

拟黑多刺蚁肌细胞增强因子 2 (MEF2) 生物信息学分析*

郭新军^{1,2} 奚耕思^{1**}

(1. 陕西师范大学生命科学院 西安 710062; 2. 西安文理学院生命科学系 西安 710065)

Bioinformatics analysis of MEF2 gene of *Polyrhachis vicina*. GUO Xin-Jun^{1,2}, XI Geng-Si^{1**} (1. College of Life Science, Shaanxi Normal University, Xi'an 710062, China; 2. Department of Life Science, Xi'an University of Arts and Science, Xi'an 710065, China)

Abstract The analytical tools NCBI, ExPASy and CBS were adopted to analyze the bioinformatic properties of the PvMEF2 protein, such as sequence characteristics, physicochemical properties, structure and functional domains. The results show that PvMEF2 has conserved MADS and MEF2 domains and that its physicochemical properties, secondary and tertiary structures are similar to those of the MEF2 protein of *Drosophila melanogaster*. PvMEF2 protein may play an important role in myogenesis in *Polyrhachis vicina* Roger.

Key words *Polyrhachis vicina*, myocyte enhancer factor 2 (MEF2), bioinformatics analysis

摘 要 应用 NCBI 上的常用程序、ExPASy 在线核苷酸序列分析工具、CBS 生物学序列分析工具及 SABLE 在线分析软件等对拟黑多刺蚁 *Polyrhachis vicina* Roger 肌细胞增强因子 2 (PvMEF2) 进行了生物信息学分析, 获得了 PvMEF2 因子的序列特征及理化性质, 并对其结构和功能结构域进行了预测。结果表明 PvMEF2 因子具有与已知种类 MEF2 因子较高一致性的 MADS 和 MEF2 结构域, 并且理化性质和二级结构、三级结构等与果蝇该因子类似, 反映了 PvMEF2 可能是参与拟黑多刺蚁肌肉发生调控的重要因子。

关键词 拟黑多刺蚁, 肌细胞增强因子 2 (MEF2), 生物信息学

肌细胞增强因子 2 (MEF2) 属 MADS 家族成员, 其氨基端具有高度保守的 MADS 和 MEF2 结构域。MADS 结构域能够与其他因子形成二聚体并具有 DNA 的结合活性, 可识别富含 A/T 的肌肉组织特异性基因相关序列; MEF2 结构域影响与 DNA 的结合, 具有与其它蛋白质相互作用的功能^[1]。C 末端含有磷酸化位点, 是多种激酶的靶点, 与启动基因表达有关, 也是 MEF2 因子活性调节及发挥功能的主要区域。相对 N 末端, C 末端的氨基酸序列变化较大, 由选择性拼接产生。因此, 无论从它的基因结构或其蛋白质的形式都决定了 MEF2 功能上的复杂性。转录因子 MEF2 的 DNA 结合位点是一段保守的 DNA 序列 CTA (A/T)₄ TAG^[2], 该序列广泛存在于肌肉组织特异性

表达基因的调控区, 因此, MEF2 在肌肉发生中发挥重要作用。MEF2 家族调节心肌、骨骼肌和平滑肌中肌肉相关基因的转录^[3]。

在无脊椎动物中, 如果蝇 *Drosophila melanogaster*、线虫 *Caenorhabditis elegans*、水母 *Podocoryne carnea* 和海葵 *Nematostella vectensis* 中都存在着 Mef2 基因, 且果蝇和水母的 Mef2 基因已证明在肌细胞系中表达^[4]。而关于该基因在蚂蚁中的研究未见报道, 了解其生物信息学知识, 可以为研究蚂蚁等社会性昆虫 Mef2 基因的调控奠定基础。

