DOI:10.16016/j.1000-5404.2004.08.017

论著

文章编号:1000-5404(2004)08-0700-04

Channel catfish 重组激活基因克隆与鉴定

公衍文,张 云,黄 庆,张 雪,府伟灵 (第三军医大学西南医院检验科,重庆 400038)

提 要:目的 尝试用简并引物扩增序列未知的channel catfish基因组DNA中*rag*1基因的片段并测序,为*rag*1基因的早期进化模式和遗传学变异情况的研究提供直接的证据。方法 基于*rag*1基因的高度保守设计简并引物,PCR扩增channel catfish基因组DNA中*rag*1基因的片段,TA克隆并双向测序。将所得序列资料拼接,用多种生物信息学方法分析其ORF、内含子、与其它物种*rag*1基因及Rag1蛋白的相似性,并做多序列比较和系统进化树分析。结果 扩增出3条channel catfish *rag*1基因的DNA片段。从拼接后的序列中找到一2 235 bp的ORF和一293 bp的内含子,所得序列与其它物种*rag*1基因的相似程度高(多大于70%),去除内含子后的channel catfish *rag*1基因序列与zebrafish、rainbow trout、bull shark的*rag*1基因cDNA全序列碱基一致的位点占43.9%,自第1 352位碱基至2 925位碱基为53.1%,而自第1位碱基至1 351位碱基为33.2%,有非常显著的差异($P<0.01$)。结论 用简并引物成功扩增出channel catfish *rag*1基因的DNA片段。序列在GenBank中的登记号是:AY423858(去除内含子序列),AY423859。*rag*1基因进化缓慢,高度保守,不同物种*rag*1基因的变异相对集中在氨基末端的区域。系统进化树分析显示了13种硬骨鱼在进化上关系的亲疏。

关键词:重组激活基因;V(D)J重排;聚合酶链反应;简并引物;生物信息学
中图法分类号:Q75;Q785 文献标识码:A

Cloning and identification of the *rag*1 gene in channel catfish

GONG Yan-wen, ZHANG Yun, HUANG Qing, ZHANG Xue, FU Wei-ling (Department of Clinical Laboratory, Southwest Hospital, Third Military Medical University, Chongqing 400038, China)

Abstract: **Objective** To amplify and sequence the fragments of recombination activating gene 1 (*rag*1) from the channel catfish genomic DNA so as to provide new insights into the early evolutionary patterns and the mechanisms of genetic diversity of *rag*1 gene. **Methods** The degenerate primers, designed based on the high conserved characteristics of *rag*1 gene, were used to amplify the fragments of *rag*1 gene from the channel catfish genomic DNA by PCR. The fragments were cloned into a T/A plasmid. Clones were randomly chosen for sequencing and all sequences were obtained on both strands. Several kinds of bioinformatic methods were used to analyze the open reading frames, introns, and identity percentage with those of other species. Selected sequences were also analyzed using the program "CLUSTAL W" for the multiple alignment and unrooted phylogenetic tree. The obtained sequences were submitted to GenBank using the program "Bankit" to apply for the sequence numbers. **Results** Three fragments of *rag*1 gene were obtained with a total length of 3 218 bp. An ORF of 2 235 bp and an intron of 293 bp within the sequence were identified. The percentage identity of the sequence with *rag*1 gene of other species was more than 70%. The multiple alignment of the cDNA sequence with *rag*1 complete cDNA sequences in zebrafish, rainbow trout, and bull shark indicted that the proportion of identical base sites was 43.9%. The proportion of identical base sites from the 1st to the 1 351st was 33.2%, and the proportion of identical base sites from the 1 352nd to the 2 925 th was 53.1%. There was a significant diversity between these two regions ($P<0.01$). **Conclusion** The fragments of channel catfish *rag*1 gene are obtained by PCR using degenerate primers. The registered numbers of those sequences are AY423858, AY423859, and AY423860. The *rag*1 gene is highly conserved in different species. The dissimilarities in *rag*1 gene in different species are concentrated to the amino-terminal region. The phylogenetic tree indicates the relationships of 13 kinds of bony fishes in the course of evolution.

