

# 棉铃虫膨胀线粒体基因组的鉴定与分析

张 岫<sup>1</sup>, 谷少华<sup>1</sup>, 李显春<sup>1,2\*</sup>

(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. 亚利桑那大学昆虫系, 图森 85721)

**摘要** 以已公布的棉铃虫线粒体 DNA 序列对来自 4 头棉铃虫雄蛹的 DNA 的三代测序数据进行筛选, 获得了 11 条与线粒体 DNA 有同源性的三代 read 序列, 并根据其中的 read 66003 鉴定出了一种膨胀的线粒体基因组。该线粒体基因组大小为 27 113 bp, 其保守区域包含 13 个蛋白编码基因、2 个 rRNA 基因、22 个 tRNA 基因以及 1 个 AT 富集区, 与已公布的棉铃虫线粒体基因组的结构相似。膨胀区域位于 *cox1* 基因编码区内部, 大小为 11 467 bp, 经预测含有一个完整的真核基因(依赖 ATP 的 RNA 解旋酶)以及多种转座元件的片段, 但与线粒体 DNA 无同源性, 也无 I 类或 II 类内含子存在的证据。对田间和室内棉铃虫 DNA 样品的 PCR 扩增未能检测到膨胀线粒体基因组的存在。以上结果表明膨胀片段可能是细胞核 DNA 序列通过偶然的水平转移事件而整合到线粒体基因组中的, 且该种膨胀方式的发生概率极低。本文报道的膨胀线粒体基因组为日后动物线粒体基因组学的研究提示了一种独特的变异方式。

**关键词** 棉铃虫; 线粒体基因组; 膨胀; RNA 解旋酶; 转座子

**中图分类号:** S 435. 622 **文献标识码:** A **DOI:** 10. 16688/j. zwbh. 2018223

## Identification and analysis of an expanded mitochondrial genome in *Helicoverpa armigera*

ZHANG Shen<sup>1</sup>, GU Shaohua<sup>1</sup>, LI Xianchun<sup>1,2</sup>

(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, University of Arizona, Tucson 85721, USA)

**Abstract** BLAST search against the whole genome PacBio sequencing reads derived from four *Helicoverpa armigera* male pupae by using the published mitochondrial genome of this insect retrieved 11 reads, one of which, read 66003, represented an expanded mitochondrial genome of *H. armigera*. Its size is 27 113 bp and, like the published mitochondrial genome of *H. armigera*, contained all the 13 protein-coding genes, two rRNA genes, 22 tRNA genes and one A/T-rich region in the same order. In addition, this genome contained a 11 467 bp expanded region inserting into the coding sequence of *cox1*. The insertion encodes a ATP-dependent RNA helicase, containing fragments of several different transposable elements but lacking group I and group II introns. Sequence-specific PCR failed to find the presence of this expanded mitochondrial genome in multiple individuals. These results indicated that this expanded mitochondrial genome rarely occurred through a recent horizontal transfer event of nuclear DNA sequence into the normal mitochondrial genome and its frequency remained extremely low. The expanded mitochondrial genome described here represents a distinct variation of mitochondrial genome.

**Key words** *Helicoverpa armigera*; mitochondrial genome; expansion; RNA helicase; transposon

线粒体是真核细胞中一种最基本的细胞器, 一般认为其起源于细菌在真核细胞中的内共生作用<sup>[1]</sup>, 且保留了一套独立于细胞核基因组之外的遗传系统, 能够进行自我复制和表达。在植物和动物

中, 尽管线粒体承担的生物学功能相似, 但其基因组却向两个极端对立的方向演化: 植物线粒体基因组复杂、庞大, 变异程度较大, 富含大量非编码序列和内含子; 而动物的线粒体基因组较为保守, 基因排布

收稿日期: 2018-05-25 修订日期: 2018-07-07

基金项目: 国家自然科学基金(31772164, 31401737, 31171874); 北京市青年拔尖人才计划(2015000021223ZK29)

\* 通信作者 E-mail: lxc@email.arizona.edu

紧凑,非编码区含量较少,一般不含内含子<sup>[2]</sup>。动物线粒体基因组大小通常在 15~20 kb 之间,包含 13 个蛋白编码基因、22 个转运 RNA(tRNA)基因、2 个核糖体 RNA(rRNA)基因和 1 个 AT 富集区(控制区)<sup>[3]</sup>。尽管动物线粒体基因组在基因组成和结构上的保守性较高,但变异事件亦会偶尔发生,主要表现为基因的复制或缺失、基因重排以及非编码区的扩张或收缩等方式。

