

中华按蚊 *AsAcel* 基因的鉴定及生物信息学分析^{*}

邬雪梅, 张玉娟, 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:利用同源性搜索,在中华按蚊(*Anopheles sinensis*)基因组中鉴定出 1 条乙酰胆碱酯酶(Ace)家族序列,命名为 *AsAcel* (GenBank 登录号:KU900233)。该基因序列的全长为 5 592 bp,编码区序列长度为 2 253 bp,编码 750 个氨基酸。在理化性质、二级结构、三级结构、系统发生、基因结构等方面对该基因及所编码蛋白进行了预测和分析。结果表明 *AsAcel* 蛋白在按蚊中具有很高的保守性;该蛋白是亲水性蛋白,分子量、等电点分别为 82.41 kD,5.92,二级结构主要以不规则卷曲为主,无跨膜结构域和信号肽;亚细胞定位显示该蛋白位于细胞质中;基因结构为两种相位(1 位和 2 位内含子)。研究结果为 *AsAcel* 基因生物学功能的挖掘奠定了理论基础。

关键词:中华按蚊;*AsAcel* 基因;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(2016)06-0019-07

疟疾是热带和亚热带地区最普遍的以按蚊为传播媒介的蚊媒传染病,是人类面临的主要健康问题^[1-2]。WHO 指出,2015 年全球约有 2.14 亿人感染疟疾,43.8 万人因疟疾死亡^[3]。中华按蚊(*Anopheles sinensis*)是中国及东南亚地区最主要的疟疾传播媒介之一,从阿富汗、日本到印度尼西亚都有分布^[4-6];该物种广泛且高密度的地理分布给人们带来了巨大的危害,因此对它的控制是人类亟待解决的重要问题。

乙酰胆碱酯酶(Ace)是脊椎动物和非脊椎动物神经系统中的一种关键酶,它能调节神经递质乙酰胆碱的传递水平和终止神经冲动^[7]。在昆虫中,由于 Ace 是有机磷和氨基甲酸酯这两种抗胆碱类农药的主要靶标,它们能和 Ace 的部分催化三联体的活性位点丝氨酸残基发生共价反应^[8],所以该基因主要涉及在杀虫剂抗性方面的研究。Ace 的生化和分子特性在大多数昆虫中已经被报道,例如黑腹果蝇(*Drosophila melanogaster*)^[9]、斯氏按蚊(*Anopheles stephensi*)^[10]、马铃薯甲虫(*Leptinotarsa decemlineata*)^[11]、家蝇(*Musca domestica*)^[12]、麦二叉蚜(*Schizaphis graminum*)^[13]等。大多数节肢动物的 Ace 有两个即 Ace1 和 Ace2。例如,德国小蠊(*Blattella germanica*)^[14]、臭虫(*Cimex lectularius*)^[15]、小菜蛾(*Plutella xylostella*)^[16]、烟青虫(*Helicoverpa assulta*)^[17]、家蚕(*Bombyx mori*)^[18]、意大利蜜蜂(*Apis mellifera*)^[19]等都有两个 Ace。但是,只有 *Acel* 基因编码突触 Ace,且与有机磷和氨基甲酸酯杀虫剂的不敏感机制相关^[16-19]。然而也有例外,微小牛蜱(*Boophilus microplus*)有 3 个 Ace,且氨基酸序列之间差异较大,表现为不同的生化特性,其中 Ace1 和 Ace2 都与有机磷抗性相关^[20]。与上述几种物种不同,在黑腹果蝇、铜绿蝇(*Lucilia cuprina*)、家蝇等昆虫中只存在 Ace2 基因,这个基因涉及突触的传递并且与有机磷和氨基甲酸酯杀虫剂抗性机制相关^[21-23]。目前,在中华按蚊中有关 *Acel* 基因的抗药性研究还未见报道。

本研究基于中华按蚊基因组数据,通过同源搜索得到 1 条 *Acel* 基因序列,并对该基因的结构以及所编码蛋白的理化性质、二级结构、三级结构、系统发生等方面进行了预测和分析。有关结果为进一步挖掘和验证中华按蚊 *Acel* 基因生物学功能提供了有力的理论基础。

1 材料与方法

1.1 数据来源

中华按蚊基因组数据由重庆师范大学昆虫分子生物学研究所提供,冈比亚按蚊(*Anopheles gambiae*),催命

^{*} 收稿日期:2016-03-09 修回日期:2016-03-28 网络出版时间:2016-11-02 13:34

资助项目:美国国立卫生研究院 NIH 项目(No. R01AI095184);国家自然科学基金(No. 31372265;No. 31071968);国家科技基础性工作专项重点项目(No. 2015FY210300);国际原子能机构 CRP 项目(No. 18268/R0);“两江学者”计划专项经费

作者简介:邬雪梅,女,研究方向为昆虫分子生物学,E-mail:1204844264@qq.com;通信作者:陈斌,教授,E-mail:bin.chen@cqnu.edu.cn

网络出版地址:http://www.cnki.net/kcms/detail/50.1165.N.20161102.1334.056.html

按蚊(*Anopheles funestus*)、微小按蚊(*Anopheles minimus*)、致倦库蚊(*Culex quinquefasciatus*)等物种的 Ace1 蛋白序列下载于 NCBI 的 GenBank 数据库(<http://www.ncbi.nlm.nih.gov/>)。

