

2010 年,豌豆蚜基因组测序完成^[57],这是首个半变态昆虫基因组的破译,也是公布较早的农业害虫基因组.豌豆蚜基因组中存在广泛的基因复制,如单性生殖基因、信号传导基因和对病毒媒介起重要作用的膜运输相关基因等,同时也存在免疫系统 IMD 通路相关基因的缺失事件,这表明蚜虫可能通过放弃对异体微生物的免疫防御能力,使其成功地与各种微生物形成共生.豌豆蚜中涉及新陈代谢,如氨基酸生物合成等的基因并未缺失,而很有可能是与其共生菌 *Buchnera* 相辅相成,构成完整的代谢体系.一些从细菌祖先演化而来的基因在豌豆蚜类菌体中大量表达,可能与调控共生菌有关.该研究首次从基因组水平上阐述了刺吸式农业害虫与其共生菌互利共生、协同进化的分子机制.

自家蚕基因组完成后,我国昆虫基因组学逐步兴起,2009 年至今,我国科学家相继开展了小菜蛾、东亚飞蝗、褐飞虱等多种重要农业害虫的基因组学研究并取得了丰硕成果.2013 年,福建农业大学领衔完成了小菜蛾基因组的破译^[58],其基因组大小约 343 Mb,预测出 18071 个蛋白质编码基因,其中 1412 个是小菜蛾特有的.通过不同种群个体基因组检测,研究人员鉴别了测序小菜蛾 Fuzhou-S 品系的全基因组水平多态性,这可能是小菜蛾具有广泛适应性的重要遗传学基础.结合转录组学的研究发现,小菜蛾中肠内的硫代葡萄糖苷硫酸酯酶 (glucosinolate sulfatase) 和硫酸酯酶修饰因子 1 (sulfatase modifying factor 1) 在幼虫时期的协同表达是决定小菜蛾能够取食十字花科蔬菜的关键.对于小菜蛾强大的植物解毒能力与抗药性,除细胞色素 P450、谷胱甘肽转移酶和羧基酯酶三大水解酶家族外,ATP 结合转运子的扩张可能具有更加重要的意义.小菜蛾基因组的研究,阐明了小菜蛾对植食性的适应以及昆虫与寄主植物协同进化的分子机理,解释了小菜蛾耐药性强的机制,为后续与抗药性和食性等生长发育密切相关基因的功能基因组学和遗传学研究奠定了基础.

东亚飞蝗是重大迁飞性害虫,且存在群居型与散居型两种可相互转换的形态,散居型转为群居型是其聚集起飞、爆发为害的前提^[59].为解释其二型性转变机制,研究人员通过构建 BAC 文库和 WGS 测序,得到约 6.3 Gb 大小的飞蝗基因组框架图,这是迄今已公布的最大的昆虫基因组^[60].飞蝗基因组中有大量的重复序列,其中 DNA 转座子和反转录转座子居多,分别占基因组的 24% 和 17%,且它们的丢失率较低.蝗虫基因组中遍布重复序列的意义可有有俩点: (i) 直翅目在进化上处于比较基部的位点,重复序列受到的进化选择压较小; (ii) 出于适应环境的需要而采取的策略,它们保留一些序列以便保持一定的可塑性,更好地应对环境变化.东亚飞蝗之所以能够对农业造成重大危害,主要原因在于蝗群的长途迁飞能力和为迁飞供能的大量摄食的能力,而两种形态的飞蝗中,群居型的迁飞能力显著强于散居型^[61].通过分析东亚飞蝗基因组信息,发现了多个与脂肪酸代谢相关的基因,这些基因组成了高效的能量供应体系,保证迁飞对能量的需求; UGT 和酯酶 (esterase) 基因家族显著膨胀,这有助于东亚飞蝗快速降解禾本科植物中的次生代谢物,提高摄食量.对飞蝗群居型和散居型个体脑组织进行的转录组和甲基化对比分析,发现二者脑部基因表达具有显著差异,与神经发育和可塑性调控相关的多个基因呈现不同的甲基化模式;与激酶、信号转导蛋白、微管蛋白等相关的基因转录以及可变剪接情况也多有不同.