基因的复制或缺失是动物中相对常见的线粒体基因组变异方式。例如,某些腔肠动物及箭虫的线粒体基因组中缺失了几个乃至全部的 tRNA 基因<sup>[4-5]</sup>,在蛛形纲部分物种和瘿蚊类昆虫的线粒体基因组中则存在 tRNA 缩短的现象<sup>[6-7]</sup>,而基因复制现象在螨、蓟马乃至脊椎动物的线粒体基因组中均有发现<sup>[8-10]</sup>。进一步地,线粒体基因复制和缺失事件的积累可能导致基因重排现象的产生<sup>[11]</sup>。邱忠营<sup>[12]</sup>对 153 种直翅目昆虫的线粒体基因组进行了比较基因组学分析,总结了其中 12 个物种的基因重排模式,其重排事件发生的概率较低,且多数为 tRNA 基因位置的颠换。线粒体基因重排极端化的一种表现是线粒体的片段化。鞭毛虫、水母和虱目等的线粒体中均发现了片段化的现象,最典型的例子是体虱 *Pediculus humanus*,其线粒体基因组由 18 个微型染色体(minichromosome)组成,每个染色体长 3~4 kb,携带 1~3 个基因及一个控制区<sup>[13-15]</sup>。

线粒体基因组非编码区的变异主要源自内含子或重复元件的复制或丢失。线粒体基因组中的内含子分为 I 类内含子(group I intron)和 II 类内含子(group II intron)两类,二者不同于真核细胞核基因组内的剪接体内含子,它们具备核酶的催化功能,能够在不依赖剪接体的情况下完成自我剪接<sup>[16-18]</sup>。I 类内含子多见于真菌线粒体,而 II 类内含子则为植物线粒体基因组内含子的主要形式。内含子在动物线粒体中较为罕见,目前仅在海葵、珊瑚等低等后生动物中有报道。Beagley 等<sup>[4,19]</sup>在绣球海葵 *Metridium senile* 线粒体细胞色素 C 氧化酶亚基 I(*cox1*)和 NADH 脱氢酶亚基 5(ND5)编码基因内部分别发现了一个 I 类内含子,其中前者能够自我成环并编码一个核酸内切酶,而后者内部则含有编码 ND1 和 ND3 的基因,且内含子自我剪接后二者能够以双顺反子的形式进行转录。进一步的研究表明,线粒体 ND5 内含子在六射珊瑚纲的物种内普遍

存在,且其内部的蛋白编码基因复制数目在演化中逐步增加<sup>[20]</sup>。

相较于内含子,控制区或其他重复序列的变异在动物线粒体基因组中更常见且形式更丰富。Tang 等<sup>[21]</sup>在寄生线虫 *Thaumamermis cosgrovei* 中鉴定出两种线粒体基因型,其差异源于基因或非编码区的变异。软体动物线粒体基因组的非编码序列的扩张最为普遍,且个体间差异明显。深海扇贝 *Placopecten magellanicus* 线粒体基因组的大小从 30.7 kb 至 40.7 kb 不等,基因组内有 3 个位点存在结构变异,每个位点至少存在 6 种不同形式的重复元件,导致了线粒体基因组在个体间的巨大变异<sup>[22-24]</sup>。类似地,在另一种扇贝 *Pecten maximus* 的线粒体基因组中存在一类拷贝数可变的长约 1.6 kb 的重复元件,其重复次数的不同致使基因组大小在 20.0 kb 到 25.8 kb 之间波动<sup>[25]</sup>。赤贝 *Scapharca broughtonii* 的线粒体基因组大小约为 50 kb,是后生动物中已知最长的线粒体 DNA,其基因组内含有一个串联重复单元以及一个可复制的非编码区,二者的变异是基因组扩张的原因<sup>[26]</sup>。Raimond 等<sup>[27]</sup>利用 RFLP 技术测定了鼠妇 *Armadillidium vulgare* 的线粒体基因组大小(20~42 kb),结合电镜观察提出一种膨胀的线粒体基因组的存在形式:约 14 kb 的线粒体 DNA 单体发生复制和重排形成 3 个单体,其中一个以线性形式存在,另外两个则连接成环。Boyce 等<sup>[28]</sup>在 3 种象甲 *Pissodes* spp. 的线粒体 DNA 中发现了扩张的 AT 富集区(9~13 kb),同时在其侧翼还存在数目不等的重复元件,这成为该属昆虫线粒体基因组变异的原因。综上所述,动物线粒体在基因组成和结构上相对保守,变异事件形式有限、发生概率较小,其中基因重排和内含子插入事件多为可遗传的,已成为物种固定的遗传属性,而重复元件的变异通常存在多样性和随机性,在物种内个体间存在差异。

棉铃虫 *Helicoverpa armigera* 属鳞翅目 Lepidoptera 夜蛾科 Noctuidae,是世界范围内的一种重要农业害虫,其线粒体基因组已在 2010 年公布,总长度为 15 347 bp,具备典型的动物线粒体基因组结构<sup>[29]</sup>。本文利用三代 DNA 测序技术,鉴定出一条异常的棉铃虫线粒体 DNA 序列,其 *cox1* 基因内部插入了一段 11 467 bp 的疑似核基因组的片段,导致线粒体基因组显著膨胀,为动物线粒体基因组学的

