论文

印度谷螟线粒体基因组测定及螟蛾总科系统发育分析

唐培安^①, 吴海晶^①, 王娟^②, 郭仲龙^②, 袁明龙^{②*}^① 南京财经大学食品科学与工程学院, 江苏省现代粮食流通与安全协同创新中心, 南京 210023;^② 兰州大学草地农业科技学院, 草地农业生态系统国家重点实验室, 兰州 730020

* 联系人, E-mail: yuanml@lzu.edu.cn

收稿日期: 2015-06-29; 接受日期: 2015-08-02; 网络发表日期: 2015-09-16

粮食公益性行业科研专项(批准号: 201413007-2)、国家科技支撑计划(批准号: 2013BAD17B01)、国家自然科学基金(批准号: 31000828)和江苏省高校优势学科建设工程资助项目

doi: 10.1360/N052015-00193

摘要 线粒体基因组广泛应用于昆虫系统进化、种群遗传学及谱系地理学等众多研究领域. 印度谷螟(*Plodia interpunctella*)属鳞翅目螟蛾科, 是世界性重要仓储害虫. 为了深入了解螟蛾总科昆虫的系统进化关系, 本研究测定并分析了印度谷螟的全线粒体基因组. 印度谷螟线粒体基因组全长 15264 bp, 包括 13 个蛋白编码基因、22 个转运 RNA 基因、2 个核糖体 RNA 基因和一个大的非编码区(线粒体控制区). 比较线粒体基因组学分析表明, 线粒体基因组的基因排序、碱基组成、密码子使用、控制区的序列结构以及转运 RNA 基因的二级结构, 在已测的 25 种螟蛾总科昆虫中高度保守. 线粒体蛋白编码基因的进化模式分析表明, 13 个基因的进化速率在螟蛾科和草螟科间存在明显差异. 基于 13 个蛋白编码基因对螟蛾总科的 9 个亚科的系统发育分析显示, 贝叶斯树和最大似然树均高度支持螟蛾科和草螟科为 2 个单系群. 在螟蛾科中, 斑螟亚科与丛螟亚科具有更近的亲缘关系. 草螟科包括 2 个分支, 一支由野螟亚科和斑野螟亚科构成, 另一支包括草螟亚科、水螟亚科、苔螟亚科和禾螟亚科等 4 个亚科.

关键词昆虫
鳞翅目
螟蛾科
草螟科
线粒体基因组
系统发育

昆虫线粒体基因组是一个闭合双链环状 DNA 分子, 通常编码 37 个基因, 即 13 个蛋白编码基因, 2 个核糖体 RNA 基因(*rrnL* 和 *rrnS*)和 22 个转运 RNA 基因^[1,2]. 此外, 线粒体基因组通常还包含一个大的非编码区, 叫做控制区^[1,3]. 由于高的 A+T 含量, 昆虫控制区也叫做 A+T 富集区. 与单个线粒体基因组相比, 昆虫线粒体基因组不仅包含更多的序列信息, 还

包含一系列组学水平上的结构信息, 如基因重排和 RNA 二级结构^[4-7]. 由于具有母系遗传、多拷贝、缺少遗传重组以及进化速率快等特点, 线粒体基因组序列已广泛应用于分子进化、系统发育重建、谱系地理学及种群遗传学等众多领域, 成为研究昆虫遗传与进化最重要的分子标记^[2,8-10].

螟蛾总科(Pyraloidea)包括螟蛾科(Pyralidae)和草

引用格式: 唐培安, 吴海晶, 王娟, 等. 印度谷螟线粒体基因组测定及螟蛾总科系统发育分析. 中国科学: 生命科学, 2015, 45: 890-900

Tang P A, Wu H J, Wang J, et al. Complete mitochondrial genome of *Plodia interpunctella* and phylogenetic analysis of Pyraloidea (Insecta: Lepidoptera). SCIENTIA SINICA Vitae, 2015, 45: 890-900, doi: 10.1360/N052015-00193

螟科(Crambidae), 是鳞翅目(Lepidoptera)昆虫中的第三大类群, 物种数量仅次于夜蛾总科(Noctuoidea)和尺蛾总科(Geometroidea)^[11]. 目前, 全世界已描述的螟蛾总科昆虫超过 16000 种^[11], 中国已记录 2000 余种^[12]. 其中, 很多种类是重要的农林牧业害虫, 如迁飞性的草地螟(*Loxostege sticticalis*), 钻蛀性的二化螟(*Chilo suppressalis*)、三化螟(*Tryporyza incertulas*)、桃蛀螟(*Dichocrocis punctiferalis*)、亚洲玉米螟(*Ostrinia nubilalis*)和印度谷螟(*Plodia interpunctella*)等. 尽管螟蛾具有重要的经济意义, 但有关其系统进化关系的研究还很薄弱. 前面少数几项研究采用形态学数据对螟蛾总科进行了支序分析, 但不同的研究者提出了不同的系统发育假说, 且大多亚科间的系统发育关系仍不明确^[13]. Regier 等人^[13]基于 19 个核基因对螟蛾总科 21 个亚科中的 18 个进行了系统发育研究, 但结果很大程度上与前面基于形态学数据提出的系统发育假说不一致. Cao 等人^[14]基于线粒体基因组数据对螟蛾总科进行了系统发育分析, 但该研究仅包含 8 个物种, 且仅有 1 种来自螟蛾科. 截止 2015 年 6 月, GenBank 数据库中共收录了 24 种螟蛾的完整或部分线粒体基因组序列, 这为基于线粒体基因组数据研究螟蛾总科的系统进化提供了良好材料. 但另一方面, 目前已测线粒体基因组的螟蛾中, 大多数种类属于草螟科, 而仅有 3 种来自螟蛾科.

