

doi: 10.3969/j.issn.1674-0858.2016.02.19

# 黄毛纺蚋线粒体基因组全序列测定与分析

王 强<sup>1</sup>, 陈丽芬<sup>1</sup>, 丁俊杰<sup>1</sup>, 刘仰青<sup>2</sup>, 王建国<sup>1\*</sup>

(1. 江西农业大学农学院, 南昌 330045; 2. 江西省疾病预防控制中心, 南昌 330038)

**摘要:** 利用 PCR 步移法对黄毛纺蚋的线粒体基因组全序列进行了测定和分析。黄毛纺蚋线粒体基因组全长 15904 bp (GenBank 序列号 KP793690), 包括 13 个蛋白编码基因、22 个 tRNA 基因、2 个 rRNA 基因以及长度为 939 bp 的非编码区。A、T、C、G 碱基含量分别为 39.1%、35.8%、10.4%、14.7%。9 个蛋白编码基因和 14 个 tRNA 基因在 J 链编码, 其余 4 个蛋白编码基因和 8 个 tRNA 基因在 N 链编码, 基因排列顺序与其它已知双翅目昆虫相同。13 个蛋白编码基因中除 COI 以 TTG 作为起始密码外, 其余蛋白质基因均以 ATN 作为起始密码子, 终止密码子多数为典型的 TAA、TAG, 只有 COI 和 ND4L 以单独的 T 作为终止密码子。在所测得的 22 个 tRNA 基因中, 除 tRNA<sup>Ser</sup> (AGN) 缺少 DHU 臂外, 其余 tRNA 均能形成典型的三叶草结构。

**关键词:** 黄毛纺蚋; 双翅目; 线粒体基因组; 密码子

**中图分类号:** Q966; S433

**文献标志码:** A

**文章编号:** 1674-0858 (2016) 02-0360-10

## Sequence and analysis of complete mitochondrial genome of *Simulium aureohirtum*

WANG Qiang<sup>1</sup>, CHEN Li-Fen<sup>1</sup>, DING Jun-Jie<sup>1</sup>, LIU Yang-Qing<sup>2</sup>, WANG Jian-Guo<sup>1\*</sup> (1. Department of Plant Protection, College of Agriculture, Jiangxi Agriculture University, Nanchang 30045, China; 2. Jiangxi Center for Disease Control and Prevention, Nanchang 330085, China)

**Abstract:** The complete mitochondrial genome of *Simulium aureohirtum* was sequenced and analyzed by using PCR amplification and prime walking. The complete mitochondrial genome of *S. aureohirtum* is a circular molecule of 15904 nucleotides (GenBank Accession No. KP793690), including 22 tRNA genes, 2 rRNA genes, 13 protein-coding genes and 939 bp non-coding region. The nucleotide composition of A, T, C, and G were 39.1%, 35.8%, 10.4%, and 14.7%, respectively. Nine protein-coding genes and 14 tRNA genes were encoded on the J strand, and the other 4 protein-coding genes and 8 tRNA genes were encoded on the N strand. The arrangement of genes in *S. aureohirtum* is consistent with that of other Diptera species. Except the initiation codons of COI gene is a TTG codon, the other protein coding genes start with a typical ATN codon. Most of the 13 PCGs in *S. aureohirtum* have complete termination codons (TAA, TAG), except the termination codons in COI and ND4 that have incomplete T. Twenty one tRNAs have typical cloverleaf structure, except for tRNA<sup>Ser</sup> (AGN) without the dihydrouridine (DHU).

**Key words:** *Simulium aureohirtum*; Diptera; mitochondrial genome; codon

黄毛纺蚋 *Simulium aureohirtum* 是双翅目 Diptera 蚋科 Simuliidae 昆虫, 由 Burnetti (1911) 根据采自印度阿萨姆的蚋种而命名 (Chen *et al.*,

2007)。黄毛纺蚋主要分布于东南亚, 在我国广东、福建和贵州、江西等地有均分布, 但是国内对该昆虫研究却不多 (安继尧和阎丙申, 1997; 陈汉斌,

基金项目: 国家自然科学基金 (31160380, 31360457)

作者简介: 王强, 女, 1988 年生, 硕士研究生, 研究方向昆虫分类与进化, E-mail: taotie12011@sina.cn

\* 通讯作者 Author for correspondence, E-mail: jgwang@jxau.edu.cn

收稿日期 Received: 2015-06-08; 接受日期 Accepted: 2015-07-19

2002)。雌蚋叮刺人畜吸血,会造成人的皮肤红肿、过敏性皮炎,严重者会过敏性休克导致死亡,危害较大(蔡茹,2005)。Thanwisai 等(2006)和王宇平等(2010)利用核基因 ITS2 对泰国黑蝇进行了分子系统发育分析,认为该基因片段可以区分黄毛纺蚋的不同地理种群。Phayuhasena 等(2010)利用线粒体基因 COI, ND4 和 16S rRNA 等序列片段分析泰国蚋的系统发育关系,得到的结果与 Thanwisai 等(2006)的结果相同。Pramual 等(2014)利用 DNA 条形码对泰国的蚋进行了研究分析,证明黄毛纺蚋存在较高的种内遗传差异,暗示该物种可能是一个复合种(Thanwisai *et al.*, 2006; Pramual *et al.*, 2014),同年 Thaijarern 等对泰国 3 个不同地理分布的黄毛纺蚋进行了分析,认为遗传差异的大小与其地理分布有关,但地理分布并不是导致黄毛纺蚋复合种出现的根本原因(Thaijarern *et al.*, 2014)。为清晰厘定双翅目蚋科昆虫物种和种下水平,我们尝试利用线粒体全基因组为标记对黄毛纺蚋种团内的近缘种进行系统发育研究。