研究提示了一种特殊的变异方式。

1 材料与方法

1.1 样品及数据来源

棉铃虫线粒体基因组序列获取自二代和三代全基因组 DNA 测序 read。DNA 样品来自 4 头蛹期的棉铃虫,对其进行全基因组 DNA 抽提和建库测序。二代测序平台为 Illumina Hiseq 2000,三代测序平台为 PacBio SMART RS II。原始测序数据下机后经过质量控制和融合,以二代数据对三代数据进行校正,校正后的数据用于筛选线粒体相关的 read 序列。

1.2 线粒体基因组的注释

线粒体基因组首先以在线工具 MFannot 和 RNAweasel<sup>[16]</sup> (<http://megasun.bch.umontreal.ca/RNAweasel/>)进行蛋白编码基因、rRNA 基因以及 tRNA 基因的预测,随后从 Yin 等<sup>[29]</sup>的报道中提取注释信息,将之补充并整合到上述预测结果中,得到最终的线粒体注释结果。线粒体基因组结构图的绘制使用在线工具 CGView Server V 1.0<sup>[30]</sup> ([http://stothard.afns.ualberta.ca/cgview\\_server/](http://stothard.afns.ualberta.ca/cgview_server/))。

1.3 线粒体膨胀区域的分析

膨胀区域的分析通过 BLASTn 和 BLASTx 分别在 NCBI 的 Nucleotide collection(Nt)和 Non-redundant(Nr)数据库中进行检索。基因结构的作图使用在线工具 Gene Structure Display Server 2.0<sup>[31]</sup> (<http://gsds.cbi.pku.edu.cn/>)。转座元件的预测使用 Repbase 数据库中的在线工具 CENSOR<sup>[32]</sup> (<https://www.girinst.org/censor/>)。I 类或 II 类内含子的预测采用 RNAweasel。

2 结果与分析

2.1 棉铃虫膨胀线粒体 DNA 的鉴定

以已公布的棉铃虫线粒体基因组序列(GenBank 登录号 GU188273)作为 query 序列,检索校正后的三代 read,从中筛选出 11 条与线粒体 DNA 有同源性的 read 序列(表 1)。序列比对结果显示,11 条 read 中有 8 条对 query 的覆盖度较低(<60%),而另外 3 条超过 90%,其中,仅有 read 66003 完整覆盖线粒体基因组且首尾存在重叠,由此得到了一条膨胀的环状线粒体 DNA 序列用于后续分析。

表 1 与棉铃虫线粒体 DNA 同源的 11 条三代 read<sup>1)</sup>

| Table 1 Eleven PacBio reads homologous to <i>Helicoverpa armigera</i> mitochondrial DNA |                              |                                |                                                     |                                  |                     |
|-----------------------------------------------------------------------------------------|------------------------------|--------------------------------|-----------------------------------------------------|----------------------------------|---------------------|
| Read 编号<br>Read no.                                                                     | Read 长度/bp<br>Length of read | Read 覆盖度/%<br>Coverage of read | Query 匹配区间<br>Matched region of query               | Query 覆盖度/%<br>Coverage of query | 相似度/%<br>Similarity |
| 2281                                                                                    | 21 586                       | 15                             | 11 005—14 322                                       | 22                               | 95                  |
| 36137                                                                                   | 13 017                       | 46                             | 1—3 176,12 607—15 347                               | 39                               | 96                  |
| 66003                                                                                   | 41 476                       | 66                             | 15 347—2 376,2 352—400,357—1                        | 100                              | 97                  |
| 99261                                                                                   | 4 612                        | 100                            | 4 399—185                                           | 30                               | 87                  |
| 119177                                                                                  | 13 381                       | 100                            | 4 690—15 344,1—3 313                                | 91                               | 94                  |
| 147013                                                                                  | 20 101                       | 42                             | 15 347—7 527,500—1                                  | 54                               | 97                  |
| 53580                                                                                   | 16 957                       | 91                             | 12 557—5 487,4 802—10 697,10 692—12 554,4 008—4 606 | 100                              | 95                  |
| 134466                                                                                  | 13 624                       | 24                             | 3 992—7 216                                         | 21                               | 97                  |
| 81475                                                                                   | 19 737                       | 26                             | 14 125—9 317,297—1                                  | 33                               | 97                  |
| 147410                                                                                  | 18 383                       | 25                             | 1 937—6 497                                         | 30                               | 95                  |
| 27297                                                                                   | 12 161                       | 35                             | 2 113—6 321                                         | 28                               | 94                  |

1) Query 指代已公布的棉铃虫正常线粒体 DNA。  
Query indicates the published normal mitochondrial genome of *H. armigera*.

