

# 中华按蚊 *CYP4G17* 基因的鉴定及生物信息学分析\*

闫正文, 张玉娟, 周 勇, 陈 斌

(重庆师范大学 昆虫与分子生物学研究所 重庆市动物生物学重点实验室, 重庆 401331)

**摘要:** 基于中华按蚊(*Anopheles sinensis*)转录组, 通过同源性搜索鉴定出一条中华按蚊 *CYP4* 家族序列, 生物信息学分析将该基因命名为 *AsCYP4G17* (GenBank 登录号: KP004246), 该序列全长 1 962 bp, 其中编码区 1 671 bp, 编码 556 个氨基酸。同源性分析表明该基因与冈比亚按蚊(*Anopheles gambiae*)*CYP4G17* 氨基酸序列相似性最高(一致率为 89%, 相似率为 94%)。该基因编码的蛋白质相对分子质量为 63.48 kD, 等电点为 7.70。该蛋白第 20~39 位氨基酸为疏水区, 蛋白质亚细胞定位显示该蛋白质位于细胞质中。基因结构分析显示, 该基因含有两个相位 2 型内含子。本研究为进一步揭示中华按蚊 *CYP4G17* 基因的功能奠定了基础, 对阐明昆虫杀虫剂抗性机理具有一定的科学意义。

**关键词:** 中华按蚊; *CYP4G17*; 鉴定; 生物信息学分析

**中图分类号:** Q966

**文献标志码:** A

**文章编号:** 1672-6693(2015)03-0000-00

疟疾是一种以按蚊为传播媒介的虫媒传染病, 是最重要的媒介传播疾病, 是全球主要的健康问题<sup>[1]</sup>。仅 2012 年, 全球就有 2.07 亿人感染疟疾, 造成约 62.7 万人死亡。蚊媒的控制是防止疟疾传播的有效方法。目前蚊媒控制主要通过室内滞留喷洒和杀虫剂处理过的蚊帐<sup>[2]</sup>。然而随着杀虫剂的广泛使用, 许多蚊媒已经对杀虫剂产生了抗药性。这对蚊媒传播疾病的控制造成了重大的威胁, 尤其是对疟疾的控制。因此, 蚊媒对杀虫剂的抗性机制已经成为一个迫切需要研究的课题。

中华按蚊(*Anopheles sinensis*)是我国和东南亚地区主要的疟疾传播媒介<sup>[3]</sup>, 研究表明其已对多种杀虫剂产生了一定程度的抗药性。而且在中国和韩国, 中华按蚊已对拟除虫菊酯表现出高水平的耐药性<sup>[4]</sup>。昆虫对杀虫剂的抗性机制是一个非常复杂的问题, 主要包括行为抗性和生理抗性, 如蚊虫倾向于回避杀虫剂, 表皮穿透性的降低, 靶标位点敏感度降低和解毒酶活性增强。而在蚊虫中抗性主要与靶标位点敏感度降低和解毒酶活性增强有关<sup>[5]</sup>。昆虫体内主要包括 3 类解毒酶: 细胞色素 P450(CYPs)、谷胱甘肽-S-转移酶(GSTs)和羧酸酯酶(CCE)<sup>[6]</sup>。其中, 参与杀虫剂代谢的细胞色素 P450 是昆虫产生抗药性的主要机制, 该机制也造成大多数媒介昆虫对杀虫剂产生高水平抗性和交互抗性<sup>[3]</sup>。就目前广泛使用的菊酯类杀虫剂而言, 昆虫对该类杀虫剂产生抗药性的主要原因为细胞色素 P450 表达量上升而介导的杀虫剂代谢作用<sup>[5, 7]</sup>。通常认为, *CYP6* 家族成员与昆虫抗药性密切相关, 但近年来研究表明 *CYP4* 家族成员也可能与昆虫抗药性相关<sup>[8]</sup>。已有研究表明, 阿拉伯按蚊(*An. arabiensis*)中 *CYP4G16* 和 *CYP4H24* 与溴氰菊酯抗性相关, 致倦库蚊(*Culex quinquefasciatus*)中 *CYP4H34* 与氯菊酯抗性相关, 尖音库蚊(*Culex pipiens*)中 *CYP4H21* 与溴氰菊酯抗性相关<sup>[5]</sup>。目前只有少量与中华按蚊抗性相关的基因被报道。

本研究基于中华按蚊转录组数据, 通过同源性搜索得到一条 *CYP4* 家族序列进行了深入研究, 分析了其序列特征、蛋白质结构, 构建了系统发育树(neighbor-joining method, NJ), 并对其基因结构进行预测。本研究为进一步研究中华按蚊 *CYP4G17* 基因的功能奠定了理论基础, 对阐明昆虫杀虫剂抗性机理具有潜在的科学意义。

## 1 材料与方法

### 1.1 数据来源

中华按蚊转录组数据(SRA 登录号: SRA073189)来自于重庆师范大学昆虫与分子生物学研究所, 该物种的基因组数据和其同源物种的 P450 序列分别自 NCBI(<http://www.ncbi.nlm.nih.gov/>)和 VectorBase(<https://www.vectorbase.org/>)中下载。

### 1.2 序列鉴定

以冈比亚按蚊的 *CYP4G17* 氨基酸序列作为 query 序列, TBLASTN(E-value  $\leq 1e-5$ )搜索中华按蚊转录组数据库。搜索得到的候选基因进行手工校对后, NCBI 在线 BLASTP 检索 nr 库, 进一步完成序列的验证。

### 1.3 序列分析

鉴定中华按蚊 *CYP4G17* cDNA 序列的开放阅读框并翻译成氨基酸序列使用软件 DNAMAN7.0; 使用软件 BioEdit 统计 cDNA 序列的碱基组成并使用其中的 BLAST 工具进行序列同源性检索; 利用 CodonW 软件分析密码子使用率; 使用 ClustalW([http://www.phylogeny.fr/version2/cgi/one\\_task.cgi?task\\_type=clustalw](http://www.phylogeny.fr/version2/cgi/one_task.cgi?task_type=clustalw))对中华按蚊