2014 年,褐飞虱及其 2 种内共生菌代谢互补基因组得到解析^[62].褐飞虱专食水稻 (*Oryza sativa*) 汁液,摄食营养不均衡,且其基因组内许多涉及重要生化途径的基因发生缺失,而其共生菌基因组恰好携带互补基因,在长期进化中与宿主形成互利共生的营养关系,帮助褐飞虱在水稻上爆发成灾.该研究为通过切断褐飞虱-共生菌营养途径防治害虫提供了新视野,也给后续的种群变异、迁飞途径和抗药性的研究提供了分子基础^[63].

3.5 其他昆虫的基因组学

除了以上提到的模式昆虫、卫生害虫、社会性昆虫和农业昆虫,昆虫学研究人员还完成了诸多具有特殊生物学现象的昆虫的基因组解析工作.例如,帝王蝶基因组学及比较基因组学研究揭示了定向迁飞的分子基础,重新明确了迁飞行为的演化模式^[64,65];

摇蚊(*Chironomus tentans*)基因组鉴定了多线染色体相关的 *Balbani ring*(BR)基因和 BR 位点的组织形式^[66]; 南极摇蚊(*Belgica antarctica*)基因组展示了极度紧凑的基因组及其对极端环境的适应^[67]; 竹节虫两种生态型的比较基因组揭示了寄主植物对昆虫平行物种进化的作用^[17]; 榕小蜂(*Ceratosolen solmsi*)基因组揭示了榕小蜂-榕树特殊的协同进化^[68]; 37 种苍蝇的比较基因组学研究揭示了性染色体在双翅目昆虫中的多态性及其核型的演化^[69]; 蛱蝶属 3 个近缘种的比较基因组学研究了与拟态相关的基因组区域的频繁交流, 推测近缘种之间的杂交和基因交流可能在适应性扩散过程中具有重要作用^[70]. 这些昆虫基因组学的研究加深了人们对昆虫中一些独特表型的决定基因及其进化机制的认识.

4 总结与展望

综上所述, 昆虫基因组学的研究使研究者们能够更加系统地发掘昆虫中重要的基因家族、解析分子调控网络和研究昆虫的系统进化, 极大地促进了昆虫学科的发展. (i) 果蝇、家蚕等模式生物的基因组学研究为生物学开辟了新的方向和研究方法, 促进了昆虫分子生物学、遗传进化乃至整个生命科学领域的发展, 提高了人们对昆虫遗传进化和生命本质的认知水平; (ii) 卫生害虫的基因组学研究解释

了诸多卫生害虫的生物学、生理生化、行为和入侵扩散问题, 以及宿主与病原物长期友好共存的现象, 为卫生害虫的治理、疾病的防控和改善人类健康状况提供指导; (iii) 农业昆虫基因组学的研究揭示了重大农业害虫爆发为害的机理, 深化了对昆虫的环境适应性和表观可塑性的认识, 为害虫综合治理提供了新思路和新手段; (iv) 具有特殊生物学习性的昆虫的基因组学研究, 阐述了昆虫独特行为, 如真社会性、迁飞等的分子基础和进化模式, 加深了对昆虫多样性及其演化过程的理解.

近年来, 在组学兴起的时代背景下, 昆虫基因组学的研究取得了诸多令人瞩目的成就, 日新月异的测序技术和生物信息学也持续推动着昆虫基因组学的进步. 未来一段时期内, 昆虫基因组学将在组学的浪潮中延续蓬勃发展的势头, 并侧重功能基因组学和比较基因组学的研究, 着重解决以下问题: (i) 更多具有重大经济意义和生态意义的昆虫的基因组解析; (ii) 近缘物种之间比较基因组学及其演化过程; (iii) 重要表型, 如抗性、迁飞、共生等的控制基因及基因家族; (iv) 调控元件以及非编码 RNA 的调控机制等. 与此同时, 各种组学, 如转录组学、蛋白质组学、代谢组学、表观遗传组学等, 以及基因芯片、免疫沉淀、基因编辑等新技术, 将进一步与昆虫基因组学相互融合, 推动昆虫学研究进入新的纪元.