1.2 序列鉴定

以冈比亚按蚊 Ace1 蛋白的氨基酸序列为参考序列,通过 TBLASTN(E 值小于或等于 1×10^{-5})搜索中华按蚊基因组数据库,搜索得到的结果用 FGENESH^[24] (<http://linux1.softberry.com/berry.phtml>)预测可能的编码区序列和氨基酸序列。序列采用在线 BLASTP 比对 NCBI 的 nr 库进行验证。

1.3 序列分析方法

同源序列搜索使用 BLAST 工具(<http://www.ncbi.nlm.nih.gov/BLAST/>);BioEdit 软件统计编码区序列的碱基组成;CodonW 软件分析密码子使用频率;ClustalX 软件对中华按蚊 Ace1 蛋白和其他昆虫 Ace1 蛋白序列进行多重序列比对;ProtParam 软件(<http://web.expasy.org/protparam/>)预测中华按蚊 Ace1 蛋白的基本理化性质;ProtScale 软件(<http://www.expasy.org/cgi-bin/protscale.pl>)预测蛋白质的疏水性;TMHMM 软件(<http://www.cbs.dtu.dk/services/TMHMM/>)预测蛋白质的跨膜区;TargetP 软件(<http://www.cbs.dtu.dk/services/TargetP/>)进行亚细胞定位^[25];SignalP4.1 软件(<http://www.cbs.dtu.dk/services/SignalP/>)预测蛋白质的信号肽;SSPro4.0 软件(<http://scratch.proteomics.ics.uci.edu/>)预测蛋白质的二级结构;PSI-BLAST 搜索 PDB(Protein Data Bank)数据库,找到与中华按蚊 Ace1 蛋白相似性较高的蛋白;利用 SWISS-MODEL(<http://swissmodel.expasy.org/>)进行同源模建和 3D 结构预测^[25]。

1.4 系统发育分析

利用 MEGA5 软件^[26]自带的 Clustal W 对中华按蚊 Ace1 蛋白序列与冈比亚按蚊、催命按蚊、微小按蚊、致倦库蚊的 Ace1 蛋白进行比对,并用邻接法(Neighbor-joining)构建系统进化树,重复次数设为 1 000。

2 结果与分析

2.1 编码区序列分析

经过同源序列比对,得到 1 条与冈比亚按蚊 Ace1 基因的同源性最高的序列,长度为 5 592 bp,序列一致性为 79%。基于生物信息学分析,此序列编码区为 2 253 bp,编码 750 个氨基酸,是一条具有起始子(ATG)和终止子(TAA)的全长编码区序列(图 1)。因此将这条中华按蚊的 Ace1 基因序列命名为 AsAce1 (Gen-Bank 登录号: KU900233)。

对 AsAce1 基因编码区序列的碱基组成分析结果得出:GC 含量为 62.63%,AT 含量为 37.37%,表明 AsAce1 基因在整个编码区序列中 GC 含量大于 AT 含量。GC 偏斜为 0.04,AT 偏斜为 0.01,说明碱基组成适度的偏好碱基 G 和碱基 A。

利用 Bioedit 软件对 AsAce1 基因的相对同义密码子使用率(RSCU)进行预测。结果表明,在 AsAce1 基因的密码子中,有 29 种密码子的 RSCU 值大于 1,为 AsAce1 基因的偏好密码子,其中有 27 种密码子都是以 G 或 C 结尾,偏好性较强的有 UAA, CUG,CGG,CCG,AUC,GUG (RSCU 值大于 2),编码甲硫氨酸的密码子 AUG 和编码色氨酸的密码子 UGG 的 RSCU 值等于 1,表明 AsAce1 基因对 AUG 和 UGG 的使用无偏好(表 1)。

表 1 AsAce1 基因编码区序列相对同义密码子

Tab. 1 Relative synonymous codon usage in AsAce1 gene coding sequence

密码子	RSCU	密码子	RSCU	密码子	RSCU	密码子	RSCU
UAA	3.00	ACC	1.58	GGG	0.61	AGG	0.37
CUG	2.84	CAC	1.58	CCA	0.60	AUA	0.36
CGG	2.19	AAC	1.57	UGU	0.60	UCU	0.35
CCG	2.19	UCC	1.53	AUG	0.59	GAA	0.35
AUC	2.16	UCG	1.53	CAA	0.53	UUG	0.32
GUG	2.07	GAC	1.52	CUA	0.49	GCU	0.31
AAG	1.91	CAG	1.47	GAU	0.49	UAU	0.31
CUC	1.86	UGC	1.4	CGU	0.49	UCA	0.24
ACG	1.77	GUC	1.29	AUU	0.48	GUA	0.21
AGC	1.76	GGU	1.21	ACA	0.47	ACU	0.19
GGC	1.76	GCG	1.13	GUU	0.43	CCU	0.15
CGC	1.71	CCC	1.06	AAU	0.43	AGA	0.12
GCC	1.69	AUG	1.00	CAU	0.42	AAA	0.09
UAC	1.69	UGG	1.00	GGA	0.42	UUA	0.08
GAC	1.65	CGA	0.98	CUU	0.41	UAG	0.00
UUC	1.63	GCA	0.88	UUU	0.38	UGA	0.00

