

中华按蚊 *AsAce2* 基因的鉴定及生物信息学分析^{*}

罗钱春, 邬雪梅, 余海蓉, 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:【目的】对中华按蚊(*Anopheles sinensis*)基因组中的乙酰胆碱酯酶(Ace)基因进行鉴定和生物信息学分析。【方法】通过同源性搜索得到1条乙酰胆碱酯酶(Ace)基因序列,命名为*AsAce2*,分析了该基因的基因结构以及所编码蛋白的理化性质、亚细胞定位、二级结构、三级结构和系统发生关系。【结果】*AsAce2*基因全长为4 008 bp,编码区序列长度为1 941 bp,共编码647个氨基酸;*AsAce2*蛋白与其他蚊虫的Ace2蛋白具有较高的同源性,在系统发育关系上与冈比亚按蚊(*Anopheles stephensi*)的Ace2蛋白和不吉按蚊(*Anopheles gambiae*)的Ace2蛋白关系更近;*AsAce2*基因结构中含有3个1位和2个2位内含子;*AsAce2*蛋白分子式为C₃₂₂₂H₄₉₂₅N₈₇₅O₉₄₇S₃₂,为亲水性蛋白,蛋白定位在细胞周质,在13~32位氨基酸为跨膜区,无疏水区,信号肽,二级结构主要以不规则卷曲为主。【结论】丰富了相关基础数据,为进一步研究该基因的功能奠定了基础。

关键词:中华按蚊;*AsAce2*基因;鉴定;生物信息学

中图分类号:Q966

文献标志码:A

文章编号:1672-6693(2017)01-0023-07

按蚊属(*Anopheles*)是疟疾、淋巴丝虫病以及少数虫媒病毒的传播媒介,是按蚊亚科(*Anophelinae*)的第一大属^[1];中华按蚊(*Anopheles sinensis*)作为中国最主要的传疟媒介之一,在中国、越南、阿富汗、马来西亚、尼泊尔、泰国、日本以及印度尼西亚等地均有分布^[2-4]。近些年来,疟疾得到较好的控制,但在2015年,全球仍有2.14亿人感染此病,43.8万人因此病而死亡^[5]。因此对疟疾的控制仍然是人类面临的巨大挑战。

乙酰胆碱酯酶(Ace)是神经突触传递过程中一种重要的酶,它能促进神经元发育和神经再生以及参与细胞的发育和成熟。该酶的主要功能是降解乙酰胆碱从而终止神经递质对突触的兴奋作用,是维持体内胆碱冲动的一种重要的丝氨酸水解酶^[6]。该酶的结构中包含有由丝氨酸、谷氨酸和组氨酸组成的催化三联体,二硫键连接的4~6个半胱氨酸以及“FGESAG”特征基序^[7]。在昆虫中,Ace多被认为与靶标抗性相关,但是通常是由*Acel*基因编码突触Ace。然而近些年来,Kim和Lee等人研究发现,在100个昆虫中有33个是由*Ace2*基因编码的催化酶^[8]。例如,黑腹果蝇(*Drosophila melanogaster*)^[9]、油橄榄果实蝇(*Bactrocera oleae*)^[10]、家蝇(*Musca domestica*)^[11]等昆虫中只存在*Ace2*基因,且这个基因在这几个物种中被报道与有机磷(OP)和氨基甲酸酯(CB)相关。但目前对*Ace2*基因的研究仍然只有较少报道。随着中华按蚊基因组数据库的完成,对*Ace2*基因的详细分析成为可能。

本研究基于中华按蚊基因组数据,通过同源性搜索得到了1条*Ace2*基因序列,通过生物信息学方法预测并分析了该序列的结构特征、碱基组成以及所编码蛋白的理化特性、二级结构、三级结构、同源性等,并以黑腹果蝇*Ace2*基因的氨基酸序列为外群,通过邻接法构建了系统发育树。本研究中对中华按蚊*Ace2*基因的鉴定以及生物信息学分析,为进一步研究该基因的生物学功能提供了理论数据。

1 材料与方法

1.1 数据来源

中华按蚊基因组数据来自重庆师范大学昆虫与分子生物学研究所,不吉按蚊(*Anopheles funestus*)、微小按

^{*} 收稿日期:2016-03-28 修回日期:2016-11-16 网络出版时间:2017-01-12 11:34

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

第一作者简介:罗钱春,女,研究方向为昆虫分子生物学, E-mail:1243167273@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.20170112.1134.022.html>

蚊(*Anopheles minimus*)、致倦库蚊(*Culex quinquefasciatus*)、冈比亚按蚊(*Anopheles gambiae*)、家蚕(*Bombyx mori*)、意大利蜜蜂(*Apis mellifer*)、黑腹果蝇等物种的 *Ace1* 和 *Ace2* 基因序列来自 NCBI 的 GenBank(<http://www.ncbi.nlm.nih.gov/>)数据库。