参考文献

- 1 贺福初. 大发现时代的“生命组学”(代序). 中国科学: 生命科学, 2013, 43: 1-15
- 2 Adams M D, Celniker S E, Holt R A, et al. The genome sequence of *Drosophila melanogaster*. Science, 2000, 287: 2185-2195
- 3 郝甜甜, 李强飞, 李国治, 等. 测序技术的研究进展. 畜牧与饲料科学, 2014, 35: 78-81
- 4 韦贵将, 邹秉杰, 陈之遥, 等. 新一代测序技术的研究进展. 现代生物医学进展, 2012, 12: 3789-3793
- 5 Hagen T, Fabian G, Donald S, et al. Defining a personal, allele-specific, and single-molecule long-read transcriptome. Proc Natl Acad Sci USA, 2014, 111: 9869-9874
- 6 VanBuren R, Bryant D, Edger P, et al. Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. Nature, 2015, 527: 508-511
- 7 Greer E L, Blanco M A, Gu L, et al. DNA methylation on N6-adenine in *C. elegans*. Cell, 2015, 161: 868-878
- 8 Clarke J, Wu H C, Jayasinghe L, et al. Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol, 2009, 4: 265-270
- 9 Bolisetty M T, Rajadinakaran G, Graveley B R. Determining exon connectivity in complex mRNAs by nanopore sequencing. Genome Biol, 2015, 16: 1-12
- 10 Hiroiyuki T, Tomoji K. Partial sequencing of a single DNA molecule with a scanning tunneling microscope. Nat Nanotechnol, 2009, 4: 518-522
- 11 Schlebusch S, Illing N. Next generation shotgun sequencing and the challenges of *de novo* genome assembly. S Afr J Sci, 2012, 108: 62-70