\* 收稿日期: 修回日期:

资助项目: 美国国立卫生研究院 NIH 项目(R01AI095184); 国家自然科学基金(31071968, 31372265); 重庆市科技攻关重点项目(CSTC2012GG-YYJSB80002); “两江学者”计划专项经费; 重庆市科委项目(CSTC2011BB1002)

作者简介: 闫正文, 男, 研究方向为昆虫分子生物学; 通讯作者: 陈斌, 教授, E-mail: bin.chen@cqnu.edu.cn

CYP4G17 和其他昆虫同源 P450 序列进行多重序列比对, 并用 Color Align Conservation 软件 ([http://www.bioinformatics.org/sms2/color\\_align\\_cons.html](http://www.bioinformatics.org/sms2/color_align_cons.html)) 进行染色; 利用 ProtParam 软件 (<http://web.expasy.org/protparam/>) 预测蛋白的基本理化性质; 蛋白质疏水性分析使用 ProtScale 软件 (<http://www.expasy.org/cgi-bin/protscale.pl>); 蛋白质跨膜区分析使用 TMHMM 软件 (<http://www.cbs.dtu.dk/services/TMHMM/>); 蛋白质亚细胞定位使用 TargetP (<http://www.cbs.dtu.dk/services/TargetP/>) 软件<sup>[6]</sup>; 信号肽预测使用 SignalP4.1 软件 (<http://www.cbs.dtu.dk/services/SignalP/>); 蛋白的二级结构预测使用软件 SSPro4.0 (<http://scratch.proteomics.ics.uci.edu/>); 蛋白质三级结构预测需先通过 PSI-BLAST 搜索 PDB (Protein Data Bank) 数据库, 找到与中华按蚊 CYP4G17 相似性高的蛋白后, 再采用 SWISS-MODEL (<http://swissmodel.expasy.org/>) 进行同源模建及 3D 结构预测<sup>[6]</sup>。

#### 1.4 系统发育分析

根据同源性分析结果选择有代表性的昆虫 P450 序列, 以黑腹果蝇 CYP4G1 序列为外群, 采用邻接法 (neighbor-joining method, NJ) 分析中华按蚊 CYP4G17 序列与其它代表性的昆虫 P450 序列的系统发育关系, 使用 MEGA5.0 软件<sup>[9]</sup> 构建系统发育树, 1000 次重复计算自展分析值<sup>[6]</sup>。

#### 1.5 基因结构分析

对搜索得到的 4 种蚊虫的 *CYP4G* 基因, 首先使用 FGGENESH<sup>[10]</sup> (<http://linux1.softberry.com/berry.phtml>) 进行基因结构预测, 然后使用本地 GeneWise 软件分析 4 个基因的内含子类型。