2.2 同源性分析

利用 NCBI 在线 BLASTP 搜索功能,将 AsAce1 蛋白的氨基酸序列作为参考序列进行搜索,结果表明冈比亚按蚊的 Ace1 与该序列的一致率最高,达到 85%;催命按蚊、微小按蚊、致倦库蚊与该序列的一致率则依次逐渐减小(表 2)。

将以上几条序列和 AsAce1 蛋白进行序列比对分析。结果表明 AsAce1 含有 Ace 的特征序列,也即是 3 个氨基酸残基组成的催化三联体 (S200, E327, H440), 3 个保守的二硫键连接的 6 个半胱氨酸 (C216~C244, C403~C416 和 C551~C695) 以及活性丝氨酸残基周围的特征基序“FGESAG”^[27] (图 2)。

2.3 系统发育分析

根据同源性分析的结果,选取 5 条氨基酸序列与 AsAce1 蛋白构建系统发育树。以致倦库蚊 Ace1 蛋白序列为外群。根据系统进化树可以得知,AsAce1 蛋白与斯氏按蚊 Ace1 蛋白的距离最近,bootstrap 值高达 100% (图 3)。

2.4 编码蛋白质的特性分析

2.4.1 理化性质分析 *AsAcel* 基因编码的蛋白质分子式为 $C_{3684}H_{5680}N_{1026}O_{1093}S_{18}$, 相对分子质量为 82.41 KD, 理论等电点为 5.92, 不稳定系数为 38.27, 由此可知 *AsAcel* 为稳定蛋白。在氨基酸组成中, 亮氨酸所占比例最高, 为 9.9%, 甲硫氨酸所占比例最低, 为 1.1% (表 3)。该蛋白的亲水性平均系数为 -0.276, 故此蛋白为亲水性蛋白。负电荷残基总数(天冬氨酸+谷氨酸)以及正电荷残基总数(精氨酸+赖氨酸)分别为 83, 71, 脂溶指数为 -0.276。

