

昆虫化学感受蛋白基因的进化分析

柳晓磊 翁群芳 胡美英 李亚楠 易欣

华南农业大学昆虫毒理研究室/天然农药与化学生物学教育部重点实验室, 广州 510642

摘要 从基因库中调取已完成测序的昆虫化学感受蛋白(chemosensory proteins, CSPs)基因序列和氨基酸序列,共涉及昆虫7个目23个种的59个基因。应用MEGA 5.0形成了这59个基因的系统发育关系树,并计算CSPs基因在昆虫中的核苷酸替换速率。对非同义替换速率和同义替换速率进行比较的结果显示,CSPs基因在昆虫中总体上受到负选择作用。用PAML软件位点一特异模型检测到6个受到正选择作用的位点,其中5个位于功能重要的区域。

关键词 昆虫; 化学感受蛋白; 基因; 进化
中图分类号 Q 963; Q 966 **文献标识码** A **文章编号** 1000-2421(2011)02-0177-05

昆虫化学感受蛋白(chemosensory proteins, CSPs)或称为感受器官蛋白(sensory appendage proteins, SAPs),是存在于化学感受器中一类主要的可溶性结合蛋白,其相对分子质量较小,平均仅约13 ku,多肽链全长约100~110个氨基酸,有4个保守的半胱氨酸位点,形成2对二硫键、6个 α -螺旋以及蛋白中心的1个亲水性孔道^[1-3]。CSPs广泛存在于昆虫不同器官中,不仅在感受器官中(触角、喙、下颚须和跗节)分布^[4],而且在身体其他部分的感受器中也有CSPs的表达^[5]。近年来,关于昆虫CSPs的功能研究发现,CSPs不仅可以溶解运输不同化学感受器亲脂性配体^[3,6],而且具有部分嗅觉功能、调节生长发育、参与体内免疫及调节生理节律等多种生理功能^[4,7-9]。这些均反映了昆虫在长期进化过程中对大量不同类型环境化学分子刺激的适应和进化^[10]。可见,CSPs在昆虫中可能具有复杂的进化过程。为了弄清CSPs基因在昆虫中进化的详细情况,笔者对目前已知不同种CSPs基因的有关信息进行较全面的系统分析,旨在为其他物种的相关研究提供理论依据。

7个目23个物种的59个化学感受蛋白基因和蛋白序列(前面为基因名称,括号内为注册号)。

烟芽夜蛾(*Heliothis virescens*) CSPs: *HvirCSP1*(AY101512), *HvirCSP2*(AY101511), *HvirCSP3*(AY101513); 甘蓝夜蛾(*Mamestra brassicae*) CSPs: *MbraCSPA1*(AF211177), *MbraCSPA2*(AF211178), *MbraCSPA4*(AF211180), *MbraCSPA5*(AF211181), *MbraCSPA6*(AF255918), *MbraCSPB1*(AF211182), *MbraCSPB2*(AF211183), *MbraCSPB3*(AF255919), *MbraCSPB4*(AF255920); 烟草天蛾(*Manduca sexta*) CSPs: *MsexSAP1*(AF117574), *MsexSAP2*(AF117592), *MsexSAP3*(AF117585), *MsexSAP4*(AF117599), *MsexSAP5*(AF117594); 粉纹夜蛾(*Trichoplusia ni*) CSPs: *TniCSP*(AY456191); 云杉色卷蛾(*Choristoneura fumiferana*) CSPs: *CfumCSP1*(AY426538), *CfumCSP2*(AY426539); 家蚕(*Bombyx mori*) CSPs: *BmorCSP1*(AF509239), *BmorCSP2*(AF509238); 小地老虎(*Agrotis ipsilon*) CSPs: *AipsCSP*(AY301978); 仙人掌螟(*Cactoblastis cactorum*) CSPs: *CcacCLP1*(U95046); 棉铃虫(*Helicoverpa armigera*) CSPs: *HarmCSP*(AF368375); 美洲棉铃虫(*Helicoverpa zea*) CSPs: *HzeaCSP*(AF448448); 黑腹果蝇(*Drosoph-*

