文章编号:1000-5404(2005)09-0847-03

论著

小鼠 peroxiredoxin 基因家族的生物信息学分析

章 波[1],向渝梅[2],白 云[1],许雪青[1],王 燕[1] (第三军医大学:[1]基础医学部分子遗传学教研室,[2]科研部,重庆 400038)

提 要:目的 分析小鼠 peroxiredoxin 基因家族的基因组结构和进化。**方法** 利用已经公布的小鼠基因组数据库,采用 BLAST 程序检索该基因家族各成员的编码基因和假基因,并利用 Cluster W 软件进行序列联配,绘制其分子进化树。**结果** 分析表明该家族半数成员具有多个假基因序列,为返转座类型假基因。这些假基因的形成时间有先后,所具有返转座假基因的典型结构也各不相同。分析该家族的进化表明 Prx Ⅰ-Ⅳ有共同的祖先,归为一个亚类,而 Prx Ⅴ和Ⅵ与前者的相似性较差,归为另一个亚类。**结论** 该家族每个成员长期进化所形成的多样性提示其功能具有独特性。

关键词: peroxiredoxin;基因家族;进化;生物信息学
中图法分类号: R318.04; R394 **文献标识码:** A

Bioinformatic analysis of mouse peroxiredoxin family

ZHANG Bo[1], XIANG Yu-mei[2], BAI Yun[1], XU Xue-qing[1], WANG Yan[1] ([1]Department of Molecular Genetics, College of Medicine, [2]Devision of Scientific Research Affairs, Third Military Medical University, Chongqing 400038, China)

Abstract: Objective To determine the genomic structure of mouse peroxiredoxins and to analyze its molecular evolution. **Methods** Based on the published database of mouse genome, the coding genes and pseudogenes of all peroxiredoxin members were retrieved by using BLAST software and their genomic structures were analyzed. All the sequences were multi-aligned with ClusterW software and the molecular evolution tree was drawn. **Results** Half of all Prx members had pseudogenes that were the retrotransposons of coding genes and produced according to chronologicaol order. The poly-A strings and terminal repeats of some pseudogenes had changed, therefore the typical pseudogene structures were different. The evolution of this gene family showed that Prx Ⅰ-Ⅳ were more homologous, which indicated they had common ancestor. Prx Ⅴ and Ⅵ were less homologous and were clustered as another subgroup. **Conclusion** As a result of long time evolution, the diversity of this family suggested that each member might have its specific functions.

Key words: peroxiredoxin; gene family; evolution; bioinformatics

peroxiredoxin(Prx)基因是新近发现的一类编码过氧化物酶的基因,其编码蛋白属于抗氧化蛋白超家族,广泛存在于原核生物和真核生物中[1,2],其催化活性和蛋白序列与其它抗氧化剂完全不同,它除了在清除活性氧中具有重要作用以外,还具有其它多种功能。哺乳动物的 peroxiredoxin 家族包括 6 个成员:Prx Ⅰ、Ⅱ、Ⅲ、Ⅳ、Ⅴ和Ⅵ。研究表明:该家族高等真核生物的同源基因在低等真核生物中仍具有功能,这显示该家族基因在进化上的保守性[3]。该家族所有蛋白均在 N-端具有保守的 Cys 残基,有的成员在 C-端还具有保守的 Cys[4]。

目前,小鼠基因组数据库已经公布,分析 Peroxiredoxin 基因家族的染色体结构已经成为可能。本研究利用已知的 peroxiredoxin 基因家族 mRNA 和蛋白质序列,分析了其亲缘关系和基因组结构,并对其可能的进化模式进行了推测。

1 材料与方法

1.1 材料

小鼠 peroxiredoxin 基因家族的 mRNA 和蛋白质序列记录 ID 号见表 1。

表 1 小鼠 peroxiredoxin 家族的序列记录 ID 号

Tab 1 ID number of the coding sequences of mouse peroxiredoxin family

	Prx Ⅰ	Prx Ⅱ	Prx Ⅲ	Prx Ⅳ	Prx Ⅴ	Prx Ⅵ
mRNA	6 754 975	6 755 757	6 680 689	7 948 998	14 198 230	6 671 548
Protein	547 923	6 755 758	6 680 690	13 097 150	6 755 114	3 719 451

1.2 方法

1.2.1 基因组比对 采用 BLAST 程序,以 mRNA 序列搜索

作者简介:章 波(1972-),男,重庆市人,博士,讲师,主要从事基因表达与调控方面的研究。电话:(023)68753670, E-mail: bo-zhang@yahoo.com
通讯作者:白 云,电话:(023)68752258
收稿日期:2004-10-10;修回日期:2005-01-01