Key words: recombination activating gene; V(D)J rearrangement; polymerase chain reaction; degenerate primer; bioinformatics

作者简介:公衍文(1971-),男,山东省蒙阴县人,硕士,主治医师,主要从事临床微生物与免疫学检验方面的研究。电话:(023)65460802
通信作者:府伟灵,电话:(023)65460802
收稿日期:2003-11-18;修回日期:2004-03-12

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

重组激活基因(recombination activating genes, RAGs)包括一对在染色体上紧邻的基因(*rag*1 和 *rag*2),它们的编码产物共同构成重组酶(recombinase),能特异地识别 Ig 和 TCR 的 V、DH、J 基因片段两侧的重组信号序列(recombination signal sequence, RSS),催化 DNA 双链断裂,引起 V(D)J 重排,在 Ig 和 TCR 库多样性的形成过程,即淋巴细胞发育成熟过程中起着关键性的作用[1],是追踪淋巴组织发育的有用标志[2]。Channel catfish 作为硬骨鱼类的一种,是最早进化出明显 DH 片段的脊椎动物之一[3~6],它的 B 细胞和 T 细胞能分别表达 Ig 和表达 TCR,而且具有透明胚胎,繁殖能力强,遗传背景简单明确,是研究免疫系统进化的一种良好的模式生物。RAGs,尤其是 *rag*1 基因,带有一些有利于种系发育研究的特点,如进化缓慢,第 3 位密码子变化最小、不同物种之间的碱基组成近乎恒定等[7],我们利用 GenBank 中其它物种,尤其是硬骨鱼类(如斑马鱼、虹鳟鱼、鲨鱼)的 *rag*1 基因及 Rag1 蛋白的序列比对资料来设计简并引物,扩增 channel catfish *rag*1 基因的片段,做 TA 克隆并测序,以及用多种生物信息学方法对所获序列进行分析。

1　材料与方法

1.1　实验材料

1.1.1　channel catfish 基因组 DNA　　由 Dr. Craig J. Lobb 惠赠,提取自 46 月龄成年 channel catfish 的前肾组织。

1.1.2　试剂　　TA Clone Kit、X-Gal、IPTG、Ex-*Taq* 酶购自宝生物工程(大连)有限公司;大肠杆菌 DH5α 由本室保存;E. Z. N. A. Gel Extraction Kit、E. Z. N. A. Plasmid Miniprep Kit 为 Omega 公司产品。

1.1.3　仪器　　PCR 扩增仪为 PE 公司 GeneAmp2400 型;紫外分光光度计为 Smart spectro 3000(Bio-Rad);凝胶成像分析仪为 Bio-profic(French)。

1.2　实验方法

1.2.1　简并引物的设计　　参照文献[7],基于不同物种间 *rag*1 基因和 Rag1 蛋白氨基酸序列的高度保守性,根据已经发表的特别是亲缘关系较近的物种序列资料的比对结果设计而来,由宝生物工程(大连)有限公司合成,序列如下:转录起始位点:5′-GTTTCRAGTTCGTRCATT-3′;421 下游:5′-MGI GCNCARAARCAYMG-3′;511 上游:5′-TACATYTTRTGRTAYTGRCT-3′;511 下游:5′-AGYCARTAYCAYAARATGTA-3′;619 上游:5′-ACRTCICCCATNCCRTCRCA-3′;619 下游:5′-TGYGAYGGNATGGGNGAYGT-3′;687 上游:5′-TCRTGRTCISWYTCRTCIACRAAC-3′;817 上游:5′-GCRTTNCCDATRTCRCARTG-3′;817 下游:5′-CAYTGYGAYATHGGNAAYGC-3′;994 上游:5′-TGICKIGCRTTCATYTTNC-3′(注:Y=C/T;K=G/T;R=A/G;M=A/C;W=A/T;S=C/G;D=A/G/T;H=A/C/T;N=A/T/C/G;I-脱氧次黄嘌呤)。根据克隆、测序的结果,逐步明定这些简并引物的确切序列,再由宝生物工程(大连)有限公司合成其特异引物。