* 资助项目: 陕西省农业攻关 (2008K02-40)。

** 通讯作者, E-mail: xigengsi@snnu.edu.cn

收稿日期: 2010-01-02, 修回日期: 2010-03-18

1 材料与方法

1.1 材料

拟黑多刺蚁 *Polyrhachis vicina* Roger, 实验室养殖经几代繁殖后根据实验需要随机选取个体材料。

1.2 方法

随机选取第 2 龄幼虫 15 只, 用无菌水清洗后用液氮迅速冷冻, 按照总 RNA 提取试剂盒 (RNAsio Plus) (TAKALA) 使用手册操作, 提取总 RNA。设计简并引物对目的片断扩增, 扩增片断经纯化, 克隆, 测序。根据测序结果设计 3' RACE 和 5' RACE 引物, 并进行末端片断的扩增, 将测序结果拼接获得拟黑多刺蚁 Mef2 基因 cDNA 序列。所有测序工作委托上海生物工程技术服务有限公司完成。

将测序结果上传美国国立生物技术信息中心 (NCBI) GenBank 数据库 (<http://www.ncbi.nlm.nih.gov/Database/index.html>)。应用 NCBI 在线工具 ORF finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) 搜索其开放阅读框。利用在线工具 ProtParam (<http://au.expasy.org/tools/protparam.html>) 预测其氨基酸序列^[5]。并分析其理化特征。

利用 CBS 在线分析服务器软件 (<http://www.cbs.dtu.dk/services/>) 分析序列翻译后修饰^[6,7]。

将预测到的氨基酸序列在 NCBI 上进行

PSI 和 PHI-BLAST, 对非冗余数据库搜索, 与 GenBank 中已知蛋白序列进行同源性比较。并且根据不同物种 MEF2 蛋白全长序列用 Mega4.0 软件计算遗传距离, 并应用邻接法建立系统进化树^[8-10]。

采用 SABLE 工具预测的 PvMEF2 蛋白的二级结构, 在同源建模服务器 SWISS-MODEL (<http://swissmodel.expasy.org/>) 上选用全自动模式 (Automated Mode), 对序列的三级结构进行预测^[11]。

2 结果与分析

2.1 蛋白序列预测与分析

根据测序结果获得了拟黑多刺蚁 Mef2 基因 (PvMef2) 的两种剪接体 (isoforms) 的 cDNA 序列, 并将结果上传至美国国立生物技术信息中心 (NCBI) GenBank 数据库 (登陆号分别为 GQ386849 和 GQ386850)。两条序列仅在 5' 末端存在区别, 其相同的开放阅读框 (ORF) 长为 1 419 bp, 编码 472 氨基酸残基蛋白。

PvMEF2 蛋白的理论等电点和分子量分别是 6.83 和 5.02 ku, 其氨基酸组成见表 1, 从表 1 可知丝氨酸残基最多, 有 71 个, 占 15%; 脯氨酸次之, 有 58 个, 占 12.3%; 半胱氨酸和色氨酸各 2 个, 占 0.4%, 含量最低。其中酸性氨基酸残基 (Asp + Glu) 38 个, 碱性氨基酸残基 (Arg + Lys) 36 个。其分子式为 C₂₁₆₈H₃₃₉₂N₆₂₈O₇₁₄S₁₆, 原子总数为 6 918。

表 1 PvMEF2 蛋白各氨基酸组成

名称	数量	百分比 (%)	名称	数量	百分比 (%)	名称	数量	百分比 (%)
Ala (A)	20	4.2	Gly (G)	43	9.1	Pro (P)	58	12.3
Arg (R)	18	3.8	His (H)	19	4.0	Ser (S)	71	15.0
Asn (N)	28	5.9	Ile (I)	18	3.8	Thr (T)	33	7.0
Asp (D)	18	3.8	Leu (L)	29	6.1	Trp (W)	2	0.4
Cys (C)	2	0.4	Lys (K)	18	3.8	Tyr (Y)	17	3.6
Gln (Q)	16	3.4	Met (M)	14	3.0	Val (V)	23	4.9
Glu (E)	20	4.2	Phe (F)	5	1.1			

通过 <http://www.cbs.dtu.dk/services/NetOGlyc/> 对 PvMEF2 的 O 型糖基化修饰进行预测发现: 自 167 氨基酸残基后, 共有 40 个可