线粒体(mitochondrion)是几乎存在于各类真核生物细胞内的重要细胞器,自从 20 世纪 60 年代 Nase 等人发现线粒体 DNA (mitochondrial DNA) 后,人们对该基因的组成结构、基因功能等方面做了深入的研究(Crosskey, 1997; Kumazawa *et al.*, 1998; Phayuhasena *et al.*, 2010; Pramual and Adler, 2014)。线粒体基因是核基因以外的 DNA 遗传信息,能编码自身所需的蛋白质、tRNA 及 rRNA,该基因结构十分简单,分子量非常小,并具有很高的专一性,易于检测,被广泛的应用于种群结构、系统地理以及不同分类水平上的系统发育关系的研究(Boore, 1999; Hebert *et al.*, 2004; Cameron *et al.*, 2007; Li *et al.*, 2010)。

双翅目 Diptera 是昆虫五大目之一。近年来随着线粒体基因组研究方法的成熟,使得双翅目昆虫线粒体基因组研究飞速发展。自 Wolstenholme 和 Clary (1985) 第一个报道了雅克布果蝇 *Drosophila yakuba* mtDNA 全序列以来(Mindell *et al.*, 1999),多种病媒昆虫基因组装配、注释和访问方式被持续开发,截止到 2015 年 5 月 31 日,共有 75 条双翅目昆虫线粒体基因组序列提交到了 IMGd (昆虫线粒体基因组数据库 <http://www.imgd.org/>) 中,占昆虫纲线粒体全基因组序列比例的 29.4%。蚋是重要的病媒昆虫,在叮刺骚扰人畜的同时会传播

多种人兽共患病,但目前蚋科昆虫尚无线粒体全基因组报道,本研究选择该科分布最广泛、最被关注的黄毛纺蚋为代表,对其进行线粒体全基因组测序和分析,为蚋科其他种类线粒体基因组研究打下基础,同时为以线粒体基因组为大尺度分子标记进行蚋科复合种系统发育分析奠定基础。

## 1 材料与方法

### 1.1 供试虫源和标本保存

黄毛纺蚋采自 2014 年 7 月江西省吉安市遂川县阡陌村前小溪里的幼虫,经实验室饲养至成虫,成虫羽化后进行形态鉴定,确定物种后饥饿 1 - 2 d,浸泡于无水酒精中保存。

### 1.2 实验流程

基因组提取→引物的设计→PCR 扩增→PCR 产物电泳→PCR 产物纯化→PCR 产物测序→数据的拼接。

#### 1.2.1 基因组提取

取单头标本的头部,采用 AXYGEN 基因组 DNA 小量试剂盒提取基因组结合酚氯仿法提取样品总 DNA,抽提完毕后 1.0% 琼脂糖凝胶电泳检测其纯度及浓度,并于 -20℃ 保存备用。

#### 1.2.2 引物的设计

参考张乃心等(2013)发表的双翅目昆虫线粒体全基因组通用引物序列和目前已发表的双翅目相近物种引物进行首轮扩增,其他未能得到的序列通过在 GenBank 中下载种源关系相近的昆虫部分线粒体基因序列进行 Clustal X (Bae *et al.*, 2004) 对比寻找相关保守区,利用引物设计软件 Primer Premier (Ver. 5.0) (Clary 和 Wolstenholme, 1987) 遵循引物设计原则(Lessinger *et al.*, 2004; 张乃心等, 2013) 设计相关引物。实验共用引物 25 对(表 1),覆盖了黄毛纺蚋线粒体基因组所有序列。

#### 1.2.3 PCR 扩增

按照引物序列扩增各个片段,所有 PCR 反应均在冰上操作。扩增结束后,采用 1% 的琼脂糖凝胶电泳检测 PCR 产物。检测后利用 AxyPrep DNA 凝胶回收试剂盒纯化 PCR 产物。

(1) PCR 扩增体系 (20 μL):

|                  |        |
|------------------|--------|
| 模板               | 1.0 μL |
| 2 x LAbuffer     | 2.0 μL |
| Primer F (10 mM) | 0.4 μL |

Primer R ( 10 mM) 0.4 μL  
dNTP ( 2.5 mM) 0.5 μL  
LATa 酶 ( 5 U/uL) 0.4 μL  
ddH<sub>2</sub>O 15.3 μL

(2) PCR 扩增条件:

|     |          |
|-----|----------|
| 95℃ | 3 min    |
| 95℃ | 30 s     |
| 50℃ | 45 s     |
| 68℃ | 1 min/kb |
| 68℃ | 10 min   |
| 15℃ | ~        |

共 32 个循环

1. 2. 4 序列测定

获得目的 PCR 扩增产物均通过直接测序法来获得目的片段信息。

(1) 测序反应体系 ( 10 μL) :

|             |      |
|-------------|------|
| 纯化后的 PCR 产物 | 3 μL |
| 水           | 4 μL |
| 测序引物        | 1 μL |
| BigDye      | 2 μL |

(2) 测序反应条件

|     |            |
|-----|------------|
| 96℃ | 2 min      |
| 96℃ | 10 s       |
| 50℃ | 10 s       |
| 60℃ | 3 min 10 s |
| 15℃ | ~          |