1.2 序列的鉴定

在 GenBank 数据库中下载冈比亚按蚊 *Ace2* 蛋白的氨基酸序列,通过 TBLASTN 对中华按蚊基因组数据库进行搜索,其中 *E*-值小于或等于 1×10^{-5} 。将搜索得到的序列通过 FGENSESH^[12](<http://linux1.softberry.com/berry.phtml>)进行预测。将预测得到的氨基酸和编码区序列在 NCBI 的 nr 库中通过 BLASTP 比对,完成对 *Ace2* 的验证。

1.3 序列分析

利用 Blast 软件(<http://www.ncbi.nlm.nih.gov/BLAST/>)进行同源性搜索;利用 MEGA 5.0^[13] 软件对序列的碱基构成、密码子的使用频率以及序列进行多重序列比对分析;蛋白质的分子质量、等电点等理化特性通过 ProtParam 软件(<http://web.expasy.org/protparam/>)进行预测;中华按蚊的 *Ace2* 蛋白含有的疏水性、跨膜区、信号肽、二级结构以及其亚细胞定位分别通过 ProtScale 软件(<http://www.expasy.org/cgi-bin/protscale.pl>)、TMHMM 软件(<http://www.cbs.dtu.dk/services/TMHMM/>)、SignalP4.1 软件(<http://www.cbs.dtu.dk/services/SignalP/>)、SSPro4.0 软件(<http://scratch.proteomics.ics.uci.edu/>)和 TargetP 软件^[14](<http://www.cbs.dtu.dk/services/TargetP/>)进行预测分析;将中华按蚊的 *Ace2* 在 Protein Data Bank(PDB)数据库中进行 PSI-BLAST 搜索,找到与之相似性较高的蛋白,利用 SWISS-MODEL(<http://swissmodel.expasy.org/>)进行同源建模和三级结构预测^[14]。

1.4 系统发育分析

将中华按蚊基因组中的 *Ace2* 蛋白序列与本文 1.1 部分提及的其他物种的 *Ace2* 序列通过 MEGA5 软件^[13] 比对并构建系统发育树,并以黑腹果蝇 *Ace2* 为外群,采用邻接法(Neighbor-joining method, NJ)进行构树,将重复次数设为 1 000。