1.2.2　扩增目的片段(以扩增片段 511-817 为例)　　反应体系:10× Ex Buffer 5.0 μl;$MgCl_2$(25 mmol/L) 4.0 μl;dNTP(10 mmol/L) 1.0 μl;degenerate Primers(20 μmol/L) 2.0 μl;channel catfish Genomic DNA(150 ng/μl) 1.0 μl,Ex *Taq*(5 U/μl) 1.0 μl;H_2O 34.0 μl。反应条件:94 ℃变性 5 min;94 ℃变性 30 s,40 ℃(42 ℃;44 ℃;46 ℃;48 ℃;50 ℃)退火 1 min,74 ℃延伸 30 s,共 4 个循环;72 ℃延伸 5 min;然后再 94 ℃变性 30 s,52 ℃退火 1 min,72 ℃延伸 30 s,共 30 个循环;72 ℃延伸 8 min,4 ℃保存。

选取条带单一、清楚明亮的扩增产物做胶回收。以胶回收产物作模板,反应体系同上,按以下循环参数扩增片段。94 ℃变性 5 min;94 ℃变性 30 s,58 ℃退火 1 min,72 ℃延伸 30 s,共 35 个循环;72 ℃延伸 8 min,4 ℃保存。电泳后切取目的带回收。

1.2.3　TA 克隆,序列测定　　采用宝生物公司的 TA Clone Kit(pMD 18-T Vector),置 PE2400 中,16 ℃过夜。将全部连接产物转化感受态大肠杆菌 DH5α,挑取单个白色菌落至 2 ml 含 Amp 的 LB 培养基(Amp^+)中,37 ℃培养约 12 ~ 14 h 后,采用 E. Z. N. A. Plasmid Miniprep Kit 抽提质粒。序列测定由宝生物工程(大连)有限公司完成,所有序列都是通过双向测序获得。

1.2.4　生物信息学方法分析 channel catfish *rag*1 基因 DNA 序列　　在 NCBI 网站中利用"ORF Finder"(www.ncbi.nlm.nih.gov/gorf/gorf.html)、BLAST-N、BLAST-X 程序做开放阅读框架(open reading frame, ORF)分析、相似性/同源性分析;用"NetGene 内含子/外显子剪接位点识别工具"(www.cbs.dtu.dk/services/NetGene2)识别内含子序列;用 clustalw(www.ebi.ac.uk/clusalw)将所得去除内含子的 channel catfish *rag*1 基因 DNA 序列与 GenBank 中斑马鱼(*Danio rerio*, zebrafish, U71093.1, GI: 2078513)、虹鳟鱼(*Oncorhynchus mykiss*, rainbow trout, U15663.1, GI: 558917)、牛鲨(*Carcharhinus leucas*, bull shark, U62645.1, GI: 1470116)的 *rag*1 基因 cDNA 全序列做多序列比较,以及大西洋鳕鱼(*Gadus morhua*, Atlantic Cod, AF369064.1, GI: 15991422)、河鲈(*Lates calcarifer*, Barramundi Perch, AF369066.1, GI: 15991426)、胭脂鱼(*Mugil cephalus*, flathead mullet, AF369065.1, GI: 15991424)、鲆鱼(*Pseudopleuronectes americanus*, winter flounder, AF369067.1, GI: 15991428)、蝴蝶鱼(*Pantodon buchholzi*, freshwater butterflyfish, AF369061.1, GI: 15991416)、密西西比浆鱼(*Polyodon spathula*, Mississippi paddlefish, AF369057.1, GI: 15991408)、翻车鱼(*Mola mola*, ocean sunfish, AF137205.1, GI: 5870746)、北方狗鱼(*Esox lucius*, northern pike, AF137197.1, GI: 5870730)、牛奶鱼(*Chanos chanos*, milkfish, AF137195.1, GI: 5870726)等的 *rag*1 基因 cDNA 部分序列,一共 13 种硬骨鱼类做系统进化树分析。