2.2 棉铃虫膨胀线粒体基因组的注释

膨胀的棉铃虫线粒体基因组大小为 27 113 bp,其基因组结构如图 1 所示。与已公布的棉铃虫线粒体基因组的比对结果显示,膨胀线粒体基因组可分为保守区域和膨胀区域两部分。保守区域中二者核苷酸序列相似度达到 97%。通过注释,在保守区域鉴定出了 13 个蛋白编码基因(*cox1*,*cox2*,*cox3*,*cob*,*nad1*,*nad2*,*nad3*,*nad4*,*nad4L*,*nad5*,*nad6*,*atp6*,

*atp8*)、2 个 rRNA 基因(*rnaL*,*rnaS*)、22 个 tRNA 基因(亮氨酸-tRNA 和丝氨酸-tRNA 分别有 2 个拷贝,其余 18 种氨基酸分别对应 1 个 tRNA)以及 1 个 AT 富集区。其基因组成和排布方式与已公布的棉铃虫线粒体基因组基本一致,不同之处在于 *cox1* 基因被打断为两部分且缺失了约 23 bp 的编码序列。膨胀区域是一个连续的、长度为 11 467 bp 的片段,与正常线粒体无同源性,存在于 *cox1* 基因编码区之中。

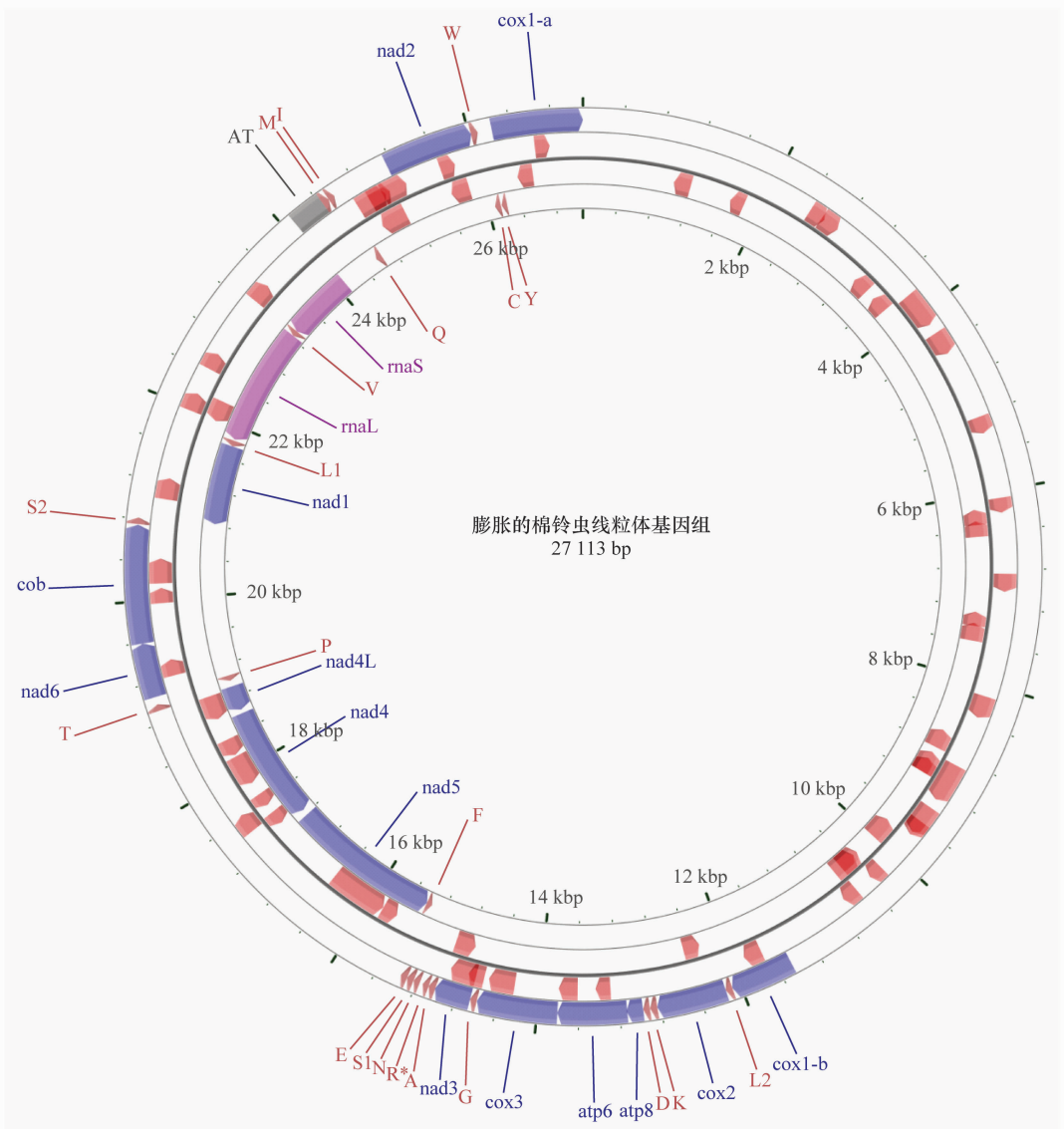


图 1 棉铃虫膨胀线粒体基因组的结构

Fig. 1 Structure of the expanded mitochondrial genome of *Helicoverpa armigera*

2.3 棉铃虫线粒体基因组膨胀区域的分析

蛋白编码基因检索的结果显示,膨胀区域内包含一个依赖 ATP 的 RNA 解旋酶基因(ATP-dependent

RNA helicase,GenBank 登录号 XM\_021342945)。该基因具有真核基因的结构,编码区位于轻链的 3 931—10 784 bp 区间,包含 5 个外显子,基因结构图如图 2 所示。

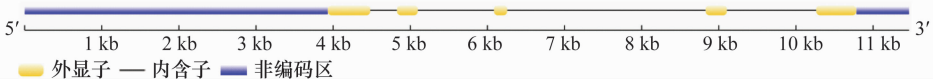


图 2 膨胀区域内 RNA 解旋酶的基因结构

Fig. 2 Gene structure of the RNA helicase in the expanded region

NCBI Nt 数据库的检索结果显示,与膨胀区域匹配度最高的 subject 为棉铃虫细菌人工染色体(bacterial artificial chromosome, BAC)序列(GenBank 登录号 FO082297)。进一步地,使用 Repbase 预测膨胀区域转座元件,结果如表 2 所示。其中较完

整的一个元件注释为非长末端重复反转录转座子(Non long terminal repeat retrotransposable element, NonLTR/RTE)RTE-5,在基因组中分为 4 个片段,总长度约 2 800 bp,其主要部分位于膨胀区域的 6 500—8 800 bp 之间,即 RNA 解旋酶基因第三个内含子

内部。其他转座元件还涉及多种 DNA 转座子、LTR 反转录转座子以及散布重复序列等,但均不超过 1 kb,不具备完整的转座子结构。此外,内含子预

测的结果显示,膨胀区域内既不含有 I 类/Ⅱ类内含子的保守结构,也没有编码逆转录酶或核酸内切酶的基因,无法找到内含子存在的证据。

表 2 膨胀区域内转座元件的预测结果

Table 2 Prediction of transposable elements (TE) in the expanded region

| 位置<br>Location | 方向<br>Direction | 转座子名称<br>Name of TE | 转座子类别<br>Class of TE | 转座子匹配区间<br>Matched region in TE | 相似度/%<br>Similarity |
|----------------|-----------------|---------------------|----------------------|---------------------------------|---------------------|
| 239—279        | +               | EnSpm2_MT           | DNA/EnSpm/CACTA      | 3289—3326                       | 85                  |
| 494—597        | +               | Tetris_Dvir         | DNA                  | 858—957                         | 73                  |
| 789—843        | —               | Gypsy1—L_AO         | LTR/Gypsy            | 1206—1255                       | 84                  |