2 结果与分析

2.1 序列分析

在中华按蚊基因组数据库中通过同源性比对搜索到 1 条与冈比亚按蚊相似性最高的序列(一致性为 90.27%),长度为 4 008 bp,其中编码区长为 1 941 bp,共编码 647 个氨基酸,且该编码区的起始密码子和终止密码子分别为 ATG 和 TAA,确定它为一条全长的编码区序列(图 1)。并将该基因命名为 *AsAce2*。

```

1  ATGACTAGCAACCACTCCGCTCTGCCGCGATCGTGCACCAACGCTGCTGGTCTGCTGCTCGGTGCGACCGCTGATATGT
1  M T S N H S V C R R S C A N V L V L L G A T V I C
79  CCGGCGTACGGCATCATCGACCGCTGGTGGTGCAGACCGAGTGGACCAATACCGCGCGCTCGACCATGGTCCAG
27  P A Y G I I D R L V V Q T S S G P I R G R S T M V Q
157  GGGCGCGAGGTGCATGTCTTCAATGGGGTCCGTTCCGCAAAACCGCGGTCGACGGGTGCGGTTCAAGAACCGGTC
53  G R E V H V F N G V P F A K P P V D G L R F K K P V
235  CCCGCCGAGCCGTGGCAGGAGTTCTCGATGCAACCAAGATTACCTCCATCGTGCAATTCAGGAACGGTATGAGTATTTT
79  P A E P W H G V L D A T R L P P S C I Q E R Y E Y F
313  CCCGCTTCGCTGGCGAAGATGTGGAATCCCAATACCAACGTTTCTGAAGATTGTCTGTATCTGAACATCTGGGTG
105  P G F A G E E M W N P N T N V S E D C L Y L N I W V
391  CCAACCAAACTCGTCTCCGACATGGCAGGGTCTGAACCTTGGCAGCAACGACTACTTCCAGGATGACGAGGACTTT
131  P T K T R L R H G R G L N F G S N D Y F Q D D E D F
469  CAGCGCCAGCAACCAATCGAAGGGTGGTCTGGCGATGCTAGTGTGGATCTACGGGGTGGCTTCATGAGTGGCACCTCC
157  Q R Q H Q S K G G L A M L V W I Y G G G F M S G T S
547  ACGCTGGACATCTACAACCGGAGATACTGGCGCGCTCGTAACTCGTACGCTCGATGCGATGCGCTGGCGG
183  T L D I Y N A E I L A A V G N V I V A S M Q Y R V G
625  GCGTTCGCTTCTCTACCTGGCACCCTACATCAACGGCTACGAGGAGGACCTCCGGGAACATGGGTTTGGGAT
209  A F G F L Y L A P Y I N G Y E E D A P G N M G L T P R D K
703  CAGGCGTAGCGATACGGTGGCTGAAGGATAACGCCAAGCGTTCGGTGGCGACCGGATCTCATACGCTGTTCCGGT
235  Q A L A I R W L K D N A K A F G G D P D L I T L F G
781  GAGTCGGCTGGCGTAGCTCGCTCAGTCTGCACTGATGTCGCCCTGACACGAGGACTGTCCGCGCTGGCATTTCTG
261  E S A G G S S V S L H L M S P V T R G L S R R G I L
859  CAGTGGGCGAGCTGAACGCCCGTGGAGCCACATGACCGCGCAAAAGCGTGCACATTGCCAGGGCTAATCGAC
287  Q S G T L N A P W S H M T A E K A L H I A E G L I D
937  GACTCGCACTGTAACCTCAGATGCTGAAGGACTCGCCAGCAGCATGCGAGTGCATGCGGAATGTCGACGCCAAA
313  D C N C N L T M L K D S P S T V M Q C M R N V D A K
1015  ACAATCTCCGTGCGAGTGGAACTCTACTCCGGCATACTGGGATTCCTCGTCCGCGGACCATCGATGGAGTCTTC
339  T I S V Q Q W N S Y S G I L G F P S A P T I D G V F
1093  ATGACCGCGGATCGATGACCATGCTGCGTGGGCGAAGTGGAGGAATTGACATCCTGGTTCGGTAGCAATCCGGAT
365  M T A D P M T M L R E A N L E G I D I L V G S N R D
1171  GAAGGACATACTTCTCTGTACGATTTATCGATTACTTTGAGAAGGATGCGGCCACTCGCTACCCAGGGACAAG
391  E G T Y F L L Y D F I D Y F E K D A T S A L T M P R D K
1249  TTTTGGAGATCATGAACACCATCTACAACAAAGCCTCCGAACCGGAACGCGAAGCCATCATATCCAGTATACGGGT
417  T L L E I M N T I Y N K A S E P E R E A I I C T Y T G
1327  TGGGAGAGTGGAACGACGGCTACCAGAACCAGCAGGAGTGGCGGGCGGTGCGGCGACCACTTCTCATCTGCGCCG
1443  W E S G N D G Y Q N Q H Q V G R A V G D H F F I C P
1405  ACCAACGAGTTCGCCCTCGGCTGACGGAACGCGGCGGTGCTCCACTATTCTTCTACGCGACCGCACCAGTACC
469  T N E F A L G L T E R G A S V H Y Y F F T H R T S T
1483  TCCCTCTGGGCGAATGGATGGCGCTTTTGACGGTGATGAGGTGGAGTACATCTTCGGCCAGCGCATGAACCGCTCG
495  S L W G E W M G V L H G D E V E Y I F G G C P M N A S
1561  CTGCAGTATCGCCAGCGGGAGCGGACCTGAGCCGGCTATGGTCTGTAGTACGCGAGTTCGACGCGGAGGAAAC
521  L Q Y R Q R E R D L S R R M V L S V S E F A R T G N
1639  CCGGCACTCGAGGGGAACTGGCCGCTCTACACGCGGAAATCCGATCTACTTCACTTCAACGCGGAAGGGAG
547  P A L E G E H W P L Y T R E N P I Y F I F N A E G E
1717  GACGATTCGCGCGGAGAGATGAGCCGGTCCGATGGCCACCTCGTGGCCCTTCTGGAACGACTTCTGCGGAGG
573  D D L R G E K Y G R G P M A T S C A F W N D F L P R
1795  TTGCGCGCTGGTCCGTTCCCTCGAAGACGCGTGAACCTGGTGGAGGCATCTCGCAGGCGAGCAGAACCAGGGCG
599  L R A W S V P S K T P C N L V E A I S Q G S R T G A
1873  ACGTGTCTGTACCTGATGCTGGTCTTAGTACTGCTCCCTGCGCAACCGCGCTGTGAACTAA
625  T L L Y L V A M L V L V L S L R Q P A L V N *

```

注:两个方框分别代表该基因的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和氨基酸序号。

图 1 *AsAce2* 基因的编码区序列以及编码的氨基酸序列

Fig. 1 The coding sequence in *AsAce2* gene and its amino acids

对 *AsAce2* 基因编码区的碱基组成进行分析(表 1),结果表明其中总体的 A,T,G,C 含量分别为 20.5%,19.7%,30.4%,29.4%。除了密码子第二位的 AT 含量(56.4%)比 GC 含量(43.6%)略高,总体、密码子第一位以及密码子第三位均表现出 GC 含量高于 AT 含量,尤其是密码子第三位上的 GC 含量达到了 78.5%,远高于 AT 含量(21.5%);由 AT-偏斜和 GC-偏斜可知,总体和密码子第一位上的偏斜均为正,表明在碱基的使用上均偏好使用 A 和 G,而在密码子的第二位和密码子第三位上偏斜均为负,表明在碱基使用上更偏好使用 T 和 C。