ila melanogaster) CSPs; *DmelOS-D* (U02546); 飞蝗 (*Locusta migratoria*) CSPs; *LmigCSPI-1* (AY149648), *LmigCSPI-5* (AY149652), *LmigCSPII-10* (AY149658), *LmigOS-D1* (AJ251075), *LmigOS-D2* (AJ251076), *LmigOS-D3* (AJ251077), *LmigOS-D4* (AJ251078), *LmigOS-D5* (AJ251079); 沙漠蝗 (*Schistocerca gregaria*) CSPs; *SgreOS-D* (AY301979), *SgreCSP1* (AF070961), *SgreCSP2* (AF070962), *SgreCSP3* (AF070963), *SgreCSP4* (AF070964), *SgreCSP5* (AF070965); 竹节虫 (*Eurycantha calcarata*) CSPs; *EcalCSP1* (AF139196), *EcalCSP2* (AF139197), *EcalCSP3* (AF139198); 欧洲马蜂 (*Polistes dominulus*) CSPs; *PdomCSP1* (AY297027); 美洲大蠊 (*Periplaneta americana*) CSPs; *PameP10* (AF030340); 阿根廷蚂蚁 (*Linepithema humile*) CSPs; *LhumCSP* (AY129672); 冈比亚按蚊 (*Anopheles gambiae*) CSPs; *AgamIR7* (AJ697730), *AgamSAP1* (AJ697731); 意大利蜜蜂 (*Apis mellifera*) CSP; *AmelASP3C* (AF481963); 赤拟谷盗 (*Tribolium castaneum*) CSPs; *TcasCSP1* (DQ855488)。

小菜蛾 (*Plutella xylostella*) CSPs; *PxylCSP1* (EF186791), *PxylCSP2* (EF186792), *PxylCSP3* (EF202828), *PxylCSP4* (EF202829) 基因由笔者所在实验室克隆得到, 甜菜夜蛾 (*Spodoptera exigua*) CSPs; *SexiCSP1* (EF186793), *SexiCSP2* (EF186794), *SexiCSP3* (EF186795) 基因由笔者所在实验室克隆得到, 斜纹夜蛾 (*Spodoptera litura*) CSPs; *CSPslit* (DQ007458) 基因由笔者所在实验室克隆得到。

1.2 研究方法

用 Clustal X 2.0 程序^[11]对 CSPs 基因编码序列进行多重序列比对。将比对结果去除起始和终止密码子后进行下一步分析。使用整合在 MEGA 5.0^[12]中的不平衡指数 (disparity index) 分析^[13]对比对结果进行核苷酸替换同质性检测 (重复 8 000 次的 Monte-Carlo 检测)。最适核苷酸替换模型通过 Modeltest 3.7^[14]对多个模型进行比较获得。根据选定的最适替换模型, 使用 DAMBE V. 4. 5. 2^[15]对核苷酸替换 (转换/颠换) 的饱和性 (substitution saturation) 进行检测; 使用 MEGA 5.0 构建 NJ 树, 并进行 1 000 次 Bootstrap 检验。用 PAML 4. 4^[16]

来分析 CSPs 基因在进化过程中选择压力的变化。分析使用 3 种模型: I. 支一特异 (branch-specific) 模型^[17]。它允许非同义替换和同义替换的比值 ω 在不同支系上有变化, 其中单一比率 (one ratio) 模型假定所有进化支的 ω 值都是一致的, 而自由比率 (free ratio) 模型假定各支的 ω 值各不相同。比较这 2 个模型的 2 倍似然率 (likelihood ratio test, LRT) 对数差的 χ^2 分布可以检验 ω 值在不同支上是否有变化。自由度等于这 2 个模型所包含参数数目的差。II. 位点一特异 (site-specific) 模型^[18]。它假定各进化支有固定的 ω 值, 但不同位点之间可以有变化。采用的 2 对成对比较模型分别是 M1a 和 M2a 以及 M8a 和 M8^[19]。III. 支一位点一特异 (branch-site-specific) 模型。它将进化树上的支系分成前景支 (foreground) 和后景支 (background) 两类, 允许 ω 在两类支系上有变化且前景支 ω 大于后景支 ω , 再假定在每一类支系上有固定的 ω 值, 但不同位点之间可以有变化^[20]。

