

黄振, 郭琼霞. 瘤胫果实蝇线粒体全基因组分析与系统发育 [J]. 福建农业学报, 2024, 39 (4): 482-491.

HUANG Z, GUO Q X. Mitochondrial Genome and Phylogeny of *Bactrocera* (B.) *tuberculata* [J]. *Fujian Journal of Agricultural Sciences*, 2024, 39 (4): 482-491.

瘤胫果实蝇线粒体全基因组分析与系统发育

黄 振^{1,2}, 郭琼霞^{2*}

(1. 福州长乐机场海关, 福建 福州 350209; 2. 福州海关技术中心, 福建 福州 350001)

摘要:【目的】明确瘤胫果实蝇 [*Bactrocera tuberculata* (Bezzi)] 线粒体全基因组序列特征, 探究其系统发育地位, 为瘤胫果实蝇分子标记、物种鉴定和进化遗传学研究提供参考依据。【方法】采用高通量测序技术对瘤胫果实蝇的线粒体全基因组进行测序、组装、拼接和注释, 获得基因组全序列, 并对基因组基本结构进行分析; 选择 NCBI 公布的果实蝇属近缘种 20 种实蝇的线粒体全基因组序列, 基于最大似然法 (Maximum likelihood, ML) 进行系统发育分析。【结果】瘤胫果实蝇线粒体基因组序列总长度为 15 854 bp; 具 37 个基因, 其中包含 13 个蛋白编码基因、22 个 tRNA 基因、2 个 rRNA 基因和 1 个非编码控制基因。瘤胫果实蝇线粒体基因组全序列 A+T 含量为 73.2%; 13 个蛋白质编码基因共有密码子 3 755 个, 在蛋白质所有 22 种氨基酸密码子中, UUA (leucine) 的使用频率 N 及相对密码子 RSCU 使用频率均最高, N (RSCU) 为 387 (3.79); tRNA 基因二级结构中, 除苯丙氨酸 (F) 以及苏氨酸 (T) 缺少假尿嘧啶 (T) 环和丝氨酸 (S1) 缺少二氢尿嘧啶 DHU 环, 其余 19 个 tRNA 基因均能折叠形成典型的三叶草二级结构, 22 个 tRNA 基因中, 共存在 178 处碱基对错配, 错配碱基对均为 G-U。通过对其近缘种线粒体全基因组的系统发育分析, 瘤胫果实蝇与桔小实蝇 (*Bactrocera dorsalis*) 和杨桃实蝇 (*Bactrocera carambolae*) 聚在一起, 表明 3 种实蝇的亲缘关系较近, 与果实蝇亚属的其他种类聚在同个分支上, 结果与形态鉴定一致。【结论】首次报道了瘤胫果实蝇线粒体全基因组, GenBank 登录号 MW 892 726, 测定和分析的瘤胫果实蝇线粒体基本结构特征, 核苷酸组成、系统进化分析, 支持瘤胫果实蝇属于果实蝇亚属, 为物种诊断、进化生物学和防治研究提供依据。**关键词:** 瘤胫果实蝇; 线粒体全基因组; 序列分析; 相对密码子使用频率; tRNA 基因二级结构; 系统发育

中图分类号: Q969

文献标志码: A

文章编号: 1008-0384 (2024) 04-0482-10

Mitochondrial Genome and Phylogeny of *Bactrocera* (B.) *tuberculata*

HUANG Zhen^{1,2}, GUO Qionxia^{2*}

(1. Fuzhou Changle Airport Customs, Fuzhou, Fujian 350209, China; 2. Fuzhou Customs Technical Center, Fuzhou, Fujian 350001, China)

Abstract: 【Objective】 Mitochondrial genome and phylogeny of *Bactrocera* (B.) *tuberculata* were studied for advancement on the molecular markers design, species identification, and evolutionary genetics relating to the destructive fruit fly. 【Methods】 High-throughput sequencing was applied to determine, assemble, join, and annotate the complete genome of *B. (B.) tuberculata* mitochondria. With the sequences on selected 20 species published on NCBI, the phylogeny of fruit flies was analyzed using the maximum likelihood method (ML). 【Results】 The total mitochondrial genome sequence was 15 854 bp long containing 13 protein coding genes, 22 tRNA genes, 2 rRNA genes, and one non-coding control gene with 73.2% A+T. There were 3 755 codons in the 13 protein-coding genes. Of the 22 amino acid codons in the protein, UUA (leucine) had the highest frequency (N) of 387 and relative codon (RSCU) of 3.79. Aside from the phenylalanine (F) and threonine (T) that lacked pseudouracine (T) rings and the serine (S1) without a dihydrouracil DHU ring, the secondary structures of the remaining 19 tRNA genes shaped typically like a canonical cloverleaf. And the 22 genes had 178 mismatched G-U base pairs. Based on the mitochondrial analysis, the phylogeny of *B. (B.) tuberculata*, *B. (B.) dorsalis*, and *B. (B.) carambolae* were closely related and in the same branch as other subgenera. The result agreed with what was revealed by the morphological observation. 【Conclusion】 For the first time, the complete mitochondrial genome of *B. (B.) tuberculata* was obtained with a GenBank accession number of MW 892726. The information secured on the structure and nucleotide composition of the mitochondria

收稿日期: 2023-12-05 修回日期: 2024-04-01

作者简介: 黄振 (1985—), 男, 硕士, 高级农艺师, 主要从事植物检疫、农业昆虫与害虫防治研究, E-mail: hz15880010585@126.com

* 通信作者: 郭琼霞 (1954—), 女, 研究员, 主要从事植物检疫研究, E-mail: gqxfjciq@126.com

基金项目: 福建省自然科学基金项目 (2011J01066、2012J01061); 福州海关科研项目 (fk2022-09)

and the phylogenetic relation with other subgenera would aid further studies on the species identification, evolutionary biology, and pest control of the devastating pest.

