

稻纵卷叶螟线粒体基因组及细胞色素 C 氧化酶基因分析

尹湧华, 司胜娟, 涂飞云, 张修月, 岳碧松*

(四川大学生命科学学院, 成都 610064)

摘要:线粒体基因是研究生物系统进化的重要分子标记, 本文对迁飞昆虫稻纵卷叶螟 *Cnaphalocrocis medinalis* 线粒体 DNA 进行扩增及注释, 结果表明: 稻纵卷叶螟线粒体基因组全长 15 377 bp, AT 含量为 82%, 基因组结构为鳞翅目所特有的 CR-M-I-Q 结构。以细胞色素 C 氧化酶基因(COX)为分子标记, 对有翅类昆虫 22 个物种进行系统树构建, 发现与昆虫翅的发生高度吻合。对存在于 PCGs 终止密码子判定的争议予以阐明并提出注释建议, 有利于今后动物线粒体基因组序列注释的统一规范。

关键词: 稻纵卷叶螟; 线粒体基因组; 细胞色素 C 氧化酶基因; 终止密码子注释; 系统进化

中图分类号: Q966 **文献标识码:** A **文章编号:** 1000-7083(2013)01-0040-04

A Complete Mitochondrial Genome of *Cnaphalocrocis medinalis* (Lepidoptera: Crambidae) and its COX Gene Analysis

YIN Yong-hua, SI Sheng-juan, TU Fei-yun, ZHANG Xiu-yue, YUE Bi-song*

(College of Life Sciences, Sichuan University, Chengdu 610064, China)

Abstract: Mitochondrial genes represent important molecular clocks for phylogenetic analysis. In present study, the complete mitochondrial genome of *Cnaphalocrocis medinalis* was sequenced, which is a circular molecule with 15 377 bp length (AT 82%). The genome structure shows a lepidopteron-unique CR-M-I-Q arrangement. Based on COX gene, we speculated phylogenetic relationship among 22 pterygotan insects species. The result suggests the evolution of insects was highly related to the divergence of their wings. Three types of model are proposed for the first time, to determine the termination codon of the animal mtDNAs.

Key words: *Cnaphalocrocis medinalis*; mitochondrial genome; COX gene; termination codon; phylogenetic analysis

线粒体基因在过去的二十年中逐渐被作为一个中性的分子标记, 广泛地应用于系统进化及种群遗传学的研究(Shen *et al.*, 2009)。线粒体是真核细胞主要的供能器官, 并拥有一套独立于细胞核的遗传系统。其中, 细胞色素 C 氧化酶基因(COX)作为细胞氧化磷酸化反应的重要组件, 参与线粒体电子传递链的末端氧化, 被认为是线粒体 DNA(mtDNA) 最保守的蛋白编码基因(Nicc & Shapiro, 2001), 为研究生物线粒体基因进化提供了稳定、有效的分子标记(任竹梅等, 2002; 黄朝晖, 王金福, 2003; 王戎疆等, 2004)。

鳞翅目昆虫种类超过 18 万种(Kristensen, 1999)。鳞翅目昆虫 mtDNA 具有显著的碱基偏倚性, 平均 AT 含量为 77.84% ~ 82.66%, 在控制区可达 89.17% ~ 98.25%(Liao *et al.*, 2010)。此外, 鳞

翅目线粒体基因组具有昆虫纲中所独有的 CR-M-I-Q 结构, 这些都为引物设计、PCR 扩增及扩增产物测序带来了困难, 因此, 核酸数据库(GenBank) 中记录的鳞翅目昆虫线粒体基因组全序列并不多。

稻纵卷叶螟 *Cnaphalocrocis medinalis* 属鳞翅目草螟科, 是广泛分布于亚洲、东非及澳洲的水稻害虫。作为重要的迁飞昆虫, 稻纵卷叶螟能够随季风气流进行持续、远距离的迁飞(高月波等, 2008), 显示了这一物种高度的能量代谢水平。本文对稻纵卷叶螟线粒体基因组进行测序, 对细胞色素 C 氧化酶基因进行分析, 为研究迁飞昆虫能量代谢及有翅类昆虫系统进化提供了重要分子依据。

1 材料和方法

1.1 基因组 DNA 的提取

收稿日期: 2012-11-16 接受日期: 2012-12-06 基金项目: 国家自然科学基金项目(No. 31272039)