共 30 个循环

表 1 黄毛纺蚋线粒体基因组全序列 PCR 扩增引物一览表

Table 1 List of PCR primers for *Simulium aureohirtum* mtDNA complete genome

| 扩增引物名称<br>Primer codes | 扩增引物序列<br>Sequence       | 目的基因<br>Target gene  |
|------------------------|--------------------------|----------------------|
| HY_MT_F9               | GCCTGATRAAAAAGGRTTACCT   | ND2                  |
| HY_MT_R736             | AATCCTCCAAWWGCTCCAAT     |                      |
| HY_MT_365F             | TAATTTACTATCCACTGAAGCTTC | ND2                  |
| HY_MT_1096R            | TGCAGAATAAGTAATCCGCA     |                      |
| HY_MT_405F             | ACCCAAGCATTAGCCTCTGC     | ND2 to COI           |
| HY_MT_2771R            | ATGCAGTATAGGCATCAGGA     |                      |
| HY_MT_1011F            | GTAATTCAATCTATCAGCAA     | COI                  |
| HY_MT_2889R            | TTACAGTTCGATGAGAAATC     |                      |
| HY_MT_1011F            | GTAATTCAATCTATCAGCAA     | COI                  |
| HY_MT_2875R            | TTGGATAAATTACAGTTCGATGAG |                      |
| HY_MT_F2637            | AGCAGGWTTTTRTYCAYTGAT    | COI to COII          |
| HY_MT_R3590            | CTCCTAAAGCWGGKAYTGTT     |                      |
| HY_MT_3054F            | CAACATGATCAAATTTAGGT     | COII                 |
| HY_MT_4749R            | TAATCTACTAGATGAAATGG     |                      |
| HY_MT_3350F            | GGTCACCAATGATACTGAAG     | COII to ATP8 to ATP6 |
| HY_MT_4749R            | TAATCTACTAGATGAAATGG     |                      |
| HY_MT_F4518            | CGACCWGGAACWTTAGCWGT     | ATP6 to COⅢ          |
| HY_MT_R5523            | TAYCCTCCTCATCARTAAAT     |                      |
| HY_MT_4937F            | ACTATTGGACTTCGATGAGG     | COⅢ to ND5           |
| HY_MT_7747R            | TTGGGTTAGTTTCTTATTGC     |                      |
| HY_MT_F7512            | TAGCWGCWGGTAATCAAGA      | ND5                  |
| HY_MT_R8476            | GCTCCTCCWACWTTTRAATT     |                      |

续上表

| 扩增引物名称<br>Primer codes | 扩增引物序列<br>Sequence      | 目的基因<br>Target gene         |
|------------------------|-------------------------|-----------------------------|
| HY_MT_7981F            | TTAATTCTAACTACCTCTCA    | ND5 to ND4                  |
| HY_MT_8860R            | TAATTTTAGCTGGAGTATTG    |                             |
| HY_MT_F8158            | TCATATCAYTRACACCACA     |                             |
| HY_MT_R9156            | GAGGKTATCARCCWGAACG     | ND4                         |
| HY_MT_8979F            | CTGATTGTTCTTTAATAAAAT   | ND4、ND4L                    |
| HY_MT_9557R            | AATTCGAACTCATGGAAATG    |                             |
| HY_MT_F8893            | TTATAGAMCCAGAAACWGG     | ND4 to ND4L                 |
| HY_MT_R9876            | TTGRTTTACAAGACCAATG     |                             |
| HY_MT_F9620            | GCHCCTTCACAWACTCTAAAWGT | ND4 to ND6                  |
| HY_MT_R10813           | CRTAATAWATTCTCGTCCTA    |                             |
| HY_MT_10006F           | GCTATAGGATTCATACTATT    | ND6 toCytB                  |
| HY_MT_11490R           | CTACTGGGCGAGCTCCAATT    |                             |
| HY_MT_F9620            | GCHCCTTCACAWACTCTAAAWGT | ND toCytB                   |
| HY_MT_R10813           | CRTAATAWATTCTCGTCCTA    |                             |
| HY_MT_F11268           | TATYCCWGCTAAYCCWTTAG    | CytB to ND1                 |
| HY_MT_R12253           | GGTTTRCGRGCTGTRGCTCA    |                             |
| HY_MT_11566F           | CACAATGATGAGATTCATTA    | ND1                         |
| HY_MT_12520R           | TTAGGCTATATCCAGATTCCG   |                             |
| HY_MT_F12286           | AACCAGCWAYTATWACWGTAT   | ND1                         |
| HY_MT_R13316           | AGTATTTTTRACTGTGCAAAGG  |                             |
| HY_MT_12669F           | GGCACTAATCTGCCAAAATA    | 1 – rRNA ( 16S)             |
| HY_MT_13333R           | ATTTATTATAAGGTCTGACC    |                             |
| HY_MT_F13152           | ATYTATAGGGTCTTCTCGTCT   | 1 – rRNA ( 16S)             |
| HY_MT_R14217           | AATATGYACACATCGCCCGT    |                             |
| HY_MT_14051F           | GCACAACATCACTTCAATG     | s – rRNA ( 12S) to D – loop |
| HY_MT_270R             | TTCCTCTAATTAGCGTTGTT    |                             |
| HY_MT_14051F           | GCACAACATCACTTCAATG     | D – loop                    |
| HY_MT_203R             | CAGTCTACTACTAGGGGAAG    |                             |

### 1. 2. 5 序列的拼接与注释

得到所有序列片段测序结果后,利用 Staden Package ( Thompson *et al.* , 1997) 软件对所有序列进行拼接。线粒体全基因组各基因的定位及注释参考已测得的双翅目昆虫荒川库蠓 *Culicoides arakae* ( 序列号 NC009809) 和黑腹果蝇 *Drosophila melanogaster* ( 序列号 NC001709) 全线粒体基因组 ( Singh *et al.* , 1998; 张新宇和高燕

宁, 2004) 注释结果进行, 通过 MEGA 5.0 ( Liao *et al.* , 2010) 比对以及 Blast 在线搜索分析相结合来依次注释 13 个编码蛋白基因、22 个 tRNA 基因、2 个 rRNA 基因和 AT 丰富区的位置 ( Matsumoto *et al.* , 2009; 李雪娟等, 2013) 。tRNA 基因二级结构的预测主要是通过 tRNAscan – SE 1.21 ( Thompson *et al.* , 1997) 在线预测并进行人工校正。