表 1 *AsAce2* 基因编码区序列的碱基组成

Tab.1 The base composition of the coding sequence in *AsAce2* gene

<i>AsAce2</i> 的编码区	A/%	T/%	C/%	G/%	AT/%	GC/%	AT-偏斜	GC-偏斜
总体	20.5	19.7	29.4	30.4	40.2	59.8	0.020	0.017
密码子第一位	24.1	18.7	24.0	33.2	42.8	57.2	0.126	0.161
密码子第二位	28.0	28.4	22.3	21.3	56.4	43.6	-0.007	-0.023
密码子第三位	9.4	12.1	41.9	36.6	21.5	78.5	-0.126	-0.068

通过 MEGA5.0 预测 *AsAce2* 基因密码子的使用频率并计算相对同义密码子使用率(RSCU)。由表 2 可知,在 64 个密码子中 UUC,CUG,CCG,AAC,GAG,GGC 的使用频率较大,均大于 22 次;而除了终止密码子 UAA 仅含有 1 个外,密码子 UGA,UAG,UCA,UCU,CUU,CCU 的使用频率也小于或等于 1 次;密码子 CUG,UCG,CCG 和 ACC 的 RSCU 值均大于 2.00,而 CUU,UCA,CCU,UAG,CAA 和 UGA 的此值均小于 0.20。这些结果表明在 *AsAce2* 基因编码过程中偏好使用以 G 或 C 结尾的密码子,而较少使用以 A 或 T 结尾的密码子,这与编码区序列碱基组成中密码子第三位上具有较高的 GC 含量结果一致。

表 2 *AsAce2* 基因编码区序列的相对同义密码子统计表

Tab.2 Relative synonymous codon usage of the coding sequence in *AsAce2* gene

密码子	数量/个	RSCU	密码子	数量/个	RSCU	密码子	数量/个	RSCU	密码子	数量/个	RSCU
UUU	5	0.32	UCU	1	0.16	UAU	8	0.62	UGU	3	0.55
UUC	26	1.68	UCC	11	1.76	UAC	18	1.38	UGC	8	1.45
UUA	2	0.20	UCA	1	0.16	UAA	1	2.00	UGA	0	0.00
UUG	7	0.69	UCG	16	2.56	UAG	0	0.00	UGG	14	2.00
CUU	1	0.10	CCU	1	0.12	CAU	2	0.29	CGU	4	0.48
CUC	12	1.18	CCC	7	0.82	CAC	12	1.71	CGC	15	1.82
CUA	4	0.39	CCA	3	0.35	CAA	2	0.18	CGA	2	0.24
CUG	35	3.44	CCG	23	2.71	CAG	20	1.82	CGG	12	1.45
AUU	4	0.32	ACU	2	0.23	AAU	6	0.36	AGU	5	0.80
AUC	21	1.68	ACC	19	2.17	AAC	27	1.64	AGC	10	1.60
AUA	6	0.44	ACA	2	0.23	AAA	5	0.67	AGA	3	0.48
AUG	21	1.56	ACG	12	1.37	AAG	10	1.33	AGG	3	0.48
GUU	4	0.40	GCU	3	0.26	GAU	13	0.79	GGU	16	1.08
GUC	18	1.80	GCC	20	1.74	GAC	20	1.21	GGC	27	1.83
GUA	3	0.30	GCA	4	0.35	GAA	14	0.76	GGA	9	0.61
GUG	15	1.50	GCG	19	1.65	GAG	23	1.24	GGG	7	0.47