作者简介: 尹湧华, 男, 博士研究生, 研究方向: 分子系统进化及生物信息学, E-mail: yhyin.scu@gmail.com

* 通讯作者 Corresponding author, E-mail: bsyue@scu.edu.cn

稻纵卷叶螟成虫富含脂类、蛋白质、几丁质及鳞片,对 DNA 的提取质量有一定影响,最好选用胸部及上段腹部为 DNA 提取材料。对总基因组的提取采用改进后的 SDS-酚氯仿抽提法(Sambrook & Russell, 2001; 林万华等, 2006)。

1.2 PCR 扩增及测序

1.2.1 短片段扩增策略

鳞翅目昆虫与其他昆虫相比,线粒体基因组 AT 含量高、富含短序列重复片段,长片段 PCR 产物测序多不理想。为此,本研究采用短片段扩增设计,即每段序列 500 ~ 1500 bp 的短片段扩增策略。共包括 18 对引物,其中 14 对筛选自 Simon 等(2006)发布的适用于昆虫类群 mtDNA 扩增的保守引物,其余 4 对引物为针对稻纵卷叶螟 mtDNA 序列而特别设计。

1.2.2 克隆测序

由于昆虫线粒体基因 AT 碱基的偏倚性及富含类似微卫星的元件,PCR 扩增产物往往形成复杂的高级结构,使直接测序比较困难。本文采用 pMD[®]19-T (TaKaRa)载体,将原液稀释 5 ~ 10 倍,以减少连接过程中载体的自连。连接时间依据片段长度而略有不同(4 ~ 10 h)。连接体系为:

Solution I 2.5 μ L, 胶回收产物 2 μ L, Vector 0.5 μ L。

1.2.3 序列注释

基因序列的拼接应用 MEGA 4 + Primer 5 + Word 2003 (Lalitha, 2000; Tamura *et al.*, 2007; Microsoft Corp.)。MEGA 用于基因间重叠区域的确认,Primer 用于核苷酸序列的格式化,Word 完成对序列的标注及文档保存。对稻纵卷叶螟线粒体基因组序列的注释依据同源序列比对 (Genes) 及软件扫描 (tRNAs) 完成。

1.3 COX 基因系统进化分析

参与构建系统发育树的 22 个昆虫物种选自有翅类不同的分类单元(表 1),这些物种 COX 基因序列的检索及获取应用计算机软件 GenScalpel 1.0 (Yin *et al.*, 2012) 完成。对经 MEGA 比对后的 COX 1 ~ 3 基因进行拼接,拼接处理后序列长度为 2943 bp。分子系统树的构建采用贝叶斯推断法 (BI),由 MrBayes 3.1.2 (Huelsenbeck & Ronquist, 2001) 完成。分析世代为 3 000 000 代,最佳模型为 GTR + I + G (Posada & Buckley, 2004)。ML 树的构建由 RAxML-7.0.3 (Stamatakis *et al.*, 2008) 完成。

表 1 参与系统发育构建的昆虫物种及其线粒体基因组序列记录
Table 1 The insects species for phylogenetic analysis and their mitogenomes