2 结果与分析

2.1 氨基酸序列分析

昆虫 59 个 CSPs 的氨基酸序列比对结果显示, 对于 CSPs 结构维持必需的 4 个半胱氨酸残基 (Cys 29, 36, 55, 58) 在几乎所有序列中高度保守, 仅有 PdomCSP1、LhumCSP 和 MsexSAP1 中第 Cys36 被 Val36、Met36 和 Val36 分别代替; 所有序列均保持了 “---YTD-D---IL-N-R---Y---C---GKC---KELK---C---C---Q---G---W---L---DP---” 的类似结构模式。

2.2 核苷酸替换模型的选择

对 59 条核苷酸序列进行不平衡指标分析, 结果显示所有序列在替换方式上不存在显著差异。如果出现替换方式上存在明显差异的序列, 在构建进化树及其分析时应该予以考虑, 并通过比较在删除和保留该序列的情况下所得到的结果; 如果出现不同的序列, 则在分析中采用剔除该序列后所获得的结果。在使用 Modeltest 3.7 选择最适替换模型时, 阶层式似然率检测 (hierarchical likelihood ratio tests, hLRTs) 和 Akaike 信息量准则 (Akaike information criterion, AIC) 所得到的结果不一致。由于 CSP 基因序列较短, 为了避免更多的参数估计带来的误差, 选用参数较少的模型, 即采用 hLRTs 的结果, 选用 Kimura 二参数 (K80) 为最适模型。使用

DAMBE V. 4. 5. 2 对核苷酸替换(转换/颠换)的饱和性进行检测,替换模型采用 K80,结果表明核苷酸替换并未饱和(图 1),说明获得的基因可靠度高,用这些基因得到的进化树是可靠的。