能位点 (图 1)。

通过 <http://www.cbs.dtu.dk/services/NetNGlyc/> 对 PvMEF2 的 N 型糖基化修饰进行

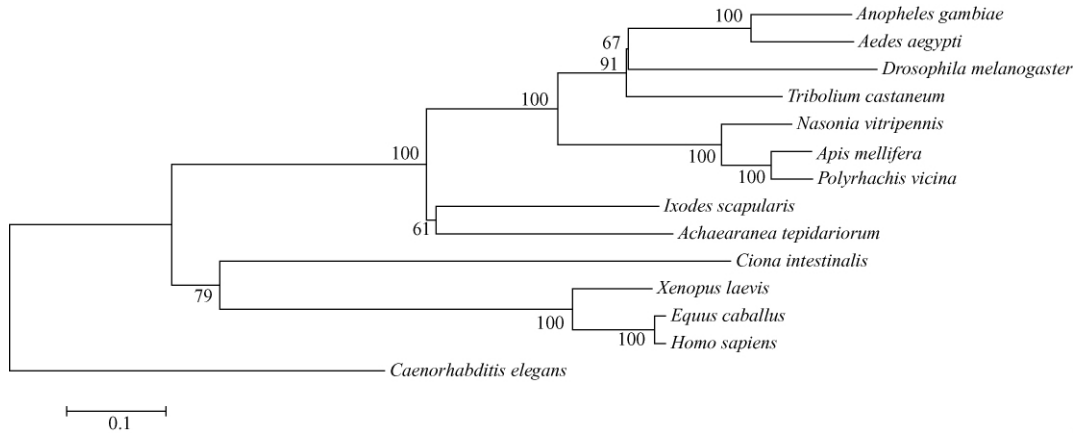


图 5 应用 Mega4.0 构建 NJ 树

2.3 蛋白质结构分析

采用 SABLE 在线分析软件 (<http://sable.cchmc.org/>), 并用在线视图工具 POLYVIEW

将预测结果转化为图形形式输出 (图 6)。



图 6 采用 SABLE 工具预测的 PvMEF2 蛋白的二级结构

代表螺旋; 代表折叠; 代表卷曲; 表示预测的准确性

该蛋白二级结构为混合型, N-末端包括了螺旋、折叠和无规则卷曲等多种形式, 后端区域多为无规则卷曲。

在同源建模服务器 (SWISS-MODEL) 上选用 SwissModel 全自动模式 (First Approach Mode), 对果蝇 MEF2 蛋白的三级结构进行预测。通过最佳模型搜索发现, 与其蛋白序列最匹配的模板是 1n6jA, 最佳比对区域位于其序列的第 2 ~ 93 位氨基酸残基, 序列一致性达 79.348 %, e 值为 8.97×10^{-40} 。预测结果以 PDB 格式输出, 并用相应 3D 微观立体结构视图软件 RASMOL 分析后, 将其保存为普通图形文件。共包括三个螺旋和两个折叠区域 (图 7)。



图 7 预测的 PvMEF2 保守域的三级结构图

3 结论与讨论

PvMEF2 基因在拟黑多刺蚁体内以不同剪接体的形式存在,通过 NCBI 在线工具 ORF finder 分析发现其不同剪接体具有相同开放阅读框,由此反映其 mRNA 5' 末端对 MEF2 在动物体的表达具有不同调控作用。预测的 PvMEF2 蛋白含 472 个氨基酸残基,其 N 末端具高度保守的 MADS 和 MEF2 结构域。并且其理化性质、氨基酸组成、修饰位点及跨膜分析等方面与果蝇 MEF2 极为相似,这一点也是 PvMEF2 在发育过程中具有相关生理功能的基础。在其保守的 MADS 和 MEF2 结构域未发现可能的糖基化位点,其可能的糖基化位点均位于 PvMEF2 的不保守区域。而 MEF2 的 C 末端结构域可被多种激酶与磷酸酶在多位点进行不同程度的修饰,调节 MEF2 的活性^[1],PvMEF2 可以通过钙调素-钙调磷酸酶通路调节自身的活性。推测翻译后 C 末端的修饰可能是维持 MADS 和 MEF2 结构域在肌肉发生中发挥作用的重要保障。