2.2 同源性分析

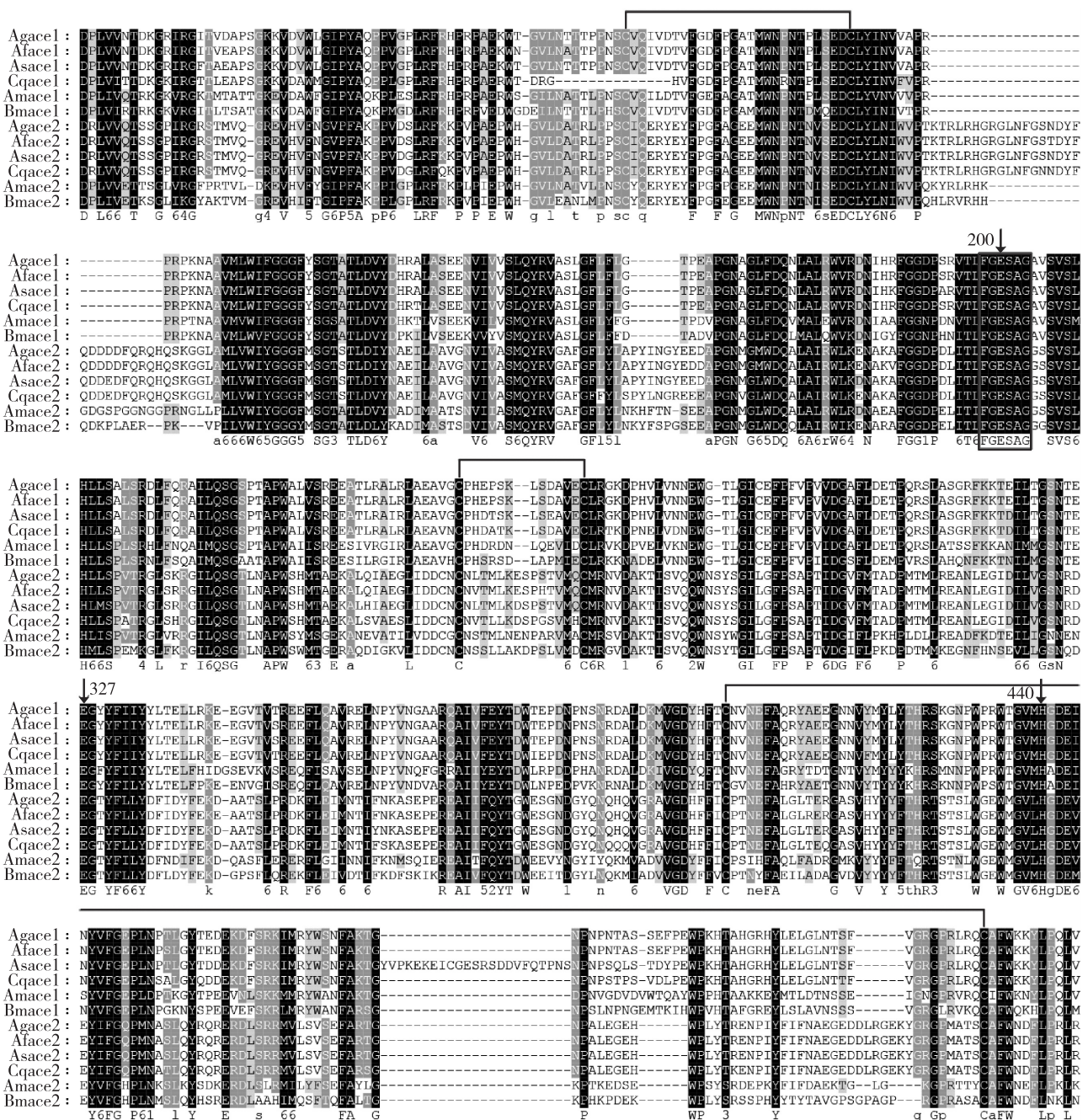
用 *AsAce2* 蛋白氨基酸序列为查询序列在 NCBI 进行在线 BLASTP,搜索结果表明,该蛋白与三带喙库蚊(*Culex tritaeniorhynchus*)的 Ace2 蛋白一致率最高,高达 92%;其次为达氏按蚊(*Anopheles darlingi*)、致倦库蚊、尖音库蚊(*Culex pipiens*)和白纹伊蚊(*Aedes albopictus*)。同时,该蛋白与达氏按蚊的 Ace2 蛋白相似率最高,高达 98%;其次是白纹伊蚊,致倦库蚊,尖音库蚊和三带喙库蚊(表 3)。从 NCBI 数据库中下载具有代表性的 Ace2 蛋白序列以及这些物种的 Ace1 蛋白序列,并与 *AsAce2* 蛋白进行比对分析,图 2 显示 *AsAce2* 蛋白中含有

构成 Ace 的特殊结构,包括催化三联体(由 S200, E327, H440 等 3 个氨基酸残基组成),6 个半胱氨酸(由 C216-C244, C403-C416, C551-C695 等 3 个保守的二硫键连接)和特征基序“FGESAG”^[7](位于活性丝氨酸残基周围)。

表 3 AsAce2 蛋白与其他蚊虫 Ace2 蛋白氨基酸序列的一致率和相似率

Tab. 3 The identity and similarity between AsAce2 and its homologous Ace2 from other mosquito species

基因名称	物种	基因登录号	一致率/%	相似率/%
AsAce2	中华按蚊 <i>A. sinensis</i>	—	—	—
AdAce2	达氏按蚊 <i>A. darlingi</i>	ETN67804	90	98
CqAce2	致倦库蚊 <i>C. quinquefasciatus</i>	XP_001842227	90	96
CpAce2	尖音库蚊 <i>C. pipiens</i>	CAJ43752	90	96
CtAce2	三带喙库蚊 <i>C. tritaeniorhynchus</i>	BAD06209	92	94
AaAce2	白蚊伊蚊 <i>A. albopictus</i>	BAE71347	88	97



