

科技前沿

蜜蜂全基因组出笼前后

黄智勇^{*}

(美国密执根州大学昆虫系 48824)

The honey bee genome: the untold stories. Zachary Huang (*Department of Entomology, Michigan State University, East Lansing, MI 48824, USA*)

Abstract For honey bee biologists around the world, the “post-genomic era” started with the publication of the honey bee genome in the British journal *Nature* on Oct 26, 2006. Costing about 8 million US dollars, and about two years to complete, this is a major undertaking by the Baylor College of Medicine. The actual annotation of the genome was done by 170 scientists representing 15 countries and 64 laboratories and took almost another two years. The waiting is worthwhile; finally with the genome published, any biologists with internet access can do their own “data-mining” and test their own hypotheses using the publicly accessible resources. In fact, more than half of the 50 or so papers published the same week as the *Nature* bee genome paper in the companion journals, which include *Science*, *Proceedings of National Academy of Sciences*, *Genome Research* and *Insect Molecular Biology*, are based on bioinformatics. Honey bees share many common features as humans: division of labor, elaborate communication systems including languages, guarding and defense systems, architecture and finally self sacrifice to defend their common good. The publication of honey bee genome will enable scientists to learn about honey bee biology as well as our own biology, because of these similarities.

Key words honey bee, genome, sequencing

摘 要 在 2006 年 10 月 26 日, 期待已久的蜜蜂的基因组序列在英国杂志《自然》上发表了。这一事件对于全世界蜜蜂生物学家来说, 标志着蜜蜂研究“后基因组时代”的开始。这一重大项目由倍乐医学院完成, 耗资 800 万美元, 历时 2 年。而实际上对于基因组的注释又经过近 2 年的时间, 由代表 15 个国家、来自 64 个实验室的 170 名科学家完成。这一等待是值得的: 基因组序列最终发表后, 生物学家通过互联网能做他们自己的“数据采掘”, 用这一可得的公共资源验证自己的假说。事实上, 在蜜蜂基因组文章在《自然》上发表的当周, 其姊妹刊上发表的约 50 篇相关论文中, 有一半以上是基于生物信息学的研究, 这些刊物包括《科学》、《基因组研究》、《美国国家科学院院报》及《昆虫分子生物学》。蜜蜂与人类享有一些共性: 劳动分工, 复杂的通信系统包括语言, 发达的保卫和防御系统, 精妙的建筑以及为了共同利益自我牺牲的特点。由于有这些共性, 蜜蜂基因组的发表将使科学家不仅能够深入研究蜜蜂生物学, 而且能深入了解我们人类自身的生物学。

关键词 蜜蜂, 全基因组, 测序

10 月 26 日是一个不平常的日子。1906 年 10 月 26 日, 德国的一个养蜂家约翰·齐尔仲 (Johann Dzierzon) 在 96 岁的高龄逝世。齐尔仲当过神父, 但他最大的发现是证明蜜蜂中蜂王产的未受精卵可以发育成雄蜂, 这种现象被称为产雄单性生殖 (arrhenotoky)。在齐尔仲逝世

后刚好 100 周年, 2006 年 10 月 26 日, 蜜蜂的全基因组在《自然》杂志发表^[1], 从而使蜜蜂生物学也进入所谓的“后基因组”时代。尽管国内的

^{*}E-mail: bees.msu@gmail.com.cn

收稿日期: 2006-12-20, 接受日期: 2006-12-22

媒体大都宣称蜜蜂的基因组是昆虫中的第 3 个,事实上蜜蜂是继果蝇、冈比亚按蚊、家蚕、果蝇后的第 5 种昆虫(表 1A)。另外还有 20 多种昆虫的全基因也还在测序中(表 1B)。

表 1 不同昆虫基因组测定情况^[3]