通过与其它物种 MEF2 比对发现,PvMEF2 保守的 MADS 和 MEF2 结构域与其它昆虫该区域的一致性甚至可以高达 100%,与其它类群物种一致性也可以达到 86% 以上,结构上的保守性与功能结构域的搜索分析反映了 MEF2 在功能上的保守性。根据不同动物 MEF2 氨基酸序列构建的系统进化树较为正确的反映了各类群的进化关系,说明尽管其 C 末端在各类群间存在很大的差异,但这种变化仍符合进化过程中变异规律的,也说明了进化过程中氨基酸变异与保守的统一。

PvMEF2 二级结构预测结果表明其 N 末端存在螺旋、折叠和无规则卷曲等多种形式,这可能与 MEF2 通过 N 末端的结构形式与其它蛋白结合而发挥一定生理功能有关,如 MEF2 可以和 MRFs 等结合形成复合体而进行调控作用。而 C 末端形式相对单一,单一的结构形式是否与影响 N 末端的功能有关需进一步的研究证明。对其三级结构的预测,搜索到的模板只与

N 末端保守区相匹配,其它部分序列没有合适模板,由此反映了 N 末端的 MADS 和 MEF2 结构域在生理功能上起重要作用。其与果蝇 MEF2 三级结构类似,可能与 N 末端的功能作用发挥有关。

MEF2 是动物肌肉发育中的重要调控因子,PvMEF2 在理化性质及结构等各方面与其它动物类群的相关因子存在相似之处,由此反映该因子在动物类群中的主要生理功能与肌肉的发生相关。其保守的 N 末端和不保守的 C 末端对 PvMEF2 的功能分别有着不同作用。生物信息学分析为相关研究提供一定参考,在获得了其相关序列信息及结构信息后,可进一步通过分子生物学及组织学等方面研究,确定其在肌肉发生过程中的具体调控形式。

参 考 文 献

- 1 Lyons G. E., Micales B. K., Schwarz J., *et al.* Expression of MEF2 genes in the mouse central nervous system suggests a role in neuronal maturation. *J. Neurosci.*, 1995, **15**: 5 727 ~ 5 738.
- 2 程震龙,朱大海,张志谦. MEF2 与肌肉发生. *遗传*, 2002, **24**(5): 581 ~ 585.
- 3 Black B. L., Molkentin J. D., Olson E. N. Multiple roles for the MyoD basic region in transmission of transcriptional activation signals and interaction with MEF2. *Mol. Cell Biol.*, 1998, **18**: 69 ~ 77.
- 4 张颖,王利凤,邵明,等. 文昌鱼 *AmphiMef2* 基因的特征及其发育表达. *中国科学 C 辑*, 2007, **37**(4): 422 ~ 426.
- 5 Gasteiger E., Hoogland C., Gattiker A., *et al.* Protein identification and analysis tools on the ExPASy Server; In: John M. W. (ed.). *The Proteomics Protocols Handbook*, Humana Press, 2005. 571 ~ 607.
- 6 Blom N., Gammeltoft S., Brunak S. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.*, 1999, **294**(5): 1 351 ~ 1 362.
- 7 Julenius K., Mølgaard A., Gupta R., *et al.* Prediction, conservation analysis and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology*, 2005, **15**: 153 ~ 164.
- 8 Kumar S., Dudley J., Nei M., *et al.* MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinformatics*, 2008, **9**: 299 ~ 306.
- 9 Tamura K., Dudley J., Nei M. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.*, 2007, **24**: 1 596 ~ 1 599.
- 10 Zuckerkandl E., Pauling L. Evolutionary divergence and convergence in proteins. In: Bryson V., Vogel H. J. (eds.). *Evolving Genes and Proteins*, New York: Academic Press, 1965. 97 ~ 166.
- 11 Arnold K., Bordoli L., Kopp J., *et al.* The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 2006, **22**: 195 ~ 201.