物种	中文名	分类(目:科)	GenBank 登录号
<i>Damaster mirabilissimus</i>	-	Coleoptera: Coleoptera	NC_016469
<i>Rhopaea magnicornis</i>	-	Coleoptera: Scarabaeidae	NC_013252
<i>Chironomus tepperi</i>	-	Diptera: Chironomidae	NC_016167
<i>Drosophila yakuba</i>	-	Diptera: Drosophilidae	NC_001322
<i>Cydistomyia duplonotata</i>	-	Diptera: Tabanidae	NC_008756
<i>Ephemera orientalis</i>	东方蜉蝣	Ephemeroptera: Ephemeridae	NC_012645
<i>Laodelphax striatellus</i>	灰飞虱	Hemiptera: Delphacidae	NC_013706
<i>Apis cerana</i>	东方蜜蜂	Hymenoptera: Apidae	NC_014295
<i>Coptotermes formosanus</i>	家白蚁	Isoptera: Rhinotermitidae	NC_015800
<i>Bombyx mori</i>	家蚕	Lepidoptera: Bombycidae	NC_002355
<i>Cnaphalocrocis medinalis</i>	稻纵卷叶螟	Lepidoptera: Crambidae	JQ305693
<i>Phthonandria atrilineata</i>	桑尺蠖	Lepidoptera: Geometridae	NC_010522
<i>Kallima inachus</i>	枯叶蛱蝶	Lepidoptera: Nymphalidae	NC_016196
<i>Teinopalpus aureus</i>	金斑喙凤蝶	Lepidoptera: Papilionidae	NC_014398
<i>Pieris rapae</i>	菜粉蝶	Lepidoptera: Pieridae	NC_015895
<i>Manduca sexta</i>	-	Lepidoptera: Sphingidae	NC_010266
<i>Neopanorpa pulchra</i>	丽新蝎蛉	Mecoptera: Mecoptera	NC_013180
<i>Chrysoperla nipponensis</i>	日本通草蛉	Neuroptera: Chrysopidae	NC_015093
<i>Davidius lunatus</i>	新月戴春蜓	Odonata: Gomphidae	NC_012644
<i>Schistocerca gregaria</i>	沙漠蝗	Orthoptera: Acrididae	NC_013240
<i>Atractomorpha sinensis</i>	短额负蝗	Orthoptera: Pyrgomorphidae	NC_011824
<i>Pteronarcys princeps</i>	-	Plecoptera: Pteronarcyidae	NC_006133

2 结果和讨论

2.1 稻纵卷叶螟线粒体基因组

稻纵卷叶螟线粒体基因组全长 15 377 bp, 共编码 NADH 脱氢酶亚基 (ND1 ~ 5; ND4L) 等 13 个蛋白编码基因, tRNA^{Met} 等 22 个转运 RNA 基因及 2 个核糖体 RNA。非编码区域包括了基因间隙与长度为 340 bp 的控制区。基因组 AT 含量为 82%, 控制区达到 95.9%, 这一比例显示出有翅类昆虫对 AT 碱基的偏倚, 在已报道的鳞翅目物种中远高于平均水平 (Liao *et al.*, 2010)。

编码基因之间、编码基因与 tRNA 之间共在 8 个区域发生有不同程度的序列重叠, 显示了后生动物线粒体基因组结构功能的极大简约。控制区位于基因 s-rRNA 与 tRNA^{Met} 之间, 存在由元件 ATAG 引导的主 poly (T) 延伸、由 ATTTA 引导的短序列重复片段 (AT)₁₂ 等亚结构, 这些不断在鳞翅目物种中发现的控制区片段, 被推测与基因组复制及转录的调控有重要关联 (Cameron & Whiting, 2008)。

2.2 终止密码子注释

在 mtDNA 序列注释中, 蛋白编码基因 (PCGs)

起始、终止密码子的确定是一大难题。在已报道的鳞翅目昆虫线粒体基因中, 起始密码子除 COX1 为 CGA 外大多为 ATN, GTG 与 TTG 也有报道 (Hong *et al.*, 2008; Salvato *et al.*, 2008), 这可以作为起始密码子判断的一个重要依据。

无脊椎动物线粒体基因的终止密码子是的, 即 TAR。然而受 mtDNA 功能与结构极大简约所致, 许多基因表现为不完整的终止信号, 碱基形式为 T-或 TA-。这一现象的分子机理已经阐明 (Boore, 1999), 但在实际的序列处理中仍无任何依据规范对终止密码子的注释。针对一些长期存在于终止密码子确定的争议, 本文建议遵循以下原则。在 mtDNA 中, 拥有不完整终止密码子的基因其 3' 端往往紧邻一个 tRNA 基因: 这时, 如果基因以 T-结尾, 则该基因完整的终止密码子由其 mRNA 转录后形成的 poly (A) 结构来补充完整 (图 1, A); 如果基因以 TA-结尾, 且 tRNA5' 端的第一位碱基为 R (A 或 G), 则定义这一基因的终止密码子是完整的, 其第三密码子与 tRNA 的第一位碱基 A 或 G 重叠 (图 1, B); 如果基因以 TA-结尾, 但 tRNA5' 端的第一位碱基不为 R 时, 则按第一类处理 (图 1, C)。