Key words: *Bactrocera* *Bactrocera tuberculata* (Bezzi); mitochondrial genome; sequence analysis; relative codon usage frequency; secondary structure of tRNA genes; phylogeny

0 引言

【研究意义】瘤胫果实蝇 [*Bactrocera* *Bactrocera tuberculata* (Bezzi)], 隶属双翅目 (Diptera) 实蝇科 (Tephritidae) 果实蝇属 (*Bactrocera*) 果实蝇亚属 (*Bactrocera*)。目前主要分布在缅甸^[1]、老挝^[2]、不丹^[3-4]等国家和地区, 我国在云南与缅甸、老挝等国家边境也开展了诱捕监测^[1-2]。瘤胫果实蝇是危害果蔬的重要害虫, 为我国进境植物检疫性害虫。近年来, 中边贸易频繁, 与缅甸、老挝等国家进境果蔬种类和数量都在不断增加, 频繁的果蔬贸易给检疫性瘤胫果实蝇的传入带来了潜在的危险^[5]。从口岸进境果蔬检疫的情况看, 截获到实蝇类的害虫种类多, 形态相似, 给物种鉴定造成困难^[6]。尤其是在传统形态分类学的基础上, 结合分子生物学技术, 采用分子标记及数据重建, 系统发育关系, 解决物种鉴定、确定分类地位以及同属种近缘遗传进化等问题都具有重要意义。【前人研究进展】线粒体基因组具有结构简单、进化速率较快、含有丰富的遗传进化信息, 不同生物类群的线粒体基因组有其各自独特的结构特征、进化规律和方式、DNA 测序技术以及线粒体全基因组测序策略, 所以, 线粒体基因已成为研究动物起源、进化及群体遗传分化的理想对象, 目前被广泛应用于昆虫不同分类系统的进化、分析系统发育关系、物种诊断、谱系地理学等领域的研究^[7]。近年来, 伴随现代各种分子生物学软件的不开发和应用, 以及测序技术的不断成熟和发展, 许多学者也对实蝇科不同属线粒体基因进行了研究。在测序方面, 要获得线粒体基因组全序列, 学者们有的根据 GenBank 上公布的实蝇的 COI、COII、COIII 等部分序列, 设计扩增实蝇的 PCR 引物; 有的是根据公开报道的实蝇 mtDNA 全序列中的同源片段序列, 设计引物; 之后进行 PCR 扩增、产物纯化、克隆等方法进行测序, 拼接组装获得线粒体基因全系列^[8-9]。这种传统的线粒体基因组测序, 效率相对低下, 试验引物 20 多对, 测试比对程序繁琐, 费时费力, 尤其是由于控制区 AT 含量高和连续重复碱基较多的特点, 导致一代测序, 即 Sanger 测序获取全长序列比较困难。随着高通量测序技术的发展, Illumina Miseq 平台为代表的二代测序技术, 以其通量高、周期短、价格低的特点, 迅速占领了 DNA 测序领域的主流市场^[10], 高通量测序

技术也被逐步应用于线粒体基因组测序, 成为获取线粒体基因组序列的更有力手段, 所以, 越来越多实蝇类物种的线粒体基因组全序列被注释发表。经查阅 GenBank 数据库中有对实蝇科昆虫线粒体基因组序列研究共计 39 条, 涉及 3 属 14 种, 其中包括果实蝇属 (*Bactrocera*) 4 亚属 12 个物种 37 条^[11]; 在进化发育与其亲缘关系研究中, 多是采用线粒体基因组的某个基因, 基于 DNA 条形码技术^[12], 通过对 mtDNA COI、COXII 单独基因片段进行系统发育分析, 比对与近缘种之间的相似距离, 分析果实蝇近缘种之间的亲缘关系^[13-14]; 但单一基因进行系统进化方面研究仍有局限性, 如单一采用 COI 基因不适合用于桔小实蝇 (*Bactrocera dorsalis*)、木瓜实蝇 (*Bactrocera papayae*)、杨桃实蝇 (*Bactrocera carambolae*) 和菲律宾实蝇 (*Bactrocera philippinensis*) 等复合种团的分析^[8]。要解决类似桔小实蝇复合团等进化问题, 还需要进化速度不等的多个基因或线粒体基因进行遗传和系统发育分析。【本研究切入点】对瘤胫果实蝇的线粒体全基因序列进行测定, 采用线粒体全基因序列探究各基因的结构特征、核苷酸的大小和排列位置, 明确其分类地位, 及与其他果实蝇种类的亲缘关系等的相关研究还有待深入进行。【拟解决的关键问题】基于 Illumina Hiseq PE150 平台, 采用高通量双向测序技术, 测定瘤胫果实蝇线粒体基因组全序列, 对其线粒体全基因组基本结构特征、重叠和间隔、碱基组成和基因排序, 偏移率, 蛋白质编码基因的密码子的起止、相对同义密码子使用频率, tRNA 基因的二级结构特征等进行分析, 明确瘤胫果实蝇的结构特征; 选取 Genbank 数据库中公布的果实蝇属的线粒体基因组序列与瘤胫果实蝇线粒体基因组序列, 基于最大似然法 (Maximum likelihood, ML) 构建系统发育树, 进行系统发育分析。比较果实蝇各物种间的系统发育及亲缘关系, 明确瘤胫果实蝇的分类地位, 为瘤胫果实蝇分类鉴定、分子标记基因的选取和分子系统发育研究提供参考依据。

1 材料与方法

1.1 供试材料

瘤胫果实蝇 [*Bactrocera* (B.) *tuberculata*] 来自中国云南口岸, 成虫采用形态分类的方法进行鉴定,

置于-20℃冰箱保存。

1.2 方法

瘤胫果实蝇 DNA 提取采用试剂盒 (rapid animal genomic dna isolation kit), 按照说明书要求进行, 提取的 DNA 进行质检、纯化与文库的构建。采用 Illumina Hiseq PE150 平台进行双向测序 (委托上海生工生物工程技术服务有限公司进行), 使用序列拼接软件 (Spades, v3.12.0) 对测序的序列进行组装拼接, 用 GeSeq 软件对线粒体基因组进行基因注释, 使用 Win32 CodonW 对蛋白质基因密码子的使用个数以及同义相对密码子 RSCU 值的使用频率进行分析, 采用 MITOS WebServer 和 tRNAscan-SE 2.0 软件, 对 tRNA 的二级结构进行分析, 基于最大似然法使用 PhyloSuite 软件重建系统发育树, 进行系统发育分析, 研究近缘种之间的亲缘关系^[15]。

