

鳞翅目昆虫比较线粒体基因组学研究进展

王维^{①②}, 孟智启^①, 石放雄^②, 李风波^{①*}

① 浙江省植物有害生物防控重点实验室——省部共建国家重点实验室培育基地, 浙江省农业科学院蚕桑研究所, 杭州 310021;

② 南京农业大学动物科技学院, 南京 210095

* 联系人, E-mail: fengboli@gmail.com

2013-02-05 收稿, 2013-04-26 接受

浙江省农业科学院科技创新能力提升工程项目(2012R06Y01E03)和浙江省蚕桑产业科技创新团队(2011R50028)资助

摘要 鳞翅目昆虫, 无论作为农林业主要害虫, 抑或作为传粉媒介和经济昆虫, 对人类社会都产生了深远影响. 昆虫线粒体基因组具有基因组成稳定、基因排列相对保守等性质, 因而广泛用于进化和系统发育研究. 目前已有 63 种鳞翅目昆虫的线粒体基因组序列被测定, 本文比较了其中已知的 58 条参考序列的基本特征; 同时构建了鳞翅目昆虫系统发育树, 发现其与传统分类具有较好的一致性. 本工作为鳞翅目昆虫的进化研究奠定了基础, 并提示线粒体比较基因组学在后基因组时代对于构建昆虫系统发育关系有着巨大潜力.

关键词

鳞翅目
线粒体基因组
比较基因组
进化
分子系统学

鳞翅目属于节肢动物门(Arthropoda)昆虫纲(Insecta), 包括蝶(butterfly)、蛾(moth) 2 类, 为全变态昆虫, 1 个世代经历卵、幼虫、蛹、成虫等 4 阶段. 鳞翅目昆虫种类繁多, 世界各地分布约 30 万种(约 95% 以上为蛾类, <http://www.discoverlife.org/>), 其中已描述的种类有 157424 种^[1]. 不论是作为农林业主要害虫, 抑或作为传粉媒介和经济昆虫, 鳞翅目昆虫对人类社会都产生了深远影响. 深入了解鳞翅目昆虫的起源、进化和系统发育关系, 对鳞翅目虫害防治、经济昆虫种质资源的利用具有重要意义^[2~5].

动物线粒体基因组(mitochondrial genome)一般为双链、闭环的 DNA 分子^[6], 具有突变率高、进化速度快、很少发生重组、易于 PCR 扩增等特点, 已成为一种重要的分子标记^[6,7], 并广泛用于包括鳞翅目昆虫在内的许多物种的遗传进化、分类鉴定和系统发育等研究^[7~10]. 根据对 NCBI 数据库的统计, 截止到 2012 年 11 月, 完成线粒体基因组测序的鳞翅目物种已达 63 种, 品种数达 112 个(仅次于双翅目的 372 个), 获得参考序列 58 条(为昆虫纲最高). 本文通过比较分析现有的 58 条参考序列, 总结归纳鳞翅目线

粒体基因组碱基组成、基因排列、tRNA 和 rRNA 基因、蛋白编码基因、基因重叠区、基因间隔区的特点, 不但有助于了解鳞翅目线粒体基因组的基本特征, 而且可以为高级分类阶元(科、目等)线粒体基因组的比较分析提供参考.

随着国际生命条形码计划(<http://ibol.org/>)的启动, 线粒体基因 *cox1*(cytochrome c oxidase subunit 1)作为公认的动物界标准 DNA 条形码基因, 已用于几乎所有动物的分类鉴定等研究^[9]. 与 *cox1* 基因相比, 目前线粒体基因组序列则主要在探讨动物(包括鳞翅目昆虫)的系统发育关系、起源、进化等问题中发挥着重要作用^[7~11]. 为此, 本文对线粒体基因组序列在鳞翅目昆虫分子系统学、性别控制以及物种起源与进化等领域的研究进展和应用潜力作一总结与展望.

1 鳞翅目昆虫线粒体基因组的比较分析

1.1 概况

鳞翅目中, 线粒体基因组第一个测序并公布在 NCBI 上的是家蚕(*Bombyx mori*), 迄今被收录为参考

引用格式: 王维, 孟智启, 石放雄, 等. 鳞翅目昆虫比较线粒体基因组学研究进展. 科学通报, 2013, 58: 3017–3029

Wang W, Meng Z Q, Shi F X, et al. Advances in comparative mitogenomic studies of Lepidoptera (Arthropoda: Insecta) (in Chinese). Chin Sci Bull (Chin Ver), 2013, 58: 3017–3029, doi: 10.1360/972012-1721

序列的物种包括 29 种蝶类和 29 种蛾类, 分布在鳞翅目蛱蝶科、凤蝶科、灰蝶科、粉蝶科、弄蝶科和蚕蛾科、大蚕蛾科、天蛾科、蝙蝠蛾科、夜蛾科、舟蛾科、灯蛾科、毒蛾科、尺蛾科、潜蛾科、卷叶蛾科、螟蛾科、草螟科等 18 科, 分属凤蝶总科、弄蝶总科、蚕蛾总科、蝙蝠蛾总科、夜蛾总科、尺蛾总科、巢蛾总科、卷叶蛾总科和螟蛾总科等 9 个总科(表 1). 鳞翅目昆虫线粒体基因组大小约 15~16 kb, 平均 15381 bp. 人支蝠蛾^[12] 线粒体 DNA(mitochondrial DNA, mtDNA) 由于 A+T 富集区较大, 长达 16173 bp, 是已知线粒体基因组最大的鳞翅目物种. 鳞翅目昆虫 mtDNA 碱基组成表现出典型 AT 偏好, 碱基 A+T 含量范围为 77.8%~82.7%, 平均 80.7%. 其中 A+T 含量最小的物种是 *O. lunifer*, 最大的物种是朝灰蝶.

1.2 基因排列

线粒体在好氧真核细胞的细胞代谢过程中发挥着至关重要的作用, 后生动物线粒体基因组的基因组成是高度保守的, 一般包括 37 个基因, 即 13 个蛋白编码基因(protein-coding gene, PCG: NADH dehydrogenase subunits 1-6, *nad1-6*; cytochrome oxidase subunits 1-3, *cox1-3*; ATPase subunits 6, 8, *atp6*, *atp8*; cytochrome *b*, *cytb*), 2 个核糖体 RNA 基因(rRNA gene: large ribosome RNA, *rrnL*; small ribosome RNA, *rrnS*)和 22 个转移 RNA 基因(transfer RNA gene, tRNA gene)^[6]. 已知的鳞翅目昆虫线粒体基因组的 13 个蛋白编码基因中 9 个(*nad2*, *cox1*, *cox2*, *atp8*, *atp6*, *cox3*, *nad3*, *nad6*, *cytb*)位于主链(majority-coding strand, J 链), 4 个(*nad5*, *nad4*, *nad4L*, *nad1*)位于次链(minority-coding strand, N 链). 除此之外, 还有 1 个非编码部分——A+T 富集区, 由于该区域含有许多参与转录、复制的重要调控元件, 因此被认为是 mtDNA 控制区(control region)^[43]. 线粒体 DNA 上的基因排列紧凑, 基因间的非编码序列非常短小, 时常可见由少数核苷酸形成的基因重叠^[12,16]. 这使得 mtDNA 上的基因排列顺序相对固定^[44], 基因重排主要发生在 tRNA 基因^[6]. 绝大多数鳞翅目昆虫线粒体基因的排列顺序, 与第一个被测序的鳞翅目昆虫家蚕一致, 而不同于第一个被测序的节肢动物果蝇(*Drosophila yakuba*)^[45]. 例外发生在人支蝠蛾和云南蝠蛾^[12]、小赭弄蝶和梳翅弄蝶: 人支蝠蛾、云南蝠蛾线粒体基因组有 22 个 tRNA 基因, 但是排序不同于家蚕(*trnM*-

trnI-trnQ) 而类似果蝇(*trnI-trnQ-trnM*)^[12,46]; 小赭弄蝶、梳翅弄蝶基因排序类似于家蚕, 却分别有 24 个和 23 个 tRNA 基因——即出现了基因排序异常、基因数目异常两种情形(图 1). tRNA 基因易位在昆虫 mtDNA 进化过程中比较普遍^[43,47], 基因重排越明显的物种分子进化速率越快^[48], 因此可用于研究物种演化过程. 人支蝠蛾和云南蝠蛾是目前唯一已完成线粒体基因组测序的非双孔亚目(nonditrysia)类群^[12], 其特殊的基因顺序或许将为分析鳞翅目高级阶元双孔亚目、非双孔亚目的分歧事件提供参考. 梳翅弄蝶和小赭弄蝶 *trnS*, *trnL* 基因个数分别为“3+3”和“2+3”, 不同于常见的“2+2”, 因此颠覆了动物 mtDNA 只有 37 个基因的传统观念^[6]. 另外, 梳翅弄蝶额外的 *trnS*, *trnL* 基因为串联重复, 归纳为弄蝶 I 型; 小赭弄蝶额外的 *trnL* 基因位于 A+T 富集区, 归纳为弄蝶 II 型, 朝灰蝶^[22]、苧麻珍蝶^[19]情形分别与之类似(表 1, 图 1).