基因组序列，分析基因组中编码基因和假基因。假基因的判断采用武断标准：基因组序列中，与 mRNA 序列相似性大于 80% 的序列可判断为假基因，并参照数据库检索结果。

1.2.2 基因、假基因及蛋白质的系统学分析 本研究使用网络资源 Clustal W 软件（http://www.es.embnet.org/cgi-bin/clustal.cgi）对全部分析序列进行多重序列排列（multiple sequence alignment），取其同源性较高区段用于建立系统树，利用 PAUP（phylogenetic analysis using parsimony）软件计算其遗传距离，用 NJ 法构建其系统发育树（neighbor-joining, NJ tree），并进行了1 000次置信度分析（bootstrap analysis）。

2 结果

2.1 peroxiredoxin 基因家族的基因组结构

小鼠基因组序列已经公布，可以利用 Blast 程序对其进行序列对比，寻找存在于基因组中的假基因。我们以 peroxiredoxin 基因家族的 mRNA 进行基因组序列对比发现，半数成员具有假基因，且数目各不相同（表 2）。具有假基因的成员为 Prx Ⅰ、Ⅱ和Ⅵ。其中，假基因数目最多的成员是 Prx Ⅰ，有 6 个假基因，分别位于 8、16、17、18 号染色体，而其编码基因位于 4 号染色体。

表 2 peroxiredoxin 基因家族的基因组结构

Tab 2 The genomic structure of peroxiredoxin family

Family member	Coding gene	Pseudogene					
Prx Ⅰ							
Chr[a]	4	8	16	16	8	17	18
Rec[b]	NT-039264.2	NT-078575.1	NT-039624.2	NT-082364.1	NT-039460.2	NT-078815.1	NT-082382.1
Sim[c]		96%	93%	93%	86%	90%	84%
Prx Ⅱ							
Chr[a]	8	9	1	17	4	—	—
Rec[b]	NT-039467.2	NT-078599.1	NT-078297.1	NT-039649.2	NT-039267.2	—	—
Sim[c]		100%	95%	92%	89%	—	—
Prx Ⅵ							
Chr[a]	1	2	4	—	—	—	—
Rec[b]	NT-039185.2	NT-039208.2	NT-039258.2	—	—	—	—
Sim[c]		89%	85%	—	—	—	—

a: Chromosome number, b: Record accession number, c: Similarity

值得注意的是 Prx Ⅱ 的编码基因，8 和 9 号染色体 DNA 均可拼接出该基因 mRNA 序列，且其 DNA 序列从编码区上游 2 000 bp延伸至编码区下游2 000 bp的范围内均完全相同。然而 9 号染色体 DNA 含有多余的 1 个第 4 外显子部分，也可以认为其为该基因的编码基因。然而，Prx Ⅱ 的这 2 个基因是否都能转录还需要进一步实验。

2.2 peroxiredoxin 家族编码基因的分子进化

以该家族编码基因的蛋白质序列进行多重序列联配，结果表明：Prx Ⅰ 和 Ⅱ 的相似性最高，相同残基为 72.1%。Prx Ⅰ－Ⅳ 的相似性大于 50%，可以认为有共同的祖先，归为一个亚类；而 Prx Ⅴ 和 Ⅵ 与前者的相似性较差，可以归为另一个亚类（表 3）。以 mRNA 为对象进行序列对比的结果类似。

表 3 小鼠 peroxiredoxin 家族蛋白质序列相似性的比较

Tab 3 The similarity comparison of mouse peroxiredoxin family based on protein sequences

	Prx Ⅰ	Prx Ⅱ	Prx Ⅲ	Prx Ⅳ	Prx Ⅴ	Prx Ⅵ
Prx Ⅰ	—	72.1	59.3	64.3	12.6	24.6
Prx Ⅱ	72.1	—	61.9	66.5	14.2	23.9
Prx Ⅲ	59.3	61.9	—	51.4	19.5	17.4
Prx Ⅳ	64.3	66.5	51.4	—	15.2	19.2
Prx Ⅴ	12.6	14.2	19.5	15.2	—	11.0
Prx Ⅵ	24.6	23.9	17.4	19.2	11.0	—