## 2 结果与分析

### 2.1 黄毛纺蚋 mtDNA 基因组结构

黄毛纺蚋线粒体基因组全长为 15904 bp, 为一环形的双链 DNA。黄毛纺蚋线粒体基因组全序列已提交 GenBank 数据库, 序列号 KP793690。黄毛纺蚋线粒体全基因组包含 22 个 tRNA 基因、2 个 rRNA (16 S/rrnL 和 12 S/rrnS) 基因、13 个蛋白质编码基因 (PCGs) 共 37 个编码基因, 12S rRNA 基因和 tRNA - Ile 基因之间的 939 bp 的序列片段为 A + T 富含区, 或称为控制区 (Boore, 1999) (图 1)。在这 37 个编码基因中, 有 9 个蛋白质编码基因 (ND3, ND6, ATP6, ATP8, CYTB, COI - III, ND2) 和 14 种 tRNA 基因 (tRNA - Ile, Met, Trp, Leu, Lys, Asp, Gly, Ala, Arg, Asn, Ser, Glu, Thr, Ser) 位于 J 链上, N 链编码剩余的 4 种蛋白质 (ND1, ND4, ND4L, ND5) 和 2 种线粒体核糖体 (12S rRNA 和 16S rRNA) 以及 8 种 tRNA 基因 (tRNA - Gln, Cys, Tyr, Phe, Pro, Leu, His, Val)。黄毛纺蚋线粒体基因组中基因间隔区 IGS (Intergenic spacer region) 共有 20 处, 最长一处位于 tRNA-Met 和 ND2 之间, 间隔序列长度为 45 bp; 基因重叠区共 4 处, 最长一处位于 tRNA-Trp 和 tRNA-Cys 之间最长为 8 bp, 既无间隔又无重叠的区域共有 13 处 (表 2)。值得注意的是, 黄毛纺蚋的 ATP8 /ATP6 重叠了 7 个核苷酸 (ATGNTAA), 符合典型的昆虫 mtDNA 结构 (Moritz *et al.*, 1987; Cannone *et al.*, 2002; Boore, 2006), 但是 ND4/ND4L 基因只重叠了 2 个基因 (AT), 与大多数昆虫的 mtDNA 结构不同。

黄毛纺蚋线粒体基因组与已在 GenBank 上发布的其他双翅目昆虫相比较发现, 它的基因排列方式与假想昆虫祖先的排列方式 trnI-trnQ-trnM 相一致, 符合双翅目昆虫的序列的一般排序方式。

### 2.2 黄毛纺蚋 mtDNA 基因的碱基组成

采用 MEGA 5.0 软件对黄毛纺蚋线粒体基因的碱基组成进行了分析, 分别统计了线粒体基因组的各种碱基含量、AT 偏斜和 GC 偏斜并与 6 种已知的典型的双翅目昆虫 mtDNA 序列进行了对比分析。

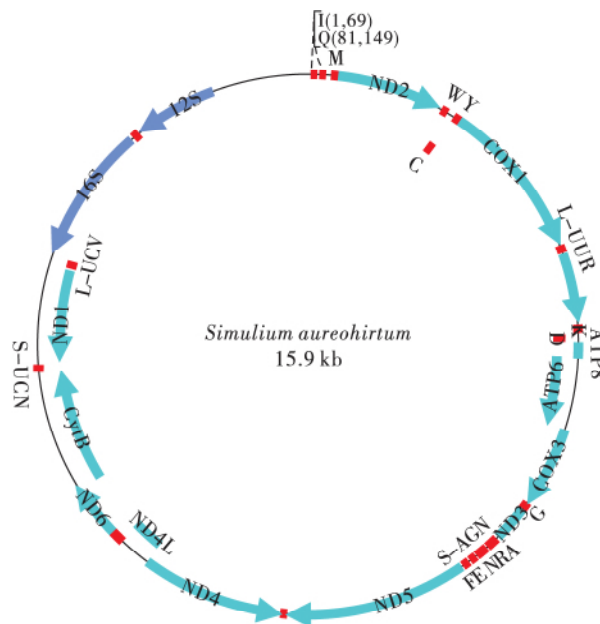


图 1 黄毛纺蚋线粒体基因组结构图

Fig. 1 Gene map of the *Simulium aureohirtum* mitogenome

注: 黄毛纺蚋线粒体基因圈图, 标记了预测基因和颜色。蛋白质编码基因, 蓝色; tRNA, 红色; rRNA 基因, 紫色; 控制区, 灰色单线; 基因名称标记在圈外。各个基因大小起止密码子均在 (表 2) 中列出。Note: The circular map of *S. aureohirtum* mtDNA. The predicted genes are labeled and color-coded; protein coding genes, blue; tRNA genes, red; rRNA genes, purple and control region (CR), grey. The gene names are labeled outside the circle.

在黄毛纺蚋线粒体基因组全序列中, 各碱基的总含量分别是 A: 39.1%, C: 14.7%, G: 10.4%, T: 35.8%, AT 总含量高达 74.88%, 符合典型的昆虫线粒体基因组特征: A + T 偏好性 (Yang *et al.*, 2009; Liao *et al.*, 2010; Tian *et al.*, 2012), 其中 13 个蛋白编码基因的 AT 总含量为 72.5%, 22 个 tRNA 中 AT 含量为 76.5%, rRNA 中为 79.6%, A + T - rich 区中为 88.2% (表 3)。与其他 6 种典型的双翅目昆虫相比, 其 A + T 总含量略高于桔小实蝇, 但是比其他 5 种含量要低, 非编码区的长度为 939 bp, 该区 A + T 含量最高, 达到 88.2%, 比桔小实蝇略高, 比其他 5 种双翅目昆虫相比含量都低 (表 4)。