1.3 tRNA 与 rRNA 基因

鳞翅目昆虫线粒体基因组中, 22 个 tRNA 基因一般包括^[49] 2 个 *trnL* [*trnL*(UUR)和 *trnL*(CUN), *trnL1* 和 *trnL2*], 2 个 *trnS* [*trnS*(AGN)和 *trnS*(UCN), *trnS1* 和 *trnS2*]以及其他 18 种 tRNA 基因各 1 个. 通常, 对鳞翅目昆虫而言, 除 *trnS*(AGN)外, 其余 21 个 tRNA 基因二级结构均为典型的三叶草结构(typical cloverleaf secondary structure)(图 2); 而 *trnS*(AGN)缺少二氢尿嘧啶臂(DHU stem), 形成一个简单的 D 环^[14,16](图 2). 当然也有例外, 如樟蚕的 *trnS*(AGN), *trnS*(UCN)二级结构均缺少 DHU 臂^[27], 而茶小卷叶蛾^[33]等物种的 *trnS*(AGN)则拥有完整的三叶草结构.

碱基错配(base pair mismatch)普遍存在于鳞翅目昆虫 tRNA 二级结构^[14]. 例如, 碱基错配数云南蝠蛾为 18^[12], 柞蚕为 24^[15], 残锷线蛱蝶为 22^[16], 黑纹粉蝶为 19^[23], 樟蚕为 24^[27], *O. lunifer* 为 32^[35], 美国白蛾为 23^[36]. 错配类型包括 G-U, A-C, A-G, A-A 和 U-U 错配, 最常见的是 G-U 错配等. 碱基错配可发生在 tRNA 二级结构的二氢尿嘧啶臂、反密码子臂、TΨC 臂、氨基酸接受臂, 其中氨基酸接受臂错配频率较高.

rRNA 基因包括 *rrnL* 基因(16S rRNA 或 *lrRNA* 基因)和 *rrnS* 基因(12S rRNA 或 *srRNA* 基因)^[6], 其大小和位置均比较保守, 碱基组成表现出明显 AT 偏好^[35]. 其中, *rrnL* 位于 *trnL*(CUN)和 *trnV* 之间, 大小 1311~

表 1 鳞翅目昆虫线粒体基因组的 58 条参考序列^{a)}

科	种名	GenBank 登录号	大小(bp)	基因排列	A+T (%)	参考文献
蛱蝶科 Nymphalidae	大紫蛱蝶 <i>Sasakia charonda</i>	NC_014224	15244	家蚕型	79.9	未发表
	大紫蛱蝶 <i>Sasakia charonda kuriyamaensis</i>	NC_014223	15222	家蚕型	79.9	未发表
	珠蛱蝶 <i>Issoria lathonia</i>	NC_018030	15172	家蚕型	81.1	未发表
	庆网蛱蝶 <i>Melitaea cinxia</i>	NC_018029	15170	家蚕型	80.0	未发表
	残锖线蛱蝶 <i>Athyma sulpitia</i>	NC_017744	15268	家蚕型	81.9	[16]
	蟾福蛱蝶 <i>Fabriciana nerippe</i>	NC_016419	15140	家蚕型	80.9	[17]
	朴喙蝶 <i>Libythea celtis</i>	NC_016724	15164	家蚕型	81.2	未发表
	异型紫斑蝶 <i>Euploea mulciber</i>	NC_016720	15166	家蚕型	81.5	未发表
	斐豹蛱蝶 <i>Argynnis hyperbius</i>	NC_015988	15156	家蚕型	80.8	未发表
	柳紫闪蛱蝶 <i>Apatura ilia</i>	NC_016062	15242	家蚕型	80.5	未发表
	细带闪蛱蝶 <i>Apatura metis</i>	NC_015537	15236	家蚕型	80.5	[18]
	大卫绢蛱蝶 <i>Calinaga davidis</i>	NC_015480	15267	家蚕型	80.4	未发表
	苎麻珍蝶 <i>Acraea issoria</i>	NC_013604	15245	弄蝶 II 型	79.7	[19]
	枯叶蛱蝶 <i>Kallima inachus</i>	NC_016196	15150	家蚕型	80.3	[20]
	仁眼蝶 <i>Hipparchia autonoe</i>	NC_014587	15489	家蚕型	79.1	[21]
凤蝶科 Papilionidae	台湾宽尾凤蝶 <i>Papilio maraho</i>	NC_014055	16094	家蚕型	80.5	未发表
	金斑喙凤蝶 <i>Teinopalpus aureus</i>	NC_014398	15242	家蚕型	79.8	未发表
	金凤蝶 <i>Papilio machaon</i>	NC_018047	15185	家蚕型	80.3	未发表
	碧凤蝶 <i>Papilio bianor</i>	NC_018040	15340	家蚕型	80.6	未发表
	红珠绢蝶 <i>Parnassius bremeri</i>	NC_014053	15389	家蚕型	81.3	未发表
灰蝶科 Lycaenidae	朝灰蝶 <i>Coreana raphaelis</i>	NC_007976	15314	弄蝶 I 型	82.7	[22]
	华西琉璃灰蝶 <i>Celastrina hersilia</i>	NC_018049	15302	家蚕型	81.5	未发表
	塔银线灰蝶 <i>Spindasis takanonis</i>	NC_016018	15349	家蚕型	82.3	[8]
	祖灰蝶 <i>Protantigius superans</i>	NC_016016	15248	家蚕型	81.7	[8]
粉蝶科 Pieridae	黑纹粉蝶 <i>Pieris melete</i>	NC_010568	15140	家蚕型	79.8	[23]
	菜粉蝶 <i>Pieris rapae</i>	NC_015895	15157	家蚕型	79.7	[24]
弄蝶科 Hesperidae	绢粉蝶 <i>Aporia crataegi</i>	NC_018346	15140	家蚕型	81.3	[25]
	小赭弄蝶 <i>Ochlodes venata</i>	NC_018048	15622	弄蝶 II 型	82.0	未发表
蚕蛾科 Bombycidae	梳翅弄蝶 <i>Ctenoptilum vasava</i>	NC_016704	15468	弄蝶 I 型	80.6	[26]
	野桑蚕 <i>Bombyx mandarina</i>	NC_003395	15928	家蚕型	81.7	[13]
	家蚕 <i>Bombyx mori</i>	NC_002355	15643	家蚕型	81.3	未发表
大蚕蛾科 Saturniidae	樟蚕 <i>Eriogyna pyretorum</i>	NC_012727	15327	家蚕型	80.8	[27]
	合目大蚕蛾 <i>Saturnia boisduvalii</i>	NC_010613	15360	家蚕型	80.6	[28]
	天蚕 <i>Antheraea yamamai</i>	NC_012739	15338	弄蝶 II 型	80.3	[29]
	柞蚕 <i>Antheraea pernyi</i>	NC_004622	15566	家蚕型	80.1	[15]
	蓖麻蚕 <i>Samia cynthia ricini</i>	NC_017869	15384	家蚕型	79.7	[30]
	天蚕蛾 <i>Actias selene</i>	NC_018133	15236	家蚕型	78.9	[31]
	烟草天蛾 <i>Manduca sexta</i>	NC_010266	15516	家蚕型	81.8	[37]
天蛾科 Sphingidae	人支蝠蛾 <i>Thitarodes renzhiensis</i>	NC_018094	16173	蝠蛾型	81.3	[12]
蝙蝠蛾科 Hepialidae	云南蝠蛾 <i>Ahamus yunnanensis</i>	NC_018095	15816	蝠蛾型	82.3	[12]
夜蛾科 Noctuidae	大螟 <i>Sesamia inferens</i>	NC_015835	15413	家蚕型	80.3	未发表
	棉铃虫 <i>Helicoverpa armigera</i>	NC_014668	15347	家蚕型	80.9	[32]
舟蛾科 Notodontidae	<i>Ochrogaster lunifer</i>	NC_011128	15593	家蚕型	77.8	[35]
	苹掌舟蛾 <i>Phalera flavescens</i>	NC_016067	15659	家蚕型	80.9	[38]
灯蛾科 Hypercompe	美国白蛾 <i>Hyphantria cunea</i>	NC_014058	15481	家蚕型	80.4	[36]
毒蛾科 Lymantriidae	舞毒蛾 <i>Lymantria dispar</i>	NC_012893	15569	家蚕型	79.9	未发表
尺蛾科 Geometridae	桑尺蠖 <i>Phthonandria atrilineata</i>	NC_010522	15499	家蚕型	81.0	[39]
潜蛾科 Lyonetiidae	旋纹潜蛾 <i>Leucoptera malifoliella</i>	NC_018547	15646	家蚕型	82.6	未发表
卷叶蛾科 Tortricidae	茶小卷叶蛾 <i>Adoxophyes honmai</i>	NC_008141	15680	家蚕型	80.4	[33]
	顶梢卷叶蛾 <i>Spilonota lechriaspis</i>	NC_014294	15368	家蚕型	81.2	[34]
	梨小食心虫 <i>Grapholita molesta</i>	NC_014806	15776	家蚕型	80.9	[40]
	黄斑长翅卷叶蛾 <i>Acleris fimbriana</i>	NC_018754	15933	家蚕型	81.0	未发表
	外米缀蛾 <i>Corcyra cephalonica</i>	NC_016866	15273	家蚕型	80.5	未发表
螟蛾科 Pyralidae	亚洲玉米螟 <i>Ostrinia furnacalis</i>	NC_003368	14536	家蚕型	80.4	[14]
草螟科 Crambidae	欧洲玉米螟 <i>Ostrinia nubilalis</i>	NC_003367	14535	家蚕型	80.2	[14]
	二化螟 <i>Chilo suppressalis</i>	NC_015612	15395	家蚕型	80.6	[41]
	稻纵卷叶螟 <i>Cnaphalocrocis medinalis</i>	NC_015985	15388	家蚕型	82.0	[41]
	小蔗螟 <i>Diatraea saccharalis</i>	NC_013274	15490	家蚕型	81.1	[42]