2 结果与分析

2.1 线粒体全基因组基本结构

瘤胫果实蝇 [*B. (B.) tuberculata*] 线粒体全基因组序列全长为 15943 bp (GenBank 登录号: MW892726)。包含 13 个蛋白编码基因 (PCGs), 22 个 tRNA 基因, 2 个 rRNA 基因及 1 个非编码控制区 (non-coding region)。测定的全基因组 37 个基因结构紧凑, 排列方式与已发表的果蝇 (*Drosophila yakuba*)、桔小实蝇等实蝇科实蝇的线粒体全基因组一样, 具有代表性和相似性^[16]。

全基因组在 J 链编码的基因有 24 个, 其中包括 9 个蛋白编码基因 ND2、COX1、COX2、ATP8、ATP6、COX3、ND3、ND6、CYTB 和 14 个 tRNA 基因 *trnI*-GAU、*trnM*-CAU、*trnW*-UCA、*trnL*-UAA、*trnK*-CUU、*trnD*-GUC、*trnG*-UCC、*trnA*-UGC、*trnR*-UCG、*trnN*-GUU、*trnS*-GCU、*trnE*-UUC、*trnT*-UGU、*trnS*-UGA, 以及 1 个非编码控制区 (non-control region)。在 N 链编码基因有 14 个, 包括 4 个蛋白编码基因 ND5、ND4、ND4L、ND1 和 8 个 tRNA 基因 *trnQ*-UUG、*trnC*-GCA、*trnY*-GUA、*trnF*-GAA、*trnH*-GUG、*trnP*-UGG、*trnL*-UAG、*trnV*-UAC, 以及 2 个 rRNA 基因 l-rRNA、s-rRNA。全基因组在 16 个基因连接处共发现 113 bp 碱基重叠, 重叠长度在 1~22 bp, 重叠最长在 *trnS*-UGA 与 ND1 和 *trnL*-UAG 与 l-rRNA 之间, ATP8/ATP6 和 ND4/ND4L 之间均重叠 7 bp。共 1 182 bp 分布在 14 个基因间隔中, 间隔长度在 2~965 bp, 碱基间隔最大在 ND1 与 *trnL*-UAG 之间。全基因组既无重叠又无间隔基因有 6 个, 实蝇基因组的基本结构和基因

排列顺序是祖先昆虫的线粒体全基因组典型集合, 与古昆虫线粒体全基因组的排列相同^[17]。全基因组既没有重叠又没有间隔的基因有 6 个, 表明瘤胫果实蝇线粒体全基因组结构紧密, 基本结构、碱基组成, 也与已发表的其他实蝇科实蝇的同源序列比较具有代表性和相似性。测定的瘤胫果实蝇的线粒体全基因组结构特征表与基因组基因结构图见表 1 和图 1。

2.2 碱基组成和含量

瘤胫果实蝇全基因组核苷酸碱基组成: 39.1% A、16.4% C、10.4% G、34.1% T, 为 A > T > C > G, A+T 总含量高达 73.2%; 其中蛋白质编码基因 PCGs、rRNA 基因、tRNA 基因、非编码控制区 A+T 的含量分别为 70.7%、77.4%、74.7% 和 88.4%, 其中控制区基因 A+T 的含量高达 80% 以上。表明瘤胫果实蝇全基因组核苷酸碱基组成具有明显的 A+T 偏向性, 表现出昆虫线粒体全基因组中著名的“A+T”偏好性^[18-19]。

通过偏移率: $AT\ skew = (A - T) / (A + T)$ 、 $GC\ skew = (G - C) / (G + C)$ 测算, 线粒体基因 AT skew 为 0.068, 高于 GC skew -0.224, 存在 AT 偏移; PCGs J 链的 AT skew 为 -0.067, 高于 N 链 -0.277; GC Skew N 链为 0.308, 高于 J 链 -0.191; tRNA 基因 AT Skew J 链为 0.005, 高于 N 链 -0.042, GC Skew N 链为 0.300, 高于 J 链 0.019。所以, PCGs 和 tRNA 基因均存在 AT 向 J 链偏移, GC 向 N 链偏移, 与 J 链和 N 链碱基组成偏向性结果一致。这种两条链之间存在碱基组成不对称性的现象表现了后生动物线粒体全基因组的显著特征^[20]。瘤胫果实蝇线粒体全基因组碱基组成见表 2。

2.3 蛋白质编码基因

瘤胫果实蝇线粒体全基因组蛋白质编码基因序列总长 11 302 bp。A+T 总含量为 70.7%, J 链和 N 链 A+T 含量分别为 68.6% 和 73.9%, J 链碱基含量大小: T > A > C > G, N 链碱基含量大小: T > A > G > C。

在 13 个蛋白质编码基因的起始密码子, 起始密码子中, COX1 基因起始启动密码子为预期 TCG 外, ATP8 为 GTG, ND1 为 ATA, 其余启动密码子 ND2、ND3、ND5、ND6 为 ATT, COX2、ATP6、COX3、ND4、ND4L、CYTB 为 ATG; 终止密码子除 CYTB 为 ATT 终外, 其余基因均为 TAA 停止, 没有发现不完整的终止密码子。

瘤胫果实蝇共有密码子 3 755 个, J 链含 2 290 个, N 链含 1 465 个; 使用频率最高的密码子为: 第

表 1 瘤胫果实蝇线粒体基因组特征
Table 1 Characteristics of mitochondrial genome of *B. (B.) tuberculata*