## 2 结果

### 2.1 cDNA 序列分析

通过同源性搜索得到 ID 号为 Unigene14353\_5 的中华按蚊转录组基因序列, 其与冈比亚按蚊 *CYP4G17* 基因的一致率为 88.06%, 同源性最高, 并且该 cDNA 推断的氨基酸序列 (图 2.1) 中含有昆虫 P450 蛋白的 5 个保守结构域 (WxxxR、GxE/DTT/S、ExLR、PxxFxFE/DRF、PFxxGxRxCxG/A)<sup>[11]</sup>。该序列共包含 1962 bp, 通过生物信息学分析发现, 其中编码区 1671 bp, 具备全长序列的基本保守区, 具有起始子 (ATG) 和终止子 (TGA), 编码 556 个氨基酸, 且在该氨基酸序列的 354-366 位发现 CYP4 家族的特征序列 EVDTFMFEGHDTT<sup>[12]</sup>, 故将该 cDNA 命名为 *AsCYP4G17* (GenBank 登录号: KP004246)。

```

-165                                     GTACCATCGGTGGTGCATCCG
-144 AGTGTTACATATCGCTCCTTATCCAGCGCTAGTGCAGTGTGAAGGTGTTTCAGTTCAGTGTCCAGTTTGG
-72 GAGTGTACCTTAGGGTGTGAATCCTGGAAGTAAACCGGTACGGAGAGTAACCAGCATCAATCAAACCAGA
1   ATCAGTTTCGAAACCATCCCGGACGACTTCCGGTCGAGGGTGGTGCAGTGTGGGTACTGCCAGTGCAGTG
1   M S F E T I P E R L P V E G G A S W V L P V T V
73 GCGACCGTGTGGTGTAGCAGCAGTCACTGTTCCACCTATGGATGCAGACGCGCCGCTACGTGCAGCTG
25  A T V L V L A A V T L F H L W M Q T R R Y V Q L
145 GGCAACCTAATACCGGGCCCGTGCCTATCCGCTCGTGGGCAACGCCAATGCTGCTCGGCAAGTCACAC
49  G N L I P G P V A Y P L V G N A N M L L G K S H
217 AATCAAATCATGGAGAAGCGATGGAGCTGAGCTACATCTACGGACCGGTGGCACGCGGCTGGATTGGGTAC
73  N Q I M E K A M E L S Y I Y G P V A R G W I G Y
289 CACQTGGTGGTGTCTCACCAGCCGTCGACGTCGAGATCATCTGAACAGCTACGTGCACCTGGAGAAG
97  H L V V F L T E P S D V E I I L N S Y V H L E K
361 TCGAGCGAGTACCGTTTCTCAAGCCCTGGCTCGGTGATGGCTGCTCATCAGCTCTGGTACAAGTGGCGC
121 S S E Y R F F K P W L G D G L L I S S G D K W R
433 TCGCACCGGAAGCTGATCGCGCCCGCTTCCACCAGAACGTGCTGAAGACGTTTCATCGACGTGTTCAACGAC
145 S H R K L I A P A F H Q N V L K T F I D V F N D
505 AACAGTCTGGCGGTGAAGCGGATGGCTCGGAAGTTGGCCACGAATTCGACTGCCACGATTATATGAGC
169 N S L A V V K R M A R E V G H E F D C H D Y M S
577 GAGGTGACGGTCGACATTCTACTGAAACGGCAATGGGCTCCAACCGGACGGGCGAGAACAAGGAAGTTTT
193 E V T V D I L L E T A M G S N R T G E N K E G F
649 GAGTACCCATGGCGTTATGAAGATGTGCGATATTCTACACAAGCGTCAGCTGAAGATTCACCTCCGCTG
217 E Y A M A V M K M C D I L H K R Q L K I H L R L
721 GACCCGCTGTTCAATCTGACTGGTGTGAAGAAGGAACAGGAGCGCTACTGAACATCATCCACAGCCTACC
241 D P L F N L T G V K K E Q E R L L N I I H S L T
793 CGCAAGGTGGTGAAGCAGAAGAAGCAATCTTTGAGCGCAACCGGCGGAGGAAAGCTACCGTCCGCTCG
265 R K V V K Q K K A I F E R N A A E G K L P S P S
865 CTCTCGGACATCATCGGAAGGAGGAGAAGCCGGGCGAGAGTCTGCTTTCATCTCGCAGGGCTCGATGTTG
289 L S D I I G K E E K P G E S P A F I S Q G S M L
937 CGCGACGATCTGGACGACAACGACGAGAACGACATCGGCGAGAAGCGGCGGCTCGCCTTCTCGACCTCATG
313 R D D L D D N D E N D I G E K R R L A F L D L M
1009 ATCGAGTCGGCGAAGGGTGGTGCGAATATCAGCGACGAAGATCAAGGAGGAGGTGGACACCATCATGTT
337 I E S A K G G A N I S D E E I K E E V D D T I M F
1081 GAGGGTCACGATACCACCGCGCGCTCCAGCTTCGTGCTCTGCTGCTCGGCATTACAGGACGTCGACG
361 E G H D T T A A G S S F V L C L L G I H Q D V Q
1153 GAGCGTGTGTACGCCGAGATGCGGCAGATCTTCGGTACTCCAACCGCAAGGCCACGTTCCGCCGACACGCTG
385 E R V Y A E M R Q I F G D S N R K A T F A D T L
1225 GAGATGAAGTACCTAGAGCGGGTATCTTCGAGACGCTCCGGATGTTCCACCGGTACCGATGATCGCACGC
409 E M K Y L E R V I F E T L R M F P P V P M I A R
1297 AAAGTCAACGAAGATGTGCAGCTCGCATCGAAGAAGTACACCATCCCGGACGGGACCGCTCGTCATCGGG
433 K V N E D V Q L A S K K Y T I P A G T T V I G
1369 ACGTACAAGATCCACCGTCGGGAGGACCTTTACCCGACGCGGAGAAGTTCAACCCGGAACCTTCTGCC
457 T Y K I H R R E D L Y P Q P E K F N P D N F L P
1441 GAGCGACCCAGAATCGTCACTACTATAGCTACATTCGGTTCAGCGCGGGTCTCGCAGTGCCTCGGGAGG
481 E R T Q N R H Y Y S Y I P F S A G P R S C V G R
1513 AAGTACGCAATGCTCAAGCTGAAGTTCTGTTGCCACCTTCTCCGGAATTACCGCGTCCGGTCCAACATT
505 K Y A M L K L K V L L S T I L R N Y R V R S N I
1585 ACCGAGAAAGACTTCAAGCTGACGGCCGACATCATTTGAAGCGACCGATGGGTTCCGTATCCAGCTGGAG
529 T E K D F K L Q A D I L K R T D G F R I Q L E
1657 CCACGGATGGCATGAGACGCAGCCAGCCACCACCCACTATGCGACGGGACACTCAGTCCCCTATCTGCC
553 P R M A *
1729 CGAACCGCGCGCTAGCCATCATGAGATAACAGAGAAACAACAGCCAAAATAGTAATAACCGCTTCCGT

```

注: 图中方框表示翻译起始密码子和终止密码子, 左边数字表示核苷酸和氨基酸编号。阴影部分表示保守结构域, 下划线表示 CYP4 家族的特征序列。

Note: The rectangle denoted start and stop codon, The positions of nucleotides and amino acids on the sequences were marked on the left. Residues in conserved P450 motifs are shaded, and characteristics sequence of CYP4 family is underlined.

图 2.1 中华按蚊 *CYP4G17* cDNA 序列和推断的蛋白质序列

Fig.2.1 The sequence of *AsCYP4G17* cDNA and deduced protein

中华按蚊 *CYP4G17* cDNA 序列的 GC 和 AT 含量分别为 58.65% 和 41.35%, AT 和 GC 偏斜分别为 0.10 和 0.02, 说明该基因碱基组成偏好 G 和 A。

RSCU 值反映密码子在所编码的氨基酸的同义密码子中出现的频率<sup>[3]</sup>, 该值越大, 表明相应的密码子出现的频率越大, 反之则越小。结果表明 UGA、CUG、CCG 等密码子出现的频率较高, 而 CUA、AGC、UCU 等密码子出现的频率较小 (表 2.1)。