a) 鳞翅目昆虫的分类地位和拉丁名依据 NCBI 标准确定

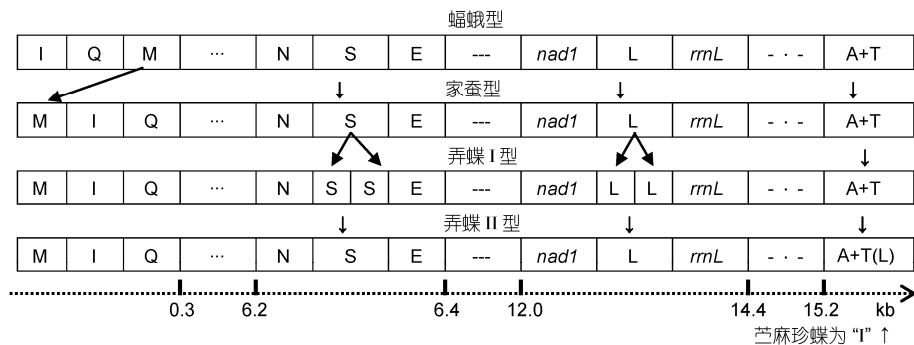


图 1 鳞翅目昆虫 mtDNA 的 4 种基因排列方式

tRNA 基因依据 IUPAC-IUB 用氨基酸单字母表示; A+T 为 A+T 富集区; ... 为基因排列 *nad2-W-C-Y-cox1-L-cox2-K-D-atp8-atp6-cox3-G-nad3-A-R*; --- 为基因排列 *F-nad5-H-nad4-nad4L-T-P-nad6-cytb-S*; - - - 为基因排列 *V-rrnS*. 标尺方向表示 5'→3', 位点信息基于家蚕 mtDNA (GenBank 登录号: NC_002355)

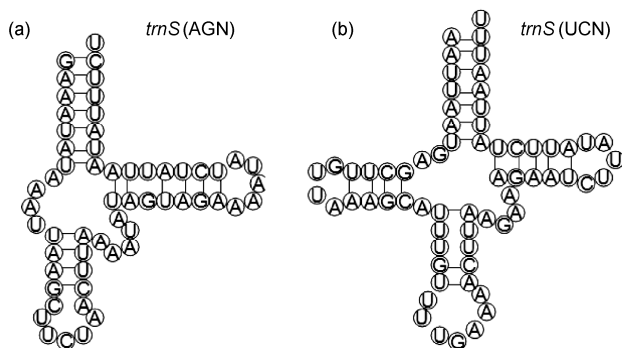


图 2 金凤蝶 tRNA 基因二级结构示意图

(a) 缺少 DHU 臂; (b) 标准二级结构. 金凤蝶线粒体基因组 GenBank 登录号为 NC_018047

1421 bp, 平均 1350 bp, 其中大紫蛱蝶 (*S. charonda kuriyamaensis*) 的长度最小, 黄斑长翅卷叶蛾长度最大. 鳞翅目 *rrnL* 基因的 A+T 含量为 81.5%~86.0%, 平均 84.4%, 其中 *O. lunifer* 最低, 云南蝠蛾最高. 相较 *rrnL* 而言, 鳞翅目昆虫的 *rrnS* (位于 *trnV* 和 A+T 富集区之间) 大小更为保守^[12,14,15,41], 范围为 739~830 bp (不包括欧洲玉米螟的 434 bp 和亚洲玉米螟的 435 bp), 平均 779 bp, 其中祖灰蝶长度最小, 苹掌舟蛾长度最大. 鳞翅目 *rrnS* 的长度虽然远小于 *rrnL*, 但是其 A+T 含量却略高于 *rrnL*^[23,29]: *rrnS* 的 A+T 含量 82.1%~87.1%, 平均 85.1%, 其中欧洲玉米螟最低, 旋纹潜蛾最高. 一般来说, rRNA 基因的 A+T 含量高于蛋白编码基因和 tRNA 基因^[15].

1.4 蛋白编码基因

PCG 的碱基组成表现出明显 AT 偏好^[15,28,32,37],

且不同鳞翅目物种的 PCG 数目、排列顺序完全一致, PCG 大小在不同物种中相差不大 (图 1). PCG 没有内含子, 这与整个 mtDNA 表现的“非编码区稀少”特征类似, 可能与线粒体进化有关^[6,50]. PCG 最常见的起始密码子为 AUN^[12], 编码 Met. *cox1* 基因起始密码子较复杂, 例如, 家蚕为 UUUUAG^[13], 亚洲玉米螟、欧洲玉米螟为 UAUUAG^[14], 苎麻珍蝶为 UUG^[19], 朝灰蝶为 UUAG^[22]. 最近一项来自不吉按蚊转录本的证据支持 UCG^[51], 而大多数鳞翅目物种^[12,16,23,42], 包括来自豆荚螟的表达序列标签数据都支持 CGA. 因此, 认同 CGA 作为 *cox1* 基因起始密码子的研究占大多数^[12]. PCG 终止密码子为 UAA, 某些基因则以不完全的 U 或 UA 作为转录终止信号^[13]. 例如, 人支蝠蛾 *nad2*, *cox1*, *cox2*, *nad3*, *nad5* 基因为 U, 而其 *atp6*, *nad4L*, *nad6* 基因为 UA; 云南蝠蛾 *cox1*, *cox2*, *nad3*, *nad5* 基因为 U, 而 *atp6*, *nad4L*, *nad6* 基因为 UA^[12]; 野桑蚕和家蚕 *cox1*, *cox2* 基因为 U^[13]; 亚洲玉米螟和欧洲玉米螟 *cox2*, *atp6* 基因分别为 U, UA^[14]; 柞蚕 *cox1*, *cox2*, *nad3*, *nad5* 基因为 U, *atp6* 基因为 UA^[15]; 残鐸线蛱蝶 *nad4* 基因为 U^[16]; 黑纹粉蝶 *cox1*, *cox2*, *nad2*, *nad5* 基因为 U, *nad3* 基因为 UAG^[23]. 一般认为, 不完全终止子 U, UA 在转录后经聚腺苷酸化 (post-transcriptional polyadenylation) 形成完整的 UAA 终止子^[52,53].

1.5 密码子使用与进化

(1) 密码子使用. 利用 Codon Usage Database 数据库^[54]收录的 12 种以及未发表的 20 种 (表 1), 合计 32 种鳞翅目昆虫 mtDNA (共 110990 个完全密码子),

统计 PCG 密码子使用情况和碱基组成, 分析密码子偏好性. 统计结果经汇总整理见表 2, 从表中可见鳞翅目 PCG 密码子使用存在以下特征: (i) 鳞翅目昆虫有 2 种完全终止密码子 UAA 和 UAG, UAA 使用频率远高于 UAG; (ii) UUA, UUG 和 CUN 等 6 种密码子均对应氨基酸 Leu, UUA 使用频率最高; (iii) UUU 和 UUC 对应氨基酸 Phe, UUU 使用频率更高; (iv) AUU 和 AUC 对应氨基酸 Ile, AUU 使用频率更高; (v) AUA 和 AUG 对应氨基酸 Met, AUA 使用频率更高; (vi) AAU 和 AAC 对应氨基酸 Asn, AAU 使用频率更高; (vii) 鳞翅目昆虫 PCG 密码子碱基组成偏好 AT, 例如 32 种鳞翅目昆虫 A+T 平均含量为 79.4%; (viii) 密码子第 3 位偏爱 AT, 所分析的这 32 个物种 PCG 的 A+T 含量平均值为 92.5%, 而朝灰蝶、天蚕等物种 A+T 含量甚至超过 95%. 另外, 通过与已发表但未被 Codon Usage Database 数据库收录的其他物种如人支蝠蛾和云南蝠蛾^[12]、残钨线蛱蝶^[16]、*O. lunifer*^[35]、美国白蛾^[36]、二化螟和稻纵卷叶螟^[41]、小蔗螟^[42]等密码子使用的分析数据比较可知, 表 2 中的密码子使用频率与上述数据具有良好的一致性.

(2) 密码子进化. 生物进化模式的差异使得多种密码子系统得以形成: 不仅是细胞核与细胞器(线粒体、叶绿体)之间, 而且细胞核与细胞核、细胞器与细胞器之间的遗传密码也不尽相同^[55]. 例如在鳞翅目昆虫线粒体基因组中, 密码子 UGA, AUA, AGA 分别对应 Trp, Met, Ser; 而在通用密码子表中它们分别对应终止子, Ile, Arg, 这 3 个密码子的变异情况也

适用于所有节肢动物和绝大部分无脊椎动物^[56,57]. AGG 在无脊椎动物线粒体密码子系统中对应 Ser, 而在少数节肢动物线粒体中 AGG 对应 Lys^[55]. 不同密码子系统的出现, 可能因为遗传密码进化可以带来直接的选择优势^[58]. 影响翻译机器的因素都可能造成遗传密码变异, 其中作用较大的有碱基修饰^[59]、tRNA 突变和 tRNA 编辑^[57]. 例如, 在鳞翅目昆虫 mtDNA 中 AGA 对应 Ser 而非 Arg, 这是其反密码子 GCU 经甲基化修饰形成 m⁷GCU 所致^[57].