基因 Gene	位置 Location/bp	基因长度 Gene length/bp	编码链 Strand J(+), N(-)	基因间隔 Intergenic sequence/bp	起始密码子 Start codon	终止密码子 Stop codon
<i>trnI</i> -GAU	1~66	66	J			
<i>trnQ</i> -UUG	64~132	69	N	-3		
<i>trnM</i> -CAU	201~269	69	J	68		
ND2	270~1 292	1 023	J	0	ATT	TAA
<i>trnW</i> -UCA	1 303~1 371	69	J	10		
<i>trnC</i> -GCA	1 364~1 426	63	N	-8		
<i>trnY</i> -GUA	1 473~1 539	67	N	46		
COX1	1 538~3 076	1 539	J	-2	TCG	TAA
<i>trnL</i> -UAA	3 072~3 137	66	J	-5		
COX2	3 142~3 831	690	J	4	ATG	TAA
<i>trnK</i> -CUU	3 836~3 906	71	J	4		
<i>trnD</i> -GUC	3 910~3 976	67	J	2		
ATP8	3 977~4 138	162	J	0	GTG	TAA
ATP6	4 132~4 809	678	J	-7	ATG	TAA
COX3	4 809~5 597	789	J	-1	ATG	TAA
<i>trnG</i> -UCC	5 607~5 671	65	J	9		
ND3	5 672~6 025	354	J	0	ATT	TAG
<i>trnA</i> -UGC	6 024~6 088	65	J	-2		
<i>trnR</i> -UCG	6 096~6 159	64	J	7		
<i>trnN</i> -GUU	6 171~6 235	65	J	11		
<i>trnS</i> -GCU	6 236~6 303	68	J	0		
<i>trnE</i> -UUC	6 304~6 370	67	J	0		
<i>trnF</i> -GAA	6 392~6 456	65	N	21		
ND5	6 437~8 176	1 740	N	-20	ATT	TAG
<i>trnH</i> -GUG	8 192~8 257	66	N	15		
ND4	8 258~9 598	1 341	N	0	ATG	TAG
ND4L	9 592~9 888	297	N	-7	ATG	TAA
<i>trnT</i> -UGU	9 891~9 950	60	J	-8		
<i>trnP</i> -UGG	9 971~10 036	66	N	20		
ND6	10 039~10 563	525	J	2	ATT	TAA
CYTB	10 563~11 697	1 135	J	-1	ATG	ATT
<i>trnS</i> -UGA	11 697~12 725	1 029	N	-1	ATA	TAA
ND1	11 704~11 770	67	J	-22		
<i>trnL</i> -UAG	12 736~12 800	65	N	965		
l-rRNA	12 778~14 135	1 358	N	-22		
<i>trnV</i> -UAC	14 133~14 204	72	N	-3		
s-rRNA	14 204~14 995	791	N	-1		
控制区 Control region	14 996~15 943	947	J	0		

表 2 瘤胫果实蝇线粒体全基因组碱基组成
Table 2 Bases in mitochondrial genome of *B. (B.) tuberculata*

区域 Region	A/%	C/%	G/%	T/%	A+T/%	AT偏移 AT Skew	GC偏移 GC Skew	基因长度 Gene length/bp
全基因组 Whole genome	39.1	16.4	10.4	34.1	73.2	0.068	-0.224	15943
PCGs	29.9	14.9	14.4	40.8	70.7	-0.154	-0.017	11302
PCGs J	32.0	18.7	12.7	36.6	68.6	-0.067	-0.191	6895
PCGs N	26.7	9.0	17.0	47.2	73.9	-0.277	0.308	4407
ND2	34.1	16.8	9.1	40.0	74.1	-0.080	-0.297	1023
COXI	30.1	18.9	16.2	34.8	64.9	-0.072	-0.077	1539
COX2	33.5	18.4	14.1	34.1	67.6	-0.009	-0.132	690
ATP8	37.7	17.9	8.6	35.8	73.5	0.026	-0.351	162
ATP6	30.4	19.9	11.8	37.9	68.3	-0.110	-0.256	678
COX3	29.7	19.4	15.3	35.6	65.3	-0.090	-0.118	789
ND3	31.9	16.7	9.9	41.5	73.4	-0.131	-0.256	354
ND5	27.8	9.0	17.1	46.0	73.8	-0.247	0.310	1740
ND4	26.3	8.9	16.9	47.8	74.1	-0.290	0.310	1341
ND4L	26.6	7.1	14.5	51.9	78.5	-0.322	0.343	297
ND6	37.0	16.6	6.9	39.6	76.6	-0.034	-0.413	525
CYTB	30.9	20.7	13.5	34.9	65.8	-0.061	-0.211	1135
ND1	25.5	9.7	17.6	47.2	72.7	-0.298	0.289	1029
tRNA	36.9	11.1	14.1	37.8	74.7	-0.012	0.119	1462
tRNA J	37.2	12.7	13.2	36.8	74.0	0.005	0.019	929
tRNA N	36.4	8.4	15.6	39.6	76.0	-0.042	0.300	533
rRNA	35.7	8.0	14.6	41.7	77.4	-0.078	0.292	2149
1-rRNA	36.7	7.1	13.8	42.5	79.2	-0.073	0.321	1358
s-rRNA	34.1	9.5	16.1	40.3	74.4	-0.083	0.258	791
控制区 Control region	46.9	6.6	5.0	41.5	88.4	0.061	0.138	947

rRNA 两个基因核苷酸长度分别为 1 358 bp 和 791 bp，A+T 含量分别为 79.2% 和 74.4%，碱基含量大小均为 T> A> G> C。AT Skew 分别为 -0.073 和 -0.083<GC Skew 0.321 和 0.258，存在 GC 偏移。

2.6 非编码控制基因

非编码控制基因在 J 链上。位于 s-rRNA 后，碱基长度 947 bp，A、C、G、T 含量分别为 46.9%、6.6%、5.0%、41.5%，A+T 含量为 88.4%，为 A+T 富含区。非编码控制基因 AT Skew 0.061<GC Skew 0.138，存在 GC 偏移。

2.7 瘤胫果实蝇与其同属近缘种类的系统发育分析

将本研究测定的瘤胫果实蝇的线粒体全基因组序列作为目标序列，利用 GenBank 中已有的果实蝇

亚属（*Bactrocera*）9 种及自测的锈红实蝇线粒体全基因组序列全长 15 972 bp（GenBank 登录号：MW 892729）、簇果实蝇亚属（*Zeugodacus*）7 种和属外群地中海实蝇（*Ceratitis capitata*），共有 19 种实蝇的线粒体全基因序列，基于最大似然法（ML）构建系统发育树，结果表明：瘤胫果实蝇、桔小实蝇和杨桃实蝇聚成一支，表明 3 种实蝇的亲缘关系较近；锈红实蝇、番石榴实蝇及桃实蝇以及果实蝇亚属的其他种类聚成一支，果实蝇亚属与簇果实蝇亚属的种类分成两大独立的支系，亲缘关系距离较远；从系统发育分析结果表明瘤胫果实蝇的分类地位、亲缘关系与传统形态鉴定结果一致，支持瘤胫果实蝇属于果实蝇亚属的结论（图 3）。