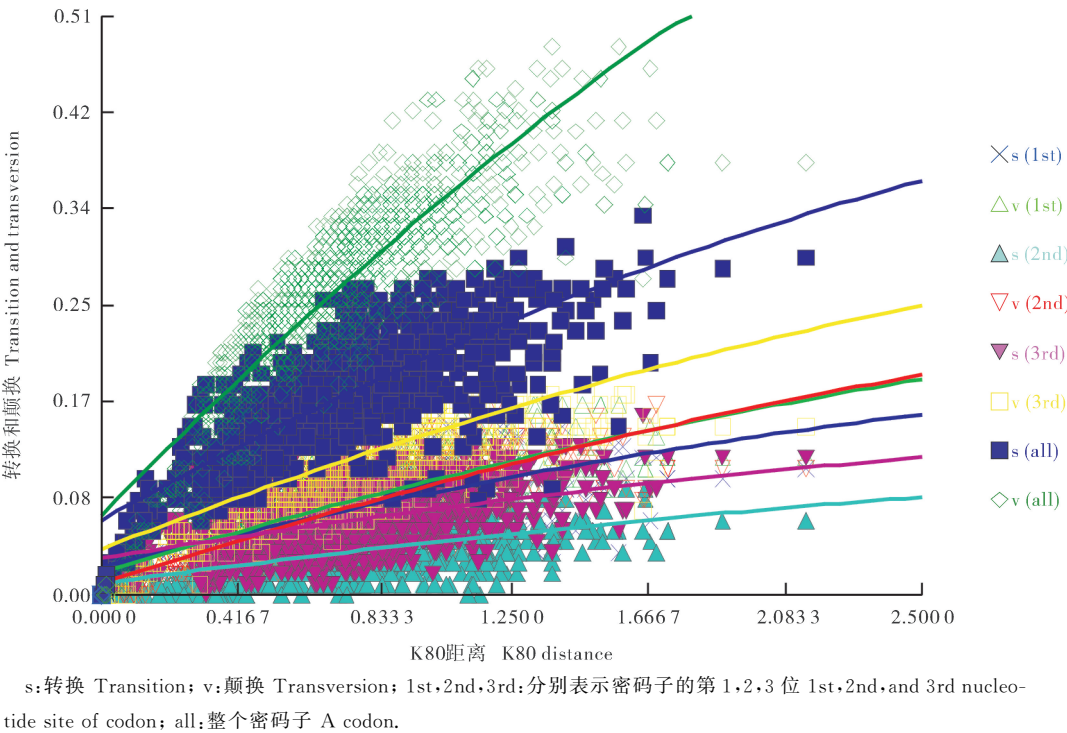


图 1 转换/颠换饱和性分析结果(K80 距离)

Fig. 1 The result of substitution saturation (K80 distance)

2.3 进化分析

利用 MEGA 5.0 构建 NJ 进化树结果表明,目前已知的昆虫 CSPs 大致划分为 3 个不同的群体,即 CSP1、CSP2 和 CSP3 (图 2)。CSP1 和 CSP2 群体主要以鳞翅目昆虫为主,CSP3 群体以直翅目昆虫为主要代表,这一结果与 Picimbon 提出的蛾类 CSPs 遗传进化关系一致^[21]。使用整合在 MEGA 5.0 中的 Pamilo-Bianchi-Li 法^[22]计算所获得昆虫的 CSPs 基因成熟蛋白编码区的每非同义位点的非同义代替数 $K_n=0.503$,每同义位点的同义替换数 $K_s=1.355$, K_n/K_s 显著小于 1 (Z -test, $P<0.05$),说明昆虫 CSPs 基因在进化中总体上体现受到负选择。但是传统的“特定(*ad hoc*)”方法对达尔文正选择的检测过于严格,在长期较强负选择和中性漂变的掩盖下,使得正选择作用难以检测出来;而使用基于最大似然法(the method of maximum likelihood)估计的正选择检测方法使得检测的灵敏性提高,可检测出处于上述一定掩盖下的正选择位点^[23]。

正选择作用可能仅作用于物种进化的某一阶段,即 CSPs 基因在昆虫某些支系的进化过程中可

能呈现插曲式的进化模式。采用支一特异模型进行检测,未能有效检测出受到正选择的支系,说明 ω 值在各支系上的分布并非差异显著。另外,由于正选择可能仅作用于部分位点,用位点一特异模型进行检测,其中 M1 和 M2 以及 M7 和 M8 组模型分别检测出 28 个和 6 个潜在的正选择位点,并且后一组结果包含于前一组的结果之中。为了减少假阳性的问题,采取了 M7 和 M8 一组的结果,氨基酸序列对比分析表明,6 个位点 (11I, 15E, 39D, 73L, 81W 和 94Y) 均位于成熟蛋白区段。

值得注意的是,氨基酸 Glu(E) 和 Asp(D) 是常见的电子受体和供体基团,这些位点的改变常会影响分子间的相互作用,因此位点 15E 和 39D 可能对 CSPs 在不同物种、同种物种的不同部位中的功能产生影响。其中,位点 15E 位于 α_2 螺旋内,39D 位于 α_3 螺旋内,73L 和 81W 位于 α_5 螺旋内,94Y 位于 α_6 螺旋内,这些位于 α 螺旋被认为是 CSPs 活性结构域的关键组成部分,在与 Rhodojaponin III 和 C14-C18 烷基链化合物^[3,24] 结合时起到重要的作用。而通过对昆虫 CSPs 基因序列的比较,发现位