| 2028—2139      | —               | RTE—5_BM            | NonLTR/RTE           | 351—462                         | 67                  |
| 2266—2327      | —               | RTE—1_DAn           | NonLTR/RTE           | 8—69                            | 76                  |
| 2506—2619      | —               | Copia—5_Mad—LTR     | LTR/Copia            | 42—143                          | 67                  |
| 2785—2909      | —               | Gypsy—1_DFi—I       | LTR/Gypsy            | 373—498                         | 73                  |
| 3222—3281      | —               | MuDR—14_GM          | DNA/MuDR             | 12490—12547                     | 77                  |
| 4503—4599      | —               | EnSpm1_HM           | DNA/EnSpm/CACTA      | 1580—1674                       | 77                  |
| 4607—4685      | +               | EnSpm6_SB           | DNA/EnSpm/CACTA      | 6366—6453                       | 76                  |
| 4793—4840      | +               | Penelope—1_DK       | NonLTR/Penelope      | 1331—1382                       | 78                  |
| 5248—5302      | —               | REP—2_PBa           | Interspersed_Repeat  | 876—924                         | 80                  |
| 5377—5415      | +               | Gypsy—1_LMi—I       | LTR/Gypsy            | 456—495                         | 82                  |
| 6085—6281      | +               | RTE—5_BM            | NonLTR/RTE           | 286—484                         | 72                  |
| 6430—8466      |                 |                     |                      | 498—2550                        |                     |
| 8467—8861      | +               | RTE—5_Hmel          | NonLTR/RTE           | 2551—2946                       | 63                  |
| 9138—9812      | +               | RTE—17_Hmel         | NonLTR/RTE           | 1—682                           | 66                  |
| 9875—9964      | —               | EnSpm3_DR           | DNA/EnSpm/CACTA      | 6700—6799                       | 77                  |