中文	拉丁名	基因组大小 (千碱基)	完成年份
A: 已完成			
果蝇	<i>Drosophila melanogaster</i>	312	2000
冈比亚按蚊	<i>Anopheles gambiae</i>	264	2002
家蚕	<i>Bombyx mori</i>	530	2004
果蝇	<i>Drosophila pseudoobscura</i>	139	2005
蜜蜂	<i>Apis mellifera</i>	262	2006
伊蚊	<i>Aedes aegyptii</i>	1 310	2005
(未发表)			
B: 未完成			
一种非州蝴蝶	<i>Biocyclus anynana</i>	500	进行中
豌豆蚜	<i>Acyrtosiphon pisum</i>	525	进行中
淡色库蚊	<i>Culex pipiens</i>	540	进行中
果蝇(10种)	<i>Drosophila</i>	~ 135	进行中
白蛉(2种)	<i>Lutzomyia longipalpis</i> <i>Phlebotomus papatasi</i>	170 ~ 300	进行中
采采蝇	<i>Glossina morsitans</i>	590	进行中
金小蜂	<i>Nasonia vitripennis</i> <i>N. giraultii</i> , <i>N. longicornis</i>	345	进行中
人虱	<i>Pediculus humanus</i>	107	进行中
吸血蜡象	<i>Rhodnius prolixus</i>	670	进行中
赤拟谷盗	<i>Tribolium castaneum</i>	158	进行中

1 要测序,先竞标

故事还要回到 2002 年,在美国,除了私营公司自己掏钱的除外,所有公费的生物基因组测序都要经过“国立人类基因组研究所”(NHGRI,美国国立卫生研究下属的一个所)的评估再“排队”(prioritize)。2002 年得到优先排队的有家鸡、大猩猩、几种真菌、海胆和蜜蜂。每种竞标的生物都必须写一个类似于课题计划一样的“白皮书”。蜜蜂的基因组白皮书是由吉因·罗宾森(Gene Robinson)为首的 9 个作者撰写而成^[2]。罗宾森不仅在蜜蜂基因组的测定中有着举足轻重的作用,而且还首先建立了蜜蜂表达序列标签文库,并提出社会基因组学(sociogenomics)的概念,同时他还频频在《科学》和《自然》杂志上发表蜜蜂基因表达的文章。

2005 年,他成为美国科学院惟一专门研究蜜蜂的院士。

白皮书认为蜜蜂的全基因组测序可以有以下好处:

1.1 给人类健康带来益处

白皮书认为蜂巢内的高温(35℃),高湿(95%)及高密度(相当 24 m² 的地方住上 15 个成人),是助长疾病的温床,但蜜蜂虽然有不少病害,仍顽强生存并繁衍昌盛,说明有不少抵抗病菌的机制。蜜蜂是如何抵抗细菌性病原的?蜜蜂又是通过何种机制来抑制大蜂螨的繁殖?科学家认为通过将蜜蜂基因组和其它昆虫(如蚊子和果蝇)的相比较,也许可以找出分子机制,从而增加人类对付疾病的知识。例如,在蜜蜂的蜂毒中,已发现有些成分可以治疗人的癌症,改善记忆、睡眠等等。也许通过基因组能找到蜂毒的其它成分,对寻找新的生物药物和治疗对蜂毒的过敏,都有好处。人类有的精神病,例如“自闭症”,与社会交往的异常有关。而蜜蜂是高度社会化群体,若对其行为进行深入研究,基因组的测定将会对以后基因如何影响社会行为的研究提供重要工具。蜜蜂还可以作为探测环境(农药、放射性元素或炸药)的生物探针,而基因组序列可以将嗅觉受体的基因找出来,帮助蜜蜂成为更好的生物探测器。

1.2 增加对人类社会生物学的了解

蜜蜂社会的复杂性与人类社会的相比,有许多相似之处,如:发达的通信系统(包括语言),由社会调控的行为发育^[4],社会分工,发达的防御与战争手段,复杂的建筑,以及为了群体的利益而牺牲自我的特性,这些性状大多都是本能(不是通过学习来的)。因为很多实验不能在人类中实行,因此用蜜蜂作为工具来研究基因与本能的关系就方便得多。蜜蜂有很强的学习能力,最近甚至发现可以有“类似”和“不同”的概念。除了众所周知的舞蹈语言外,守卫蜂还可以辨别同巢蜂和外巢蜂。尽管蜜蜂脑子的神经元只比果蝇的多 4 倍,但其认知能力却可与脊椎动物相媲美。通过研究蜜蜂的行为复杂性,神经元的个数及基因数量之间的关