- [J]. Science, 2005, 309(5732): 311-314.
- [7] MALESZKA J, FORET S, SAINT R, et al. RNAi-induced phenotypes suggest a novel role for a chemosensory protein CSP5 in the development of embryonic integument in the honeybee (*Apis mellifera*) [J]. Dev Genes Evol, 2007, 217(3): 189-196.
- [8] SANDOZ J C, GALIZIA C G, MENZEL R, et al. Side-specific olfactory conditioning leads to more specific odor representation between sides but not within sides in the honeybee antennal lobes [J]. Neuroscience, 2003, 120(4): 1137-1148.
- [9] MCDONALD M J, ROSBASH M. Microarray analysis and organization of circadian gene expression *Drosophila* [J]. Cell, 2001, 107(5): 567-578.
- [10] VOGT R G, CALLAHAN F E, ROGERS M E, et al. Odorant binding protein diversity and distribution among the insect orders, as indicated by LAP, an OBP-related protein of the true bug *Lygus lineolaris* (Hemiptera, Heteroptera) [J]. Chem Senses, 1999, 24(5): 481-495.
- [11] THOMPSON J D, GIBSON T J, PLEWNIAC F, et al. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools [J]. Nucleic Acids Res, 1997, 25(24): 4876-4882.
- [12] KUMAR S, DUDLEY J, NEI M, et al. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences [J]. Brief Bioinform, 2008, 9(4): 299-306.
- [13] KUMAR S, GADAGKAR S R. Disparity index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences [J]. Genetics, 2001, 158(3): 1321-1327.
- [14] POSADA D, CRANDALL K A. MODELTEST: testing the model of DNA substitution [J]. Bioinformatics, 1998, 14(9): 817-818.
- [15] XIA X, XIE Z. DAMBE: software package for data analysis in molecular biology and evolution [J]. J Hered, 2001, 92(4): 371-373.
- [16] YANG Z. PAML 4: a program package for phylogenetic analysis by maximum likelihood [J]. Mol Biol Evol, 2007, 24(8): 1586-1591.
- [17] YANG Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution [J]. Mol Biol Evol, 1998, 15(5): 568-573.
- [18] YANG Z, NIELEN R. Synonymous and nonsynonymous rate variation in nuclear genes of mammals [J]. J Mol Evol, 1998, 46(4): 409-418.
- [19] YANG Z, WONG W S, NIELSEN R. Bayes empirical bayes inference of amino acid sites under positive selection [J]. Mol Biol Evol, 2005, 22(4): 1107-1118.
- [20] YANG Z, NIELSEN R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages [J]. Mol Biol Evol, 2002, 19(6): 908-917.
- [21] PICIMBON J F. Biochemistry and evolution of OBP and CSP proteins [M]//BLOMQUIST G J, VOGTRG. Insect pheromone biochemistry and molecular biology. London: Elsevier Academic Press, 2003: 539-566.
- [22] LI W H. Unbiased estimation of the rates of synonymous and nonsynonymous substitution [J]. J Mol Evol, 1993, 36(1): 96-99.
- [23] YANG Z. Inference of selection from multiple species alignments [J]. Curr Opin Genet Dev, 2002, 12(6): 688-694.
- [24] 罗倩. 甜菜夜蛾和小菜蛾化学感受蛋白的研究 [D]. 广州: 华南农业大学资源环境学院, 2008.

Evolution analysis of the insect chemosensory protein gene family

LIU Xiao-lei WENG Qun-fang HU Mei-ying LI Ya-nan YI Xin

Laboratory of Insect Toxicology, South China Agricultural University/Key Laboratory of Natural Pesticide and Chemical Biology, Ministry of Education of China, Guangzhou 510642, China

Abstract The phylogenetic trees of the insect chemosensory proteins (CSPs) gene from 59 sequences of CSPs in 23 species under 7 orders published in GeneBank were generated by MEGA 5.0. Complete comparisons of CSPs sequences were used to determine the rates of synonymous (K_s) and nonsynonymous (K_n) codon substitutions and the ratio of K_n/K_s was calculated to estimate of the selection pressure on CSPs genes in insects. The K_n/K_s ratio was far less than 1, suggesting that CSPs may be evolving under negative selection. By the site-specific model in PAML 4.4, 6 amino acid sites undergoing positive selection were detected, 5 of them were in the important function regions.

Key words insect; chemosensory proteins; gene; evolution