3 讨论

本文从棉铃虫全基因组三代测序原始数据中,筛选得到了与线粒体相关的序列。其中,read 99261 和 read 119177 的 read 覆盖度为 100%,表明二者可能是正常线粒体 DNA 的片段;read 66003 完整覆盖 subject,且在二代测序数据的辅助下能拼接成环,因此我们借助该 read 鉴定出了膨胀的棉铃虫线粒体基因组;而其余 read 对 subject 的覆盖度较低或存在片段倒位(表 1),由于原始数据未经过拼接组装,不存在拼接错误,因此我们推测这些序列并非是线粒体源的 DNA,而是包含线粒体假基因的核基因组序列。线粒体基因组向核基因组迁移的现象广泛存在。在动物中,线粒体基因通常以假基因形式存在于核基因组中,分布形式多变<sup>[33-34]</sup>,其中昆虫基因组中的线粒体假基因已在蚱蜢、蚜虫、天牛等类别的物种中得到研究<sup>[35-37]</sup>。植物中线粒体 DNA 迁入核基因组的事件比动物更为普遍,核基因组能获得线粒体基因并保留其功能<sup>[38]</sup>,且线粒体基因组能以大片段乃至完整形式插入核基因组<sup>[39]</sup>。在本研究中,线粒体相关 read 的鉴定暗示棉铃虫核基因组中可能存在诸多的线粒体假基因,且假基因与真基因的相

似度较高,但假基因的种类、拷贝数以及分布方式等问题尚未明确,仍有待进一步研究。Song 等<sup>[40]</sup>的研究表明,细胞核基因组中的线粒体假基因会干扰 DNA 条形码鉴定技术的准确性,这也提示我们未来使用 DNA 条形码定量检测棉铃虫种群时需注意排除线粒体假基因的干扰。

后生动物线粒体基因组通常编码 37 个基因,长度一般不超过 20 kb,非编码序列含量很低。棉铃虫线粒体基因组前人已有报道<sup>[29]</sup>,符合动物线粒体基因组的这些保守特征。本文通过高通量测序发现了一种不同于该报道的、异常膨胀的棉铃虫线粒体基因组,其大小达到 27 113 bp。与已公布的序列相比,基因组内部发生了显著的膨胀,膨胀原因主要是 *cox1* 基因编码区中间插入了一段 11 467 bp 的序列。线粒体基因组的膨胀主要通过基因复制或非编码区扩张的方式实现。为此,我们分析了膨胀序列的来源以及该区域内编码基因、重复元件或内含子存在的可能性。如 2.3 所述,该片段在 NCBI Nt 数据库中能够匹配到棉铃虫 BAC 序列,却与线粒体 DNA 无任何匹配,表明其并非线粒体基因组自身的复制或扩张,而更可能来源于核基因组。通过功能基因的注释,发现膨胀区域包含一个完整的核基因

组蛋白编码基因(依赖 ATP 的 RNA 解旋酶),且具有外显子-内含子的真核基因结构。这进一步表明膨胀片段可能是一段核基因组序列通过水平转移被整合进线粒体基因组的产物。

I 类内含子和 II 类内含子不同于核基因组内的剪接体内含子,二者不依赖复杂的剪接体即可完成自我剪接,后者通常还具备逆转录酶、成熟酶或核酸内切酶的编码区<sup>[16]</sup>。本研究中,利用核酸内切酶、逆转录酶保守结构域以及内含子预测工具均未在膨胀区域发现内含子存在的证据,这表明棉铃虫线粒体基因组的膨胀并非由内含子插入导致。我们进一步使用 Repbase 预测了膨胀线粒体基因组内转座元件存在的可能性。结果表明,该区域内存在一个较完整的 NonLTR 类反转录转座子以及诸多转座元件的片段。这些转座子序列大多不完整,且呈分散或嵌套分布,因此我们推测膨胀片段的祖先序列经历了多次转座事件,多数转座子结构被破坏,仅残留部分痕迹,而线粒体基因组中该膨胀部分的插入则可能是在上述转座事件后发生的一次细胞核 DNA 水平转移事件。