表 2 黄毛纺蚧线粒体基因组结构

Table 2 Organization of the *Simulium aureohirtum* mitochondrial genome

| 基因<br>Gene                    | 方向<br>Direction | 位置<br>Location | 大小<br>Size | IGS | 反密码子<br>Anticodon | 起始密码子<br>Start code | 终止密码子<br>Stop code | 链<br>Strand |
|-------------------------------|-----------------|----------------|------------|-----|-------------------|---------------------|--------------------|-------------|
| tRNA <sup>Ile</sup> (I)       | F               | 1 – 69         | 69         |     | GAT               |                     |                    | J           |
| tRNA <sup>Gln</sup> (Q)       | R               | 81 – 149       | 69         | 11  | TTG               |                     |                    | N           |
| tRNA <sup>Met</sup> (M)       | F               | 195 – 263      | 69         | 45  | CAT               |                     |                    | J           |
| ND2                           | F               | 264 – 1292     | 1032       | 0   |                   | ATT                 | TAA                | J           |
| tRNA <sup>Trp</sup> (W)       | F               | 1296 – 1367    | 70         | 2   | TCA               |                     |                    | J           |
| tRNA <sup>Cys</sup> (C)       | R               | 1360 – 1422    | 63         | –8  | GCA               |                     |                    | N           |
| tRNA <sup>Tyr</sup> (Y)       | R               | 1428 – 1492    | 65         | 5   | GTA               |                     |                    | N           |
| COI                           | F               | 1506 – 3039    | 1534       | 13  |                   | TTG                 | T                  | J           |
| tRNA <sup>Leu</sup> (L – UUR) | F               | 3040 – 3105    | 66         | 0   | TAA               |                     |                    | J           |
| COII                          | F               | 3110 – 3799    | 690        | 4   |                   | ATG                 | TAA                | J           |
| tRNA <sup>Lys</sup> (K)       | F               | 3802 – 3872    | 71         | 2   | CTT               |                     |                    | J           |
| tRNA <sup>Asp</sup> (D)       | F               | 3888 – 3955    | 68         | 15  | GTC               |                     |                    | J           |
| ATP8                          | F               | 3956 – 4126    | 171        | 0   |                   | ATC                 | TAA                | J           |
| ATP6                          | F               | 4120 – 4797    | 678        | –7  |                   | ATG                 | TAA                | J           |
| COIII                         | F               | 4806 – 5594    | 789        | 8   |                   | ATG                 | TAA                | J           |
| tRNA <sup>Gly</sup> (G)       | F               | 5595 – 5660    | 66         | 0   | TCC               |                     |                    | J           |
| ND3                           | F               | 5661 – 6014    | 354        | 0   |                   | ATT                 | TAA                | J           |
| tRNA <sup>Ala</sup> (A)       | F               | 6018 – 6082    | 65         | 3   | TGC               |                     |                    | J           |
| tRNA <sup>Arg</sup> (R)       | F               | 6083 – 6146    | 64         | 0   |                   |                     |                    | J           |
| tRNA <sup>Asn</sup> (N)       | F               | 6151 – 6215    | 65         | 4   | GTT               |                     |                    | J           |
| tRNA <sup>Ser</sup> (S – AGN) | F               | 6217 – 6281    | 65         | 1   |                   |                     |                    | J           |
| tRNA <sup>Glu</sup> (E)       | F               | 6283 – 6348    | 66         | 1   | TTC               |                     |                    | J           |
| tRNA <sup>Phe</sup> (F)       | R               | 6363 – 6429    | 67         | 14  | GAA               |                     |                    | N           |
| ND5                           | R               | 6430 – 8166    | 1737       | 0   |                   | ATT                 | TAA                | N           |
| tRNA <sup>His</sup> (H)       | R               | 8167 – 8231    | 65         | 0   | GTG               |                     |                    | N           |
| ND4                           | R               | 8233 – 9573    | 1341       | 1   |                   | ATG                 | TAA                | N           |
| ND4L                          | R               | 9572 – 9863    | 292        | –2  |                   | ATG                 | T                  | N           |
| tRNA <sup>Thr</sup> (T)       | F               | 9866 – 9933    | 68         | 2   | TGT               |                     |                    | J           |
| tRNA <sup>Pro</sup> (P)       | R               | 9934 – 9999    | 66         | 0   | TGG               |                     |                    | N           |
| ND6                           | F               | 10002 – 10523  | 522        | 2   |                   | ATT                 | TAA                | J           |
| CytB                          | F               | 10523 – 11659  | 1137       | –1  |                   | ATG                 | TAA                | J           |
| tRNA <sup>Ser</sup> (S – UCN) | F               | 11663 – 11729  | 67         | 3   | TGA               |                     |                    | J           |
| ND1                           | R               | 11748 – 12689  | 942        | 18  |                   | ATA                 | TAG                | N           |
| tRNA <sup>Leu</sup> (L – CUN) | R               | 12700 – 12765  | 66         | 10  | TAG               |                     |                    | N           |
| l – rRNA (16S)                | R               | 12766 – 14101  | 1335       | 0   |                   |                     |                    | N           |
| tRNA <sup>Val</sup> (V)       | R               | 14173 – 14102  | 72         | 0   | TAC               |                     |                    | N           |
| s – rRNA (12S)                | R               | 14174 – 14965  | 792        | 0   |                   |                     |                    | N           |
| D – loop                      |                 |                | 939        | 0   |                   |                     |                    |             |