印度谷螟(*Plodia interpunctella*)属螟蛾总科螟蛾科, 是一种世界性重要的仓储害虫. 该虫食性极广, 主要以幼虫对各种粮食、坚果、干果、中药材及巧克力等进行危害, 喜食粮食的胚部与表皮, 并在粮堆内部和表面吐丝结网, 连缀粮粒及排泄物, 造成粮食结块变质, 大量发生时还会引起粮食内部发热, 对储藏物的安全构成严重威胁^[15,16]. 为了进一步理解螟蛾总科的系统进化关系, 本研究测定了印度谷螟的全线粒体基因组序列, 连同前面已测的 24 种螟蛾总科昆虫进行比较线粒体基因组学分析, 并采用贝叶斯推断和最大似然法构建基于线粒体基因组数据的螟蛾总科的系统进化关系.

1 材料与方法

1.1 基因组 DNA 提取、PCR 扩增及测序

印度谷螟样本采自江苏省吴江市, 保存于 -80°C .

采用 DNA 提取试剂盒(Omega, 美国)提取单头印度谷螟的基因组总 DNA. 采用昆虫通用引物及新设计的特异引物扩增印度谷螟全线粒体基因组序列, 引物序列及来源见表 1. PCR 扩增采用 LA Taq(TaKaRa, 日本)在 $25\ \mu\text{L}$ 体系中进行, 扩增程序为: 92°C 起始变性 2 min; 92°C 变性 10 s, $48^{\circ}\text{C}\sim 55^{\circ}\text{C}$ 退火 1 min, 68°C 延伸 1~4 min, 共进行 35 个循环; 68°C 继续延伸 20 min. 所有扩增获得的片段, 在 1.5% 的琼脂糖凝胶上电泳检测. 目的片段纯化后, 送至上海生工生物工程技术服务有限公司直接进行双向测序(测序引物同 PCR 扩增), 或克隆至 pEASY-T1 vector(北京全式金生物技术有限公司)后送菌液进行双向测序.

1.2 序列分析

测序获得的 DNA 序列, 采用 BioEdit 7.0.9.0^[17] 进行校正并拼接. 印度谷螟线粒体基因组的注释包括 13 个蛋白编码基因、2 个 rRNA 基因及 22 个 tRNA 基因的鉴定, 均采用文献[18]的方法进行. 碱基组成及密码子使用采用 Mega 6.06^[19]进行分析, 同义突变(Ks)及非同义突变(Ka)采用 DnaSP 5.0^[20]进行分析. 2 条链碱基组成不对称通过以下公式进行计算: $\text{AT-偏斜} = [\text{A}-\text{T}]/[\text{A}+\text{T}]$, $\text{GC-偏斜} = [\text{G}-\text{C}]/[\text{G}+\text{C}]$ ^[21].

1.3 系统发育分析

为了探讨螟蛾总科的系统进化关系, 从 GenBank 数据库中下载了已测的 24 个螟蛾总科昆虫的线粒体基因组(表 2), 并以美国白蛾(*Hyphantria cunea*)^[22]、舞毒蛾(*Lymantria dispar*)^[23]和门源草原毛虫(*Gynaephora menyanensis*)^[24]等 3 种夜蛾总科昆虫作为外群. 为了明确线粒体基因组数据是否存在饱和性, 采用 PAUP 软件计算未校正的 P 遗传距离与一般时间可逆(general time reversible, GTR)遗传距离, 并作散点图. 结果显示, 蛋白编码基因的 3 个密码子位点均未达到显著饱和, 而 2 个 rRNA 基因均显著饱和. 因此, 13 个蛋白编码基因的所有 3 个密码子位点均用来构建螟蛾总科的系统发育关系. 采用 TranslatorX online server 的 MAFFT^[25]对 13 个蛋白编码基因分别进行序列比对, 并采用 GBlocks 去除空位及模糊位点. 将比对好的单个蛋白编码基因联合在一起, 获得一个数据集. 采用 PartitionFinder 1.1.1^[26]来选择数据集最佳的 partitioning schemes 及其相应的核苷酸和氨基酸进化模型. 输入文件中的