表 3 瘤胫果实蝇线粒体基因组蛋白质编码基因密码子使用统计
Table 3 Statistics of codon usage in mitochondria of *B. (B.) tuberculata*

密码子 Codon	N	R	密码子 Codon	N	R	密码子 Codon	N	R	密码子 Codon	N	R
UUU(F)	235	1.42	UCU (S)	114	2.68	UAU(Y)	144	1.73	UGU(C)	41	1.91
UUC(F)	96	0.58	UCC (S)	10	0.24	UAC(Y)	22	0.27	UGC(C)	2	0.9
UUA(L)	387	3.79	UCA (S)	95	2.24	UAA(*)	0	0.00	UGA(W)	95	1.88
UUG(L)	68	0.67	UCG (S)	2	0.05	UAG(*)	0	0.00	UGG(W)	6	0.12
CUU(L)	66	0.65	CCU(P)	75	2.17	CAU(H)	55	1.39	CGU(R)	15	1.02
CUC(L)	5	0.05	CCC(P)	13	0.38	CAC(H)	24	0.61	CGC(R)	3	0.20
CUA(L)	81	0.79	CCA(P)	48	1.39	CAA(Q)	74	1.87	CGA(R)	37	2.51
CUG(L)	5	0.05	CCG(P)	2	0.06	CAG(Q)	5	0.13	CGG(R)	4	0.27
AUU(I)	298	1.76	ACU (T)	79	1.57	AAU(N)	163	1.72	AGU(S)	47	1.11
AUC(I)	40	0.24	ACC (T)	19	0.38	AAC(N)	26	0.28	AGC(S)	9	0.21
AUA(M)	174	1.73	ACA (T)	101	2.01	AAA(K)	63	1.42	AGA(S)	63	1.48
AUG(M)	27	0.27	ACG (T)	2	0.04	AAG(K)	26	0.58	AGG(S)	0	0.00
GUU(V)	84	1.53	GCU(A)	116	2.43	GAU(D)	58	1.63	GGU(G)	71	1.23
GUC(V)	8	0.15	GCC(A)	22	0.46	GAC(D)	13	0.37	GGC(G)	3	0.05
GUA(V)	112	2.02	GCA(A)	50	1.05	GAA(E)	73	1.95	GGA(G)	126	2.18
GUG(V)	16	0.29	GCG(A)	3	0.06	GAG(E)	2	0.05	GGG(G)	31	0.54

N为线粒体全基因组蛋白质编码基因的密码子使用个数；R：RSCU为相对同义密码子的使用频率。*：终止密码子。
N: number of codons used for protein-coding genes of whole mitochondrial genome; R: frequency of use of relative synonymous codon (RSCU). *: stop codon.

3 讨论与结论

本研究所测的瘤胫果实蝇线粒体全基因组的大小介于已发表的实蝇科线粒体基因组全序列的15.6~16.0 kb^[11]。瘤胫果实蝇的线粒体基因组在16个连接处共发现113 bp重叠，重叠最长的碱基数为22 bp，位于 *trnS*-UGA)与ND1和 *trnL*-UAG与1-*rRNA*之间；在所有基因组中，ATP8/ATP6和ND4/ND4L两对基因之间都重叠7个碱基，与叶军等^[9]研究的果实蝇属的橘小实蝇、木瓜实蝇、菲律宾实蝇和杨桃实蝇线粒体基因组，以及Zhang等^[22]研究的葫芦寡鬃实蝇 (*Dacus bivittatus*)和埃塞俄比亚寡鬃实蝇 (*Dacus ciliatus*)的线粒体基因组的同样基因位置重叠7个碱基一致，这可能是实蝇科物种基因组序列排序的一个特征。此外，瘤胫果实蝇的线粒体基因组共发现既没有重叠又没有间隔的基因6个。表明瘤胫果实蝇线粒体全基因组结构紧密，与其他昆虫比较具有代表性和相似性；线粒体基因组所有区域均有明显的AT偏好性，全基因组核苷酸碱基含量A+T总含量高达73.2%；在线粒体全基因组核苷酸序列PCGs基因、tRNA、rRNA基因和非编码控制基因的A+T含量在线粒体基因组所有区域均明显高于G+C，均有明显的AT偏好性，这一特征和其他实蝇线粒体基因组的序列具有相似性^[17-18]。

在线粒体基因组中，PCGs蛋白质编码基因核苷酸A+T含量最低，其中COXI基因的核苷酸A+T含量又是PCGs蛋白质编码基因中的最低，为64.9%，为此相对较高的GC含量，有利于降低基因的突变率，保持基因的稳定性，这有利于物种编码其所需的蛋白质功能，所以，很多学者常将COXI的DNA条形码基因，应用于昆虫的识别鉴定^[23]。

通过线粒体基因组偏移率测算，AT Skew为0.068、GC Skew为-0.224，为AT偏移，结果与碱基组成偏向性一致，这一特征与后生动物典型的线粒体基因组AT偏移相一致^[20]。

从密码子使用的个数和相对同义密码子的使用频率统计分析，表明在所有22种氨基酸密码子中，UUA (leucine)的使用频率N及相对密码子RSCU使用频率均最高，N (RSCU)为387 (3.79)；在22个tRNA中，除苯丙氨酸 (F)以及苏氨酸 (T)缺少假尿嘧啶 (T)环和丝氨酸 (S1)缺少二氢尿嘧啶 DHU环，其余19个tRNA基因均能折叠形成典型的三叶草二级结构；共存在178处碱基对错配，错配碱基对均为G-U，这些都表明G-U配对在线粒体基因组中，很可能是一种正常的配对形式^[24-25]。

将本研究测定的瘤胫果实蝇的线粒体全基因组序列作为目标序列，利用GenBank中已有的果实蝇亚属 (*Bactrocera*)、簇果实蝇亚属 (*Zeugodacus*)种和属外群地中海实蝇 (*Ceratitis capitata*)以及自测