在植物中,核内 DNA 序列、尤其是转座元件整合进入线粒体基因组的现象较为普遍<sup>[39]</sup>,但在动物线粒体基因组中此类变异事件较为罕见,且对于简单、紧凑的线粒体基因组,转座子通常是有害的,在演化过程中易受净化选择(purifying selection)作用而被清除。针对本文中鉴定得到的棉铃虫膨胀线粒体基因组,我们曾尝试在室内及田间收集的数十个棉铃虫 DNA 样品中克隆该膨胀区域,但均未发现目的片段整合到线粒体基因组的证据。该膨胀的线粒体基因组序列是直接由 DNA 测序 read 得到的,不会掺入拼接错误,因此可以保证其存在的真实性,而其克隆证据的缺失一方面可能是因为膨胀序列的保守性较低,存在诸多变异形式,另一方面,更大的可能性在于此种线粒体膨胀事件发生的概率极低,且其破坏了 *cox1* 基因的编码区,对个体生存是不利的,很难被遗传下去,只是发生在体细胞的一种偶然突变。尽管如此,本文报道的棉铃虫线粒体基因组的膨胀,仍可为日后的研究提示一种特殊的线粒体基因组变异方式。

## 参考文献

- [1] GRAY M W. Origin and evolution of mitochondrial DNA [J]. Annual Review of Cell Biology, 1989, 5:25 - 50.
- [2] 张尚宏. 动物与植物线粒体基因组结构的差异—两种进化途径[J]. 动物学研究, 1995, 16(2):132 - 145.
- [3] 黄族豪, 刘迺发. 动物线粒体基因组变异研究进展[J]. 生命科学研究, 2010, 14(2):166 - 171.
- [4] BEAGLEY C T, OKIMOTO R, WOLSTENHOLME D R. The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): introns, a paucity of tRNA genes, and a near-standard genetic code [J]. Genetics, 1998, 148(3):1091 - 1108.
- [5] HELFENBEIN K G, FOURCADE H M, VANJANI R G, et al. The mitochondrial genome of *Paraspadella gotoi* is highly reduced and reveals that chaetognaths are a sister group to protostomes [J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(29):10639 - 10643.
- [6] MASTA S E, BOORE J L. Parallel evolution of truncated transfer RNA genes in arachnid mitochondrial genomes [J]. Molecular Biology and Evolution, 2008, 25(5):949 - 959.
- [7] BECKENBACH A T, JOY J B. Evolution of the mitochondrial genomes of gall midges (Diptera: Cecidomyiidae): rearrangement and severe truncation of tRNA genes [J]. Genome Biology and Evolution, 2009, 1(1):278 - 287.
- [8] SHAO Renfu, MITANI H, BARKER S C, et al. Novel mitochondrial gene content and gene arrangement indicate illegitimate inter-mtDNA recombination in the chigger mite, *Lep-totrombidium pallidum* [J]. Journal of Molecular Evolution, 2005, 60(6):764 - 773.
- [9] SHAO Renfu, BARKER S C. The highly rearranged mitochondrial genome of the plague thrips, *Thrips imaginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes [J]. Molecular Biology and Evolution, 2003, 20(3):362 - 370.
- [10] FUJITA M K, BOORE J L, MORITZ C. Multiple origins and rapid evolution of duplicated mitochondrial genes in parthenogenetic geckos (*Heteronotia binoei*; Squamata, Gekkonidae) [J]. Molecular Biology and Evolution, 2007, 24(12):2775 - 2786.
- [11] LAVROV D V, BOORE J L, BROWN W M. Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: duplication and nonrandom loss [J]. Molecular Biology and Evolution, 2002, 19(2):163 - 169.
- [12] 邱忠营. 直翅目昆虫线粒体基因组比较、谱系及进化研究[D]. 西安:陕西师范大学, 2016.
- [13] BURGER G, FORGET L, ZHU Yun, et al. Unique mitochondrial genome architecture in unicellular relatives of animals [J]. Proceedings of the National Academy of Sciences of the United States of America, 2003, 100(3):892 - 897.
- [14] SMITH D R, KAYAL E, YANAGIHARA A A, et al. First complete mitochondrial genome sequence from a box jellyfish reveals a highly fragmented linear architecture and insights into telomere evolution [J]. Genome Biology and Evolution, 2011, 4(1):52 - 58.
- [15] SHAO Renfu, KIRKNESS E F, BARKER S C. The single

- mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, *Pediculus humanus* [J]. *Genome Research*, 2009, 19(5):904–912.
- [16] LANG B F, LAFOREST M J, BURGER G. Mitochondrial introns: a critical view [J]. *Trends in Genetics*, 2007, 23(3): 119–125.
- [17] HAUSNER G. Fungal mitochondrial genomes, plasmids and introns [J]. *Applied Mycology and Biotechnology*, 2003, 3(3): 101–131.
- [18] BROWN G G, COLAS DES FRANCS-SMALL C, OSTER-SETZER-BIRAN O. Group II intron splicing factors in plant mitochondria [J]. *Frontiers in Plant Science*, 2014, 5(5):35.
- [19] BEAGLEY C T, OKADA N A, WOLSTENHOLME D R. Two mitochondrial group I introns in a metazoan, the sea anemone *Metridium senile*: one intron contains genes for subunits 1 and 3 of NADH dehydrogenase [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1996, 93(11):5619–5623.
- [20] MEDINA M, COLLINS A G, TAKAOKA T L, et al. Naked corals: Skeleton loss in *Scleractinia* [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2006, 103(24):9096–9100.
- [21] TANG Sha, HYMAN B C. Mitochondrial genome haplotype hypervariation within the isopod parasitic nematode *Thaumamermis cosgrovei* [J]. *Genetics*, 2007, 176(2):1139–1150.
- [22] SNYDER M, FRASER A R, LAROCHE J, et al. Atypical mitochondrial DNA from the deep-sea scallop *Placopecten magellanicus* [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1987, 84(21):7595–7599.
- [23] FULLER K M, ZOUROS E. Dispersed discrete length polymorphism of mitochondrial DNA in the scallop *Placopecten magellanicus* (Gmelin) [J]. *Current Genetics*, 1993, 23(4): 365–369.
- [24] SMITH D R, SNYDER M. Complete mitochondrial DNA sequence of the scallop *Placopecten magellanicus*: evidence of transposition leading to an uncharacteristically large mitochondrial genome [J]. *Journal of Molecular Evolution*, 2007, 65(4):380–391.
- [25] RIGAA A, MONNEROT M, SELLOS D. Molecular cloning and complete nucleotide sequence of the repeated unit and flanking gene of the scallop *Pecten maximus* mitochondrial DNA: putative replication origin features [J]. *Journal of Molecular Evolution*, 1995, 41(2):189–195.
- [26] LIU Yunguo, KUROKAWA T, SEKINO M, et al. Complete mitochondrial DNA sequence of the ark shell *Scapharca broughtonii*: an ultra-large metazoan mitochondrial genome [J]. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 2012, 8(1):72–81.
- [27] RAIMOND R, MARCADÉ I, BOUCHON D, et al. Organization of the large mitochondrial genome in the isopod *Armadilidium vulgare* [J]. *Genetics*, 1999, 151(1):203–210.
- [28] BOYCE T M, ZWICK M E, AQUADRO C F. Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy [J]. *Genetics*, 1989, 123(4):825–836.
- [29] YIN Jiao, HONG Guiyun, WANG Aiming, et al. Mitochondrial genome of the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) and comparison with other lepidopterans [J]. *Mitochondrial DNA Part A*, 2010, 21(5):160–169.
- [30] GRANT J R, STOTHARD P. The CGView Server: a comparative genomics tool for circular genomes [J]. *Nucleic Acids Research*, 2008, 36:w181–w184.
- [31] HU Bo, JIN Jinpu, GUO Anyun, et al. GSDS 2.0: an upgraded gene feature visualization server [J]. *Bioinformatics*, 2014, 31(8):1296–1297.
- [32] KOHANY O, GENTLES A J, HANKUS L, et al. Annotation, submission and screening of repetitive elements in Repbase; Repbase Submitter and Censor [J]. *BMC Bioinformatics*, 2006, 7(1):474.
- [33] HARTL D L. Mitochondrial pseudogenes: evolution's misplaced witnesses [J]. *Trends in Ecology and Evolution*, 2001, 16(6):314–321.
- [34] ZISCHLER H. Nuclear integrations of mitochondrial DNA in primates: inference of associated mutational events [J]. *Electrophoresis*, 2000, 21(3):531–536.
- [35] VAUGHAN H E, HESLOPHARRISON J S, HEWITT G M. The localization of mitochondrial sequences to chromosomal DNA in orthopterans [J]. *Genome*, 1999, 42(5):874–880.
- [36] SUNNUCKS P, HALES D F. Numerous transposed sequences of mitochondrial cytochrome oxidase I–II in aphids of the genus *Sitobion* (Hemiptera: Aphididae) [J]. *Molecular Biology and Evolution*, 1996, 13(3):510–524.
- [37] KOUTROUMPA F A, LIEUTIER F, ROUX-MORABITO G. Incorporation of mitochondrial fragments in the nuclear genome (Numts) of the longhorned beetle *Monochamus galloprovincialis* (Coleoptera, Cerambycidae) [J]. *Journal of Zoological Systematics and Evolutionary Research*, 2009, 47(2): 141–148.
- [38] PALMER J D, ADAMS K L, CHO Y, et al. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2000, 97(13):6960–6966.
- [39] 陈志文, 王玉美, 聂虎帅, 等. 植物水平基因转移研究进展 [J]. *中国农业大学学报*, 2018, 23(1):1–11.
- [40] SONG H, BUHAY J E, WHITING M F, et al. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(36):13486–13491.