表 1 印度谷螟线粒体基因组 PCR 扩增引物序列及来源

片段	引物名称	引物序列(5'→3')	退火温度(°C)	引物来源
1	C1-J1709	AATTGGWGGWTTYGGAAAYTG	53	文献[8]
	C1-N2776	GATAATCTGAGTATCGWCGNGG		
2	C1-J2756	ACATTCTTTCCTCARCAYTT	50	文献[8]
	C2-N3665	CCACAAATTTCTGAACATTG		
3	C2-J3399	ACAATTGGTCAYCAATGATAYTG	53	文献[8]
	C3-N4908	CGTGAAAYATCTCGTCATCATTG		
4	C3-J4792	GTTGATTATAGACCWTGRCC	50	文献[8]
	TN-N6160	TCAATTTTRTCATTAACAGTGA		
5	PIC-F1	CAATAATTTGCATTTAATTTTCGAC	56	本研究
	PIN-R1	TGCTGGAATATTAAGTCCCTA		
6	N5-J7572	AAAGGGAATTTGAGCTCTTTTWGT	50	文献[8]
	N4-N8727	AAATCTTTRATTGCTTATTCWTC		
7	PIN-F1	CCAGAAGAACATAATCCATGACC	56	本研究
	PIH-R1	GGCATTATCTACGGCAAATCC		
8	CB-J10933	GTTCTACCTTGAGGNCAAATRTC	53	文献[8]
	N1-N12595	GTWGCTTTTTTAACTTTATTRGARCG		
9	N1-J12261	TACTTCATAAGAAATAGTYTGRGC	50	文献[8]
	LR-N13000	TTACCTTAGGGATAACAGCGTAA		
10	LR-J12888	CCGGTTTGAACCTCARATCATGTA	55	文献[8]
	SR-N14220	ATATGYACAYATTGCCCGTC		
11	SR-J14197	TACCYCTACTTTGTACGACTT	50	文献[8]
	TM-N200	TCCTTTATATRTGAGGTATGARCC		
12	PIK-F2	AATTTAAGCTTTTGGGTTCATAC	56	本研究
	PID-R1	TTCTGCCCCATTTTCTACA		

设置为: 每一蛋白编码基因的每一位点为单独的 partition, 贝叶斯信息标准 (Bayesian information criterion, BIC), “greedy”和“unlinked”. 由 Partition-Finder 选择的最佳 partition 及进化模型用于后续系统发育分析.

系统发育分析采用最大似然法 (maximum likelihood, ML) 和贝叶斯推断 2 种方法进行, 均在 CIPRES Science Gateway 3.3^[27] 在线平台上进行. ML 分析采用 RAxML-HPC2 on XSEDE 8.0.24^[28] 进行, 采用 GTRGAMMAI 模型, 支的可靠性采用 1000 次的 bootstraps (BS) 进行评估. 贝叶斯分析采用 MrBayes 3.2.2^[29] 进行, 4 条独立的马尔可夫链 (Markov chains), 即 3 条热链 (hot chain) 和 1 条冷链 (cold chain) 同时运行 1×10^8 代 (generation). 每运行 1000 带取样一次, 当 ESS 大于 100 且 PSRF (potential scale reduction factor) 接近 1.0 时即认为 2 个分析过程趋于稳定状态^[29]. 舍弃 25% 的老化样本, 剩余样本用来构建 50% 一致树, 并计算贝叶斯后验概率 (posterior probability, PP).

2 结果与讨论

2.1 基因组大小、含量及结构

印度谷螟线粒体基因组是一个典型的闭合环状双链 DNA 分子 (GenBank 登录号: KT207942), 包含 13 个蛋白编码基因、2 个 rRNA 基因、22 个 tRNA 基因和 1 个控制区 (图 1 和表 3). 该线粒体基因组全长 15264 bp, 大小介于其他已测螟蛾总科昆虫全线粒体基因组的大小, 即 14960 (桑绢野螟 (*Glyphodes pyloalis*))~15490 bp (小蔗螟 (*Diatraea saccharalis*)) (表 3). 与节肢动物线粒体基因原始排序相比, 印度谷螟线粒体基因组中仅 *trnI-trnQ-trnM* 重排为 *trnM-trnI-trnQ* (图 1), 这与目前已测的绝大多数鳞翅目昆虫的线粒体基因排序完全一致.

印度谷螟线粒体基因组是比较紧凑的, 除控制区外, 仅有少量小的非编码区, 即基因间隔区. 基因间隔区共有 14 处, 大小为 1~41 bp, 共计 166 bp, 其