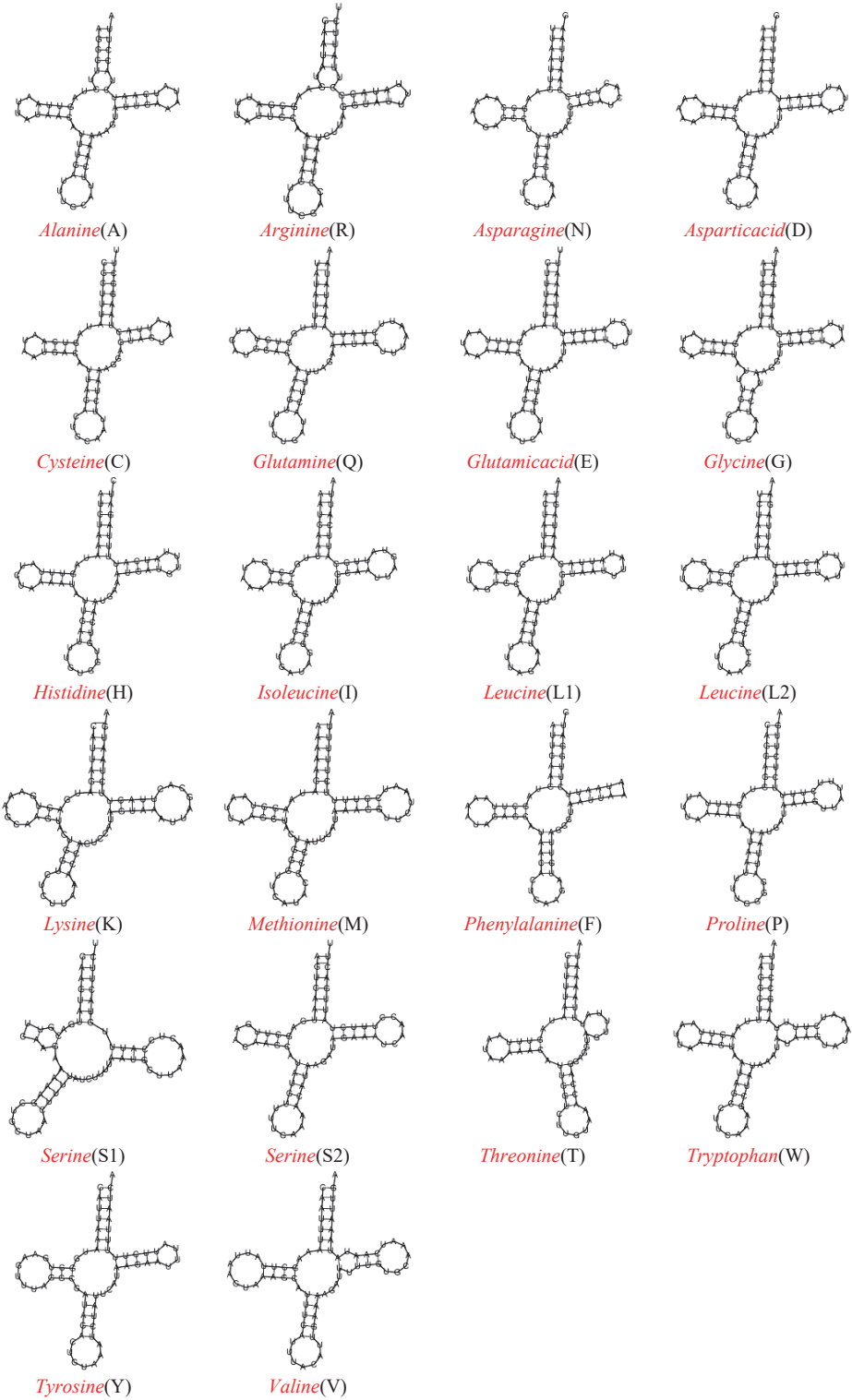


图 2 瘤胫果实蝇线粒体 22 个 tRNA 结构

Fig. 2 Cloverleaf-like structure of 22 inferred tRNAs in mitogenome of *B. (B.) tuberculata*

的锈红实蝇 [*B. (B.) rubigina*] 线粒体全基因组序列全长 15 972 bp (GenBank: MW 892 729)，采用最大似然法 (ML) 构建系统发育树，分析结果表明瘤胫果实蝇+桔小实蝇 [*B. (B.) dorsalis*]+杨桃实蝇 [*B. (B.) carambolae*] 聚成一支，亲缘关系最密切，同归属为果实蝇亚属，与形态学分类结果一致。

当前动物线粒体基因组的测序在飞速增长，但实蝇科线粒体基因组全序列的数据还明显偏少，尤其是果实蝇属种类多、发现的新种也多，而 mtDNA 基因序列研究的数量仍有限，从而影响和限制了果实蝇属实蝇分子进化的研究，所以本研究所建立瘤胫果实蝇线粒体全基因组测定与分析的结果和开展