-9 CCGGACGCGCATGAGATCCGAGGGCTCTGATGGGACGGTCTCGGTAGGTGGCCGCCAGGTGGCGCCTCTGCTGATG
-3 P D A M E I R G L L M G R R L L G G G A C G G A C G G C V A P L L S
70 CTGTGACACCTCGCCCTGCTCGCCCGCTCCCTGCTGGTGGCCGTGCTCCAGGACGGCACCAGAGCTCAATAATGGC
24 L C T L A L L A P S L L L V P V V H G R H H E L N N G
147 GCCGGTTCGCTCGGGTCGCATCAGCTGTGCGCGCGCGGTGGCGTTGGCGGCGTTGGCCTATCTCGTCCCAGCCAGCC
49 A G S L G S H Q L S A A A G G V G G V G L S S S G Q P A
225 GACTCGATCTCGGGTCGCCGCTCGCGGGGCGAGGAGGACGTGGGCGGATGTGCTGCCCTCAGCAAAAGACGCAGAT
75 D S I S G A A V A G A E E D V G R I A A L S K D A D
303 GCATTTTTTACACCATATATAGGTCACGGTGAGTCCGTACGAATTATAGATCCCAGCTGGGCACGCTCGAGCGCGAG
101 A F F T P P Y I G H G E S V R I I D P E L G T L E R E
381 CATGTCCACGGTGGAGCACACCGCTCGGCGGGTCTGACGAGACGTGAGTCAAACCTCAGATGCCAACGATAATGA
127 H V H G A T G T P R R R G L T R R E S N S D A N D N D
459 CCGTGGTGGTGAACACCGACAAGGGTCGCATCCGCGGGTTTACGGCCGAGGCACCGAGCGGCAAGAAGTGGACGTG
153 P L V V N T D K G R I R G T F T A E A P S G G K K V D V
537 TGGCTCGGATCCGCTACGCGAACCCCGCTCGGACCGTTCGCTTCCGACACCCGCGCCGCGGAGAGTGGACC
179 W L G I P Y A Q P P V G P L R F R H P R P A E K W T
615 GGTGTGCTGAACACGACGACCCCGCGGAACAGCTGCGTGCAGATCGTGGACACCGTGTTCCGTGACTTCCCGGTGCG
205 G V L N T T T P P N S C V Q I V D T V F G D F P G A
693 ACCATGTGGAACCCCAACACGCCCTCTCCGAGGACTGCCTGTACATCAACGTGGTGGCACCCCGCGCCCGTCCGAAG
231 T M W N P N T P L S E D C L Y I N V V A P R P R P K
771 AATCGGCGCTCATGCTGGATCTTCGGCGGTGGCTTCTACTCCGACCGGCCACGCTTGACGTCTACGACCACCGG
257 N A A V M L W I F G G G F Y S G T A T L D V Y D H R
849 CGCTGGCGTCCGAGGAGAAGTGATCGTGTCTGCTGCAGTACCGTGTGCTAGCTGGGTTCCTGTTCTCTCGGCG
283 A L A S E E N V I V V S L Q Y R V A S L G F L F L G
927 ACACCGAAGCACCCGGCAATCGGGACTTTTCGATCAGAACCTCGCCCTACGATGGGTTCGAGACAACATTCAAG
309 T P E A P G N A G L F D Q N L A L R L W V R D N I H K
1 005 TTCGGTGGAGATCCCGACAGGTGACACTGTTCCGGCAAAGTCTGGTCTGTTCCTGCTCGCTGACTACTGTCCTC
335 F G G D P A R V T L F G E S A G A V S V S L H L L S
1 083 GCACCTCTCGCGGACCTTTTCCAGCGGGCCATCTTGCAGAGTGGTTCGCCGACGCGACCCCTGGGCACTTGTGCTCGC
361 A L S R D L F Q R A I L Q S G S P T A P W A L V S R
1 161 GAGGAAGCAACTTCCGAGCCATTTCGGTTGGCGAAGCGTTGGATGCCCGACGACACGAGCAAGCTCAGCGAAGCG
387 E E A T L R A I R L A E A V G C P H D T S K L S E A
1 239 GTCGAGTGTCTCGCGCGCAAGACCCGACGTCGTGTCGAACACAGTGGGGTACGCTCGGCATCTGCGAGTTCCCG
413 V E C L R G K D P H V L V N N E W G T G G G C I C E F P
1 317 TCTGTGCGGTGTGACGGTTCGCTTCTGACGAGACGCGCAGCGGTCTGCGAGCGGCGCTTCAAGAAGACG
465 F V P V V D G A F L D E T P Q R S L A S G R F K K T
1 47 GACATCCTACCCGACGAAACCGAGGAGGGCTACTACTTATCATCTACTACCTGACCGAGCTGCTCGCGAAGGAG
491 D I L T G S N T E E G Y Y F I I Y Y L T E L L K K E
1 551 GAGGGCGTGACGGTTCGCGCGAGGAGTTCCTCGAGGCGTTCGCGAGCTCAACCCGTACGTCAACGGGGCCGCCCGG
517 E G V T V S R E E F L Q A V R E L N P Y V N G A A R
1 629 CAGGCGATCTGTTCGAGTACCCGACTGACCGAGCGGACACCCGAACAGCAACCGGACGCGCTGGACAAGATG
640 Q A I V F E Y T D W T E P D N P N S N R D A L D K M
1 707 GTGCGGACTACCACTTCACTGCAACGTGAACGAGTTGCCACCGGCTACGCCAGGAGGGCAACACGCTCTACAT
566 V G D Y H F T C N V N E F A Q R Y A E G G N N V Y M
1 785 TACCTGTACACGCACCGGAGCAAGGGCAACCCGCGGCTGGCCCGCTGAGCCGCGTCAATGCACGGCGACGAGATCAACTAC
592 Y L Y T H R S K G N P W P R W T G V M H G D E I N Y
1 863 GTGTTGGGGAACCGCTCAATCCCACCCTCGGCTACACCGACGACGAGAAGGACTTCAGCCGGAAGATCATGCGATAC
618 V F G E P L N P T L G Y T D E K D F S R K I M R Y
1 941 TGGTCCAACCTTGCCAAAGACTGGGTACGTACCGAAGGAGAAGGAGATCTGTGGGAGTCGAGGCTGACGATGTTTTT
644 W S N F A K T G Y V P K E K E I C G E S R S D D V F
2 019 CAAACCCCAACAGCAATCCAATCCCAGCCAACCTCAGCACCGACTATCCCAGTGGCCCAAGCATACGCCCCACGGA
670 Q T P N S N P N P S Q L S T D Y P E W P K H T A H G
2 097 CGGCACTATCTGGAGCTCGGCTGAAACACGTCCTTCTGTCGGGCGGGGCCACGGTTGAGGCAGTGTGCCTTCTGGAAG
696 R H Y L E L G L N T S F V G R G P R L R Q C A F W K
2 175 AAGTATCTACCCAGCTAGTTGCAGCTACCTCTAACCTACAAGTGGCACCACCACCAAGTGCACCGTGGCAAAGTAGC
722 K Y L P Q L G A T C T S N A L Q V A P P P S A P C A C E S S
2 253 GCCTTTTTCTACCGGCTGATCTGATCGTACTGATAGTGTGCTGCTTCCCGCCCGGCTTACCTACACACAATAA