注:3 个二硫键连接 6 个半胱氨酸(框线);短线代表“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%。Ag、Af、As、Cq、Am 和 Bm 分别表示冈比亚按蚊、不吉按蚊、中华按蚊、致倦库蚊、意大利蜜蜂和家蚕,下同。

图 2 AsAce2 蛋白与其他物种 Ace1 及 Ace2 蛋白氨基酸序列比对

Fig.2 Alignment of AsAce2 with other species Ace1 and Ace2 protein amino acid sequences

2.3 系统发育分析

通过同源性比对,选取包括中华按蚊等 7 个物种的 Ace2 序列和 Ace1 序列,以黑腹果蝇的 Ace2 为外群,用 NJ 法构建系统发育树。结果表明,中华按蚊 AsAce2 在进化关系上与冈比亚按蚊 Ace2 和不吉按蚊 Ace2 更近,Bootstrap 值为 99%(图 3)。

2.4 基因结构分析

对 *AsAce2* 基因运用 GSDS2.0 软件进行分析,结果表明该基因不含 0 位内含子,有 3 个 1 位内含子、2 个 2 位内含子及 6 个外显子(图 4)。

2.5 蛋白质的特性分析

2.5.1 理化性质分析 利用 ProtParam 软件分析 AsAce2 蛋白的理化性质, 预测结果显示: 该蛋白分子式为 $C_{3222}H_{4925}N_{875}O_{947}S_{32}$, 相对分子质量为 72.10 kD, 理论等电点为 5.41, 负电荷残基总数(Asn+Glu)、正电荷残基总数(Arg+Lys)分别为 70, 54 个, 脂溶指数为 -0.192; 此外, AsAce2 蛋白不稳定系数为 37.05, 表明该蛋白较稳定。AsAce2 基因编码的氨基酸中, 编码 Leu 的数量最多, 达到 61 次, 占总编码氨基酸次数的 9.43%; 编码 Cys 的数量最低, 仅有 11 次, 占总体的 1.70%(图 4)。

2.5.2 疏水区分析 对 AsAce2 蛋白运用在线 ProScale 软件进行疏水区分析,结果表明该蛋白中无疏水区(图 6)。

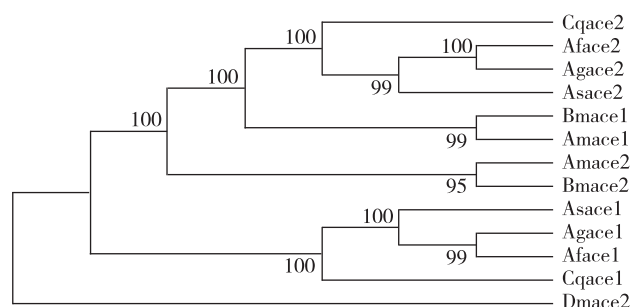
2.5.3 亚细胞定位 对 AsAce2 蛋白运用在线 TargetP 软件进行亚细胞定位分析,结果显示分泌通路蛋白分值最高,为 0.909,表明该蛋白的定位最可能在细胞周质。

2.5.4 跨膜区分析 对 AsAce2 蛋白运用 TMHMM-2.0 软件进行跨膜结构域分析,结果表明该蛋白在 13~32 位氨基酸之间有 1 个跨膜区(封二彩图 7)。

2.5.5 信号肽预测 对 AsAce2 蛋白运用 SingalP 软件进行信号肽分析,结果表明此蛋白的 C 值为 0.437,Y 值为 0.424,S 值为 0.736,无信号肽切割位点,即该蛋白无信号肽(封二彩图 8)。

2.5.6 二级结构预测 对 AsAce2 蛋白运用 SSPro4.0 软件进行二级结构分析,结果表明 AsAce2 蛋白二级结构中的无规卷曲、延伸链和 α -螺旋分别占 54.49%, 23.07% 和 22.45%,表明该蛋白的二级结构中以无规卷曲为主,其次是延伸链,而 α -螺旋最少(图 9)。