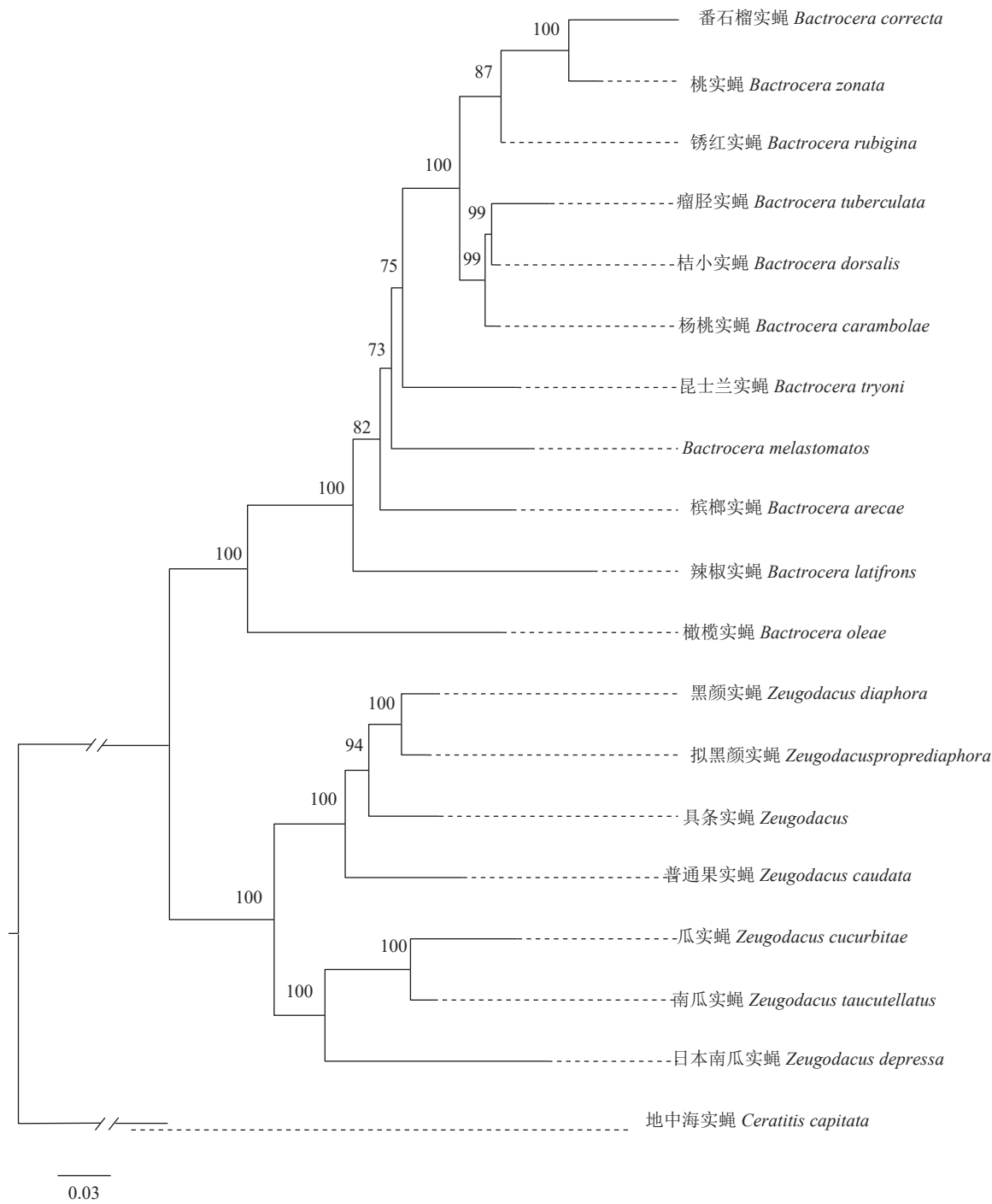


图 3 瘤胫果实蝇的最大似然法 (ML) 系统发育树
Fig. 3 ML-generated phylogenetic tree of *B. (B.) tuberculata*

更多的实蝇种类的线粒体基因组研究，可为实蝇分子标记、物种诊断、分类地位以及快速鉴定等奠定基础并提供重要参考，为后续获得实蝇科更多物种线粒体基因组及重建系统发育关系、研究基因突变提供依据。

参考文献:

[1] 邓裕亮, 李正跃, 蒋小龙, 等. 缅甸第四特区实蝇类害虫调查 [J]. 植物检疫, 2005, 19 (5): 316-318.

DENG Y L, LI Z Y, JIANG X L, et al. Investigation on fruit fly pests in the fourth special zone of Myanmar [J]. *Plant Quarantine*, 2005, 19(5): 316-318. (in Chinese)
[2] 邓裕亮, 李志红, 白永华, 等. 老挝中北部地区果实蝇属 (*Bactrocera*) 害虫种类初步调查 [J]. *植物检疫*, 2010, 24 (1): 52-53.
DENG Y L, LI Z H, BAI Y H, et al. Preliminary investigation on insect species of *Bactrocera* in central and northern Laos [J]. *Plant Quarantine*, 2010, 24(1): 52-53. (in Chinese)
[3] DREW R A I, ROMIG M C, DORJI C. Records of Dacine fruit flies and new species of *Dacus* (Diptera: Tephritidae) in Bhutan [J]. *The*