1.6 基因间隔区与基因重叠区

A+T 富集区也称控制区或 D-Loop, 是动物 mtDNA 的复制起始区和最大的基因间隔区^[43,60]. 除亚洲玉米螟、欧洲玉米螟等 2 个物种(“A+T 富集区”未确定)外, 其余 56 个物种的 A+T 富集区位置相对固定, 位于 *rrnS*, *trnM(trnI)* 基因之间. 这 56 种鳞翅目昆虫 A+T 富集区碱基平均含量: A 为 44.1%, G 为 2.2%, C 为 4.3%, T 为 49.4%. 这 56 种鳞翅目昆虫 A+T 富集区的 A+T 含量桑尺蠖最高(98.2%), 天蚕蛾最低(87.89%), 平均为 93.5%. A+T 富集区大小并不保守(311~1367 bp, 平均 464 bp), 最小的是大螟, 最大的是人支蝠蛾. 曾有学者指出, A+T 富集区是 mtDNA 进化最快的区域, 可作为动物群体遗传学的重要分子标记, 然而从核苷酸替换的角度而言, A+T 富集区可能并不是昆虫 mtDNA 变化最大的区域^[61]. 例如比较家蚕和日本野桑蚕 mtDNA 可知, 它们 A+T 富集区的核苷酸序列变异程度仅高于 *rrnS* 基因^[10]; 比较家

表 2 32 种鳞翅目昆虫 PCG 密码子使用统计表^{a)}

密码子	频次	频率(%)	密码子	频次	频率(%)	密码子	频次	频率(%)	密码子	频次	频率(%)
UUU(F)	10764	9.70	UCU(S)	3429	2.93	UAU(Y)	5301	4.78	UGU(C)	902	0.81
UUC(F)	797	0.72	UCC(S)	360	0.32	UAC(Y)	387	0.35	UGC(C)	86	0.08
UUA(L)	13784	12.42	UCA(S)	2501	2.25	UAA(Stop)	299	0.27	UGA(W)	2678	2.41
UUG(L)	592	0.53	UCG(S)	85	0.08	UAG(Stop)	21	0.02	UGG(W)	157	0.14
CUU(L)	1030	0.93	CCU(P)	1959	1.77	CAU(H)	1674	1.51	CGU(R)	464	0.42
CUC(L)	87	0.08	CCC(P)	427	0.39	CAC(H)	277	0.25	CGC(R)	32	0.03
CUA(L)	723	0.65	CCA(P)	1131	1.02	CAA(Q)	1704	1.54	CGA(R)	947	0.85
CUG(L)	16	0.01	CCG(P)	39	0.04	CAG(Q)	97	0.09	CGG(R)	74	0.07
AUU(I)	12866	11.59	ACU(T)	2277	2.05	AAU(N)	6963	6.27	AGU(S)	934	0.84
AUC(I)	788	0.71	ACC(T)	280	0.25	AAC(N)	667	0.60	AGC(S)	62	0.06
AUA(M)	7878	7.10	ACA(T)	1828	1.65	AAA(K)	3012	2.71	AGA(S)	2473	2.23
AUG(M)	723	0.65	ACG(T)	47	0.04	AAG(K)	322	0.29	AGG(S)	51	0.05
GUU(V)	2115	1.91	GCU(A)	2128	1.92	GAU(D)	1684	1.52	GGU(G)	1723	1.55
GUC(V)	75	0.07	GCC(A)	219	0.20	GAC(D)	146	0.13	GGC(G)	125	0.11
GUA(V)	1710	1.54	GCA(A)	1081	0.97	GAA(E)	1878	1.69	GGA(G)	3126	2.82
GUG(V)	132	0.12	GCG(A)	56	0.05	GAG(E)	257	0.23	GGG(G)	720	0.65

a) 密码子总数为 110990

蚕和中国野桑蚕也发现 A+T 富集区核苷酸序列差异低于蛋白编码基因 *cytb*, *cox3* 和 *nad4L*^[10]. 鳞翅目昆虫 mtDNA 的 A+T 富集区一般存在 4 个保守结构^[15,19,23,61], 包括: (1) 始于临近核糖体 RNA 基因一端第 10~30 碱基位点的 ATAGA^[20]或 ATAGT 及其后 15~20 bp 的 polyT, 即复制子^[61]所在区域. 例如, 至少 35 个物种(35/56)A+T 富集区序列中存在“ATAGA(T)”+“18 bp polyT”元件; 同时, 除人支蛱蛾、云南蛱蛾、华西琉璃灰蝶、金凤蝶、绢粉蝶外, 其余 51 种鳞翅目昆虫(51/56)A+T 富集区均含 ATAGA(T)+polyT 片段. (2) 微卫星元件(AT)_n^[23]或(TA)_n, 靠近转运 RNA 基因一端. 然而, 仅 22 个鳞翅目昆虫物种(22/56)A+T 富集区含有微卫星片段“(AT)_n”(n≥8), 仅 25 个鳞翅目昆虫物种(25/56)A+T 富集区含有“(TA)_n”(n≥8), 因此 A+T 富集区微卫星元件的保守性有待进一步研究. (3) 位于转运 RNA 基因上游的 polyA^[15,23]或 polyT, 长约 5~15 bp. (4) 贯穿整个 A+T 富集区的

重复序列^[15,23], 其大小、重复数目均不保守.

除 A+T 富集区外, mtDNA 上散在分布着数目不等(1~20 个)、大小不定(几碱基对到几百碱基对)的间隔序列. 线粒体进化趋向于基因组大小减少, 是通过消除/削减基因间隔区(intergenic spacer)实现的, 这一观点已被证实^[62,63]. 为探讨基因间隔区和基因重叠区(overlapping region)保守特征, 对 58 种鳞翅目昆虫 mtDNA 中 *trnQ*—*nad2*, *trnS2*—*nad1*, *trnW*—*trnC*, *atp8*—*atp6* 等 4 处区域进行统计分析. 统计方法: 登录 GenBank 数据库, 检索并计算 58 种鳞翅目昆虫在此 4 处区域的碱基个数, 统计结果见表 3.

统计结果表明, 鳞翅目昆虫 mtDNA 至少有 2 个相对保守的基因间隔区和 2 个相对保守的基因重叠区. (1) 基因间隔区. 间隔区 1(Spacer1)较大, 处在 *trnQ*—*nad2* 基因间, 大小 40~90 bp, 平均 53 bp; 间隔区 2(Spacer2)较小, 处在 *trnS2*—*nad1* 基因间, 大小 5~40 bp, 平均 17 bp. 相比于间隔区 2, 间隔区 1 更为

表 3 基因间隔区和基因重叠区碱基数目统计表(单位: bp)^{a)}

种名	<i>trnQ</i> — <i>nad2</i>	<i>trnS2</i> — <i>nad1</i>	<i>trnW</i> — <i>trnC</i>	<i>atp8</i> — <i>atp6</i>	种名	<i>trnQ</i> — <i>nad2</i>	<i>trnS2</i> — <i>nad1</i>	<i>trnW</i> — <i>trnC</i>	<i>atp8</i> — <i>atp6</i>
<i>B. mori</i>	65	25	-7	-7	<i>A. metis</i>	61	13	-8	-7
<i>O. nubilalis</i>	62	40	-7	-7	<i>C. suppressalis</i>	52	12	-8	-7
<i>O. furnacalis</i>	62	34	-7	-7	<i>S. inferens</i>	68	18	-8	-11
<i>B. mandarina</i>	47	25	-8	-7	<i>P. rapae</i>	46	14	-8	-7
<i>C. raphaelis</i>	56	19	-8	-7	<i>C. medinalis</i>	74	16	-8	-7
<i>A. honmai</i>	64	25	-6	-7	<i>A. hyperbius</i>	52	-2	-8	-7
<i>A. pernyi</i>	54	18	-8	-7	<i>P. superans</i>	51	19	-8	-7
<i>M. sexta</i>	54	31	-8	-7	<i>S. takanonis</i>	48	31	3	-7
<i>P. atrilineata</i>	63	18	-8	-7	<i>A. ilia</i>	49	13	-8	-7
<i>P. melete</i>	48	16	-8	-7	<i>P. flavescentis</i>	58	30	-8	-4
<i>S. boisduvalii</i>	53	22	-8	-7	<i>K. inachus</i>	51	16	-8	-7
<i>O. lunifer</i>	72	17	-8	-7	<i>F. nerippe</i>	52	-2	-8	-7
<i>E. pyretorum</i>	54	18	-8	-7	<i>C. vasava</i>	61	17	-5	-7
<i>A. yamamai</i>	53	24	-8	-7	<i>E. mulciber</i>	46	20	-8	-7
<i>L. dispar</i>	47	28	-8	-7	<i>L. celtis</i>	52	17	-8	-7
<i>D. saccharalis</i>	56	9	-8	-7	<i>C. cephalonica</i>	61	16	-8	-7
<i>A. issoria</i>	51	-2	-8	-7	<i>A. sulphitia</i>	52	-2	-8	-7
<i>P. bremeri</i>	40	16	-8	-7	<i>S. cynthia ricini</i>	54	25	-8	-7
<i>P. maraho</i>	47	16	-8	-7	<i>M. cinxia</i>	46	24	-8	-7
<i>H. cunea</i>	50	34	-8	-7	<i>I. lathonia</i>	52	17	-8	-7
<i>S. charonda*</i>	87	-1	-8	-7	<i>P. bianor</i>	45	16	-8	-7
<i>S. charonda</i>	87	-1	-8	-7	<i>P. machaon</i>	55	16	-8	-7
<i>S. lechriaspis</i>	47	18	-8	-7	<i>O. venata</i>	67	17	-8	-7
<i>T. aureus</i>	47	16	-8	-7	<i>C. hersilia</i>	53	16	-8	-7
<i>H. autonoe</i>	50	6	-8	-7	<i>T. renzhienensis</i>	0	15	-8	-7
<i>H. armigera</i>	45	18	-8	-7	<i>A. yunnanensis</i>	0	15	-8	-7
<i>G. molesta</i>	62	19	-8	-7	<i>A. selene</i>	52	15	-16	-7
<i>A. crataegi</i>	49	16	-8	-7	<i>L. malifoliella</i>	43	19	-8	-7
<i>C. davidis</i>	46	-2	-8	-7	<i>A. fimbriana</i>	77	16	-8	-7