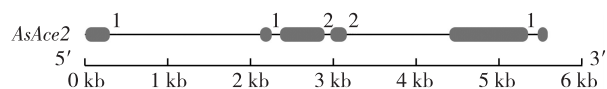
2.5.7 三级结构预测 经 PSI-BLAST 搜索表明,黑腹果



注:大于 50 的 Bootstrap 值标记在树的分支节点上。

图 3 7 个物种 Ace1 和 Ace2 蛋白的系统进化树

Fig. 3 The phylogenetic tree of the Ace1 and Ace2 in seven species



注:粗线表示外显子,细线表示内含子,粗细线上文数字表示内含子相位。

图 4 *AsAce2* 基因的结构分析

Fig. 4 The gene structure analysis of *AsAce2*

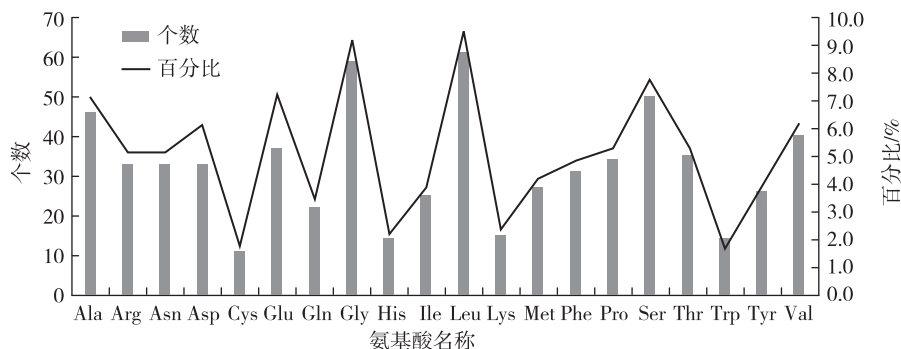


图 5 AsAce2 蛋白氨基酸组成及百分比

Fig. 5 The composition and percentage of amino acids of AsAce2

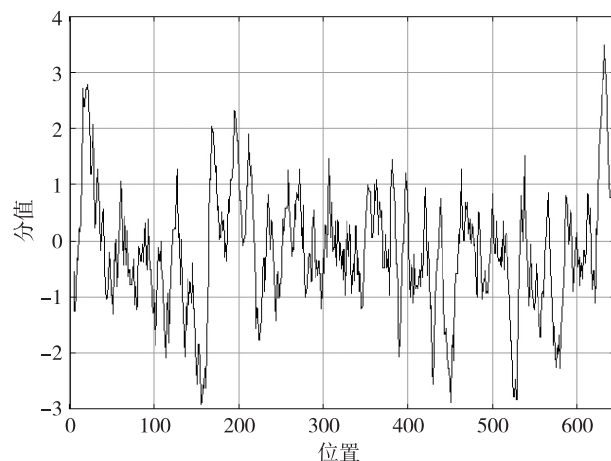


图 6 AsAce2 蛋白的疏水区预测

Fig. 6 Hydrophobic region prediction of the AsAce2

蝇的 Ace 与 AsAce2 蛋白的氨基酸相似性最高,为 70.16%。对 AsAce2 蛋白通过 SWISS-MODEL 软件进行三级结构分析,其中将黑腹果蝇 Ace2 序列选为同源建模模板,得到 AsAce2 蛋白的三级结构模型(封二彩图 10)。

3 讨论

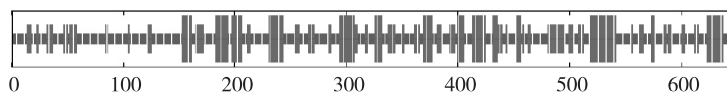
在昆虫中,Ace 作为调节神经冲动的一类关键酶,与 OP 和 CB 杀虫剂的靶标抗性密不可分^[15]。在已有报道中,对 Ace1 基因的靶标抗性研究较多^[16-17];但近些年来也有报道提出,Ace2 基因可能也与 OP 和 CB 杀虫剂抗性相关^[9,18],不过仍需要进一步研究才能证明。随着中华按蚊基因组测序的完成,为 AsAce2 基因的分析提供了条件。

本研究通过同源性搜索在中华按蚊基因组数据中得到了一条 AsAce2 基因序列,对该序列的结构及 AsAce2 蛋白与其他物种的多重序列进行了分析,并构建了系统关系树。AsAce2 的基因结构中只含有 1,2 位内含子,AsAce2 蛋白具有由 S200,E327,H440 等 3 种氨基酸残基组成的催化三联体、二硫键连接的 6 个半胱氨酸及其特征基序“FGESAG”,说明它是一种具有功能的活性丝氨酸水解酶。系统关系树的分析表明,中华按蚊 AsAce2 在系统发育关系上与冈比亚按蚊和不吉按蚊的 AsAce2 关系更近。同源性分析表明,AsAce2 与其他蚊虫的 Ace2 具有较高的同源性,与达氏按蚊、致倦库蚊、尖音库蚊、三带喙库蚊和白蚊伊蚊的一致率和相似率均高于 88%。本研究还对 AsAce2 蛋白的特性进行了分析,包括理化性质、疏水区、亚细胞定位、跨膜区、信号肽、二级结构以及三级结构。结果表明该蛋白是一种稳定蛋白,不含信号肽和疏水区,在 13~32 位氨基酸之间有 1 个跨膜区,二级结构主要以无规卷曲为主。

本研究报道的 AsAce2 基因序列进一步丰富了 Ace 的基础数据,通过和其他物种的比较,也得出了一些较好的结论,这为后续对该基因的进一步研究奠定了生物信息学基础。

参考文献:

- [1] 陆宝麟. 中国动物志,昆虫纲,第九卷,双翅目:蚊科(下)[M]. 北京:科学出版社,1997.
- [2] LU B L. Fauna Sinica, Insecta, Vol. 9, Diptera: Culicidae (2nd)[M]. Beijing: Science Press, 1997.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] World Health Organization. Disease burden[R]. Geneva: WHO, 2015.
- [7] TOUTANT J P. Insect acetylcholinesterase: catalytic properties, tissue distribution and molecular forms[J]. Progress in Neurobiology, 1989, 32(5): 423-446.
- [8] 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.
- [9] KIM Y H, LEE S H. Which acetylcholinesterase functions as the main catalytic enzyme in the Class Insecta[J]. Insect Biochemistry and Molecular Biology, 2013, 43(1): 47-53.
- [10] HALL L M, SPIERER P. The Ace locus of *Drosophila melanogaster*: structural gene for acetylcholinesterase with an unusual 50 leader[J]. EMBO J, 1986, 5(11): 2949-2954.
- [11] VONTAS J G, HEJAZI M J, HAWKES N J, et al. Resistance associated point mutations of organophosphate insensitive acetylcholinesterase, in the olive fruit fly *Bactrocera oleae* [J]. Insect Mol Biol, 2002, 11(4): 329-336.
- [12] 黄瑶, 乔传令, Williamson M S, 等. 杀虫药剂抗性家蝇品系乙酰胆碱酯酶基因的特征分析[J]. 生物工程学报, 1997, 13(3): 258-263.
- [13] HUANG Y, QIAO C, WILLIAMSON M S, et al. Characterization of acetylcholinesterase gene from insecticide-resistant houseflies *Musca domestica* [J]. Chin J Biotech, 1997, 13(3): 258-263.
- [14] SOLOVYEV V, KOSAREV P, SELEDISOV I, et al. Automatic annotation of eukaryotic genes, pseudogenes and promoters[J]. Genome Biol, 2006, 7(Suppl1): 1-12.
- [15] TAMURA K, PETERSON D, PETERSON N, et al.



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 9 AsAce2 蛋白二级结构预测

Fig. 9 Secondary structure prediction of the AsAce2

- MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods[J]. *Molecular Biology and Evolution*, 2011, 28(10): 2731-2739.
- [14] 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.
- [15] 林建国, 张传溪, 唐振华. 与昆虫抗药性相关的乙酰胆碱酯酶基因突变研究进展[J]. *农药学学报*, 2005, 7(1): 1-6.
LIN J G, ZHANG C X, TANG Z H. Gene mutation of acetylcholinesterase associated with insect resistance research progress[J]. *Chin J Pestic Sci*, 2005, 7(1): 1-6.
- [16] GAO J R, KAMBHAMPATI S, ZHU K Y. Molecular cloning and characterization of a greenbug (*Schizaphis graminum*) cDNA encoding acetylcholinesterase possibly evolved from a duplicate gene lineage[J]. *Insect Biochem Mol Biol*, 2002, 32(7): 765-775.
- [17] ANAZAWA Y, TOMITA T, AIKI Y, et al. Sequence of a cDNA encoding acetylcholinesterase from susceptible and resistant two-spotted spider mite, *Tetranychus urticae* [J]. *Insect Biochem Mol Biol*, 2003, 33(5): 509-514.
- [18] FOURNIER D, KARCH F, BRIDE J M, et al. *Drosophila melanogaster* acetylcholinesterase gene: structure, evolution and mutations [J]. *Journal of Molecular Biology*, 1989, 210(1): 15-22.

Animal Sciences

Identification and Bioinformatic Analysis of the AsAce2 Gene in *Anopheles sinensis*

LUO Qianchun, WU Xuemei, YU Hairong, CHEN Bin

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

Abstract: [Purposes] The acetylcholinesterase gene sequence was identified based on the *Anopheles sinensis* genome database and bioinformatic analysis. [Methods] The acetylcholinesterase gene sequence was identified by homology search and the sequence is designated as AsAce2. Furthermore, we analyzed the gene structure, biochemical properties, subcellular localization, secondary structure, tertiary structure and phylogenetic relationships. [Findings] The full-length sequence is 4 008 bp with a 1 941 bp coding sequence encoding 647 amino acids. The AsAce2 protein sequence was highly homologous with the Ace2 protein sequence in other mosquito species. The AsAce2 has a closer phylogenetic relationship with *Anopheles gambiae* and *Anopheles funestus*. Gene structure analysis showed that this gene contains phase 1 (three number) and phase 2 (two number) intron, the protein molecular formula was $C_{3222}H_{4925}N_{875}O_{947}S_{32}$, belonging to hydrophilic protein. AsAce2 was localized in cytoplasm. The transmembrane domain was localized between 13~32 amino acids. Secondary structure showed it has no hydrophobic region and signal peptide, and there is mainly random coils. [Conclusions] The bioinformatic analysis of AsAce2 study which enriched the basic data and laid a foundation for further study on its function.

Keywords: *Anopheles sinensis*; AsAce2 gene; identification; bioinformatic

(责任编辑 方 兴)