- Raffles Bulletin of Zoology*, 2007, 55(1): 1–21.
- [4] LEBLANC L, HOSSAIN M A, KHAN S, et al. Additions to the Fruit Fly Fauna (Diptera: Tephritidae: Dacinae) of Bangladesh, with a Key to the Species [J]. *Proceedings of the Hawaiian Entomological Society*, 2014, 46: 31–40.
- [5] 蒋小龙. 云南边境检疫性实蝇风险分析研究 [J]. *西南农业大学学报*, 2002, 24 (5): 402–405, 421.
- JIANG X L. Risk analysis of quarantine fruit flies along Yunnan border [J]. *Journal of Southwest Agricultural University*, 2002, 24(5): 402–405, 421. (in Chinese)
- [6] 黄振, 侯有明, 郭琼霞, 等. 利用物种特异性 PCR 技术快速鉴定南瓜实蝇 [J]. *福建农业学报*, 2021, 36 (2): 215–220.
- HUANG Z, HOU Y M, GUO Q X, et al. Rapid species-specific PCR identification of *Bactrocera tau* (Diptera: Tephritidae) [J]. *Fujian Journal of Agricultural Sciences*, 2021, 36(2): 215–220. (in Chinese)
- [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 [J]. *Molecular Phylogenetics and Evolution*, 2004, 32(3): 978–985.
- [8] 徐浪, 余道坚, 张润杰, 等. 桔小实蝇线粒体基因组全序列及其分析 [J]. *昆虫学报*, 2007, 50 (8): 755–761.
- XU L, YU D J, ZHANG R J, et al. Sequencing and analysis of the complete mitochondrial genome of the Oriental fruit fly, *Bactrocera dorsalis* (Hendel) (Diptera: Tephritidae) [J]. *Acta Entomologica Sinica*, 2007, 50(8): 755–761. (in Chinese)
- [9] 叶军, 房蕊, 易建平, 等. 4 种实蝇线粒体 DNA 全序列测定 [J]. *植物检疫*, 2010, 24 (3): 11–14.
- YE J, FANG R, YI J P, et al. The complete sequence determination and analysis of four species of *bactrocera* mitochondrial genome [J]. *Plant Quarantine*, 2010, 24(3): 11–14. (in Chinese)
- [10] 梅琰, 岳巧云, 贾凤龙. 双翅目昆虫线粒体基因组研究进展 [J]. *环境昆虫学报*, 2012, 34 (4): 497–503.
- MEI Y, YUE Q Y, JIA F L. Research progress on mitochondrial genomes of Dipteran insect [J]. *Journal of Environmental Entomology*, 2012, 34(4): 497–503. (in Chinese)
- [11] 姜帆, 李志红, 梁亮, 等. 实蝇科昆虫线粒体基因组研究进展 [J]. *植物检疫*, 2016, 30 (3): 12–17.
- JIANG F, LI Z H, LIANG L, et al. Research progress in mitochondrial genomes of Tephritidae insect [J]. *Plant Quarantine*, 2016, 30(3): 12–17. (in Chinese)
- [12] 梁亮, 江威, 余慧, 等. 中国果实蝇属种类的 DNA 条形码鉴定 (双翅目, 实蝇科) [J]. *动物分类学报*, 2011, 36(4): 925–932.
- LIANG L, JIANG W, YU H, et al. Identification of Chinese *bactrocera* species through DNA barcoding (Diptera, Tephritidae) [J]. *Acta Zootaxonomica Sinica*, 2011, 36(4): 925–932. (in Chinese)
- [13] 范京安, 顾海丰, 陈世界, 等. 实蝇科昆虫 mtDNA-COI 基因序列分析及系统发育研究 [J]. *西南大学学报: 自然科学版*, 2009, 31(10): 88–95.
- FAN J A, GU H F, CHEN S J, et al. Sequence analysis of mtDNA-COI gene and molecular phylogeny of Fruitfly (Tephritidae) [J]. *Journal of Southwest University (Natural Science Edition)*, 2009, 31(10): 88–95. (in Chinese)
- [14] 刘慎思, 张桂芬, 万方浩. 基于 mtDNA COI 基因的离腹寡毛实蝇属常见种 DNA 条形码识别和系统发育分析 [J]. *昆虫学报*, 2014, 57 (3): 343–355.
- LIU S S, ZHANG G F, WAN F H. DNA barcoding and phylogenetic analysis of common species of the genus *Bactrocera* (Diptera: Tephritidae) based on mtDNA COI gene sequences [J]. *Acta Entomologica Sinica*, 2014, 57(3): 343–355. (in Chinese)
- [15] KROSCHE M N, SCHUTZE M K, ARMSTRONG K F, et al. A molecular phylogeny for the Tribe Dacini (Diptera: Tephritidae): Systematic and biogeographic implications [J]. *Molecular Phylogenetics and Evolution*, 2012, 64(3): 513–523.
- [16] CLARY D O, WOLSTENHOLME D R. The mitochondrial DNA molecule of *Drosophila yakuba*: Nucleotide sequence, gene organization, and genetic code [J]. *Journal of Molecular Evolution*, 1985, 22(3): 252–271.
- [17] BOORE J L. Animal mitochondrial genomes [J]. *Nucleic Acids Research*, 1999, 27(8): 1767–1780.
- [18] CLARY D O, WOLSTENHOLME D R. *Drosophila* mitochondrial DNA: Conserved sequences in the A+T-rich region and supporting evidence for a secondary structure model of the small ribosomal RNA [J]. *Journal of Molecular Evolution*, 1987, 25(2): 116–125.
- [19] 林兴雨, 刘向阳, 宋南. 大赤隐翅虫的线粒体基因组研究及系统发育分析 [J]. *福建农业学报*, 2023, 38 (5): 607–615.
- LIN X Y, LIU X Y, SONG N. Mitochondrial genome and phylogenetics of *Philonthus spinipes* (Sharp, 1874) [J]. *Fujian Journal of Agricultural Sciences*, 2023, 38(5): 607–615. (in Chinese)
- [20] BOORE J L, LAVROV D V, BROWN W M. Gene translocation links insects and crustaceans [J]. *Nature*, 1998, 392(6677): 667–668.
- [21] 王菊平, 宣善滨, 张育平, 等. 武铠蛱蝶线粒体基因组全序列测定和分析 [J]. *昆虫学报*, 2015, 58 (3): 319–328.
- WANG J P, XUAN S B, ZHANG Y P, et al. Sequencing and analysis of the complete mitochondrial genome of *Chitoria ulupi* (Lepidoptera: Nymphalidae) [J]. *Acta Entomologica Sinica*, 2015, 58(3): 319–328. (in Chinese)
- [22] ZHANG Y, FENG S Q, ZENG Y Y, et al. The first complete mitochondrial genome of *Bactrocera tsuneonis* (Miyake) (Diptera: Tephritidae) by next-generation sequencing and its phylogenetic implications [J]. *International Journal of Biological Macromolecules*, 2018, 118(Pt A): 1229–1237.
- [23] 黄振, 郭琼霞. 基于种特异性 SS-COI 技术快速鉴定瘤胫实蝇 [J]. *生物安全学报*, 2022, 31 (2): 179–184.
- HUANG Z, GUO Q X. Rapid identification of *Bactrocera tuberculata* (Diptera: Tephritidae) by species-specific SS-COI technique [J]. *Journal of Biosafety*, 2022, 31(2): 179–184. (in Chinese)
- [24] ISAZA J P, ALZATE J F, CANAL N A. Complete mitochondrial genome of the Andean morphotype of *Anastrepha fraterculus* (Wiedemann) (Diptera: Tephritidae) [J]. *Mitochondrial DNA Part B, Resources*, 2017, 2(1): 210–211.
- [25] 周志军, 杨明茹, 常岩林, 等. 两种纺织娘线粒体基因组的比较分析 [J]. *昆虫学报*, 2013, 56 (4): 408–418.
- ZHOU Z J, YANG M R, CHANG Y L, et al. Comparative analysis of mitochondrial genomes of two long-legged katydids (*Orthoptera*: Tettigoniidae) [J]. *Acta Entomologica Sinica*, 2013, 56(4): 408–418. (in Chinese)