a) 正数为间隔碱基数目, 负数为重叠碱基数目; * 为 *S. charonda kuriyamaensis*

保守. 早期研究^[16,23,39]通过比对间隔区 1 与 *nad2* 基因核苷酸序列发现两者相似性较高, 因此推断间隔区 1 可能来自 *nad2* 基因. 间隔区 2 一般含有保守序列 ATACTAA^[36]. (2) 基因重叠区. 2 个保守的重叠区, 即 *atp8*—*atp6* 之间 7 bp 的 ATGATAA 序列(Region1) 和 *trnW*—*trnC* 之间 8 bp 的 AAGCCTTA 序列(Region2), 同时存在于绝大多数已知的鳞翅目线粒体基因组参考序列. Region1 位于蛋白编码基因, 而 Region2 序列位于 tRNA 基因, 因此根据 Knibbe 等人^[64]的研究, 可以推测 Region1 可能比 Region2 更为保守. 对序列变异情况的统计结果(表 3)表明, Region2(7/58)比 Region1 (2/58)变异更大, 证实了上述推测. (3) 与基因间隔区相比, 基因重叠区更为保守^[12]. 综上, 从碱基个数来看, 不同区域的保守程度为: 间隔区<tRNA 基因重叠区<蛋白编码基因重叠区. 此外, mtDNA 非编码区 Spacer1 和 Spacer2 的保守程度低于编码区 Region1 和 Region2, 非蛋白编码区 Region2 的保守程度略低于蛋白编码区 Region1, 提示鳞翅目线粒体蛋白编码基因受到更强的选择限制作用.

2 鳞翅目昆虫线粒体基因组的应用研究

2.1 分子系统学

DNA 条形码已广泛用于鳞翅目物种的分类鉴定研究^[65], 常见的 DNA 条形码包括 *cox1*^[65,66], *cox2*^[67], *nad5*^[68] 基因等, 它们在鳞翅目害虫的快速识别、争议物种的分类鉴定中发挥了一定作用. 另一方面, 对 DNA 条形码的质疑一直存在^[69,70], 如 NUMT(nuclear mitochondrial pseudogene)干扰^[71]、计算模型的选择^[71]、*Wolbachia* 对 mtDNA 多态性的间接选择效应^[72] 等都可能降低其结果的可靠性.

与 DNA 条形码(如 *cox1*)可广泛运用于真核生物^[9,65,66]相比, 基于线粒体基因组序列进行的系统发育研究主要集中在高级分类阶元如总科、目^[7], 如鞘翅目^[7]、双翅目^[11]、直翅目^[73]等. 在鳞翅目, Jiang 等人^[27]、Hong 等人^[28]、Feng 等人^[74]用线粒体基因组序列重建了鳞翅目不同总科物种间的系统发育关系, 其结论与传统分类方法(如形态鉴定)得出的观点一致. Bae 等人^[7]认为线粒体基因组序列作为一种分子标记, 可以较好地用于所有全变态类昆虫. 本文利用现有的 58 条鳞翅目线粒体基因组参考序列构建了鳞

翅目不同总科之间的系统发育关系, 其结果也与传统分类系统具有较好的一致性(图 3). 而 Liao 等人^[36]基于线粒体基因组序列对鳞翅目总科凤蝶总科(Papilionoidea)进行系统发育分析后, 发现与传统分类方法得出的观点不一致. 事实上, 在进行系统发育研究时, 线粒体基因组表现的一些特征(如非稳定进化等), 容易导致系统偏差^[50]. 图 3 中基于线粒体基因组核苷酸序列构建的 NJ 树, 相比基于 PCG 氨基酸序列得到的 NJ 树(与传统分类一致性较好)存在诸多不一致的分支结构, 这种不一致可能由于上述特征所导致. 解决途径之一是采取优化策略. 例如, Kim 等人^[8]在利用线粒体基因组数据重建鳞翅目昆虫的系统发育关系时, 为降低系统偏差而采取了剔除 PCG 第 3 位密码子等优化策略, 获得的结果表明^[8], 螟蛾总科与蚕蛾总科、尺蛾总科、夜蛾总科、凤蝶总科等聚为一个类群, 因此不支持“大鳞翅类(Macrolepidoptera)”的传统描述; 而灰蝶科、蛱蝶科、粉蝶科、凤蝶科等聚成一支, 则与早先对凤蝶总科的猜想一致. 总之, 线粒体基因组序列不但可以作为鳞翅目系统发育研究中有效的分子工具, 而且有望为某些争议问题提供新的解决途径.

2.2 物种起源与进化

基于线粒体基因组序列, 研究者围绕家蚕等鳞翅目昆虫的起源和进化展开了许多研究. (1) 蚕蛾科与大蚕蛾科亲缘关系. 蚕蛾科的家蚕, 大蚕蛾科的樟蚕、柞蚕都能吐出精美的蚕丝, 那么它们是近亲吗? 1999 年, Hwang 等人^[75,76]基于 mtDNA 的 RFLP 以及 *cox1*, *rrnL*, *rrnS* 核苷酸序列同源性探讨蚕蛾科和大蚕蛾科亲缘关系, 其结果均支持蚕蛾科、大蚕蛾科均为单系群(monophyly). 同时, 基于线粒体基因组序列的系统发育树也显示(图 3), 家蚕、野桑蚕与樟蚕、柞蚕等明显分为两支, 进一步证实了该结果. (2) 家蚕起源. Li 等人^[10]基于线粒体基因组核苷酸替换模式和 Li 等人^[77]基于线粒体基因组 SNP 的比较分析, Hu 等人^[78]基于线粒体全基因组序列对 4 个家蚕品系和 3 个野桑蚕品系进行的系统发育分析, 以及 Arunkumar 等人^[79]尝试利用多个线粒体基因 *rrnS*, *rrnL*, *cox1* 和控制区分别构建的系统发育树, 都发现家蚕与中国野桑蚕而非日本野桑蚕亲缘关系更近, 提示家蚕起源于中国野桑蚕. 这一观点与基于其他分子标记如 *cytb* 基因^[80]、*cox1* 基因^[81]得出的结论是