表 2.1 中华按蚊 *CYP4G17* 核酸序列相对同义密码子统计表  
Tab. 2.1 Relative synonymous codon usage (RSCU) of *AsCYP4G17*

| Codon | RSCU | Codon | RSCU | Codon | RSCU | Codon | RSCU |
|-------|------|-------|------|-------|------|-------|------|
|-------|------|-------|------|-------|------|-------|------|

|     |      |     |      |     |      |     |      |
|-----|------|-----|------|-----|------|-----|------|
| UGA | 3    | UCG | 1.69 | CUA | 0.74 | UCA | 0.19 |
| CUG | 2.84 | ACG | 1.69 | AUU | 0.65 | CGA | 0.17 |
| CCG | 2.62 | UAC | 1.68 | CCC | 0.62 | AGG | 0.17 |
| CGC | 2.5  | GAG | 1.62 | GGG | 0.61 | UUU | 0.16 |
| GUG | 2.46 | AAC | 1.57 | GAU | 0.47 | ACU | 0.15 |
| AUC | 2.27 | GAC | 1.53 | CCA | 0.46 | ACA | 0.15 |
| CGG | 2.17 | GCC | 1.41 | AAU | 0.43 | CAA | 0.13 |
| AGC | 2.06 | GGU | 1.33 | GAA | 0.38 | GCU | 0.12 |
| ACC | 2    | GCG | 1.29 | UUG | 0.32 | UUA | 0.11 |
| CAC | 2    | GCA | 1.18 | UAU | 0.32 | AAA | 0.11 |
| UGC | 2    | UCC | 1.13 | GUU | 0.31 | AUA | 0.08 |
| AAG | 1.89 | GUC | 1.03 | CCU | 0.31 | UAA | 0    |
| CAG | 1.88 | AUG | 1    | GGA | 0.24 | UAG | 0    |
| UUC | 1.84 | UGG | 1    | CUU | 0.21 | CAU | 0    |
| GGC | 1.82 | CGU | 1    | GUA | 0.21 | UGU | 0    |
| CUC | 1.79 | AGU | 0.75 | UCU | 0.19 | AGA | 0    |

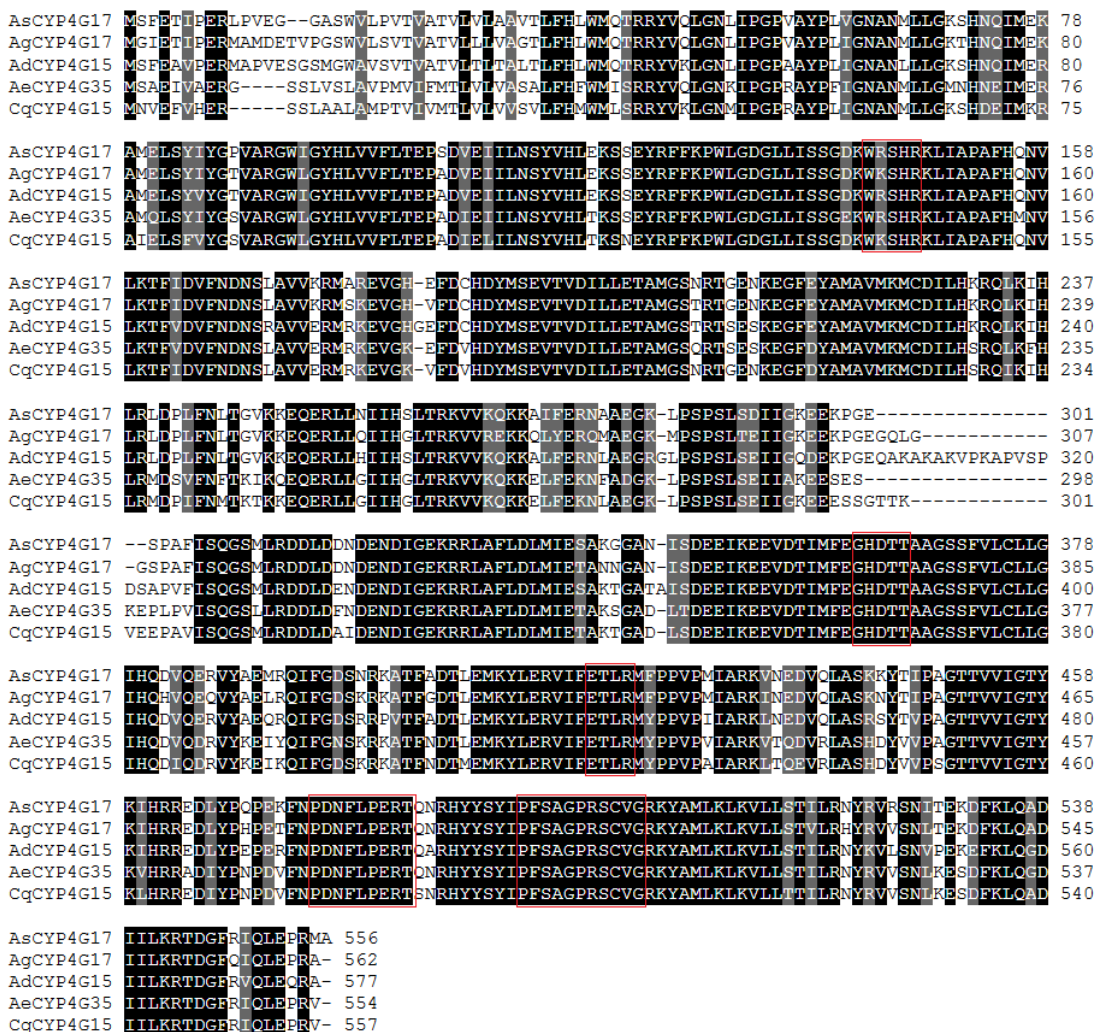
## 2.2 同源性分析

将中华按蚊 *CYP4G17* 序列作为询问序列, 在线 BLASTP 搜索得到 90 多个同源序列, 与冈比亚按蚊 *CYP4G17* 氨基酸序列相似性最高, 一致率为 89%; 其次是达林按蚊、埃及伊蚊, 其一致率分别为 86%、80% (表 2.2)。