表 2 已测螟蛾总科昆虫线粒体基因组基本特征^{a)}

科	物种	GenBank 登录号	线粒体基因组				蛋白编码基因		rrnL		rrnS		控制区	
			大小 (bp)	(A+T)%	AT-skew	GC-skew	大小 (bp)	(A+T)%	大小 (bp)	(A+T)%	大小 (bp)	(A+T)%	大小 (bp)	(A+T)%
Crambidae	<i>Chilo auricilius</i>	NC_024644	15367	82.03	-0.008	-0.197	11181	80.56	1346	84.92	789	86.44	337	94.07
Crambidae	<i>Chilo suppressalis</i>	NC_015612	15395	80.67	0.008	-0.235	11196	78.83	1383	84.24	788	86.17	348	95.40
Crambidae	<i>Cnaphalocrocis medinalis</i>	NC_015985	15388	81.94	-0.015	-0.175	11178	80.51	1389	84.88	781	86.17	339	95.87
Pyralidae	<i>Corcyra cephalonica</i>	NC_016866	15273	80.43	-0.036	-0.218	11160	78.85	1360	84.49	793	86.00	351	96.58
Crambidae	<i>Diatraea saccharalis</i>	NC_013274	15490	80.02	0.021	-0.258	11148	78.89	1355	82.95	778	85.86	335	94.93
Crambidae	<i>Dichocrocis punctiferalis</i>	NC_021389	15355	80.60	-0.025	-0.207	11172	77.83	1412	84.77	781	85.53	338	96.45
Crambidae	<i>Elophila interruptalis</i>	NC_021756	15351	80.32	-0.011	-0.229	11148	78.56	1367	84.13	786	85.50	339	93.51
Pyralidae	<i>Ephestia kuehniella</i>	NC_022476	15295	79.77	-0.049	-0.234	11145	78.12	1328	84.19	773	84.86	321	93.15
Crambidae	<i>Eudonia angustea</i>	KJ508052*	15386	-	-	-	11116	79.87	-	-	-	-	-	-
Crambidae	<i>Glyphodes pyloalis</i>	KM576860	14960	80.77	-0.016	-0.194	11166	79.59	1345	84.46	784	85.97	67	91.04
Crambidae	<i>Glyphodes quadrimaculalis</i>	NC_022699	15255	80.80	-0.007	-0.192	11154	79.23	1350	84.89	779	85.49	327	94.50
Crambidae	<i>Hellula undalis</i>	KJ636057*	14678	-	-	-	11175	78.56	1318	83.61	779	86.01	-	-
Pyralidae	<i>Lista haraldusalis</i>	NC_024535	15213	81.52	-0.007	-0.171	11172	80.00	1335	85.17	786	86.13	310	96.13
Crambidae	<i>Loxostege sticticalis</i>	KR080490	15218	80.82	0.002	-0.191	11169	79.51	1334	83.66	774	86.05	-	-
Crambidae	<i>Maruca testulalis</i>	NC_024283	15110	80.81	-0.005	-0.171	11160	79.42	1253	85.55	783	85.44	335	92.84
Crambidae	<i>Maruca vitrata</i>	NC_024099	15385	80.70	-0.002	-0.172	11316	79.32	1304	84.43	765	85.23	341	92.96
Crambidae	<i>Nomophila noctuella</i>	NC_025764	15309	81.41	0.002	-0.176	11130	79.80	1366	84.77	720	84.72	426	94.84
Crambidae	<i>Ostrinia furnacalis</i>	NC_003368*	14536	-	-	-	11156	79.37	1339	84.99	435	82.76	-	-
Crambidae	<i>Ostrinia nubilalis</i>	NC_003367*	14535	-	-	-	11154	79.10	1339	84.91	434	82.03	-	-
Crambidae	<i>Ostrinia penitalis</i>	KM395814*	12612	-	-	-	10738	78.41	589	78.27	-	-	-	-
Crambidae	<i>Paracymoriza distinctalis</i>	NC_023471	15354	82.27	-0.002	-0.155	11154	81.01	1390	84.75	784	86.35	351	95.16
Crambidae	<i>Paracymoriza prodigalis</i>	NC_020094	15326	81.53	0.002	-0.183	11151	80.06	1389	85.53	781	85.92	343	95.34
Pyralidae	<i>Plodia interpunctella</i>	KT207942	15264	80.12	-0.050	-0.233	11187	78.54	1337	83.47	782	84.91	320	95.63
Crambidae	<i>Tryporyza incertulas</i>	NC_021413	15223	77.05	0.031	-0.324	11160	74.80	1314	82.04	768	84.11	403	92.80
Crambidae	<i>Tyspanodes hypsalis</i>	NC_025569	15329	81.41	-0.017	-0.175	11163	79.92	1365	84.84	791	85.59	350	95.43

a) *表示不完整的线粒体基因组

图 1 印度谷螟线粒体基因组结构图
下画线表示该基因由负链编码

中最大的位于 *trnQ* 和 *nad2* 之间(图 1 和表 3). 此外, 还存在 5 个基因重叠区, 大小为 1~8 bp, 其中最大的位于 *trnW* 和 *trnC* 之间(图 1 和表 3).