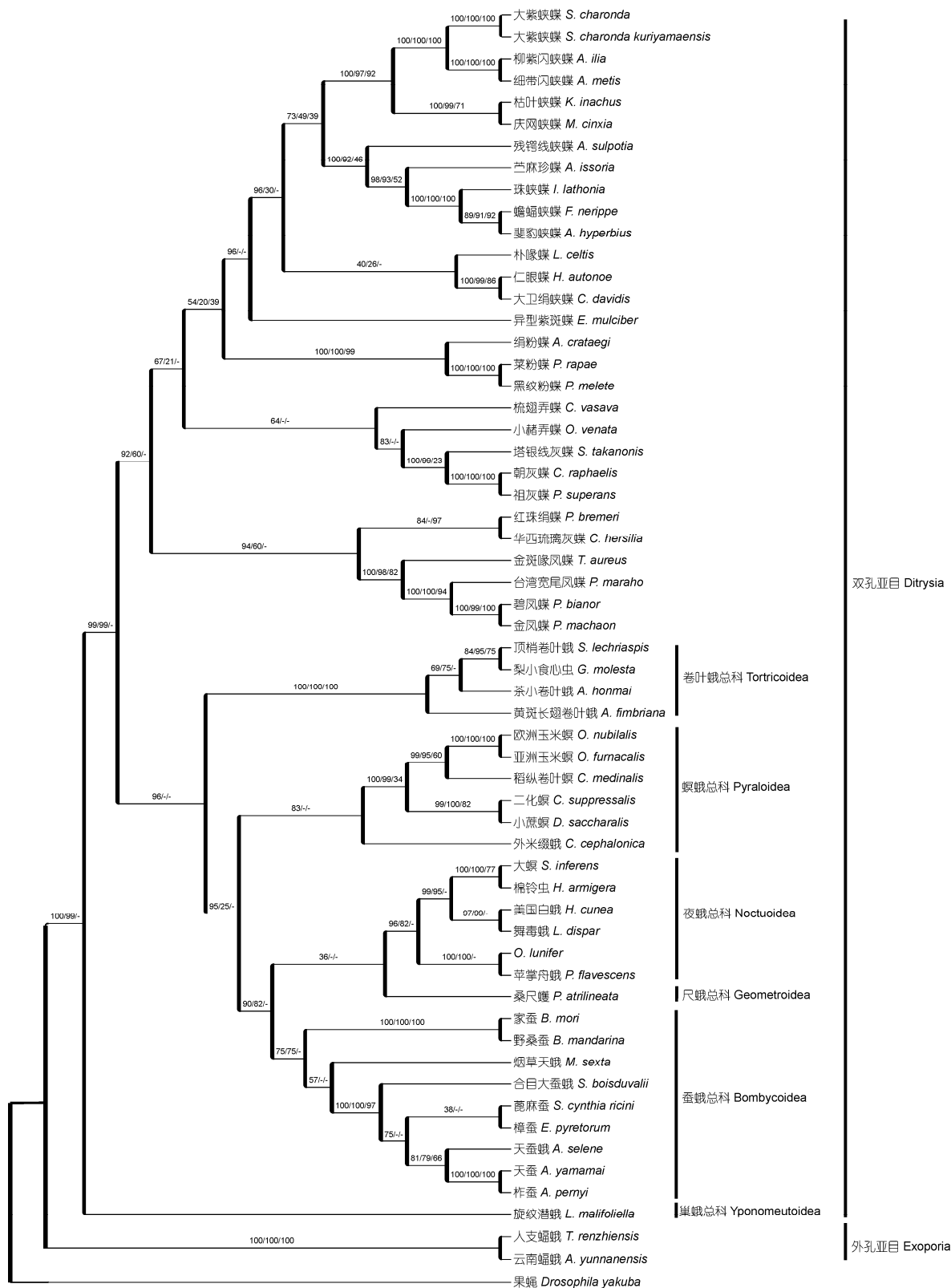


图 3 基于线粒体基因组序列构建的 58 种鳞翅目昆虫的系统发育树

节点置信度使用 bootstrap 重抽样检测 1000 次, 图中仅显示其拓扑结构. 节点处的数字依次表示 NJ 法(基于 13 个 PCGs 的氨基酸序列)、MP 法(基于 13 个 PCGs 的氨基酸序列)、NJ 法(基于线粒体基因组核苷酸序列)自展值(%). “-”表示不支持. 果蝇 *D. yakuba*^[45]作为外群

一致的。(3) 日本野桑蚕和中国野桑蚕何时发生的分歧? Yukuhiro 等人^[13]、潘敏慧等人^[82]、孙伟等人^[83]发现两者线粒体基因组序列差异较大,并分别根据线粒体基因 *nad5*, *cox1/cox2*, *cox1/A+T* 富集区核苷酸替换速率,计算得出两者分歧时间分别为 7.1 Mya (million years ago), 0.95~1.66 Mya 和 23600 Ya(years ago)。(4) 群体遗传学研究. Li 等人^[77]通过比较分析 41 个蚕类(家蚕和野桑蚕)线粒体基因组序列,揭示了家蚕驯化过程中群体结构的变化,认为家蚕和中国野桑蚕群体大小均未经经历扩张或收缩。然而,这一推断与最近基于线粒体基因和核基因联合分析得出的家蚕在约 1000 年前经历群体扩张^[83]这一推断并不一致。总之,线粒体基因组序列在鳞翅目昆虫物种起源和进化研究中发挥了重要作用;对于一些有争议的问题,增加核基因序列数据可能获得更全面的认识。

2.3 性别控制

沃尔巴克氏体(*Wolbachia*)是广泛存在于昆虫体内的一种内共生细菌,能影响宿主的生殖行为,进而控制宿主性别比例^[72]。已发现 *Wolbachia* 可影响 mtDNA 核苷酸序列(即降低或增加 mtDNA 多态性)^[72],调控线粒体蛋白编码基因的表达^[84],并最终导致孤雌生殖和雄性致死等^[85]。例如在幻紫斑蛱蝶(*Hypolimnast bolina*)^[86], *Wolbachia* 通过提升被感染雌性个体的相对适合度(relative fitness)而快速传播,进而快速提升雌蝶数量;在小菜蛾(*Plutella xylostella*),未感染 *Wolbachia* 种群雌雄比例为 1:1,感染 *Wolbachia* 种群雌雄比例为 2:1^[87]。关于 *Wolbachia* 影响 mtDNA 多态性的具体分子机制,最近的实验证据表明,两者之间不存在直接的相互作用关系:在侵入宿主过程中, *Wolbachia* 与线粒体的共同传递使得 *Wolbachia* 能够影响 mtDNA 多态性,且这种影响是一种间接效应^[88]。但是, *Wolbachia* 通过哪些途径发挥作用,以及 *Wolbachia* 作用于哪些 mtDNA 位点均有待进一步的研究。

2.4 其他

(1) 线粒体 DNA 复制. 线粒体 DNA 的复制起始点(replication region, O_R)位于 A+T 富集区,且主链、次链各有一个独立的 O_R 。Saito 等人^[89]研究

发现,尽管全变态昆虫(包括鳞翅目昆虫) O_R 具体位置不同,但它们都有一个共同的特征,即位于长约 10~20 bp 的 polyT 之后,因此认为 polyT 可能参与了全变态昆虫 mtDNA O_R 的识别。由于 Saito 等人^[89]的实验以果蝇为主,鳞翅目昆虫仅分析了家蚕,因此其研究结论是否同样适用于其他鳞翅目物种还有待检验。本文以序列(T)₁₀搜索 56 种鳞翅目昆虫 A+T 富集区,发现除旋纹潜蛾外,其余 55 个物种在靠近 *rrnS* 基因一端均存在 10 bp 及以上 polyT,因而基本支持“全变态昆虫 mtDNA 的 O_R 位于 polyT 下游”这一结论。

(2) 碱基偏好的形成机制. 鳞翅目线粒体基因组碱基组成表现明显的 AT 偏好,其形成机制是什么?一般认为可能与线粒体起源进化有关^[90],而线粒体由单源的内共生细菌进化而来^[91],提示可以从其他内共生细菌获得间接证据。例如, *Buchnera aphidicola* 是蚜虫(aphids)体内的一种内共生细菌,基因组富含 AT。Wernegreen 等人^[92]通过分析发现,在与宿主漫长的共生过程中, *Buchnera* 基因组高 AT 含量可能由于其高 AT 突变及长期遗传漂变导致。这些信息对于探讨鳞翅目线粒体基因组的 AT 偏好具有借鉴意义。

3 结语与展望

鳞翅目作为昆虫纲第二大目,不但物种资源丰富,而且对人类社会影响深远。相比鳞翅目庞大的物种数,目前已完成线粒体全基因组测序的鳞翅目种类还比较少,因此,获得更多鳞翅目物种的线粒体基因组序列显得颇为迫切。在利用各种测序方法^[93]获取线粒体基因组序列的基础上,研究者可进行后续的数据库与网站建设、线粒体基因组注释、线粒体比较基因组学、线粒体转录组学、线粒体蛋白质组学等研究。在应用研究方面,关于鳞翅目高级阶元系统发育关系如“大鳞翅类不同总科之间的关系”、“凤蝶总科不同科之间的关系”等研究还存在不少争议,线粒体基因组序列的测定有望为解决争议问题提供新的途径和手段。最后,基于线粒体基因组研究鳞翅目物种的起源与进化、*Wolbachia*, mtDNA 和鳞翅目宿主性别比例三者的具体关系等,将有望为鳞翅目农林害虫防治、经济昆虫种质资源的利用等提供新的思路和方法。