表 2.2 中华按蚊 *CYP4G17* 与其它昆虫 P450 基因氨基酸序列的一致率和相似率

| Tab. 2.2 The identity and similarity between AsCYP4G17 and its homologs P450 from other insect species |   |                       |              |                |
|--|---|-----------------------|--------------|----------------|
| 基因名  | 物种  | GenBank 登录号           | 一致率 (%)      | 相似率 (%)        |
| Gene name  | Species                                   | GenBank Accession No. | Identity (%) | similarity (%) |
| <i>CYP4G17</i>   | 冈比亚按蚊<br><i>An. gambiae</i>               | XP_555875.4           | 89%          | 94%            |
| <i>CYP4G15</i>   | 达林按蚊<br><i>An. darlingi</i>               | ETN67528.1            | 86%          | 91%            |
| <i>CYP4G35</i>   | 埃及伊蚊<br><i>Aedes aegypti</i>              | XP_001659149.1        | 80%          | 88%            |
| <i>CYP4G15</i>   | 致倦库蚊<br><i>Culex quinquefasciatus</i>     | XP_001869039.1        | 79%          | 89%            |
| <i>CYP4G33</i>   | 摇蚊<br><i>Chironomus tentans</i>           | AAW78325.1            | 64%          | 80%            |
| <i>CYP4G48</i>   | 六星灯蛾<br><i>Zygaena filipendulae</i>       | ACZ97414.1            | 61%          | 77%            |
| <i>CYP4G49</i>   | 烟草天蛾<br><i>Manduca sexta</i>              | ADE05583.1            | 62%          | 76%            |
| <i>CYP4G44</i>   | 丽蝇蛹集金小蜂<br><i>Nasonia vitripennis</i>     | NP_001165993.1        | 60%          | 76%            |
| <i>CYP4G24</i>   | 家蚕<br><i>Bombyx mori</i>                  | BAM73879.1            | 61%          | 75%            |
| <i>CYP4G23</i>   | 家蚕<br><i>Bombyx mori</i>                  | BAM73905.1            | 61%          | 76%            |
| <i>CYP4G29</i>   | 马铃薯叶甲<br><i>Leptinotarsa decemlineata</i> | AAZ94273.1            | 59%          | 77%            |
| <i>CYP4G19</i>   | 德国小蠊<br><i>Blattella germanica</i>        | AAO20251.1            | 58%          | 74%            |

利用推导的中华按蚊 *CYP4G17* 氨基酸序列与 4 种序列一致性较高的蚊虫 *CYP4* 家族的序列 (表 2.2) 进行 ClustalW 多重序列比对并建立同源序列比对图 (图 2.2)。通过同源序列比对图可知 5 条序列都具有典型昆虫 P450 蛋白的五个保守结构域 (WxxxR、GxE/DTT/S、ExLR、PxxFxPE/DRF、PFxxGxRxCxG/A), 且 Meander 区 (PxxYxPE/DRF) 发生了氨基酸的替换, 该区域第 9 位的苯丙氨酸被苏氨酸替代。



注: 图中用红框标出的为 P450 的 5 个保守结构域, 黑色表示氨基酸序列保守性为 100%。As、Ag、Ad、Ae 和 Cq 分别表示中华按蚊、冈比亚按蚊、达林按蚊、埃及伊蚊和致倦库蚊。

Note: Red rectangles denoted five conserved domains, and amino acids with 100% identity were marked with black shade. As, Ag, Ad, Ae and Cq denoted *An. sinensis*, *An. gambiae*, *An. darling*, *Aedes aegypti* and *Culex quinquefasciatus* respectively.

图 2.2 中华按蚊 *CYP4G17* 序列与 4 种蚊虫 *CYP4G* 序列比对

Fig. 2.2 Alignment of AsCYP4G17 with other 4 homologous mosquito CYP4G

### 2.3 编码氨基酸序列分析

2.3.1 基本理化性质分析 通过分析得到 *CYP4G17* 蛋白的分子式为  $C_{2850}H_{4517}N_{775}O_{817}S_{24}$ , 分子量约为 63.48 kDa, 理论等电点为 7.70; 在 *CYP4G17* 氨基酸组成中, 亮氨酸(Leu)所占比例最高, 达到 10.30%, 而半胱氨酸(Cys)所占比例最低, 为 0.70% (表 2.3); 该蛋白的脂肪指数为 92.39, 负电荷的残基(Asp+Glu)的总电荷为 72, 相应的正电荷残基(Arg+Lys)总电荷为 73; 总平均亲水性是 -0.260, 故推测其为亲水性蛋白; 预测蛋白的不稳定参数为 46.79, 故认为该蛋白为不稳定蛋白。

表 2.3 中华按蚊 *CYP4G17* 蛋白的氨基酸组成序列

Tab. 2.3 Composition of amino acids of AsCYP4G17

| 氨基酸 | 数目 | 比例     | 氨基酸 | 数目 | 比例    | 氨基酸 | 数目 | 比例    | 氨基酸 | 数目 | 比例    |
|-----|----|--------|-----|----|-------|-----|----|-------|-----|----|-------|
| Leu | 57 | 10.30% | Arg | 36 | 6.50% | Pro | 26 | 4.70% | Tyr | 19 | 3.40% |
| Glu | 42 | 7.60%  | Ala | 34 | 6.10% | Thr | 26 | 4.70% | Gln | 16 | 2.90% |
| Val | 39 | 7.00%  | Gly | 33 | 5.90% | Phe | 25 | 4.50% | His | 15 | 2.70% |
| Ile | 37 | 6.70%  | Ser | 32 | 5.80% | Asn | 23 | 4.10% | Trp | 5  | 0.90% |
| Lys | 37 | 6.70%  | Asp | 30 | 5.40% | Met | 20 | 3.60% | Cys | 4  | 0.70% |

2.3.2 疏水区分析 通过 ExPaSy 软件的 ProScale 功能分析, 得到中华按蚊 *CYP4G17* 蛋白的 N 端 20-39 位氨

氨基酸之间有一个典型的疏水区域 (图 2.3)。

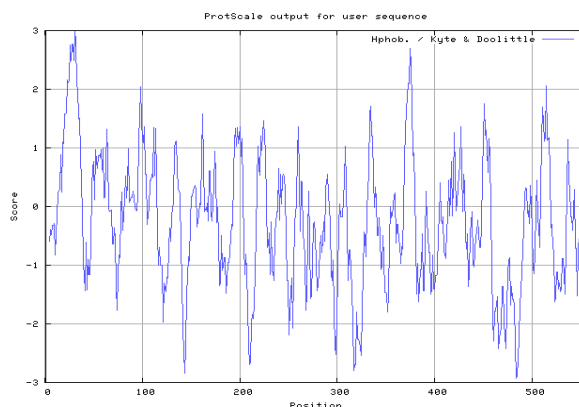


图 2.3 中华按蚊 CYP4G17 蛋白序列的疏水曲线

Fig. 2.3 Hydrophobicity profile of AsCYP4G17

2.3.3 跨膜区分析 采用 TMHMM 软件预测蛋白的跨膜区域, 发现 CYP4G17 具有两个跨膜片段, 分别位于 20-39 和 46-68 位 (图 2.4), 预测该蛋白为膜结合蛋白。