系,可以增加对人类认知的了解,并且可以对治疗神经退化方面的疾病有所帮助。蜜蜂还可以用来研究老化与长寿的机制。因为工蜂与蜂王的基因型相同,但由于幼虫时的食物不同,引起基因表达的变化,造成蜂王的寿命比工蜂长 100 多倍。可见,研究蜂王长寿的分子机制可以帮助人类提高寿命。最新的研究^[9]也确实发现蜂王中一个关键的抗氧化物基因, *mnSOD*, 比工蜂的表达高 10 倍以上。

2 测序: 800 万美金, 3 个月

2003 年 1 月国立人类基因研究所批下了 690 万美金, 美国农业部又加上 75 万。具体的测序工作交给德克萨斯州的倍乐医学院 (Baylor College of Medicine) 完成。由于工作大多由机器人完成, 倍乐当时说 3 个月即可以完成测序工作。但是蜜蜂基因中的 AT (A=腺嘌呤, T=胸腺嘧啶) 非常多, 小片段随机克隆到大肠杆菌中的基因可能 AT 丰富区都被分解, 因此基因序列出来后有很多缺失。大片段的 BAC 克隆文库中的基因也不全, 这个问题到最后都未解决, 所以这些克隆最后用的很少 (~20%)。80% 以上的基因是用完整基因组霰弹枪式测序法 (genome shotgun) 来测的。2003 年 12 月 17 日, 第 1 版的基因汇编发表, 刚好是测序开始后 1 年。第 2 版加上了第 1 版缺失的 AT 丰富区, 发表于 2005 年 1 月。第 3 版和第 4 版的基因汇编分别发表于 2005 年 5 月和 2006 年 3 月。主要的测序工作基本上在 2004 年底完成。花了 2 年的时间完成主要原因是倍乐还有别的基因组要测, 加上 BAC 的问题, 实际工作时间也就是 3 个月左右。总共测了 300 多万条核酸序列, 它们有 1 亿 8 千万个碱基, 蜜蜂的基因组为 236 000 个, 因此覆盖率为 7.5 倍。估计有 26 000 个碱基没有被测到。

3 更难的工作: 基因注释

基因序列出来后, 并不是工作就完成了, 首先是用不同电脑软件将这些随机的小序列片段组成重叠序 (bins of reads), 再构建成叠连群

(contig), 然后连成基因架 (scaffolds), 再根据染色体上的标记基因将这些基因架放在染色体的某一位置上, 成为超基因架 (superscaffolds)。这些工作完成后, 将这些数据放在网上, 由各行的科学家来注释各种基因。所谓“注释”就是根据基因的同源性, 将基因组中的序列与已知别的生物 (如果蝇或蚊子) 基因序列比较, 找出垂直同源的基因 (ortholog) 来。也就是说原来看似随机的一段 DNA, 可以知道哪一段是某一个基因。用不同生物的垂直同源基因可以建成系统树, 从而知道它们间的进化关系。2003 年底开始, 有来自世界各地的 63 个单位的 112 个专家分成若干小组, 着手注释与抗微生物、免疫、发育和代谢、基因调控等有关的基因。当时本文作者曾答应负责注释与保幼激素合成和分解的基因, 但后来由于时间原因没有参与。

4 蜜蜂全基因揭晓

从 2003 年底第 1 版全基因序列出来, 到 2005 年初第 2 版问世, 加上向《自然》杂志第 1 次投稿时遭到拒稿, 到最后 2006 年 10 月 26 号最终发表时, 可谓“饱经风霜”。最后在《自然》署名的作者共有来自 15 个国家, 64 个单位的 170 名科学家。加上《科学》、《基因组研究》、《美国国家科学院院报》及《昆虫分子生物学》各杂志的文章, 共有 50 余篇文章, 可以说是历史上头一次同时出这么多基因组的文章, 以下是《科学》杂志文章^[1]的主要结论:

4.1 蜜蜂的 A 和 T 比别的昆虫基因组要高得多, 为 67%, 而黑尾果蝇 *Drosophila melanogaster* 为 58%, 库蚊为 56%。与脊椎动物相反, 在蜜蜂基因组中 AT 丰富区中基因分布反而较多。

4.2 蜜蜂中的转座子基因 (可用在寄主染色体上跳跃; 引起突变的基因) 比别的昆虫少。

4.3 蜜蜂基因数为 10 157 个基因, 比果蝇和库蚊少 25% 左右。

4.4 垂直同源基因数: 蜜蜂和果蝇只有 10% 同源, 这比人和鸡之间共同的 85% 要少, 说明昆虫的进化速度较快。

4.5 人和蜜蜂与人和果蝇, 哪 2 个关系更近?