表 3 黄毛纺蚋线粒体基因组核苷酸组成分析

Table 3 Base compositions of the mitochondrial genome of *Simulium aureohirtum*

|                                                 | A 含量( %) | G 含量( %) | C 含量( %) | T 含量( %) | AT 含量( %) | AT 倾斜     | GC 倾斜     |
|-------------------------------------------------|----------|----------|----------|----------|-----------|-----------|-----------|
|                                                 | A        | G        | C        | T        | A + T     | AT – skew | GC – skew |
| mtDNA 线粒体基因组                                    | 39. 1    | 10. 4    | 14. 7    | 35. 8    | 74. 9     |           |           |
| tRNA genes 转运 RNA 基因                            | 38. 5    | 13. 3    | 10. 2    | 38. 0    | 76. 5     | 0. 006    | 0. 132    |
| rRNA genes 核糖体 RNA                              | 43. 4    | 7. 4     | 13. 0    | 36. 2    | 79. 6     | 0. 091    | – 0. 275  |
| A + T 富含区 A + T – rich zone                     | 46. 1    | 4. 4     | 7. 4     | 42. 1    | 88. 2     | 0. 045    | – 0. 254  |
| 蛋白质编码基因第一位密码子<br>1 <sup>st</sup> codon position | 36. 4    | 10. 1    | 16. 6    | 37. 0    | 73. 4     | – 0. 007  | – 0. 199  |
| 第二位密码子<br>2 <sup>nd</sup> codon position        | 42. 8    | 7. 5     | 12. 6    | 37. 0    | 79. 8     | 0. 072    | – 0. 254  |
| 第三位密码子<br>3 <sup>rd</sup> codon position        | 38. 2    | 13. 7    | 14. 9    | 33. 0    | 71. 2     | 0. 073    | – 0. 042  |

表 4 黄毛纺蚋与 6 种双翅目昆虫线粒体基因组基本特征比较

Table 4 Base characteristics of the mitochondrial genome of *Simulium aureohirtum* and six dipteras

| 科<br>Family             | 种<br>Specis                                 | GenBank<br>登录号 | 基因<br>全长( bp) | AT<br>含量 | AT<br>偏斜 | GC<br>偏斜 | 控制区长<br>( bp) |
|-------------------------|---------------------------------------------|----------------|---------------|----------|----------|----------|---------------|
| 蚋科<br>Simuliidae        | 黄毛纺蚋<br><i>S. aureohirtum</i>               | KP793690       | 15904         | 74. 88   | 0. 045   | – 0. 169 | 939           |
| 实蝇科<br>Tephritidae      | 桔小实蝇<br><i>Bactrocera dorsalis</i>          | NC008748       | 15915         | 73. 58   | 0. 068   | – 0. 228 | 949           |
| 瘿蚊科<br>Calliphoridae    | 黑森瘿蚊<br><i>Mayetiola destructor</i>         | GQ387648       | 14759         | 84. 12   | 0. 061   | – 0. 110 | 604           |
| 蠓科<br>Ceratopogonidae   | 荒川库蠓<br><i>Culicoides arakae</i>            | NC009809       | 18135         | 77. 38   | 0. 009   | – 0. 237 | 1437          |
| 果蝇科<br>Drosophildae     | 黑腹果蝇<br><i>Drosophila melanogaster</i>      | NC001709       | 19517         | 82. 16   | 0. 017   | – 0. 150 | 4601          |
| 细腰大蚊科<br>Ptychopteridae | 细腰蚊<br><i>Ptychoptera</i> sp.               | JN861744       | 15214         | 75. 94   | 0. 038   | – 0. 187 | 370           |
| 家蝇科<br>Muscidae         | 西方角蝇<br><i>Haematobia irritans irritans</i> | NC007102       | 16078         | 79. 07   | 0. 004   | – 0. 125 | 1259          |

2.3 蛋白质编码基因分析

13 种蛋白质编码基因总长为 11219 bp, 排列顺序与典型的双翅目昆虫的 mtDNA 排列顺序一致, 其中 9 个在 J 链上, 其余的 4 个在 N 链上。AT 含

量为 72. 5%, 低于 tRNA ( 76. 5%) 和 rRNA ( 79. 6%) 基因的 AT 含量。蛋白质编码基因的第 2 位和第 3 位密码子偏向于使用 AT ( 0. 072, 0. 073), 但是第 1 位密码子偏向于使用 TA

(0.007) (表3)。除 COI 的起始密码子为 TTG 以外,另外 12 个蛋白质编码基因的起始密码子均为 ATN;多数蛋白质编码基因使用常见的三联体终止密码子为 TAA,但是 COI 基因和 ND4L 基因的终止密码子为 T,ND1 基因的终止密码子为 TAG,符合大多数昆虫密码子的使用规律 (Tzeng *et al.*, 1992; Lee 和 Kocher, 1995; Salvato *et al.*, 2008)。编码的氨基酸,使用最频繁的氨基酸为 Leu、Ile, Phe、Lys。

#### 2.4 黄毛纺蚧线粒体基因组 tRNA 与 rRNA 结构分析

黄毛纺蚧 mtDNA 的 22 个 tRNA 中,与丝氨酸和亮氨酸对应的 tRNA 各有 2 个,其他的氨基酸都只有一个 tRNA 与之对应。22 个 tRNA 中, tRNA-ser<sup>AGN</sup> 无 D 臂,由 8 个核苷酸取代了该位置,其他 21 个 tRNA 都具有规则的三叶草结构,符合昆虫线粒体基因组的普遍特征 (Yukuhiro *et al.*, 2002; Wang *et al.*, 2011)。

黄毛纺蚧线粒体 16S rRNA 和 12S rRNA 基因的长度分别是 1335 bp 和 792 bp, A + T 含量分别为 81.1% 和 77.2%, 分别位于 tRNA-Leu 与 tRNA-Val, tRNA-Val 与 A + T 控制区之间,表现出明显的 AT 偏向性 (表3)。

#### 2.5 非编码区

在黄毛纺蚧的线粒体全基因组中,包括 A + T 富含区在内共有 21 处非编码区。最长的为 A + T 富含区,位于 12S rRNA 和 tRNA-Ile 之间,长度为 939 bp, A + T 含量为 88.2%, 其他 20 个非编码片段长度为 1–45 bp, 共 164 bp, 其中最长的 45 bp 片段位于 tRNA-Met 和 tRNA-Gln 之间。这些分散的非编码序列可能是基因重排过程中产生的假基因遗迹 (Thao 和 Baumann, 2004; Wolstenholme, 1992)。