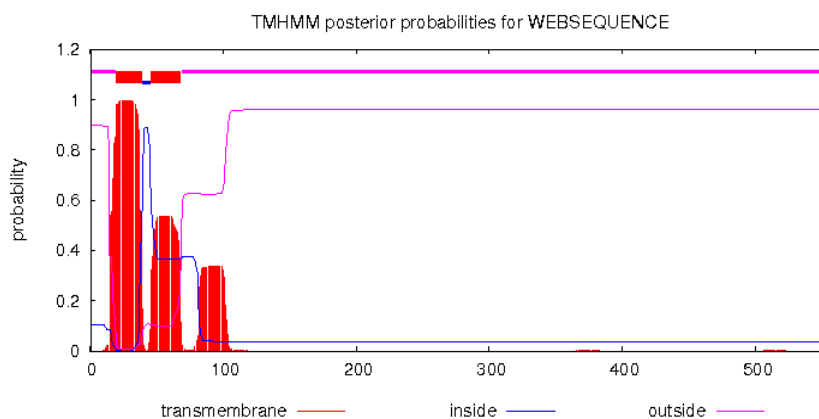


图 2.4 中华按蚊 CYP4G17 跨膜区预测

Fig. 2.4 The deduced transmembrane domain topology model of AsCYP4G17

2.3.4 信号肽预测 中华按蚊 CYP4G17 蛋白的信号肽预测使用 SignalP 软件, 该蛋白 C 值是 0.127, Y 值 0.130, S 值 0.281, 不存在信号肽切割位点, 未发现信号肽存在 (图 2.5)。

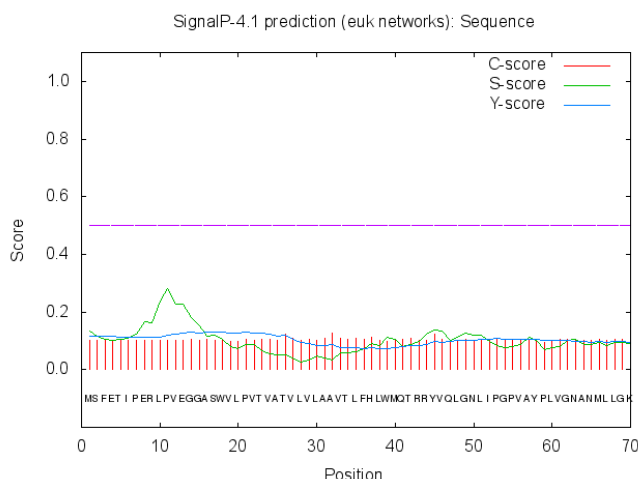


图 2.5 中华按蚊 CYP4G17 的信号肽分析

Fig. 2.5 Singal peptide prediction of AsCYP4G17

2.3.5 亚细胞定位 蛋白质亚细胞定位软件 TargetP 预测显示该蛋白为分泌通路, 即分泌到细胞周质, 故定位在细胞质。

2.3.6 蛋白的二级结构及三级结构预测 蛋白质二级结构预测使用 SSPro4.0 软件, 结果显示该蛋白二级结构中  $\alpha$ -螺旋、无规卷曲和  $\beta$ -折叠分别占 54.24%, 37.59% 和 9.17%, 可以发现  $\alpha$ -螺旋占比最多。通过 PSI-BLAST 的搜索, AsCYP4G17 与人微粒体 CYP2C19 氨基酸序列 (PDB 编号: 4gqs.pdb) 一致性为 32.0%, 被选作同源模建的模板, 通过 SWISS-MODEL 模建预测得到 AsCYP4G17 的蛋白质三级结构 (图 2.6)。

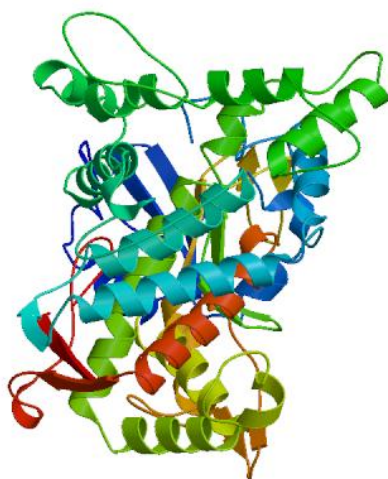
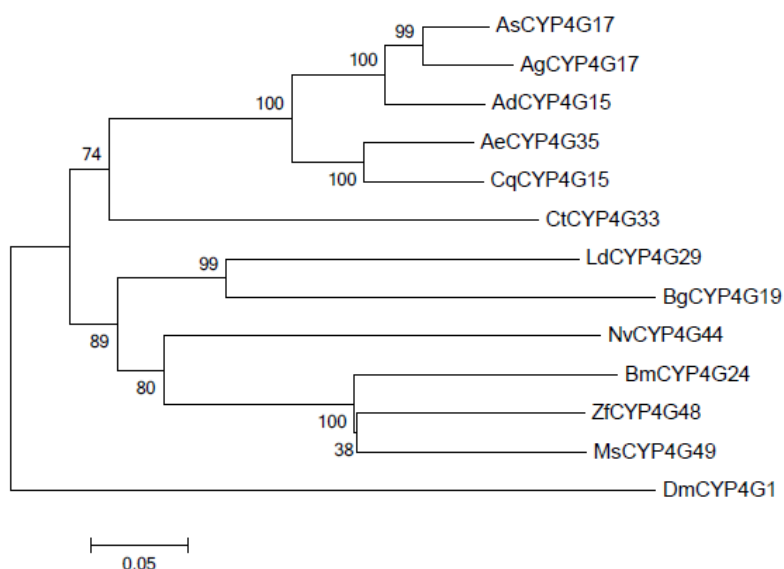