注:两个方框分别代表起始密码子和终止密码子,奇、偶数行的数字分别表示核苷酸、氨基酸序号。

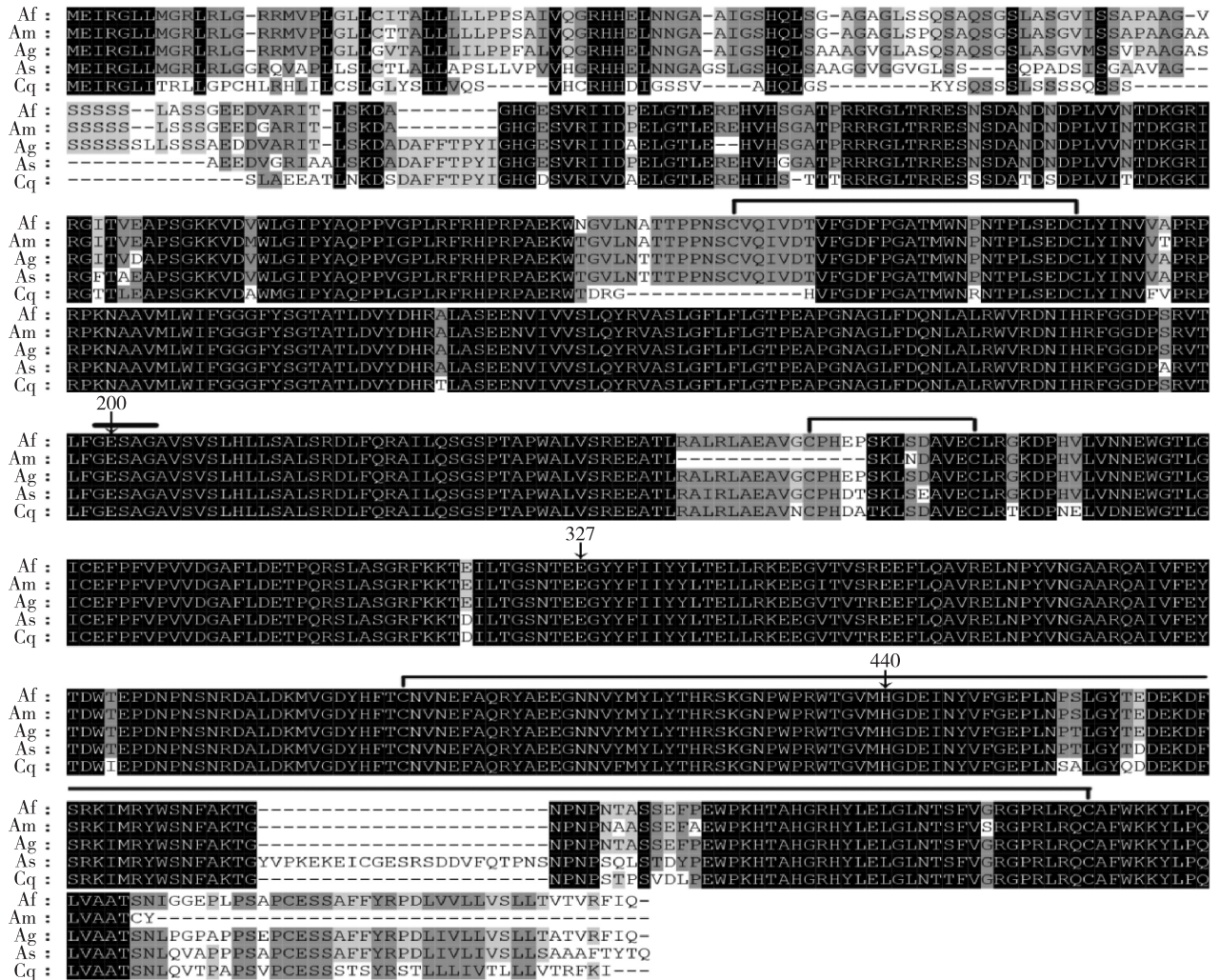
图 1 *AsAcel* 基因编码区序列及其预测的氨基酸序列

Fig. 1 The coding sequence of *AsAcel* gene and deduced amino acid sequence

表 2 AsAcel 蛋白与其他蚊虫 Acel 蛋白氨基酸序列的一致率和相似率

Tab. 2 The identity and similarity between AsAcel and its homologous Acel from other mosquito species

基因名称	物种	基因登录号	一致率	相似率
<i>AgAce1</i>	冈比亚按蚊 <i>A. gambiae</i>	AGAP001356	85	88
<i>AfAce1</i>	催命按蚊 <i>A. funestus</i>	AFUN011616	84	87
<i>AmAce1</i>	微小按蚊 <i>A. minimus</i>	AMIN000542	79	81
<i>CqAce1</i>	致倦库蚊 <i>C. quinquefasciatus</i>	CPIJ006034	78	82

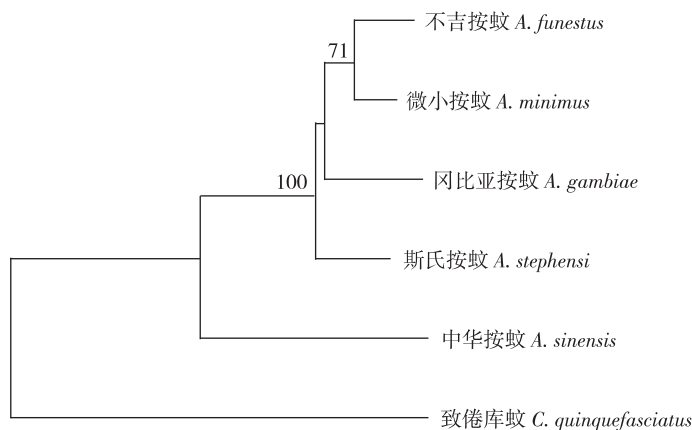


注:3个二硫键连接6个半胱氨酸(框线);短线处为“FGESAG”一致序列;组成催化三联体的3个氨基酸残基用箭头表示;黑色阴影表示氨基酸保守性为100%;Af、Am、Ag、As和Cq分别表示侏侏按蚊、微小按蚊、冈比亚按蚊、中华按蚊和致倦库蚊。