参考文献

- 1 van Nieukerken E J, Kaila L, Kitching I J, et al. Order Lepidoptera Linnaeus, 1758. In: Zhang Z Q, ed. Animal Biodiversity: An Outline of Higher-Level Classification and Survey of Taxonomic Richness. Zootaxa, 2011, 3148: 212–221
- 2 Kfir R. Origin of the diamondback moth (Lepidoptera: Plutellidae). Ann Entomol Soc Am, 1998, 91: 164–167
- 3 Goldsmith M R, Shimada T, Abe H. The genetics and genomics of the silkworm, *Bombyx mori*. Annu Rev Entomol, 2005, 50: 71–100
- 4 Kristensen N P, Scoble M J, Karsholt O. Lepidoptera phylogeny and systematics: The state of inventorying moth and butterfly diversity. Zootaxa, 2007, 1668: 699–747
- 5 Muirhead K A, Murphy N P, Sallam N, et al. Phylogenetics and genetic diversity of the *Cotesia flavipes* complex of parasitoid wasps (Hymenoptera: Braconidae), biological control agents of lepidopteran stemborers. Mol Phylogenet Evol, 2012, 63: 904–914
- 6 Boore J L. Animal mitochondrial genomes. Nucleic Acids Res, 1999, 27: 1767–1780
- 7 Bae J S, Kim I, Sohn H D, et al. The mitochondrial genome of the firefly, *Pyrocoelia rufa*: Complete DNA sequence, genome organization, and phylogenetic analysis with other insects. Mol Phylogenet Evol, 2004, 32: 978–985
- 8 Kim M J, Kang A R, Jeong H C, et al. Reconstructing intraordinal relationships in Lepidoptera using mitochondrial genome data with the description of two newly sequenced lycaenids, *Spindasis takanonis* and *Protantigius superans* (Lepidoptera: Lycaenidae). Mol Phylogenet Evol, 2011, 61: 436–445
- 9 Hebert P D, Penton E H, Burns J M, et al. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc Natl Acad Sci USA, 2004, 101: 14812–12817
- 10 Li Y P, Song W, Shi S L, et al. Mitochondrial genome nucleotide substitution pattern between domesticated silkworm, *Bombyx mori*, and its wild ancestors, Chinese *Bombyx mandarina* and Japanese *Bombyx mandarina*. Genet Mol Biol, 2010, 33: 186–189
- 11 Cameron S L, Lambkin C L, Barker S C, et al. A mitochondrial genome phylogeny of Diptera: Whole genome sequence data accurately resolve relationships over broad timescales with high precision. Syst Entomol, 2007, 32: 40–59
- 12 Cao Y Q, Ma C, Chen J Y, et al. The complete mitochondrial genomes of two ghost moths, *Thitarodes renzhiensis* and *Thitarodes yunnanensis*: The ancestral gene arrangement in Lepidoptera. BMC Genomics, 2012, 13: 276
- 13 Yukuhiro K, Sezutsu H, Itoh M, et al. Significant levels of sequence divergence and gene rearrangements have occurred between the mitochondrial genomes of the wild mulberry silkworm, *Bombyx mandarina*, and its close relative, the domesticated silkworm, *Bombyx mori*. Mol Biol Evol, 2002, 19: 1385–1389
- 14 Coates B S, Sumerford D V, Hellmich R L, et al. Partial mitochondrial genome sequences of *Ostrinia nubilalis* and *Ostrinia furnicalis*. Int J Biol Sci, 2005, 1: 13–18
- 15 Liu Y Q, Li Y P, Pan M H, et al. The complete mitochondrial genome of the Chinese oak silkworm, *Antheraea pernyi* (Lepidoptera: Saturniidae). Acta Biochim Biophys Sin, 2008, 40: 693–703
- 16 Tian L L, Sun X Y, Chen M, et al. Complete mitochondrial genome of the five-dot sergeant *Parathyma sulphitia* (Nymphalidae: Limenitidinae) and its phylogenetic implications. Zool Res, 2012, 33: 133–143
- 17 Kim M J, Jeong H C, Kim S R, et al. Complete mitochondrial genome of the nerippe fritillary butterfly, *Argynnis nerippe* (Lepidoptera: Nymphalidae). Mitochondrial DNA, 2011, 22: 86–88
- 18 Zhang M, Nie X P, Cao T W, et al. The complete mitochondrial genome of the butterfly *Apatura metis* (Lepidoptera: Nymphalidae). Mol Biol Rep, 2012, 39: 6529–6536
- 19 Hu J, Zhang D X, Hao J S, et al. The complete mitochondrial genome of the yellow coaster, *Acraea issoria* (Lepidoptera: Nymphalidae: Heliconiinae: Acraeini): Sequence, gene organization and a unique tRNA translocation event. Mol Biol Rep, 2010, 37: 3431–3438
- 20 Qin X M, Guan Q X, Zeng D L, et al. Complete mitochondrial genome of *Kallima inachus* (Lepidoptera: Nymphalidae: Nymphalinae): Comparison of *K. inachus* and *Argynnis hyperbius*. Mitochondrial DNA, 2012, 23: 318–320
- 21 Kim M J, Kim I. Complete nucleotide sequence and organization of the mitogenome of the endangered *Eumenis autonoe* (Lepidoptera: Nymphalidae). Afr J Biotechnol, 2010, 9: 735–754
- 22 Kim I, Lee E M, Seol K Y, et al. The mitochondrial genome of the Korean hairstreak, *Coreana raphaelis* (Lepidoptera: Lycaenidae). Insect Mol Biol, 2006, 15: 217–225
- 23 Hong G Y, Jiang S T, Yu M, et al. The complete nucleotide sequence of the mitochondrial genome of the cabbage butterfly, *Artogeia melete* (Lepidoptera: Pieridae). Acta Biochim Biophys Sin, 2009, 41: 446–455
- 24 毛增辉, 郝家胜, 朱国萍, 等. 菜粉蝶线粒体基因组的全序列测定和分析. 昆虫学报, 2010, 53: 1295–1304
- 25 Park J S, Cho Y, Kim M J, et al. Description of complete mitochondrial genome of the black-veined white, *Aporia crataegi* (Lepidoptera: Papilionoidea), and comparison to papilionoid species. J Asia Pac Entomol, 2012, 15: 331–341
- 26 Hao J S, Sun Q Q, Zhao H B, et al. The complete mitochondrial genome of *Ctenoptilum vasava* (Lepidoptera: Hesperidae: Pyrginae) and its phylogenetic implication. Comp Funct Genomics, 2012, 2012: 328049

- 27 Jiang S T, Hong G Y, Yu M, et al. Characterization of the complete mitochondrial genome of the giant silkworm moth, *Eriogyna pyretorum* (Lepidoptera: Saturniidae). *Int J Biol Sci*, 2009, 5: 351–365
- 28 Hong M Y, Lee E M, Jo Y H, et al. Complete nucleotide sequence and organization of the mitogenome of the silk moth *Caligula boisduvalii* (Lepidoptera: Saturniidae) and comparison with lepidoptera insects. *Gene*, 2008, 413: 49–57
- 29 Kim S R, Kim M I, Hong M Y, et al. The complete mitogenome sequence of the Japanese oak silkmoth, *Antheraea yamamai* (Lepidoptera: Saturniidae). *Mol Biol Rep*, 2009, 36: 1871–1880
- 30 Kim J S, Park J S, Kim M J, et al. Complete nucleotide sequence and organization of the mitochondrial genome of eri-silkworm, *Samia cynthia ricini* (Lepidoptera: Saturniidae). *J Asia Pac Entomol*, 2012, 15: 162–173
- 31 Liu Q N, Zhu B J, Dai L S, et al. Characterization of the complete mitochondrial genome of the wild silkworm moth, *Actias selene*. *Gene*, 2012, 505: 291–299
- 32 Yin J, Hong G Y, Wang A M, et al. Mitochondrial genome of the cotton bollworm, *Helicoverpa armigera* (Lepidoptera: Noctuidae) and comparison with other Lepidopterans. *Mitochondrial DNA*, 2010, 21: 160–169
- 33 Lee E S, Shin K S, Kim K S, et al. The mitochondrial genome of the smaller tea tortrix *Adoxophyes honmai* (Lepidoptera: Tortricidae). *Gene*, 2006, 373: 52–57
- 34 Zhao J L, Zhang Y Y, Luo A R, et al. The complete mitochondrial genome of *Spilonota lechriaspis* Meyrick (Lepidoptera: Tortricidae). *Mol Biol Rep*, 2011, 38: 3757–3764
- 35 Salvato P, Simonato M, Battisti A, et al. The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). *BMC Genomics*, 2008, 9: 331
- 36 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
- 37 Cameron S L, Whiting M F. The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta* (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. *Gene*, 2008, 408: 112–123
- 38 Sun Q Q, Sun X Y, Wang X C, et al. Complete sequence of the mitochondrial genome of the Japanese buff-tip moth, *Phalera flavescent* (Lepidoptera: Notodontidae). *Genet Mol Res*, 2012, 11: 4213–4225
- 39 Yang L, Wei Z J, Hong G Y, et al. The complete nucleotide sequence of the mitochondrial genome of *Phthonandria atrilineata* (Lepidoptera: Geometridae). *Mol Biol Rep*, 2009, 36: 1441–1449
- 40 Son Y, Kim Y. The complete mitochondrial genome of *Grapholita molesta* (Lepidoptera: Tortricidae). *Ann Entomol Soc Am*, 2011, 104: 788–799
- 41 Chai H N, Du Y Z, Zhai B P. Characterization of the complete mitochondrial genomes of *Cnaphalocrocis medinalis* and *Chilo suppressalis* (Lepidoptera: Pyralidae). *Int J Biol Sci*, 2012, 8: 561–579
- 42 Li W W, Zhang X Y, Fan Z X, et al. Structural characteristics and phylogenetic analysis of the mitochondrial genome of the sugarcane borer, *Diatraea saccharalis* (Lepidoptera: Crambidae). *DNA Cell Biol*, 2011, 30: 3–8
- 43 Tayler M F, McKechnie S W, Pierce N, et al. The lepidopteran mitochondrial control region: Structure and evolution. *Mol Biol Evol*, 1993, 10: 1259–1272
- 44 Kilpert F, Podsiadlowski L. The complete mitochondrial genome of the common sea slater, *Ligia oceanica* (Crustacea, Isopoda) bears a novel gene order and unusual control region features. *BMC Genomics*, 2006, 20: 241
- 45 Clary D O, Wolstenholme G R. The mitochondrial DNA molecular of *Drosophila*: Nucleotide sequence, gene organization, and genetic code. *J Mol Evol*, 1985, 22: 252–271
- 46 Boore J L, Lavrow D, Brown W M. Gene translocation links insects and crustaceans. *Nature*, 1998, 392: 667–668
- 47 Wolstenholme D R. Animal mitochondrial DNA: Structure and evolution. *Int Rev Cytol*, 1992, 141: 173–216
- 48 Xu W, Jameson D, Tang B, et al. The relationship between the rate of molecular evolution and the rate of genome rearrangement in animal mitochondrial genomes. *J Mol Evol*, 2006, 63: 375–392
- 49 Schattner P, Brooks A N, Lowe T M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res*, 2005, 33 (Suppl 2): W686–W689
- 50 Sheffield N C, Song H, Cameron S L, et al. Nonstationary evolution and compositional heterogeneity in beetle mitochondrial phylogenomics. *Syst Biol*, 2009, 58: 381–394
- 51 Krzywinski J, Grushko O G, Besansky N J. Analysis of the complete mitochondrial DNA from *Anopheles funestus*: An improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. *Mol Phylogenet Evol*, 2006, 39: 417–423
- 52 Anderson S, Bankier A T, Barrell B G, et al. Sequence and organization of the human mitochondrial genome. *Nature*, 1981, 290: 457–465
- 53 Ojala D, Montoya J, Attardi G. tRNA punctuation model of RNA processing in human mitochondrial. *Nature*, 1981, 290: 470–474
- 54 Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from international DNA sequence database: Status for the year 2000. *Nucleic Acids Res*, 2000, 28: 292