图 2.6 中华按蚊 *CYP4G17* 蛋白的三级结构预测分析  
Fig. 2.6 The tertiary structure of *AsCYP4G17*

### 2.4 系统发育分析

根据同源性分析结果选择了 12 个氨基酸序列 (包括 *AsCYP4G17*), 以黑腹果蝇 *CYP4G1* 作为外群, 采用邻接法构建系统发育树, 并用 bootstrap 1000 次重复对其分支作了统计检测<sup>[6]</sup> (图 2.7)。结果表明中华按蚊 *CYP4G17* 与冈比亚按蚊的 *CYP4G17* 最近, bootstrap 值达 99%, 聚成一支。



注: 使用黑腹果蝇 *CYP4G1* 作为外群, Bootstrap 值标记在树的分支节点上。各物种的中文名及其序列的基因库登录号见表 2.2。Note: The *DmCYP4G1* was used as outgroup, and bootstrap values are indicated on each branch. The Chinese names of species and their GenBank accession numbers of P450 were listed in Tab.2.2.

图 2.7 13 种昆虫 P450 系统发育树  
Fig. 2.7 The phylogenetic tree of P450s of 13 insect species

### 2.5 基因结构分析

基因结构预测发现 *AsCYP4G17* 基因有 2 个相位 2 型内含子, 与致倦库蚊的相同, 与冈比亚按蚊和达林按蚊的不相同。外显子 1、外显子 2、外显子 3、内含子 1、内含子 2 长度范围为分别为 662~680 bp、836~893 bp、161~164 bp、66~128 bp、61~98 bp (表 2.4)。

表 2.4 4 种蚊虫 *CYP4G* 亚家族基因结构分析  
Tab.2.4 The gene structure analysis among *CYP4Gs* of 4 mosquito species

| 基因名<br>Gene name | GenBank 登录号<br>GenBank Accession No. | 外显子长度 (bp)<br>Exon size | 内含子长度 (bp)<br>Intron size | 内含子相位 (0/1/2)<br>Intron phrase |
|------------------|--------------------------------------|-------------------------|---------------------------|--------------------------------|
| <i>AsCYP4G17</i> | ATLV01017222.1                       | 671/836/164             | 66/65                     | 2/2                            |
| <i>AgCYP4G17</i> | AAAB01008963.1                       | 677/851/161             | 128/74                    | 2/1                            |
| <i>AdCYP4G15</i> | ADMH02000163.1                       | 680/893/161             | 75/98                     | 2/1                            |
| <i>CqCYP4G15</i> | AAWU01039145.1                       | 662/851/161             | 83/61                     | 2/2                            |

### 3 讨论

细胞色素 P450 是最古老最庞大的超基因家族之一,包括多个基因家族,在昆虫体内各类杀虫剂和其他外源性和内源性化合物的代谢中发挥重要作用<sup>[13]</sup>。其中 CYP6 家族基因被认为与杀虫剂抗性密切相关<sup>[14]</sup>,但有研究表明 CYP4 家族的高水平转录与抗性也相关。本研究基于中华按蚊转录组数据,通过同源性搜索获得 *AsCYP4G17* 基因,该序列编码区包含 1671 bp,编码 556 个氨基酸,分子量约为 63.48 kDa,等电点为 7.70,其编码氨基酸序列的 354-366 位发现 CYP4 家族的特征序列 EVDTFMFEGHDTT,推断该基因属于 CYP4 家族基因。同源性分析发现,该基因与冈比亚按蚊 *CYP4G17* 相似性最高,达 89%,其次是与达林按蚊 *CYP4G15*,相似性为 86% (表 2.3),进一步表明该基因为 CYP4G 亚家族基因。近来,有研究表明中华按蚊抗性品系中 CYP4 家族的基因有明显的高表达<sup>[15]</sup>,其中包括 *CYP4G16*, *CYP4G17* 和 *CYP4H15*,这说明中华按蚊 *CYP4G17* 可能与抗性有关,但目前对其的研究相对较少,需进一步的研究来从功能上证明其与抗性相关。

在 *AsCYP4G17* 氨基酸序列的 354-366 位发现 CYP4 家族的特征序列 EVDTFMFEGHDTT,但其第 4 位氨基酸由异亮氨酸代替了苏氨酸(图 2.1)。与其它 4 种蚊虫 CYP4 家族基因多重序列对比发现,在 5 种蚊虫中都发生了该氨基酸的替换(图 2.3)。五个保守结构域中的 Meander 区(PxxYxPE/DRF)也发生了氨基酸的替换,该区域第 9 位的苯丙氨酸被苏氨酸代替了。以上保守区氨基酸的替换,可能是 CYP4G 亚家族的特征序列。有研究认为氨基酸替换可能与昆虫 P450 介导的抗性有关<sup>[16]</sup>,以上保守区氨基酸的替换是否与不同昆虫细胞色素 P450 功能差异相关有待进一步研究。

采用 NJ 法构建系统发育树,中华按蚊的 *CYP4G17* 与冈比亚按蚊的 *CYP4G17*、达林按蚊的 *CYP4G15* 聚在一支,bootstrap 值均大于 90%,表明它们的亲缘关系最近。而与蚊科及其它昆虫 CYP4 相似性在 60%~80% 之间。该结果进一步证明了该序列确实为 CYP4G 亚家族基因。基因结构分析发现,中华按蚊与其他 3 种蚊虫的内含子插入位置基本相同,3 个外显子的长度也基本相同,但是内含子的长度和相位略有不同。昆虫细胞色素 P450 基因内含子与进化的关系有待深入探讨<sup>[17]</sup>。