图2 中华按蚊 AsAcel1 蛋白与其他蚊虫 Acel1 蛋白氨基酸序列比对

Fig.2 Alignment of AsAcel1 with other homologous mosquito Acel1

2.4.2 疏水区分析 利用在线软件 ExPaSy 的 ProScale 功能,分析结果显示 AsAcel1 蛋白没有疏水区(图4)。



注:用致倦库蚊 Acel1 作为外群,大于50的bootstrap值标记在树的分支节点上。斯氏按蚊 Acel1 的基因库的登录号为 ASTE107510,其他物种的相关情况见表2。

图3 6种蚊虫 Acel1 蛋白的系统进化树

Fig.3 The phylogenetic tree of Acel1 of 6 mosquito species

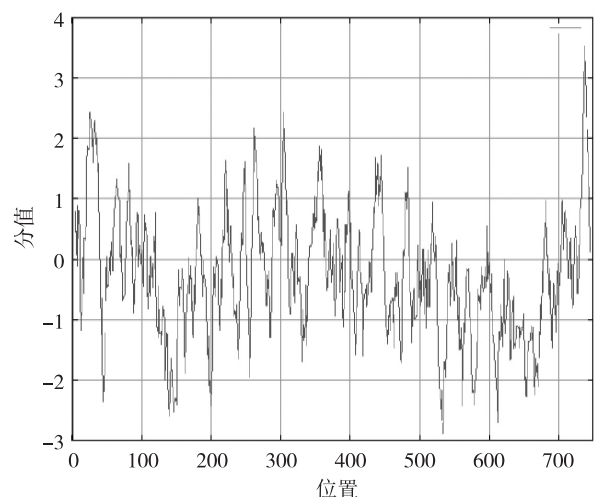


图4 AsAcel1 蛋白的疏水区预测

Fig.4 Hydrophobic region prediction of AsAcel1

表 3 中华按蚊 *AsAcel* 蛋白的氨基酸组成
Tab. 3 Composition of amino acids of *AsAcel*

氨基酸	个数	比例/%	氨基酸	个数	比例/%	氨基酸	个数	比例/%	氨基酸	个数	比例/%
亮氨酸	74	9.9	丝氨酸	51	6.8	天冬氨酸	37	4.9	组氨酸	19	2.5
甘氨酸	66	8.8	精氨酸	49	6.5	苯丙氨酸	32	4.3	谷氨酰胺	19	2.5
丙氨酸	64	8.5	谷氨酸	46	6.1	酪氨酸	26	3.5	色氨酸	13	1.7
缬氨酸	56	7.5	苏氨酸	43	5.7	异亮氨酸	25	3.3	半胱氨酸	10	1.3
脯氨酸	53	7.1	天冬酰胺	37	4.9	赖氨酸	22	2.9	甲硫氨酸	8	1.1

2.4.3 跨膜区分析 运用 TMHMM-2.0 软件对 *AsAcel* 蛋白的跨膜结构域进行预测,结果表明此蛋白不具有跨膜结构域(封二彩图 5)。

2.4.4 信号肽预测 使用 SingalP 软件对中华按蚊 *AsAcel* 蛋白进行信号肽预测,结果表明此蛋白 C 值为 0.613,Y 值为 0.511,S 值为 0.636,不存在信号肽切割位点,无信号肽(封二彩图 6)。

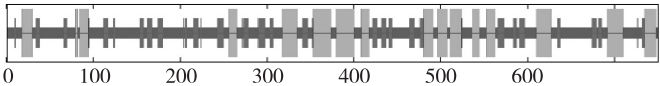
2.4.5 亚细胞定位 用 TargetP 软件对 *AsAcel* 进行亚细胞定位分析,结果显示分泌通路蛋白的分值最高,为 0.743(表 4),即最有可能分泌到细胞周质,故预测该蛋白定位在细胞周质。

表 4 *AsAcel* 蛋白的亚细胞定位分析

Tab. 4 Subcellular localization analysis of *AsAcel*

名称	长度	线粒体靶肽	分泌途径	其他	定位	可靠性等级
<i>AsAcel</i>	750	0.265	0.743	0.044	分泌蛋白	3
临界值		0.000	0.000	0.000		

2.4.6 二级结构预测 利用 SSPro4.0 软件对 *AsAcel* 蛋白的二级结构进行预测,结果显示该蛋白二级结构分组中 α -螺旋占 27.33%,延伸链占 18.27%,无规卷曲占 54.40%,表明此蛋白以无规卷曲为主, α -螺旋和延伸链分散在整个蛋白质中(图 7)。



注:图中线条由低到高依次表示无规则卷曲、延伸链、 α -螺旋。

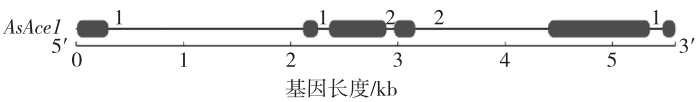
图 7 *AsAcel* 蛋白二级结构预测

Fig. 7 Secondary structure prediction of *AsAcel*

2.4.7 三级结构预测 经 PSI-BLAST 软件搜索,小鼠(*Mus musculus*)Ace 与 *AsAcel* 氨基酸的相似性达到 48.61%,将该序列选为同源建模模板,通过 SWISS-MODEL 预测得到中华按蚊 *AsAcel* 蛋白的三级结构模型(封二彩图 8)。