比较 3 种生物的同源基因后发现人和蜜蜂占基因的共同率为 47.5%，人和果蝇的为 44.5%，人和蚊子的为 46.6%。作者认为蜜蜂的基因可能进化很慢，而双翅目昆虫进化很快，从而造成这种现象。有趣的是，盖·布洛克(Guy Bloch)等也发现蜜蜂的分子钟系统竟更像人类的^[9]，而与其它昆虫有很大不同。

4.6 蜜蜂的气味受体基因比别的昆虫多，与果蝇(62)和库蚊(79)相比，蜜蜂有 163 个气味受体基因。蜜蜂有很敏锐的化学感受能力，用来探测外激素，辨别亲属，辨别花香等。有趣的是这些基因的数量与脑中触角叶(antennal lobe)的嗅小体(glomeruli)的数目(160~170)很相近。与此相反的是蜜蜂的味觉基因(10 个)比其它昆虫(50~76)少得多。

4.7 蜜蜂的免疫和抗病基因。与白皮书所预见的相反，蜜蜂的免疫和抗病有关的基因数目不是提高，而是降低了。库蚊有 209 个，果蝇有 196 个，而蜜蜂只有 71 个与免疫和抗病的基因。有可能因为蜜蜂的清洁行为，以及蜂王浆和蜂胶的抗菌特性，加上蜂群有像城堡一样的结构，使蜜蜂不需要库蚊和果蝇那么多的抗病和免疫基因。还有一个可能就是基因组漏掉了 3% 左右的基因，而免疫基因大多在那些漏掉的片段中，但这种可能性不大。

与此类似，蜜蜂对农药和毒药的解毒基因也同样少了很多(表 2)

表 2 蜜蜂与其它昆虫的解毒基因数量比数^[7]

	果蝇	库蚊	蜜蜂
谷胱甘肽硫转移酶	38	31	10
P450(细胞色素氧化酶)	85	106	46
羟基/乙酯脂碱脂酶	35	51	24
总计	157	188	80

5 蜜蜂的性别基因的新进展

以下介绍作者的一篇与蜜蜂基因组有关的文章。蜜蜂的性别是由一种“互补性决定”(complementary sex determinative, csd)的机制来决定的。双倍体的杂合子成为雌性，而单倍体成为雄性。但是，当性等位基因是纯合子时，双

倍体仍发育成为雄性。而在正常蜂群中，双倍体的雄性会被工蜂吃掉。因此，使工蜂成为杂合子的选择压力很大。几乎 100 年以来，这个理论一直是一个假说而已。直到最近，csd 的基因才被测序，并最终证明为决定性别的基因^[8]。由于强大的选择压力，csd 应该会表现出高度的种内多态性(intraspecific polymorphism)，其等位基因的年龄可能比种的年龄更老。借助 Genbank 中的数据，我们将 3 种蜜蜂(西方蜜蜂、东方蜜蜂、大蜜蜂)的 csd 和随机选择的中性基因区的基因序列进行对比，发现：(1)csd 的多型性比中性基因的多态性约高 7 倍左右；(2)基因家谱分析发现，csd 有种内多态性，而中性基因没有；(3)与低频等位基因优势(rare allele advantage)的预计相符，非同义的突变(non-synonymous mutation)在刚出现时是受到正选择压力的。我们还发现在 3 种不同的蜜蜂中，每个种都有自已的高变(hypervariable)重复区域，从而推测每个种有自已不同的、决定性等位基因特异性的机制。这些结果已发表在 2006 年 10 月的《基因组研究》上^[9]。

6 后基因组时代

美国人在完成一件事后常见的一包话是“Now, what?”,意思是，现在还有什么可以做的呢?尽管在《自然》杂志蜜蜂基因组文章的 170 名作者中至少有 13 位华人，但他们都不是从中国的工作单位参与的。现在蜜蜂的基因组等于是属于全世界的资源了，因为任何有互联网连接的人都可以上网对基因组做注释，用生物信息学的方法去验证自己的假说，同样可以出好的文章。在 10 月份发表的 50 多篇文章中，作者查核了 23 篇，发现有 35% 的文章(8 篇)完全用基因组的资料做的分析，而未做另外的实验。中国人力资源丰富，应该可以很好地利用这个公共资源，作出一些与外国科研水平接轨的文章来。估计在未来 2~3 年内，至少可以出 50~80 篇文章，纯粹用生物信息学手段来开发蜜蜂基因库这块肥沃的资源。因为研究蜜蜂舞蹈而获得诺贝尔奖的冯·弗利希曾经说过，蜜蜂世界是一个“魔井”，我