- 12 Robert E, Wolf J B W. A field guide to whole-genome sequencing, assembly and annotation. *Evol Appl*, 2014, 7: 1026–1042
- 13 Miller J R, Koren S, Sutton G. Assembly algorithms for next-generation sequencing data. *Genomics*, 2010, 95: 315–327
- 14 Yandell M, Ence D. A beginner's guide to eukaryotic genome annotation. *Nat Rev Genet*, 2012, 13: 329–342
- 15 Robinson G E, Hackett K J, Purcell-Miramontes M, et al. Creating a buzz about insect genomes. *Science*, 2011, 331: 1386
- 16 Poulsen M, Hu H F, Li C, et al. Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proc Natl Acad Sci USA*, 2014, 111: 14500–14505
- 17 Víctor S C, Zachariah G, Comeault A A, et al. Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, 2014, 344: 738–742
- 18 Goffeau A, Barrell B G, Bussey H, et al. Life with 6000 Genes. *Science*, 1996, 274: 546–567
- 19 The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*, 1998, 282: 2012–2018
- 20 Richards S, Liu Y, Bettencourt B R, et al. Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and *cis*-element evolution. *Genome Res*, 2005, 15: 1–18
- 21 *Drosophila* 12 Genomes Consortium. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*, 2007, 450: 203–218
- 22 Markow T A, O'Grady P M. *Drosophila* biology in the genomic age. *Genetics*, 2007, 177: 1269–1276
- 23 Hahn M W, Han M V, Han S G. Gene family evolution across 12 *Drosophila* genomes. *PLoS Genet*, 2007, 3: 2135–2146
- 24 Stark A, Lin M F, Kheradpour P, et al. Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature*, 2007, 450: 219–232
- 25 Xia Q, Zhou Z, Lu C, et al. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science*, 2004, 306: 1937–1940
- 26 Mita K, Kasahara M, Sasaki S, et al. The genome sequence of silkworm, *Bombyx mori*. *DNA Res*, 2004, 11: 27–35
- 27 The International Silkworm Genome Consortium. The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochem Mol Biol*, 2008, 38: 1036–1045
- 28 Xia Q, Guo Y, Zhang Z, et al. Complete resequencing of 40 genomes reveals domestication events and genes in silkworm. *Science*, 2009, 326: 433–436
- 29 Li D, Guo Y, Shao H, et al. Genetic diversity, molecular phylogeny and selection evidence of the silkworm mitochondria implicated by complete resequencing of 41 genomes. *BMC Evol Biol*, 2010, 10: 81
- 30 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
- 31 Kwiatkowski D P. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet*, 2005, 77: 171–192
- 32 Cohuet A, Harris C, Robert V, et al. Evolutionary forces on Anopheles: what makes a malaria vector? *Trends Parasitol*, 2010, 26: 130–136
- 33 Holt R A, Subramanian G M, Halpern A, et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science*, 2002, 298: 129–149
- 34 Mongin E, Louis C, Holt R A, et al. The *Anopheles gambiae* genome: an update. *Trends Parasitol*, 2004, 20: 49–52
- 35 Sharakhova M V, Hammond M P, Lobo N F, et al. Update of the *Anopheles gambiae* PEST genome assembly. *Genome Biol*, 2007, 8: R5
- 36 Torre A, Fanello C, Akogbeto M, et al. Molecular evidence of incipient speciation within *Anopheles gambiae* ss in West Africa. *Insect Mol Biol*, 2001, 10: 9–18
- 37 Lawnczak M K N, Emrich S J, Holloway A K, et al. Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science*, 2010, 330: 512–514
- 38 Marinotti O, Cerqueira G C, de Almeida L G, et al. The Genome of *Anopheles darlingi*, the main neotropical malaria vector. *Nucleic Acids Res*, 2013, 41: 7387–7400
- 39 Dan Z, Zhang D, Ding G, et al. Genome sequence of *Anopheles sinensis* provides insight into genetics basis of mosquito competence for malaria parasites. *BMC Genomics*, 2014, 15: 73–78
- 40 Neafsey D E, Waterhouse R M, Abai M R, et al. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes. *Science*, 2015, 347: 1258522–1258522
- 41 Nene V, Wortman J R, Lawson D, et al. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science*, 2007, 316: 1718–1723
- 42 Arensburger P, Megy K, Waterhouse R M, et al. Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. *Science*, 2010, 330: 86–88