2.2 碱基组成与密码子使用

与其他已测鳞翅目及螟蛾总科昆虫相似, 印度谷螟线粒体基因组的碱基组成明显偏向于 AT, 全基因组正链的 AT 含量为 80.12%(表 2). A 与 T 数目差异不大(AT-偏斜=-0.050), 而 C 明显高于 G(GC-偏斜=-0.233). 对线粒体基因组的不同部分进行比较分析发现, 在所有螟蛾总科中均是控制区的 AT 含量最高(91.04%~96.58%), 而 13 个蛋白编码基因的最低(74.80%~81.01%)(表 2). 印度谷螟蛋白编码基因密码子第 3 位点的 AT 含量(92.33%), 显著高于第 1(73.34%)和第 2 位点(69.94%).

印度谷螟线粒体基因组高的 A+T 含量及链组成偏斜性, 也反映在蛋白编码基因的密码子使用上. 相

表 3 印度谷螟线粒体基因组注释结果^{a)}

基因	位置	大小(bp)	基因间隔区	起始密码子	终止密码子	反密码子
<i>trnM</i>	1~68	68	0			CAU
<i>trnI</i>	71~134	64	2			GAU
<i>trnQ</i>	141~209	69	6			UUG
<i>nad2</i>	251~1264	1014	41	ATT	TAA	
<i>trnW</i>	1291~1356	66	26			UCA
<i>trnC</i>	1349~1415	67	-8			GCA
<i>trnY</i>	1417~1482	66	1			GUA
<i>cox1</i>	1487~3017	1531	4	CGA	T	
<i>trnL2(UUR)</i>	3018~3084	67	0			UAA
<i>cox2</i>	3085~3766	682	0	ATT	T	
<i>trnK</i>	3767~3837	71	0			CUU
<i>trnD</i>	3837~3903	67	-1			GUC
<i>atp8</i>	3904~4065	162	0	ATC	TAA	
<i>atp6</i>	4059~4738	680	-7	ATG	TA	
<i>cox3</i>	4739~5527	789	0	ATG	TAA	
<i>trnG</i>	5530~5596	67	2			UCC
<i>nad3</i>	5597~5950	354	0	ATT	TAA	
<i>trnA</i>	5963~6027	65	12			UGC
<i>trnR</i>	6027~6090	64	-1			UCG
<i>trnN</i>	6088~6152	65	-3			GUU
<i>trnS1(AGN)</i>	6158~6224	67	5			GCU
<i>trnE</i>	6225~6292	68	0			UUC
<i>trnF</i>	6296~6363	68	3			GAA
<i>nad5</i>	6364~8098	1735	0	ATT	T	
<i>trnH</i>	8099~8162	64	0			GUG
<i>nad4</i>	8163~9519	1357	0	ATA	T	
<i>nad4L</i>	9520~9810	288	3	ATG	TAA	
<i>trnT</i>	9813~9876	64	2			UGU
<i>trnP</i>	9877~9942	66	0			UGG
<i>nad6</i>	9945~10478	534	2	ATT	TAA	
<i>cob</i>	10511~11662	1152	32	ATG	TAA	
<i>trnS2(UCN)</i>	11666~11730	65	3			UGA
<i>nad1</i>	11753~12691	939	22	ATT	TAG	
<i>trnL1(CUN)</i>	12692~12758	67	0			UAG
<i>rrnL</i>	12759~14095	1337	0			
<i>trnV</i>	14096~14162	67	0			UAC
<i>rrnS</i>	14163~14944	782	0			
控制区	14945~15264	320	0			

a) 下画线表示该基因由负链编码

对同义密码子使用频率(relative synonymous codon usage, RSCU)分析表明, 螟蛾科和草螟科的密码子使用没有显著差异(图 2). 在所有已测螟蛾总科线粒体基因组中, 以 A 或 T 结尾的同义密码子出现的次数, 远大于其他同义密码子出现的次数, 表明前者被严重过量使用. 例如, 在印度谷螟线粒体基因组中, 编码甲硫氨酸(Met)的同义密码子中, 密码子 AUG 仅使用了 18 次, RSCU 为 0.13, 但密码子 AUA 使用了 269 次, RSCU 为 1.87. 在无脊椎动物的 62 种线粒体密码

子中, CUG, ACG, AGC 和 GGC 等 4 个富含 GC 的密码子未被印度谷螟线粒体基因组使用. 除三化螟外, 至少有 1 个富含 GC 的密码子在其余 23 种螟蛾线粒体基因组中未被使用.

2.3 蛋白编码基因

印度谷螟线粒体基因组的 13 个蛋白编码基因全长 11187 bp(去除终止密码子), 共编码 3729 个氨基酸, 这与其他已测螟蛾总科的非常接近(表 2). 在 13

图2 螟蛾科和草螟科密码子使用比较分析

个蛋白编码基因中, AT 含量最低的是 *cox1* (71.70%), 最高的是 *atp8* (86.79%). 除 *cox1* 使用 CGA 外, 其余 12 个蛋白编码基因的起始密码子均为 ATN, 其中 1 个基因(*nad4*)使用 ATA, 1 个基因(*atp8*)使用 ATC, 4 个基因使用 ATG(*atp6*, *cox3*, *nad4L* 和 *cob*), 6 个基因(*cox2*, *nad1*, *nad2*, *nad3*, *nad5* 和 *nad6*)使用 ATT(表 3). 印度谷螟 8 个线粒体蛋白编码基因的终止密码子为 TAA 或 TAG, 其余 5 个基因以 TA(*atp6*)或单个 T(*cox1*, *cox2*, *nad4* 和 *nad5*)作为终止密码子. 不完全的终止密码子在昆虫线粒体基因组中非常普遍, 在转录过程中通过添加 A 可以将这些不完整的终止密码子补全而并不影响翻译^[30,31].