新疆意大利蝗不同地理种群 16S rDNA 序列差异研究*

程虹¹ 王晗¹ 吴敏^{2**} 季荣^{1**}

(1. 新疆师范大学生命科学学院 乌鲁木齐 830054; 2. 浙江大学生命科学院 杭州 320058)

Calliptamus italicus 16S rDNA in different geographic populations in Xinjiang. CHENG Hong¹, WANG Han¹, WU Min^{2**}, JI Rong^{1**} (1. College of Life Sciences, Xinjiang Normal University, Urumqi 830054, China; 2. College of Life Sciences, Zhejiang University, Hangzhou 320058, China)

Abstract Partial sections of the mitochondrial 16S rDNA of *Calliptamus italicus* (L.) from different geographic populations of Xinjiang Province were compared. The results show that: 1) The percentage of A + T (67.7%) of *C. italicus* was more than G + C (32.3%); 2) Multiple sequence alignment and the UPGMA phylogenetic tree indicate that different geographic populations of *C. italicus* display almost no genetic difference.

Key words *Calliptamus italicus*, 16S rDNA, genetic differentiation, molecule phylogenetic tree

摘 要 研究新疆东部和西部不同地理种群意大利蝗 *Calliptamus italicus* (L.) 线粒体基因中 16S rDNA 序列差异。结果表明:(1) 意大利蝗 A + T 的含量(67.7%)明显高于 G + C 的含量(32.3%);(2) 多重序列比对和 UPGMA 系统发育树的结果表明新疆东、西部不同地理种群的意大利蝗的遗传变异极小。

关键词 意大利蝗, 16S rDNA, 遗传分化, 分子系统树

植食性昆虫寄主植物的选择差异及其机理较为复杂,常因地理种群分化、昆虫发育阶段、生境中优势植物种类、植被长势的不同而不同^[1]。意大利蝗 *Calliptamus italicus* (L.) 是新疆草原优势危害蝗虫种类之一,主要分布于北疆地区,可取食 17 科 45 种植物^[1]。研究发现,不同分布区域意大利蝗对寄主植物的选择存在一定差异,而目前尚不确定造成这种差异的原因是否与意大利蝗不同地理种群的遗传分化有关。本研究以新疆最东部和西部地区的意大利蝗虫种群的 16S rDNA 为研究内容,旨在弄清意大利蝗不同地理种群之间是否存在遗传差异,明确意大利蝗对寄主植物的选择差异是否与不同地理种群间遗传分化有关,并为进一步研究新疆境外迁入危害的意大利蝗种群与新疆境内的意大利蝗虫种群间的遗传差异奠定基础。

1 材料与方法

1.1 样品采集

意大利蝗于 2008 年 7 月分别采自新疆东部地区的哈密巴里坤白石头乡(42.82°N, 93.49°E, 海拔 2 200 ~ 2 400 m)和新疆西部博乐青得里乡(44.56°N, 87.68°E, 海拔 550 m)的荒漠草地。前者植被种类以针茅(*Stipa capillata* L.)、冷蒿(*Artemisia frigida* Willd.)、棘豆(*Oxytropis* sp.)、椒蒿(*Artemisia dracunculus* L.)、鸢尾(*Iris* sp.)等为主,后者主要有紫花苜蓿(*Medicago sativa* L.)、绢蒿(*Seriphidium*

* 资助项目:国家自然科学基金(30660030, 30960230)、新疆师范大学博士科研启动基金(XJNUBS0902)、新疆师范大学重点实验室基金(XJNUSYS0810)。

** 通讯作者, E-mail: jirongxj@ yahoo. com. cn; wumin@ zju. edu. cn

收稿日期:2009-11-24, 修回日期:2010-05-24