### 3 结论与讨论

黄毛纺蚧 mtDNA 是双翅目蚧科昆虫中第一个被测得线粒体全基因序列,本文提供了提取该序列所需的全部引物,为蚧科其他种类获得线粒体全基因组提供了参考。黄毛纺蚧线粒体全基因组序列全长 15904 bps, 长度在双翅目昆虫线粒体基因组

长度范围之内,最短为瘿蚊科 Cecidomyiidae 的 *Rhopalomyia pomum* 长 14503 bp (Broughton *et al.*, 2001; Bae *et al.*, 2004)。最长为黑腹果蝇长 19517 bp, 黄毛纺蚧 AT 含量为 74.88%, 略低于双翅目昆虫的平均 AT 含量。基因的排列顺序符合典型的双翅目昆虫基因的排列顺序。已经证实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 (Cannone *et al.*, 2002; Cha *et al.*, 2007; Salvato *et al.*, 2008; Beckenbach and Joy, 2009), 而与其他双翅目昆虫相比,黄毛纺蚧的非编码区长度较平均水平高,所以可认为黄毛纺蚧线粒体基因的进化水平较其他双翅目昆虫要低。

对江西省 2 类黄毛纺蚧的种群生态调查发现,这 2 类种群广泛分布于江西省吉安市、上饶市和赣州市,分布的依据是水源的特征,似与地理分布无关。A 类种群幼虫分布于细小的小溪中,水流速度较缓慢,多只聚集在一块石头或者一片叶子下; B 类种群幼虫分布于水流湍急的河水中,呈零散分布,一块石头下只有一两头,这和泰国的情况类似 (Thaijarern *et al.*, 2014)。本文扩增并测定出种群 A 的线粒体全基因组,设计出一套引物,但是该套引物在种群 B 上却不十分适用,扩增效果十分差,说明两类种群的遗传差异较大,推测我国黄毛纺蚧是一个种团,种团内有多个隐含种,对每个隐含种利用线粒体全基因组为分子标记进行系统发育分析,有利于揭示蚧科昆虫复合种的演变过程。

#### 参考文献 (References)

- An JY, Yan BS. Simuliidae insect and disease relationship [J]. *J. Med. Pest. Cont.*, 1997, 2: 51–54. 安继尧, 阎丙申. 蚧科昆虫与疾病的关系 [J]. *医学动物防制*, 1997, 2: 51–54]
- Bae JS, Kim I, Sohn HD, *et al.* The mitochondrial genome of the firefly, *Pyrocoelia rufa*: Complete DNA sequence, genome organization, and phylogenetic analysis with other insects [J]. *Mol. Phylogenet Evol.*, 2004, 32 (3): 978–985.
- Beckenbach AT, Joy JB. Evolution of the mitochondrial genomes of *Gall* midges (Diptera: Cecidomyiidae): Rearrangement and severe truncation of tRNA genes [J]. *Genome Biol. Evol.*, 2009, 1 (1): 278–287.
- Boore JL. Animal mitochondrial genomes [J]. *Nucleic Acids Res.*, 1999, 27 (8): 1767–1780.
- Boore JL. The complete sequence of the mitochondrial genome of