- 55 Schultz D W, Yarus M. On malleability in the genetic code. *J Mol Evol*, 1996, 42: 597–601
- 56 Abascal F, Posada D, Knight R D, et al. Parallel evolution of the genetic code in Arthropod mitochondrial genomes. *PLoS Biol*, 2006, 4: e127
- 57 Watanabe K, Yikobori S. tRNA modification and genetic code variations in animal mitochondria. *J Nucleic Acids*, 2011, doi: 10.4061/2011/623095
- 58 Santos M A, Cheesman C, Costa V, et al. Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in *Candida* spp. *Mol Microbiol*, 1999, 31: 937–947
- 59 Knight R D, Freeland S T, Landweber L F. Rewiring the keyboard: Evolvability of the genetic code. *Nat Rev Genet*, 2001, 2: 49–58
- 60 Clayton D A. Replication of animal mitochondrial DNA. *Cell*, 1982, 28: 693–705
- 61 Zhang D X, Hewitt G M. Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. *Biochem Syst Ecol*, 1997, 25: 99–120
- 62 Moritz C, Dowling T E, Brown W M. Evolution of animal mitochondrial DNA: Relevance for population biology and systematics. *Annu Rev Ecol Syst*, 1987, 18: 269–292
- 63 Sheffield N C, Song H, Cameron S L, et al. A comparative analysis of mitochondrial genomes in coleoptera and genome descriptions of six new beetles. *Mol Biol Evol*, 2008, 25: 2499–2509
- 64 Knibbe C, Coulon A, Mazet O, et al. A long-term evolutionary pressure on the the amount of noncoding DNA. *Mol Biol Evol*, 2007, 24: 2344–2353
- 65 Hebert P D, Cywinska A, Ball S L, et al. Biological identifications through DNA barcodes. *Proc Biol Sci*, 2003, 270: 313–321
- 66 Li Q Q, Li D Y, Ye H, et al. Using *COI* gene sequence to barcode two morphologically alike species: The cotton bollworm and the oriental tobacco budworm (Lepidoptera: Noctuidae). *Mol Biol Rep*, 2011, 38: 5107–5113
- 67 Zhu L X, Wu X B. Phylogenetic evaluation of the taxonomic status of *Papilio maackii* and *P. syfanius* (Lepidoptera: Papilionidae). *Zool Res*, 2011, 32: 248–254
- 68 Albre J, Gers C, Legal L. Molecular phylogeny of the *Erebia tyndarus* (Lepidoptera, Rhopalocera, Nymphalidae, Satyrinae) species group combining *CoxII* and *ND5* mitochondrial genes: A case study of a recent radiation. *Mol Phylogenet Evol*, 2008, 47: 196–210
- 69 Rubinoff D, Cameron S, Will K. A genomic perspective on the shortcomings of mitochondrial DNA for “barcoding” identification. *J Hered*, 2006, 97: 581–594
- 70 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. *Proc Natl Acad Sci USA*, 2008, 105: 13486–13491
- 71 Brower A V. Problems with DNA barcodes for species delimitation: Ten species of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperriidae). *Syst Biodivers*, 2006, 4: 127–132
- 72 Hurst G D, Jiggins F M. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: The effects of inherited symbionts. *Proc Biol Sci*, 2005, 272: 1525–1534
- 73 Fenn J D, Song H, Cameron S L, et al. A preliminary mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. *Mol Phylogenet Evol*, 2008, 49: 59–68
- 74 Feng X, Liu D F, Wang N X, et al. The mitochondrial genome of the butterfly *Papilio xuthus* (Lepidoptera: Papilionidae) and related phylogenetic analysis. *Mol Biol Rep*, 2010, 37: 3877–3888
- 75 Hwang J S, Lee J S, Goo T W, et al. The comparative molecular study between Bombycidae and Saturniidae based on mtDNA RFLP and cytochrome oxidase *I* gene sequences: Implication for molecular evolution. *Z Naturforsch C*, 1999, 54: 587–594
- 76 Hwang J S, Lee J S, Goo T W, et al. Molecular genetic relationships between Bombycidae and Saturniidae based on the mitochondria DNA encoding of large and small rRNA. *Genet Anal Biomol Eng*, 1999, 15: 223–228
- 77 Li D, Guo Y R, Shao H J, 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
- 78 Hu X L, Cao G L, Xue R Y, et al. The complete mitogenome and phylogenetic analysis of *Bombyx mandarina* strain Qingzhou. *Mol Biol Rep*, 2010, 37: 2599–2608
- 79 Arunkumar K P, Metta M, Nagaraju J. Molecular phylogeny of silkmoths reveals the origin of domesticated silkworm, *Bombyx mori* from Chinese *Bombyx mandarina* and paternal inheritance of *Antheraea proylei* mitochondrial DNA. *Mol Phylogenet Evol*, 2006, 40: 419–427
- 80 Li A L, Zhao Q L, Tang S M, et al. Molecular phylogeny of the domesticated silkworm, *Bombyx mori*, based on the sequences of mitochondrial cytochrome *b* genes. *J Genet*, 2005, 84: 137–142
- 81 Yukuhiro K, Sezutsu H, Tamura T, et al. Nucleotide sequence variation in mitochondrial *COI* gene among 147 silkworm (*Bombyx mori*) strains from Japanese, Chinese, European and molting classes. *Genes Genet Syst*, 2011, 86: 315–323
- 82 潘敏慧, 余泉友, 夏玉玲, 等. 中国野桑蚕线粒体基因组特征. *中国科学 C 辑: 生命科学*, 2008, 38: 751–759

- 83 孙伟, 于红松, 沈以红, 等. 蚕系统发生及进化历史分析. 中国科学: 生命科学, 2012, 42: 489–502
- 84 Strübing U, Lucius R, Hoerauf A, et al. Mitochondrial genes for heme-dependent respiratory chain complexes are up-regulated after depletion of *Wolbachia* from filarial nematodes. *Int J Parasitol*, 2010, 40: 1193–1202
- 85 Sasaki T, Massaki N, Kubo T. *Wolbachia* variant that induces two distinct reproductive phenotypes in different hosts. *Heredity*, 2005, 95: 389–393
- 86 Duploux A, Hurst G D, O'Neill S L, et al. Rapid spread of male-killing *Wolbachia* in the butterfly *Hypolimnas bolina*. *J Evol Biol*, 2009, 23: 231–235
- 87 Delgado A M, Cook J M. Effects of a sex-ratio distorting endosymbiont on mtDNA variation in a global insect pest. *BMC Evol Biol*, 2009, 9: 49
- 88 Mouton L, Henri H, Fleury F. Interactions between coexisting intracellular genomes: Mitochondrial density and *Wolbachia* infection. *Appl Environ Microbiol*, 2009, 75: 1916–1921
- 89 Saito S, Tamura K, Aotsuka T. Replication origin of mitochondrial DNA in insects. *Genetics*, 2005, 171: 1695–1705
- 90 Rodríguez-Trelles F, Tarrío R, Ayala F J. Fluctuating mutation bias and the evolution of base composition in *Drosophila*. *J Mol Evol*, 2000, 50: 1–10
- 91 Gray M W, Burger G, Lang B F. Mitochondrial evolution. *Science*, 1999, 283: 1476–1481
- 92 Wernegreen J J, Funk D J. Mutation exposed: A neutral explanation for extreme base composition of an endosymbiont genome. *J Mol Evol*, 2004, 59: 849–858
- 93 贾万忠, 闫鸿斌, 倪兴维, 等. 蠕虫线粒体基因组研究及其应用进展. 科学通报, 2011, 56: 2358–2372

Advances in comparative mitogenomic studies of Lepidoptera (Arthropoda: Insecta)

WANG Wei^{1,2}, MENG ZhiQi¹, SHI FangXiong² & LI FengBo¹

¹ State Key Laboratory of Breeding Base for Zhejiang Sustainable Pest and Disease Control, Sericulture Research Institute, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China;

² College of Animal Science and Technology, Nanjing Agricultural University, Nanjing 210095, China

Lepidopteran insects have a marked impact on human society, whether functioning as major pests of agriculture and forestry, pollinators, or a primary producer of silk. Insect mitochondrial genome (mitogenome) has a relatively conserved gene content and gene order, so it has been widely used in molecular phylogeny and evolution. To date, complete mitogenomes have been sequenced for a total of 63 species of Lepidoptera. Among the sequenced mitogenomes, however, only 58 are reference sequences available in GenBank. Here, a comparative analysis of all the mitogenome reference sequences revealed some basic features of lepidopteran mitogenomes. Furthermore, the phylogenetic relationships within Lepidoptera based on the complete mitogenome sequences proved to be similar to that based on traditional morphological characters. This work paves the way for better understanding of the evolution of lepidopteran insects, and suggests that comparative mitogenomic analysis has great potential for resolving phylogenetic relationships among the major groups of insects at the dawn of the postgenomic era.

Lepidoptera, mitochondrial genome, comparative genomics, evolution, molecular phylogeny

doi: 10.1360/972012-1721