本研究报道的中华按蚊 *CYP4G17* cDNA 序列进一步丰富了 P450 基因的基础数据,对后续研究该基因的功能具有重要意义。但该基因是否与中华按蚊抗药性相关,还需通过实验来从功能上进行验证。

#### 参考文献:

- [1] 杨永焱,王秋生.疟疾预防的现状与进展[J].海南医学,2014(04):547-549.
- [2] Yang YY,Wang QS.Status and progress of malaria prevention [J].Hainan Medical Journal, 2010(04):547-549.
- [3] World Health Organization.World malaria report [R]. Geneva:WHO,2013.
- [4] 车燕飞,张玉娟,唐尧,等.中华按蚊 *CYP6P5* 基因的生物信息学鉴定及特征分析[J].中国媒介生物学及控制杂志,2014,(03):205-210.
- [5] CHE YF,ZHANG YJ,TANG Y,et al.Bioinformatic identification and characterization of *CYP6P5* gene in *Anopheles sinensis*[J].Chinese Journal of Vector Biology and Control,2014(03):205-210.
- [6] Chen B,Zhang YJ,He Z,et al.De novo transcriptome sequencing and sequence analysis of the malaria vector *Anopheles sinensis* (Diptera: Culicidae)[J].Parasites & vectors,2014,7(1):314.
- [7] David JP,Ismail HM,Chandor-Proust A,et al.Role of cytochrome P450s in insecticide resistance: impact on the control of mosquito-borne diseases and use of insecticides on Earth[J].Philosophical transactions of the Royal Society of London Series B,Biological sciences,2013,368(1612):20120429.
- [8] 唐尧,乔梁,张玉娟,等.中华按蚊 *CYP6Y* 亚家族基因的鉴定和生物信息学分析[J].昆虫学报,2014(06):663-672.
- [9] TANG Y,QIAO L,ZHANG YJ,et al.Identification and bioinformatics analysis of genes of the *CYP6Y* Subfamily in *Anopheles sinensis* (Diptera: Culicidae)[J].Acta Entomologica Sinica,2014(06):663-672.
- [10] Chang X,Zhong D,Fang Q,et al.Multiple resistances and complex mechanisms of *Anopheles sinensis* mosquito: a major obstacle to mosquito-borne diseases control and elimination in China[J].PLoS neglected tropical diseases,2014,8(5):e2889.
- [11] 杨帆,王进军.昆虫细胞色素 P450 与抗药性关系研究进展[J].四川动物,2008,27(3):460-463.
- [12] Yang F,Wang JJ.Advances in relations between cytochrome P450 and insecticide resistance[J].Sichuan Journal of Zoology,2008,27(3):460-463.
- [13] Tamura K,Peterson D,Peterson N,et al.MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods[J].Molecular Biology and Evolution,2011,28:2731-2739.
- [14] Solovyev V,Kosarev P,Seledsov I,et alAutomatic annotation of eukaryotic genes, pseudogenes and promoters[J].Genome Biol,2006,7,Suppl1:P.10.1-10.12.
- [15] Ai J,Zhu Y,Duan J,et al.Genome-wide analysis of cytochrome P450 monooxygenase genes in the silkworm, *Bombyx mori*[J].Gene,2011,480(1-2):42-50.
- [16] Bradfield JY,Lee YH,Keeley LL.Cytochrome P450 family 4 in a cockroach: molecular cloning and regulation by regulation by hypertrehalosemic hormone[J].Proc Natl Acad Sci USA,1991,88(10):4558-4562.
- [17] Feyerisen R.Insect P450 enzymes[J].Ann Rev Entomol,1999,44:507-533.
- [18] Nikou D,Ranson H,Hemingway J.An adult-specific CYP6 P450 gene is overexpressed in a pyrethroid-resistant strain of the malaria vector, *Anopheles gambiae*[J].Gene,2003,318:91-102.
- [19] Zhu G,Zhong D,Cao J,et al.Transcriptome profiling of pyrethroid resistant and susceptible mosquitoes in the malaria vector, *Anopheles sinensis*[J].BMC Genomics,2014,15:448-461.
- [20] Maitra S,Dombrowski SM,Waters LC,et al.Three second chromosome-linked clustered Cyp6 genes show differential constitutive and barbital-induced expression in DDT-resistant and susceptible strains of *Drosophila melanogaster*[J].Gene,1996,180:165-171.
- [21] Scott JA,Collins FH,Feyerisen R.Diversity of cytochrome P450 genes in the mosquito, *Anopheles albimanus*[J].Biochem Biophys Res Commun,1994:205(2):1452-1459.



**Animal Sciences****Identification and Bioinformatic Analysis of the *CYP4G17* Gene in *Anopheles sinensis***

YAN Zhengwen, ZHANG Yujuan, ZHOU Yong, CHEN Bin

(Institute of Entomology and Molecular Biology, Chongqing Key Laboratory of Animal Biology, Chongqing Normal University, Chongqing 401331, China)

**Abstract:** In this study, a sequence of CYP4 family was identified by homology search based on the database of *Anopheles sinensis* transcriptome data. The sequence is designated as *AsCYP4G17* by the P450 nomenclature committee. The full-length cDNA is 1962 bp long with an opening reading frame (ORF) of 1671 bp encoding 556 amino acids. Homology analysis showed that this gene shares the highest sequence similarity (94%) with *CYP4G17* of *Anopheles gambiae*. The calculated molecular weight is 63.48 kDa and the estimated pI is 7.70. Protein subcellular localization showed that this protein is localized in cytoplasm. Gene structure analysis showed that this gene subfamily contains two phase “2” intron. This study lays a foundation for further revealing the function of cytochrome P450 in *Anopheles sinensis*, and has certain scientific significance to elucidate the mechanism of insecticide resistance.

**Keywords:** *Anopheles sinensis*; *CYP4G17*; identification; bioinformatics analysis