- Nautilus macromphalus* (Mollusca: Cephalopoda) [J]. *BMC Genomics*, 2006, 7 (3): 303–309.
- Broughton RE, Milam JE, Roe BA. The complete sequence of the zebra fish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA [J]. *Genome Res.*, 2001, 11 (11): 1958–1976.
- Cai R. Study of Classification on Black Flies in Some Areas in China [J]. Hefei: Anhui University of Science & Technology, 2005.
- 蔡茹. 蚋科昆虫分类研究 [J]. 合肥: 安徽理工大学硕士学位论文, 2005]
- Cameron SL, Lambkin CL, Barker SC, et al. A mitochondrial genome phylogeny of Diptera: Whole genome sequence data accurately resolve relationships over broad timescales with high precision [J]. *Syst. Entomol.*, 2007, 32 (1): 40–59.
- Cannone JJ, Subramanian S, Schnare MN, et al. The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs [J]. *BMC Bioinformatics*, 2002, 3 (1): 189–198.
- Cha SY, Yoon HJ, Lee EM, et al. The complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, *Bombus ignites* (Hymenoptera: Apidae) [J]. *Gene*, 2007, 392 (1–2): 206–220.
- Chen HB, Kang Z, Zhang CL. A new species of the genus *Simulium* from Jiangxi Province (Diptera, Simuliidae) [J]. *Acta Zoot. Sini.*, 2007, 32 (3): 579–580.
- Chen HB. Faunistic distribution and geographical divisions of black flies in China (Diptera: Simuliidae) [J]. *Zool. Syst.*, 2002, 27 (3): 624–630. 陈汉斌. 中国蚋类区系分布和地理区划 (双翅目: 蚋科) [J]. *动物分类学报*, 2002, 27 (3): 624–630]
- Clary DO, Wolstenholme DR. The mitochondrial DNA molecule of *Drosophila yakuba*: Nucleotide sequence, gene organization, and genetic code [J]. *J. Mol. Evol.*, 1987, 22 (3): 252–271.
- Crosske RW. An Annotated Checklist of the World Black Flies (Diptera: Simuliidae) [M]. London: Nat. Hist. Mus., 1997: 145.
- Hebert PD, Penton EH, Burns JM, et al. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator* [J]. *Proc. Natl. Acad. Sci. USA*, 2004, 101 (41): 14812–14817.
- Kumazawa Y, Ota H, Nishida M, et al. The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions [J]. *Genetics*, 1998, 150 (1): 13–329.
- Lee WJ, Kocher TD. Complete Sequence of a sea lamprey (*Petromyzon marinus*) mitochondrial genome: Early establishment of the vertebrate genome organization [J]. *Genetics*, 1995, 139 (2): 873–887.
- Lessinger AC, Junqueira ACM, Conte FF, et al. Analysis of a conserved duplicated tRNA gene in the mitochondrial genome of blowflies [J]. *Gene*, 2004, 339 (1): 1–6.
- Li XJ, Yang J, Wang JH, et al. Methods and software tools for mitochondrial genome assembly and annotation [J]. *Chin. J. Appl. Entomol.*, 2013, 50 (1): 298–304. 李雪娟, 杨婧, 王俊红, 等. 线粒体基因组数据的分析方法和软件 [J]. *应用昆虫学报*, 2013, 50 (1): 298–304]
- Li YP, Song W, Shi SL, et al. Mitochondrial genome nucleotide substitution pattern between domesticated silkworm, *Bombyx mori*, and its wild ancestors, Chinese *Bombyx mandarina* and Japanese *Bombyx mandarina* [J]. *Genet. Mol. Biol.*, 2010, 33 (1): 186–189.
- Liao F, Wang L, Wu S, et al. The complete mitochondrial genome of the fall webworm, *Hyphantria cunea* (Lepidoptera: Arctiidae) [J]. *Int. J. Biol. Sci.*, 2010, 6 (1): 172–186.
- Matsumoto Y, Yanase T, Tsuda T, et al. Species – specific mitochondrial gene rearrangements in biting midges and vector species identification [J]. *Med. Vet. Entomol.*, 2009, 23 (1): 47–55.
- Mindell DP, Sorenson MD, Dimcheff DEM, et al. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes [J]. *Syst. Biol.*, 1999, 48 (1): 138–152.
- Moritz C, Dowling TE, Brown WM. Evolution of animal mitochondrial DNA: Relevance for population biology and systematics [J]. *Annu. Rev. Ecol. Syst.*, 1987, 18 (1): 269–292.
- Phayuhase S, Colgan DJ, Kuvangkadilok C. Phylogenetic relationships among the black fly species (Diptera: Simuliidae) of Thailand based on multiple gene sequences [J]. *Genetica*, 2010, 138 (6): 633–648.
- Pramual P, Adler PH. DNA barcoding of tropical black flies (Diptera: Simuliidae) of Thailand [J]. *Mol. Ecol. Resour.*, 2014, 14 (2): 262–271.
- Salvato P, Simonato M, Battisti A, et al. The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae) [J]. *BMC Genomics*, 2008, 9 (12): 1908–1914.
- Singh VK, Mangalam AK, Dwivedi S, et al. Primer premier: Program for design of degenerate primers from a protein sequence [J]. *Biotechniques*, 1998, 2 (24): 318–319.
- Thaijarern J, Sopaladawan PN, Wongpakam K, et al. Phylogeography of the black fly *Simulium aureohirtum* (Diptera: Simuliidae) in Thailand [J]. *Genome*, 2014, 57: 411–417.
- Thanwisai A, Kuvangkadilok C, Baimai V. Molecular phylogeny of black flies (Diptera: Simuliidae) from Thailand, using ITS2 rDNA [J]. *Genetica*, 2006, 128 (1–3): 177–204.
- Thao ML, Baumann P. Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts [J]. *Appl. Environ.*

- Microb. (Aem.)*, 2004, 70 (6): 3401–3406.
- Thompson JD, Gibson TJ, Plewniak F, *et al.* The Clustal X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools [J]. *Nucleic Acids Res.*, 1997, 24 (2): 173–216.
- Tian LL, Sun XY, Chen M, *et al.* Complete mitochondrial genome of the five-dot sergeant *Parathyma sulphita* (Nymphalidae: Limenitidinae) and its phylogenetic implications [J]. *Zool. Res.*, 2012, 33 (2): 133–143.
- Tzeng CS, Hui CF, Shen SC. The complete nucleotide sequence of the *Crossostoma lacustre* mitochondrial genome: Conservation and variations among vertebrates [J]. *Nucleic Acids Res.*, 1992, 20 (18): 4853–4958.
- Wang S, Lei Z, Wang H, *et al.* The complete mitochondrial genome of the leafminer *Liriomyza trifolii* (Diptera: Agromyzidae) [J]. *Mol. Biol. Rep.*, 2011, 38 (2): 687–692.
- Wolstenholme DR. Animal mitochondrial DNA: Structure and evolution [J]. *Int. Rev. Cytol.*, 1992, 141 (1): 173–216.
- Yang L, Wei ZJ, Hong GY, *et al.* The complete nucleotide sequence of the mitochondrial genome of *Phthonandria atrilineata* (Lepidoptera: Geometridae) [J]. *Mol. Biol. Rep.*, 2009, 36 (6): 1441–1449.
- 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 silkmoth, *Bombyx mandarina*, and its close relative, the domesticated silkmoth, *Bombyx mori* [J]. *Mol. Biol. Evol.*, 2002, 19 (8): 1385–1389.
- Zhang NX, Zhang YJ, Yu G, *et al.* Structure characteristics of the mitochondrial genomes of Diptera and design and application of universal primers for their sequencing [J]. *Acta Entomol. Sin.*, 2013, 56 (4): 398–407. 张乃心, 张玉娟, 余果, 等. 双翅目昆虫线粒体基因组结构特点及其测序通用引物的设计 and 应用 [J]. *昆虫学报*, 2013, 56 (4): 398–407]
- Zhang XY, Gao YN. To design PCR primers with Oligo 6 and Primer Premier 5 [J]. *Bioinformatics*, 2004, 2 (4): 15–18. 张新宇, 高燕宁. PCR 引物设计及软件使用技巧 [J]. *生物信息学*, 2004, 2 (4): 15–18]