- 43 Kirkness E F, Haas B J, Sun W, et al. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc Natl Acad Sci USA*, 2010, 107: 12168–12173
- 44 Crespi B J, Yanega D. The definition of eusociality. *Behav Ecol*, 1995, 6: 109–115
- 45 牛泽清, 吴燕如, 黄大卫. 真社会性生活方式在隧蜂亚科中的分布格局. *应用昆虫学报*, 2006, 43: 578–582
- 46 The Honeybee Genome Sequencing Consortium. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature*, 2006, 443: 931–949
- 47 Boomsma J J, Baer B, Heinze J. The evolution of male traits in social insects. *Annu Rev Entomol*, 2005, 50: 395–420
- 48 Terrapon N, Li C, Robertson H M, et al. Molecular traces of alternative social organization in a termite genome. *Nat Commun*, 2014, 5: 536–538
- 49 Korb J, Poulsen M, Hu H, et al. A genomic comparison of two termites with different social complexity. *Front Genet*, 2015, 6: 9
- 50 Bonasio R, Zhang G, Ye C, et al. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science*, 2010, 329: 1068–1071
- 51 Nygaard S, Zhang G, Schiøtt M, et al. The genome of the leaf-cutting ant *Acromyrmex echinator* suggests key adaptations to advanced social life and fungus farming. *Genome Res*, 2011, 21: 1339–1348
- 52 Schultz T R, Brady S G. Major evolutionary transitions in ant agriculture. *Proc Natl Acad Sci USA*, 2008, 105: 5435–5440
- 53 Simola D F, Wissler L, Donahue G, et al. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Res*, 2013, 23: 1235–1247
- 54 Roux J, Privman E, Moretti S, et al. Patterns of positive selection in seven ant genomes. *Mol Biol Evol*, 2014, 31: 1661–1685
- 55 Kapheim K M, Pan H, Li C, et al. Genomic signatures of evolutionary transitions from solitary to group living. *Science*, 2015, 348: 1139–1143
- 56 彭露, 何玮毅, 夏晓峰, 等. 基因组学时代害虫治理的研究进展及前景. *应用昆虫学报*, 2015, 1: 1–22
- 57 International Aphid Genomics Consortium. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol*, 2010, 8: e1000313
- 58 You M, Yue Z, He W, et al. A heterozygous moth genome provides insights into herbivory and detoxification. *Nat Genet*, 2013, 45: 220–225
- 59 赵春晓, 崔为正. 飞蝗的型变及其调控机理. *中国植保导刊*, 2006, 26: 11–13
- 60 Wang X, Fang X, Yang P, et al. The locust genome provides insight into swarm formation and long-distance flight. *Nat Commun*, 2014, 5: 2957–2965
- 61 刘辉. 两型东亚飞蝗飞行能力及其相关生理机制研究. 硕士研究论文. 北京: 中国农业科学院, 2007
- 62 Xue J, Zhou X, Zhang C X, et al. Genomes of the rice pest brown planthopper and its endosymbionts reveal complex complementary contributions for host adaptation. *Genome Biol*, 2014, 15: 521
- 63 张传溪. 中国农业昆虫基因组学研究概况与展望. *中国农业科学*, 2015, 17: 3454–3462
- 64 Zhan S, Merlin C, Boore J L, et al. The monarch butterfly genome yields insights into long-distance migration. *Cell*, 2011, 147: 1171–1185
- 65 Zhan S, Zhang W, Niitpöld K, et al. The genetics of monarch butterfly migration and warning colouration. *Nature*, 2014, 514: 317–321
- 66 Kutsenko A, Svensson T, Nystedt B, et al. The *Chironomus tentans* genome sequence and the organization of the *Balbani ring* genes. *BMC Genomics*, 2014, 15: 232–233
- 67 Kelley J L, Peyton J T, Fiston-Lavier A S, et al. Compact genome of the *Antarctic midge* is likely an adaptation to an extreme environment. *Nat Commun*, 2014, 5: 4611–4611
- 68 Xiao J H, Yue Z, Jia L Y, et al. Obligate mutualism within a host drives the extreme specialization of a fig wasp genome. *Genome Biol*, 2013, 14: 1–18
- 69 Vicoso B, Bachtrog D. Numerous transitions of sex chromosomes in Diptera. *PLoS Biol*, 2015, 13: e1002078
- 70 Dasmahapatra K K, Walters J R, Briscoe A D, et al. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, 2012, 487: 94–98

Entomology moves into the genomics era

ZHANG Shen¹, GU ShaoHua¹ & LI XianChun^{1,2}

1 State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing 100193, China;

2 Department of Entomology and BIOS Institute, The University of Arizona, Tucson AZ 85721, USA

Since the genome of *Drosophila melanogaster* was completed and published in 2000, insect genomics has achieved rapid development, which in turn has revolutionized the philosophy and methodology of entomological research. This review introduces the recent progress in DNA sequencing and bioinformatics that makes insect genomics develop more rapidly, or only possible. This review also summarizes the findings and progress made in the genomics of model insects, medical insects, social insects and agricultural insects and briefly discusses the future prospects of insect genomics.

insect, genomics, comparative genomics, DNA sequencing, bioinformatics

doi: 10.1360/N052016-00222