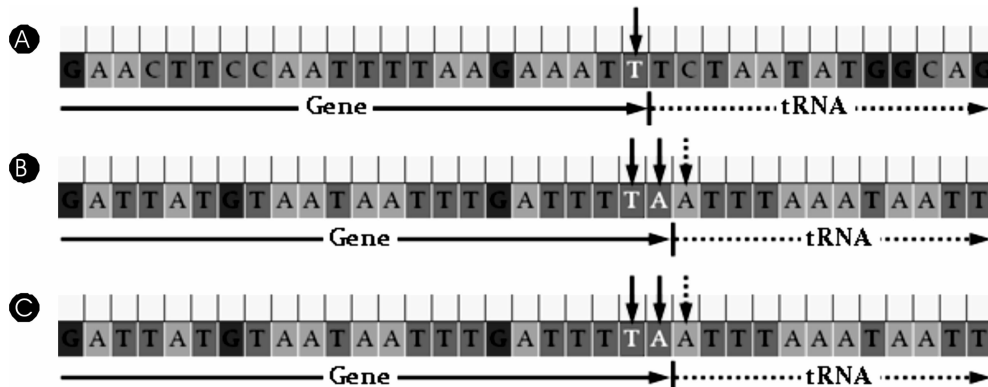


图 1 包含不完整终止密码子基因的注释
Fig. 1 The annotation proposals for genes contained incomplete termination codons

经验证, 这一原则广泛适用于动物 mtDNA 终止密码子的确定, 唯一的例外出现于部分物种的 ND4 与 ND4L 基因之间 (Clary & Wolstenholme, 1985; Yukuhiro *et al.*, 2002), 其中 ND4 拥有不完整的终止密码子 T-或 TA-。这些情况下, ND4 下游紧邻的为蛋白编码基因而非 tRNA 基因, 由于分子机理的相同, 以上的原则同样适用于对 ND4 基因不完整终止密码子的解释。

经注释后的稻纵卷叶螟线粒体基因组序列发布于 GenBank, 登录号为 JQ305693。

2.3 COX 基因及系统进化分析

稻纵卷叶螟 COX 基因 1 ~ 3 位于 J-链 tRNA^{Tyr} 与 tRNA^{Leu(UUR)}, tRNA^{Leu(UUR)} 与 tRNA^{Lys} 及 ATP6 与 tRNA^{Gly} 之间, 编码长度分别为 1531、685 及 789 bp。

由 COX 基因构建的昆虫系统发育树如图 2 所示 (得到 ML 分析结果支持)。有趣的是, 这些源自不同分类单元的昆虫物种在树状拓扑结构上, 与有翅类昆虫 Pterygota 翅型的分化 (许再福, 2009) 形成高度吻合: 东方蜉蝣与新月戴春蜓对古翅类 Palaeoptera 物种的支持; 家白蚁、沙漠蝗与短额负蝗