为了分析 13 个蛋白编码基因在螟蛾总科中的进化模式是否存在差异性, 对螟蛾科和草螟科分别计算了 K_a , K_s 和 K_a/K_s (图 3). 结果表明, 25 种螟蛾的平

均 K_s 值, 在 13 个蛋白编码基因中非常相似; 而 K_a 值在 13 个基因间变异显著. 在 13 个基因中, *cox1*~3 和 *cob* 等 4 个基因的 K_a 在螟蛾科和草螟科中均进化最慢. *nad3* 和 *nad6* 在螟蛾科中具有最快的进化速率, 而草螟科中 *atp6* 和 *nad6* 的进化速率最快. 线粒体蛋白编码基因由于参与了细胞的氧化磷酸化过程, 通常在进化过程中受到强烈的净化选择(即 $K_a/K_s < 1$)而表现出功能约束性. 然而, 在螟蛾科中, *atp8*, *nad3* 和 *nad6* 等 3 个基因的 K_a/K_s 值均大于 1, 而在草螟科中 8 个基因的 K_a/K_s 值均大于 1(图 3). 这些结果表明, 螟蛾总科的线粒体蛋白编码基因在进化过程中受到了正选择作用抑或选择松弛, 且螟蛾科和草螟科的进化模式存在明显的不同. 另一方面, 不同蛋白编码基因间及螟蛾科和草螟科间的平均 K_s 值非常相似, 故 K_a/K_s 值大于 1 也可能是由于线粒体基因的同义突

图3 螟蛾总科 13 个线粒体蛋白编码基因的进化模式

变达到饱和所致。

2.4 RNA 基因

印度谷螟线粒体基因组的 2 个 rRNA 基因 *rrnL* 和 *rrnS* 的大小分别是 1337 和 782 bp, AT 含量分别为 83.47% 和 84.91%, 这与其他已测螟蛾总科的大小及碱基组成非常相似(表 2)。

印度谷螟线粒体基因组 22 个 tRNA 基因总长 1462 bp, 其中最长的为 71 bp, 最短的仅 64 bp, 平均长度为 (66.4 ± 1.8) bp。所有 tRNA 均能形成经典的三叶草结构, 包括氨基酸接受臂、双氢尿嘧啶(dihydrouridine, DHU)臂、T ψ C 臂和反密码子臂等 4 个臂和 1 个可变环(图 4)。部分 tRNA 的 DHU 臂(*trnC* 和 *trnS1*)、氨基酸接受臂(*trnL2* 和 *trnA*)、反密码子臂(*trnL2*、*trnK* 和 *trnS2*)和 T ψ C 臂(*trnE*)存在个别碱基错配, 这种现象在昆虫线粒体基因组中非常普遍。然而, 值得指出的是, 印度谷螟线粒体 *trnS1* 的 DHU 臂, 不仅存在 A:A 错配, 且长度仅为 3 bp、环仅由 3 个核苷酸构成, 暗示该茎环结构可能并不存在。事实上, *trnS1* 缺少 DHU 臂在已测后生动物中非常普遍^[2]。在其他已测的螟蛾总科昆虫中, *trnS1* 似乎亦缺少 DHU 臂。通常, 由于 *trnS1* 序列的高度变异, 要准确的预测其二级结构是比较困难的。

2.5 控制区

印度谷螟线粒体基因组的控制区位于 *rrnS* 和 *trnM* 之间, 长度为 320 bp, AT 含量为 95.63%(表 2)。印度谷螟线粒体控制区具有以下特征: (1) 在 *rrnS* 的下游具有“ATAGA”序列; (2) 18 bp 的 poly-T 结构; (3) (AT)_n 重复序列。这 3 个结构特征在其他 19 种已测螟蛾及大多数鳞翅目昆虫全线粒体基因组中广泛存在, 是鳞翅目昆虫线粒体控制区的保守性结构特征。研

究表明, poly-T 为线粒体 DNA 的复制起始提供了必要的信号识别位点^[32]。除桑绢野螟外, 其余螟蛾总科线粒体控制区均有 poly-T 结构。值得指出的是, 桑绢野螟线粒体控制区的长度在所有已测螟蛾中最小(表 2), 长度仅为 67 bp, 推测可能是其他结构元件(如茎环二级结构)为线粒体 DNA 的复制起始提供必要的信号^[32]。