2.5 基因结构分析

运用 GSDS2.0 对 *AsAcel* 基因的结构进行分析,结果显示,该基因有 6 个外显子,外显子的长度按顺序分别为 295,141,538,198,950,122 bp;内含子长度按顺序分别为 1 814,100,70,1 239,116 bp。1 位内含子有 3 个,2 位内含子有 3 个,没有 0 位内含子(图 9)。



注:粗线表示外显子,细线表示内含子,粗、细线上方数字表示内含子相位。

图 9 *AsAcel* 基因的结构分析

Fig. 9 The gene structure analysis of *AsAcel* gene

3 讨论

Ace 是一种丝氨酸水解酶,最主要的作用是水解已经结合在胆碱能突触中的神经递质。该类酶在昆虫中多有研究,其中研究最深入的是 *Acel* 基因,通常认为该基因与靶标抗性相关^[1]。有关 *Acel* 基因涉及杀虫剂抗性的研究已在冈比亚按蚊、家蚕、德国小蠊、意大利蜜蜂等昆虫中报道。随着越来越多的昆虫全基因组测序的完成,从全基因组角度来分析和研究基因成为了可能;因此,中华按蚊基因组测序的完成为全面分析 *Acel* 基因提供了条件。

本研究通过同源搜索中华按蚊基因组数据库获得了 *AsAcel* 基因,预测了该基因的编码区序列及其氨基酸序列,并且对该基因的特性进行了详细的分析。通过几个物种氨基酸序列的比对可以得知,该基因具有其他蚊虫 *Acel* 基因共有的特性,且 *AsAcel* 蛋白具有催化三联体(S200,E327,H440)活性中心、3 个二硫键连接的 6 个

半胱氨酸、活性丝氨酸残基周围的特征基序“FGESAG”等,说明 AsAce1 是具有功能的活性酶。比对结果也可以看出该酶基因在蚊虫物种间的保守度很高。与冈比亚按蚊、催命按蚊、微小按蚊、致倦库蚊等蚊虫的 Ace1 蛋白氨基酸序列的一致率和相似率都在 78% 以上,说明不同蚊虫物种的 Ace1 基因之间的同源性很高。另外,从系统发生树的分析可以看出 AsAce1 基因在进化上与斯氏按蚊 Ace1 基因距离最近。

通过对 AsAce1 蛋白序列的理化性质、信号肽、疏水区、二级结构、三级结构等进行详细预测和分析,结果显示:AsAce1 蛋白为亲水性蛋白,二级结构主要以不规则卷曲为主。以上分析对预测该基因的潜在功能及为下一步开展 AsAce1 基因的生物学功能方面的研究奠定了基础;但该基因是否与靶标抗性相关还需通过具体实验进行验证。

参考文献:

- [1] Patarroyo M A, Calderón D, Moreno-Pérez D A. Vaccines against *Plasmodium vivax*: a research challenge[J]. Taylor & Francis Group, 2012, 11(10): 1249-1260.
- [2] Eastman R T, Pattaradilokrat S, Raj D K, et al. A class of tricyclic compounds blocking malaria parasite oocyst development and transmission[J]. Antimicrobial Agents and Chemotherapy, 2013, 57(1): 425-435.
- [3] World Health Organization. Disease burden[R]. Geneva: WHO, 2015.
- [4] Ma Y, Fan Y. Isolation and characterization of polymorphic microsatellite markers from Asian malaria mosquito *Anopheles sinensis* (Diptera: Culicidae)[J]. Molecular Ecology Resources, 2008, 8(5): 1059-1061.
- [5] Rueda L, Pecor J, Harrison B. Updated distribution records for *Anopheles vagus* (Diptera: Culicidae) in the Republic of Philippines, and considerations regarding its secondary vector roles in Southeast Asia[J]. DTIC Document, 2011, 28(1): 181-187.
- [6] Sinka M E, Bangs M J, Manguin S, et al. A global map of dominant malaria vectors[J]. Parasites & Vectors, 2012, 5(1): 181-187.
- [7] Toutant J P. Insect acetylcholinesterase: catalytic properties, tissue distribution and molecular forms[J]. Progress in Neurobiology, 1989, 32(5): 423-446.
- [8] Fournier D, Mutero A. Modification of acetylcholinesterase as a mechanism of resistance to insecticides[J]. Comparative Biochemistry and Physiology Part C: Pharmacology Toxicology and Endocrinology, 1994, 108(1): 19-31.
- [9] Hall L, Spierer P. The Ace locus of *Drosophila melanogaster*: structural gene for acetylcholinesterase with an unusual 5' leader[J]. EMBO Journal, 1986, 5(11): 2949-2954.
- [10] Hall L M, Malcolm C A. The acetylcholinesterase gene of *Anopheles stephensi*[J]. Cellular and Molecular Neurobiology, 1991, 11(1): 131-141.
- [11] Zhu K Y, Clark J M. Cloning and sequencing of a cDNA encoding acetylcholinesterase in Colorado potato beetle, *Leptinotarsa decemlineata* (Say)[J]. Insect Biochemistry and Molecular Biology, 1995, 25(10): 1129-1138.
- [12] Huang Y, Qiao C, Williamson M, et al. Characterization of the acetylcholinesterase gene from insecticide-resistant houseflies (*Musca domestica*)[J]. Chinese Journal of Biotechnology, 1996, 13(3): 177-183.
- [13] Gao J R, Kambhampati S, Zhu K. Molecular cloning and characterization of a greenbug (*Schizaphis graminum*) cDNA encoding acetylcholinesterase possibly evolved from a duplicate gene lineage[J]. Insect Biochemistry and Molecular Biology, 2002, 32(7): 765-775.
- [14] Kim J I, Jung C S, Koh Y H, et al. Molecular, biochemical and histochemical characterization of two acetylcholinesterase cDNAs from the German cockroach *Blattella germanica*[J]. Insect Molecular Biology, 2006, 5(4): 513-522.
- [15] Seong K M, Kim Y H, Kwon D H, et al. Identification and characterization of three cholinesterases from the common bed bug, *Cimex lectularius*[J]. Insect Molecular Biology, 2012, 21(2): 149-159.
- [16] Baek J H, Kim J I, Lee D W, et al. Identification and characterization of Ace1-type acetylcholinesterase likely associated with organophosphate resistance in *Plutella xylostella*[J]. Pesticide Biochemistry and Physiology, 2005, 81(3): 164-175.
- [17] Lee D W, Kim S S, Shin S W, et al. Molecular characterization of two acetylcholinesterase genes from the oriental tobacco budworm, *Heliothis virescens* (Guenée)[J]. Biochimica et Biophysica Acta (BBA)-General Subjects, 2006, 1760(2): 125-133.
- [18] Yu Q Y, Lu C, Li W L, et al. Annotation and expression of carboxylesterases in the silkworm, *Bombyx mori*[J]. BMC Genomics, 2009, 10(1): 260-262.
- [19] Claudianos C, Ranson H, Johnson R, et al. A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee[J]. Insect Molecular Biology, 2006, 15(5): 615-636.
- [20] Temeyer K B, Pruetz J H, Olafson P U. Baculovirus expression, biochemical characterization and organophosphate sensitivity of rBmAChE1, rBmAChE2, and rBmAChE3 of *Rhipicephalus (Boophilus) microplus*[J]. Vet-

- erinary Parasitology, 2010, 172(1): 114-121.
- [21] Mutero A, Pralavorio M, Bride J M, et al. Resistance-associated point mutations in insecticide-insensitive acetylcholinesterase[J]. Proceedings of the National Academy of Sciences, 1994, 91(13): 5922-5926.
- [22] Chen Z, Newcomb R, Forbes E, McKenzie J, et al. The acetylcholinesterase gene and organophosphorus resistance in the Australian sheep blowfly, *Lucilia cuprina* [J]. Insect Biochemistry and Molecular Biology, 2001, 31(8): 805-816.
- [23] Walsh S, Dolden T, Moores G, Kristensen M, et al. Identification and characterization of mutations in housefly (*Musca domestica*) acetylcholinesterase involved in insecticide resistance[J]. Biochem J, 2001, 359: 175-181.
- [24] Solovyev V, Kosarev P, Seledsov I, et al. Automatic annotation of eukaryotic genes, pseudogenes and promoters [J]. Genome Biol, 2006, 7(Suppl1): 1-12.
- [25] 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(6): 663-672. Tang Y, Qiao L, Zhang Y J, et al. Identification and bioinformatics analysis of genes of the CYP6Y Subfamily in *Anopheles sinensis* (Diptera: Culicidae) [J]. Acta Entomologica Sinica, 2014(6): 663-672.
- [26] 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(10): 2731-2739.
- [27] Kaur K, Bakke M J, Nilsen F, et al. Identification and molecular characterization of two acetylcholinesterases from the salmon louse, *Lepeophtheirus salmonis* [J]. Plos One, 2015, 10(5): e0125362.

Animal Sciences

Identification and Bioinformatic Analysis of the *AsAcel* Gene in *Anopheles sinensis*

WU Xuemei, ZHANG Yujuan, CHEN Bin

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

Abstract: In this study, a sequence of acetylcholinesterase family was identified by homology search based on the *Anopheles sinensis* genome database and the sequence is designated as *AsAcel* (GenBank accession number: KU900233). The full-length sequence is 5 592 bp with a coding sequence of 2 253 bp encoding 750 amino acids. Then the biochemical properties, secondary structures, tertiary structure, evolutionary relationship and gene structure were predicted and analyzed. The results showed that *Acel* protein is highly conserved through the *Anopheles* and it is a hydrophilic protein with calculated molecular weight 82.41 kD and estimated pI 5.92. Secondary structure is mainly irregular coils, and there are no transmembrane domain topology model and signal peptide. Protein subcellular localization analysis showed that the *AsAcel* protein is localized in cytoplasm. Gene structure analysis showed that this gene contains two phases (phase 1 and 2). This research established a bioinformatic foundation for the further function research on the *AsAcel* gene in *Anopheles sinensis*.

Key words: *Anopheles sinensis*; *AsAcel*; identification; bioinformatic analysis

(责任编辑 方 兴)