利用该家族基因的 mRNA 序列和蛋白质序列联配结果构建其系统发生树。基于 mRNA 序列的进化树表明 Prx Ⅰ 和 Ⅱ 与该家族的共同祖先关系较近，其次是 Prx Ⅲ 和 Ⅳ，而 Prx Ⅴ 和 Ⅵ 最远。该家族的演化可以推断为：由一个共同的祖先经过倍增形成 3 个同源基因，其中一个演变为 Prx Ⅳ，一个演变为 Prx Ⅰ 和 Ⅱ 的共同祖先，另外一个则演变为 Prx Ⅲ、Ⅴ 和 Ⅵ 的共同祖先，此基因再分化形成 Prx Ⅲ，以及 Prx Ⅴ 与 Prx Ⅵ 的祖先基因。利用该家族蛋白质序列构建的系统发生树推断的分子进化过程与上述过程相似。

将该家族编码基因序列和假基因序列采用 Cluster W 软件进行序列联配后绘制其进化树，结果见图 1。各个编码基因与其相应的假基因成簇聚类，关系密切。Prx Ⅰ 的 6 个假基因中，I-Chr18-2127277 形成较早，与编码基因的相似性最低为 84%；而在 Prx Ⅱ 的 4 个假基因中，最先分化的是 II-Chr4-12222623，与编码基因的相似性为 89%。

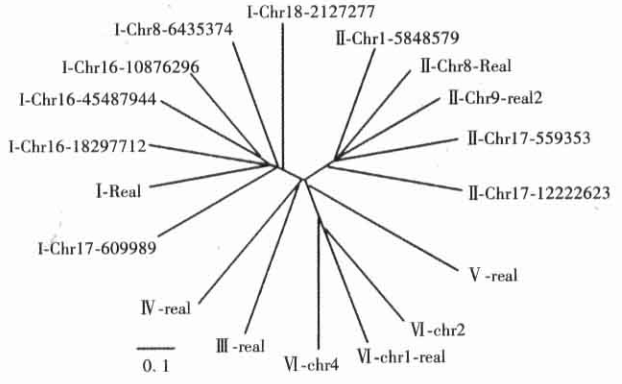


图 1 Prx 基因家族编码基因和假基因的分子系统树（假基因以成员-染色体-登记号之起始点表示，编码基因以 real 表示）

Fig 1 Molecular phylogenetic trees of mouse peroxiredoxin family based on the sequences of coding gene and pseudogene

2.3　假基因的结构特征

在 Prx 家族中，多数为返转座假基因。其典型的结构特征为：① 3′端具有 PolyA 结构；② 假基因序列两端具有 9～21 bp 的正向重复序列；③无内含子序列。在 Prx Ⅰ的所有假基因中，保留典型的结构特征者少，没有末端重复序列和 PolyA 结构，有较多的缺失部分。Prx Ⅱ的假基因则相对于前者更为典型，均有末端重复序列(表 4)。

表 4　Prx Ⅱ假基因的结构特征

Tab 4　The structural characteristics of mouse Prx Ⅱ pseudogenes

Position[1]	Flanking terminal repeat		PolyA
	5′ terminal repeat	3′ terminal repeat	
1/ NT-078297.1	TCCT AAGAAAGTTTCTTGAG	A AAGAAGGTTTCTTGATT	+
17/ NT-039649.2	AT TAAT G TCACAATATAG TT	A TAAT AT CACAATATAG AC	+
4/ NT-039267.2	AT CTTTAGCA CAATTTTT	CCCT CTTTCT CAATTGTGAG	-

1: The position of each pseudogene is showed by its chromosome number and record accession

3　讨论

在生物的长期进化中，许多基因能够形成无功能的假基因。这些假基因在没有选择压力的情况下，可发生较大的突变。已有研究证实：Prx Ⅰ有 1 个编码基因和 2 个假基因[5]。这是通过基因组文库筛选、测序所得到的结论。然而这个结论与本研究有所区别。另外，Lim 等[6]通过基因组文库筛选发现：Prx Ⅱ有 1 个编码基因和 2 个假基因，假基因可能是通过返转座方式形成。在他们的结果中没有 9 号染色体的编码序列和另一个假基因。通过基因组序列的对比可以较为全面地发现这些假基因序列。这说明了采用基因组序列对比的方法来确定基因的结构更为准确。

在基因组中，编码蛋白质的序列只占少部分，绝大多数序列被称为“垃圾 DNA”。它们包括内含子、重复序列、假基因和可移动元件[7]。假基因的形成可能有 2 种机制：基因倍增和返转座。在目前发现的假基因中，其存在形式有 2 种：保留间隔序列的假基因和处理后假基因(又称为返转座假基因)，分别与这 2 种机制相对应[8]。在该家族中，只有 Prx Ⅱ的假基因具有较为典型的结构，说明其形成相对较晚，而 Prx Ⅰ和Ⅵ的假基因基本没有假基因的典型结构，说明其分化的时间更早。