2.6 系统发育分析

基于 13 个蛋白编码基因构建的最大似然树和贝叶斯树, 两个树的拓扑结构几乎完全一致, 仅在分支的支持率上存在差异性(图 5)。系统发育结果高度支持螟蛾科和草螟科均为单系群(PP=1.0, BS=100), 这符合目前普遍接受的螟蛾总科分类^[11,13]。在螟蛾科内, 斑螟亚科(Phycitinae)与丛螟亚科(Epipaschiinae)具有更近的亲缘关系, 两者相比蜡螟亚科(Galleriinae)处于更进化的系统位置, 符合形态学和分子数据的研究结果^[11,13]。草螟科分为两大分支, 其中一支由野螟亚科(Pyraustinae)和斑野螟亚科(Spilomelinae)构成, 且这两个亚科分别被高度支持为单系群(PP=1.0, BS>91)且互为姊妹群关系(PP=1.0, BS=100), 这不符合前面基于形态学系统分类的研究结果, 但与近期采用 19 个核基因的研究结果一致^[13]。草螟科的另一个分支包括草螟亚科(Crambinae)、水螟亚科(Nymphulinae)、苔螟亚科(Scopariinae)和禾螟亚科(Schoenobiinae)等 4 个亚科, 其中前两者具有更近的亲缘关系且均为单系群。贝叶斯树适度支持菜心野螟(*Hellula undalis*)与 *Eudonia angustea* 同属于苔螟亚科, 但最大似然树并不支持。在菜心野螟线粒体基因组学分析的原始文献中^[33], 将该虫错误地归于螟蛾科。尽管菜心野螟是否属于苔螟亚科以及该亚科的系统发育位置还需进一步研究,

图 4 印度谷螟 22 个线粒体 tRNA 基因的二级结构

但本研究的系统发育结果高度支持菜心野螟属于草螟科。禾螟亚科的系统发育位置也不确定，这与前面基于形态学和核基因数据的研究结果一致^[11,13]。禾螟亚科不确定的系统发育位置，也可能是由于目前该亚科仅有一个代表性物种的线粒体基因组以及长

枝吸引所致。

3 结论

本研究测定了重要仓储害虫印度谷螟的全线粒体

图5 基于13个蛋白编码基因构建的螟蛾总科系统发育树

左边为贝叶斯后验概率(PP), 右边为最大似然法的自展率(BS); *: PP=0.1, 且 BS=100; -: PP<0.5, 或 BS<50

基因组序列, 联合已测的 24 种螟蛾总科昆虫进行比较线粒体基因组学分析, 并基于 13 个蛋白编码基因构建

了螟蛾总科 9 个亚科间的系统进化关系, 为深入理解螟蛾总科及鳞翅目昆虫的谱系进化关系提供了基础数据.

参考文献

- 1 Boore J L. Animal mitochondrial genomes. Nucleic Acids Res, 1999, 27: 1767–1780
- 2 Cameron S L. Insect mitochondrial genomics: implications for evolution and phylogeny. Annu Rev Entomol, 2014, 59: 95–117
- 3 Wolstenholme D R. Animal mitochondrial DNA: structure and evolution. Int Rev Cytol, 1992, 141: 173–216
- 4 Boore J L. The use of genome-level characters for phylogenetic reconstruction. Trends Ecol Evol, 2006, 21: 439–446
- 5 Yuan M L, Wei D D, Wang B J, et al. The complete mitochondrial genome of the citrus red mite *Panonychus citri* (Acari: Tetranychidae): high genome rearrangement and extremely truncated tRNAs. BMC Genomics, 2010, 11: 597
- 6 Masta S E. Mitochondrial rRNA secondary structures and genome arrangements distinguish chelicerates: comparisons with a harvestman (Arachnida: Opiliones: *Phalangium opilio*). Gene, 2010, 449: 9–21
- 7 Downton M, Castro L R, Austin A D. Mitochondrial gene rearrangements as phylogenetic characters in the invertebrates: the examination of genome ‘morphology’. Invertebr Syst, 2002, 16: 345–356
- 8 Simon C, Buckley T R, Frati F, et al. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annu Rev Ecol Evol Syst, 2006, 37: 545–579