Pteronarcys princeps 代表了多新翅部 Polyneoptera 的重要分支等翅目、直翅目和襁翅目;其余物种除灰飞

虱外,均归于新翅次纲 Neoptera 内生翅类 Endopterygota。

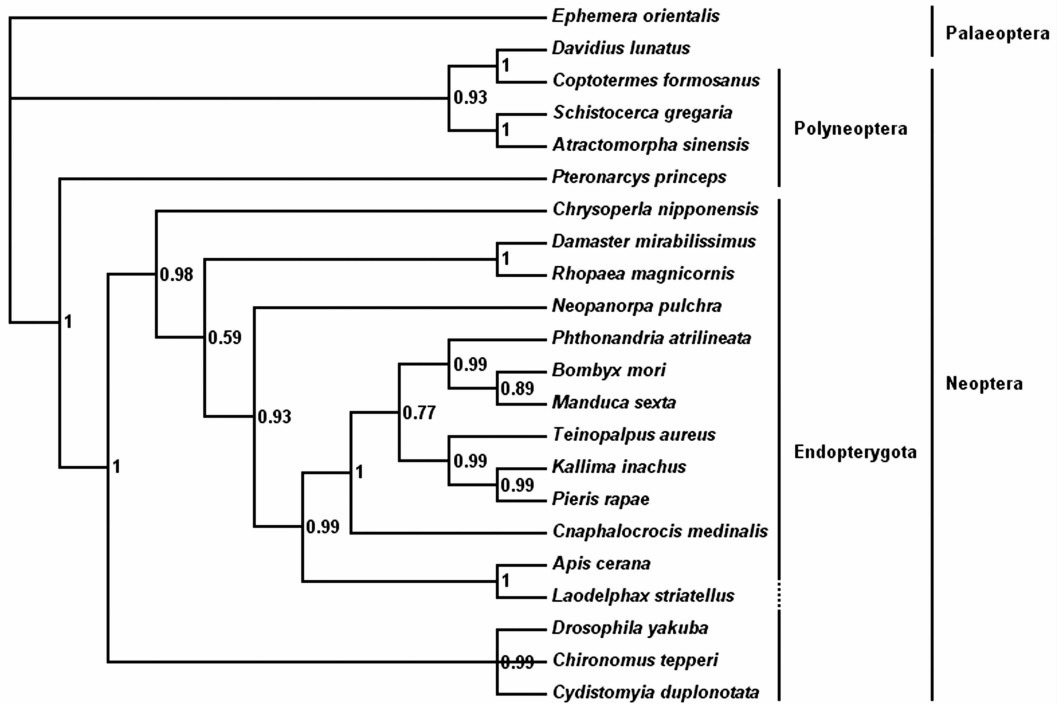


图 2 由 COX 基因构建的昆虫系统发育树(贝叶斯分析法)
Fig. 2 Insects phylogenetic tree reconstructed from COX gene using BI analysis

稻纵卷叶螟与桑尺蠖、菜粉蝶等其他鳞翅目物种形成相互姐妹群,与东方蜜蜂及灰飞虱组成独立的单系群,显示了新翅类昆虫演化中,内生翅类与副新翅类 Paraneoptera 物种的分化。内生翅类昆虫的出现,使昆虫由陆地发展到空间,是对历史上石炭纪到二迭纪南北半球复杂恶劣生境的积极适应(雷朝亮,荣秀兰,2003),被认为是昆虫发展史上的一次飞跃(吴千红等,1991)。

3 参考文献

高月波,陈晓,陈钟荣,等. 2008. 稻纵卷叶螟(*Cnaphalocrocis medinalis*)迁飞的多普勒昆虫雷达观测及动态[J]. 生态学报, 28(11): 5238~5247.
黄朝晖,王金福. 2003. 3种蚊虫线粒体CO II基因的分子进化[J]. 浙江大学学报(理学版), 30(4): 457~460.
雷朝亮,荣秀兰. 2003. 普通昆虫学[M]. 北京:中国农业出版社: 4~6, 445~446.
林万华,黎宗展,陈振鹏,等. 2006. 昆虫全基因组DNA的保存及提取[J]. 广西师范大学学报(自然科学版), 24(2): 86~88.
任竹梅,马恩波,郭亚平. 2002. 蝗总科部分种类Cyt b基因序列及系统进化研究[J]. 遗传学报, 29(4): 314~321.
王戎疆,万宏,龙玉,等. 2004. 利用线粒体CO II基因序列对中国尾蚱蝶属系统分化的研究(鳞翅目:蛱蝶科)[J]. 昆虫学报, 47(2): 243~247.

吴千红,邵则信,苏德明. 1991. 昆虫生态学实验[M]. 上海:复旦大学出版社: 38~39.
许再福. 2009. 普通昆虫学[M]. 北京:科学出版社: 220~222, 270~271.
Boore JL. 1999. Animal mitochondrial genomes[J]. Nucleic Acids Res, 27(8): 1767~1780.
Cameron SL, Whiting MF. 2008. The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta* (Insecta; Lepidoptera; Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths[J]. Gene, 408(1-2): 112~123.
Clary DO, Wolstenholme DR. 1985. The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code[J]. J Mol Evol, 22(3): 252~271.
Hong MY, Lee EM, Jo YH, et al. 2008. Complete nucleotide sequence and organization of the mitogenome of the silk moth *Caligula boisduvalii* (Lepidoptera: Saturniidae) and comparison with other lepidopteran insects[J]. Gene, 413(1-2): 49~57.
Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees[J]. Bioinformatics, 17(8): 754~755.
Kristensen NP. 1999. Lepidoptera, Moths and Butterflies [M]. New York (NY): Walter de Gruyter, Inc.
Lalitha S. 2000. Primer Premier 5[J]. Biotech Software & Internet Report, 1(6): 270~272.
Liao F, Wang L, Wu S, et al. 2010. The complete mitochondrial genome of the fall webworm, *Hyphantria cunea* (Lepidoptera: Arctiidae)[J]. Int J Biol Sci, 6(2): 172~186.

(下转第 50 页)