们发现的东西越多, 有更多的奥秘等待我们去探索。如果现实中的蜜蜂是“魔井”的话, 蜜蜂的基因组的发表, 无疑是开放了一个广阔的有魔力的油田, 全世界的科学家不分国籍, 不分贫富, 都可以分享这个丰富的资源。

7 资料

美国国家生物技术信息中心 <http://www.ncbi.nlm.nih.gov/genome/guide/bee/>

国立人类基因组研究所蜜蜂基因组 <http://www.hgsc.bcm.tmc.edu/projects/honeybee/>

德克萨斯州农工大学“蜂库” <http://racex00.tamu.edu/bee-resources.html>

依利诺大学俄般那分校“蜂路” <http://www.beespace.uiuc.edu/>

密歇根州立大学“网蜂” <http://cyberbee.msu.edu>

基因组研究十一月(所有文章可免费下载) <http://www.genome.org/content/vol16/issue11/>

昆虫分子生物学十月(所有文章可免费下载) <http://www.blackwell-synergy.com/toc/imb/15/5>

致谢 感谢王琛柱博士的邀请、中文输入和修改。

参 考 文 献

- 1 The Honeybee Genome Sequencing Consortium. *Nature*, 2006, **443**: 931~949.
- 2 Robinson G. E., Aronstein K., Evans J. E., Fahrback S. E., Johnston J. S., et al. *Weaver* 2002. http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/HoneyBee_Genome.pdf
- 3 Robinson G. E., Evans J. D., Maleszka R., Robertson H. M., Weaver D. B., et al. *Insect Mol. Biol.*, 2006 **15** (5): 535~539.
- 4 Huang Z. Y., Robinson G. E. *PNAS*, 1992, **89**: 11 726~11 729.
- 5 Corona M. Robinson G. E. *Insect Mol. Biol.*, 2006 **15** (5), 687~701.
- 6 Rubin E. B., Shemesh Y., Cohen M., Elgavish S., Robertson H. M., et al. *Gen. Res.*, 2006 **16**: 1 352~1 365.
- 7 Claudianos C., Ranson H., Johnson R. M., Biswas S., Schuler M. A., et al. *Insect Mol. Biol.*, 2006, **15** (5): 615~636.
- 8 Beye M., Hasselmann M., Fondrk M., Page R., Omholt S. *Cell* 2003, **114**: 419~429.
- 9 Cho S., Huang Z. Y., Green D. R., Smith D. R., Zhang J. Z. *Gen. Res.* 2006, **16**: 1 366~1 375.

生物行为变化研究的新模式——工蜂

曲 宁^{1*} 邓爱华¹ 孙亮先² 徐金森¹ 吴学记^{1**}

(1. 厦门大学——新加坡国立大学共建生物医学科学实验室 厦门大学生命科学院 厦门大学医学院 361005;

2. 泉州师范学院模式生物实验室 泉州师范学院 362000)

Honey bee workers are a model system to study organism behavioral changes. QU Ning^{1*}, DENG Ai-Hua¹, SUN Liang-Xian², XU Jin-Sen¹, WU Xue-Ji^{1**} (1. *Xiamen University-National University of Singapore Laboratory of Biomedical Sciences, School of Life Sciences, Medical College, Xiamen University, Xiamen 361005, China*; 2. *Model Organism Research Centre of Quanzhou Normal College, Quanzhou 362000, China*)

Abstract There are four stages in the development of honey bees, including egg, larva, pupa and adult. As honey bee workers switch from in-hive tasks to foraging, they undergo transition from constant exposure to the controlled homogenous physical and sensory environment of the hive to prolonged diurnal exposures to a far more heterogeneous environment outside the hive. The switch offers an opportunity for the integrative study of the physiological and genetic

* E-mail: qning123@126.com

**通讯作者, E-mail: xuejiwu@hotmail.com

收稿日期: 2006-04-18, 修回日期: 2006-09-13