- 9 Avise J C. Phylogeography: retrospect and prospect. *J Biogeogr*, 2009, 36: 3–15
- 10 Wang I J. Recognizing the temporal distinctions between landscape genetics and phylogeography. *Mol Ecol*, 2010, 19: 2605–2608
- 11 Solis M. Phylogenetic studies and modern classification of the Pyraloidea (Lepidoptera). *Rev Colomb Entomol*, 2007, 33: 1–8
- 12 马晓静. 河南螟蛾总科昆虫的分类、区系及分布地理研究. 硕士学位论文. 郑州: 郑州大学, 2010
- 13 Regier J C, Mitter C, Solis M, et al. A molecular phylogeny for the pyraloid moths (Lepidoptera: Pyraloidea) and its implications for higher-level classification. *Syst Entomol*, 2012, 37: 635–656
- 14 Cao S S, Yu W W, Sun M, et al. Characterization of the complete mitochondrial genome of *Tryporyza incertulas*, in comparison with seven other Pyraloidea moths. *Gene*, 2014, 533: 356–365
- 15 Hoflehner E, Binder M, Hemmer W, et al. Thioredoxin from the indianmeal moth *Plodia interpunctella*: cloning and test of the allergenic potential in mice. *PLoS One*, 2012, 7: e42026
- 16 Campos M, Phillips T W. Contact toxicity of insecticides for attract-and-kill applications against adult *Plodia interpunctella* (Hübner) (Lepidoptera: Pyralidae). *Pest Manag Sci*, 2010, 66: 752–761
- 17 Hall T A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Sym Ser*, 1999, 41: 95–98
- 18 Yuan M L, Zhang Q L, Guo Z L, et al. The complete mitochondrial genome of *Corizus tetraspilus* (Hemiptera: Rhopalidae) and phylogenetic analysis of Pentatomomorpha. *PLoS One*, 2015, 10: e0129003
- 19 Tamura K, Stecher G, Peterson D, et al. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*, 2013, 30: 2725–2729
- 20 Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 2009, 25: 1451–1452
- 21 Perna N T, Kocher T D. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J Mol Evol*, 1995, 41: 353–358
- 22 Liao F, Wang L, Wu S, et al. The complete mitochondrial genome of the fall webworm, *Hyphantria cunea* (Lepidoptera: Arctiidae). *Int J Biol Sci*, 2010, 6: 172–186
- 23 朱雅君, 周国梁, 房蕊, 等. 舞毒蛾线粒体 DNA 全序列分析. *植物检疫*, 2010, 24: 6–11
- 24 Yuan M L, Zhang Q L. The complete mitochondrial genome of *Gynaephora menyuanensis* (Lepidoptera: Lymantriidae) from the Qinghai-Tibetan Plateau. *Mitochondrial DNA*, 2013, 24: 328–330
- 25 Abascal F, Zardoya R, Telford M J. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res*, 2010, 38: W7–W13
- 26 Lanfear R, Calcott B, Ho S Y, et al. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol*, 2012, 29: 1695–1701
- 27 Miller M A, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Gateway Computing Environments Workshop (GCE)*, 2010 Nov. 14, New Orleans. New York: IEEE, 2010. 1–8
- 28 Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 2014, 30: 1312–1313
- 29 Ronquist F, Teslenko M, van der Mark P, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*, 2012, 61: 539–542
- 30 Ojala D, Montoya J, Attardi G. tRNA punctuation model of RNA processing in human mitochondria. *Nature*, 1981, 290: 470–474
- 31 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. *Mol Biol Evol*, 2002, 19: 163–169
- 32 Saito S, Tamura K, Aotsuka T. Replication origin of mitochondrial DNA in insects. *Genetics*, 2005, 171: 1695–1705
- 33 Dong W W, Feng X J, Huang G H, et al. Characterization of the mitochondrial genome of the cabbage webworm, *Hellula undalis* (Lepidoptera: Pyralidae). *Mitochondrial DNA*, 2014, doi:10.3109/19401736.2014.926491

Complete Mitochondrial Genome of *Plodia interpunctella* and phylogenetic analysis of Pyraloidea (Insecta: Lepidoptera)

TANG PeiAn¹, WU HaiJing¹, WANG Juan², GUO ZhongLong² & YUAN MingLong²

¹ Collaborative Innovation Center for Modern Grain Circulation and Safety, College of Food Science and Engineering, Nanjing University of Finance and Economics, Nanjing 210023, China;

² State Key Laboratory of Grassland Agro-Ecosystems, College of Pastoral Agricultural Science and Technology, Lanzhou University, Lanzhou 730020, China

Insect mitochondrial genome (mitogenome) has been widely used in systematic evolution, population genetics and phylogeographics. *Plodia interpunctella* is an important stored product pest worldwide. To better understand the phylogeny and evolution of Pyraloidea, we sequenced and annotated the complete mitogenome of *P. interpunctella*. This mitogenome is 15264 bp in length, and contains 13 protein-coding genes (PCGs), two ribosomal RNA genes, 22 transfer RNA genes (tRNAs), and a large non-coding region (mitochondrial control region). Comparative mitogenomic analyses reveal that gene arrangement, base composition, codon usage, the sequence and structure of control region, and secondary structures of 22 tRNAs are well conserved within Pyraloidea. Evolutionary pattern analyses show that the evolutionary rates of 13 PCGs differ between Pyralidae and Crambidae. Phylogenetic analyses of nine subfamilies based on 13 PCGs data using Bayesian inference and maximum likelihood methods strongly support the monophyly of each family within Pyraloidea. Within Pyralidae, there is a closer relationship between Phycitinae and Epipaschiinae. Crambidae is composed of two main clades: one clade contains Pyraustinae and Spilomelinae, the other includes Crambinae, Nymphulinae, Scopariinae and Schoenobiinae.

insects, Lepidoptera, Pyralidae, Crambidae, mitochondrial genome, phylogenetic analysis

doi: 10.1360/N052015-00193