关于 Prx 基因家族的功能有一点是明确的，即均具有还原过氧化物的功能。除此之外，它们还具有其它功能诸如刺激细胞增殖，参与细胞的信号转导等[9]。研究表明，它们在细胞质中的定位有所不同，因而功能各异。Prx Ⅰ和Ⅱ定位于细胞质中，在多种组织中表达，其功能主要是清除过氧化物累积，能够刺激细胞增殖，如增强 NK 细胞活性[10]。Prx Ⅲ定位于线粒体，该蛋白在内质网中合成后经过线粒体定位信号而转运至线粒体。Prx Ⅳ的 N-端含有分泌信号，该蛋白在内质网中合成后被转运至细胞外，具有结合肝素的能力[11]。因此，在内质网和细胞外基质可以检测到该蛋白。Prx Ⅴ具有长短 2 种形式，存在于细胞质、过氧化物酶体和线粒体中[12]，参与调节细胞内信号分子 H_2O_2。Prx Ⅵ也存在于细胞质中，在多种组织中均能检测到，其中肺脏中表达最高。该蛋白是一个双功能蛋白，具有过氧化物酶和磷脂酶 A2 活性[13]。Prx 基因家族在长期的进化过程中形成了多个成员，与其功能的多样性一致。

参考文献：

[1] Rhee S G, Kang S W, Chang T S, *et al*. peroxiredoxin, a novel family of peroxidases[J]. IUBMB Life, 2001, 52 (1－2): 35－41.

[2] Fujii J, Ikeda Y. Advances in our understanding of peroxiredoxin, a multifunctional, mammalian redox protein[J]. Redox Rep, 2002, 7(3): 123－130.

[3] Tien Nguyen nhu N, Knoops B. Mitochondrial and cytosolic expression of human peroxiredoxin 5 in Saccharomyces cerevisiae protect yeast cells from oxidative stress induced by paraquat[J]. FEBS Lett, 2003, 544(1-3): 148－152.

[4] Chang T S, Jeong W, Choi S Y, *et al*. Regulation of peroxiredoxin I activity by Cdc2-mediated phosphorylation[J]. J Biol Chem, 2002, 277(28): 25370－25376.

[5] Lee T H, Yu S L, Kim S U, *et al*. Characterization of mouse peroxiredoxin I genomic DNA and its expression[J]. Gene, 1999, 239(2): 243－250.

[6] Lim M J, Chae H Z, Rhee S G, *et al*. The type Ⅱ peroxiredoxin gene family of the mouse: molecular structure, expression and evolution[J]. Gene, 1998, 216(1): 197－205.

[7] Clifford R J, Edmonson M N, Nguyen C, *et al*. Bioinformatics tools for single nucleotide polymorphism discovery and analysis[J]. Ann N Y Acad Sci, 2004, 1020(2): 101－109.

[8] 崔　峰，夏家辉. 关于假基因的研究进展[J]. 生命科学研究，1999, 3(3): 203－209.

[9] Zhou Y, Kok K H, Chun A C, *et al*. Mouse peroxiredoxin V is a thioredoxin peroxidase that inhibits p53-induced apoptosis[J]. Biochem Biophys Res Commun, 2000, 268(3): 921－927.

[10] Shau H, Butterfield L H, Chiu R, *et al*. Cloning and sequence analysis of candidate human natural killer-enhancing factor genes[J]. Immunogenetics, 1994, 40(2): 129－134.

[11] Okado-Matsumoto A, Matsumoto A, Fujii J, *et al*. peroxiredoxin Ⅳ is a secretable protein with heparin-binding properties under reduced conditions [J]. J Biochem (Tokyo), 2000, 127(3): 493－501.

[12] Seo M S, Kang S W, Kim K, *et al*. Identification of a new type of mammalian peroxiredoxin that forms an intramolecular disulfide as a reaction intermediate[J]. J Biol Chem, 2000, 275(27): 20346－20354.

[13] Chen J W, Dodia C, Feinstein S I, *et al*. 1-Cys peroxiredoxin, a bifunctional enzyme with glutathione peroxidase and phospholipase A2 activities [J]. J Biol Chem, 2000, 275(37): 28421－28427.

(编辑　陈聪连)