1.2.5　提交 GenBank,申请序列号　　在 NCBI 网站中用 Bankit 直接提交序列。

1.3　统计学处理

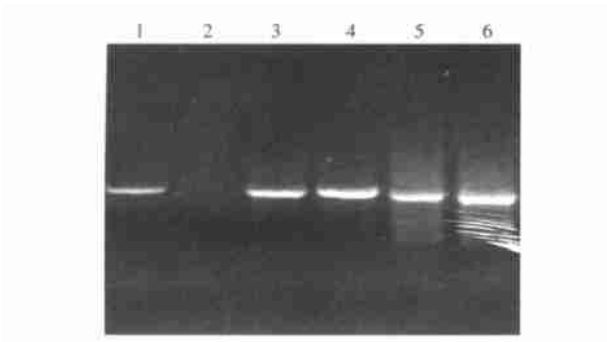
采用 χ^2 检验。

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

2 结果

2.1 片段的扩增

用简并引物从 channel catfish genomic DNA 中成功扩增出 3 段 *rag*1 基因的片段: REV-511, 511-817, 817-994。在扩增过程中发现, 低退火温度下的扩增产物常有多个条带, 随着退火温度的升高, 有些条带逐渐模糊乃至消失(以片段 511-817 为例, 见图 1), 将与目的片段大小相符的条带回收, 以之作模板, 扩增产物为清晰的单一条带。



1, 2: Product of gel extraction; 3, 4: 48 ℃; 5, 6: 50 ℃

图 1 *rag*1 片段 511-817 在不同温度退火的 PCR 产物电泳图

Fig 1 Agarose gel analysis of *rag*1 gene 511-817 PCR production

2.2 双向测序结果

片段 REV-511 有 1 718 bp; 片段 511-817 为 977 bp; 片段 817-994 为 555 bp, BLAST 分析显示它们与其它物种 *rag*1 基因的相似程度高。三段序列拼接, 共获得 3 218 bp 的 channel catfish *rag*1 基因的 DNA 序列。由于篇幅的原因, 详细的序列资料请到 GenBank 中查找: AY423858, AY423859, AY423860。

2.3 生物信息学方法分析结果

2.3.1 开放性阅读框架(ORF)分析 找到一个 2 235 bp 的 ORF, 起始密码子为第 624 位起的 ATG, 终止密码子为第 2 856 起的 TGA, G+C=49.3%, 编码 745 个氨基酸。

2.3.2 内含子分析 发现一个内含子(从第 228 位碱基至第 520 位碱基, 共 293 bp), 相比斑马鱼 *rag*1 基因的内含子(AF074330, gi: 3329499)、虹鳟鱼 *rag*1 基因的内含子(U15663, gi: 558917)要短的多(293 bp: 1 085 bp: 666 bp)。做三个内含子序列的多序列比较, 结果三者相似率为 22.9% (67/293), Channel catfish 内含子序列与 zebrafish、rainbow trout 内含子序列的两两比较, 相似率分别为 49.1% (144/293) 和 44.3% (127/293), 而 zebrafish 内含子与 rainbow trout 内含子的相似率为 50.1% (334/666), 统计学处理显示, 有显著性差异($P<0.01$)。

2.3.3 相似性/同源性分析 BLAST-X 分析显示与其它物种 *rag*1 基因序列的相似性(两两比较)多超过 70%, 而预测的氨基酸序列与 GenBank 中其它物种 *rag*1 的氨基酸序列相似性分析(BLAST-X)结果显示核心区域的相似率有些甚至大于 80%, 见表 1。

表 1 水渠鲇鱼 *rag*1 基因推导氨基酸序列与其它物种的一致性百分率

Tab 1 Identity percentage of deduced amino acid sequences of *rag*1 in the channel catfish with that of other species

Species	Identity percentage (%)	Identity percentage of core region (%)	Gap	Bit scores
Zebrafish	76	86	3	1 262
Rainbow trout	74	83	16	1 230
Takifugu	68	80	9	1 120
Chicken	56	69	14	880
Cathara skua	55	69	14	862

2.3.4 *rag*1 基因变异程度与硬骨鱼系统进行关系的分析 多序列比较[CLUSTAL W (1.82) multiple sequence alignment] 四者碱基一致的位点共 1 285 个, 占 43.9% (1 285/2925), 自第 1 352 位碱基至 2 925 位碱基(去除内含子后的 channel catfish *rag*1 序列)的相似率为 53.1% (836/1 574), 而自第 1 位碱基至 1 351 位碱基的相似率为 33.2% (449/1 351), 有非常显著的差异($P<0.001$)。图 2 直观地反映了 *rag*1 基因在氨基一侧和羧基一侧变异程度的差异。图 3 反映了 13 种硬骨鱼的亲缘关系的亲疏。

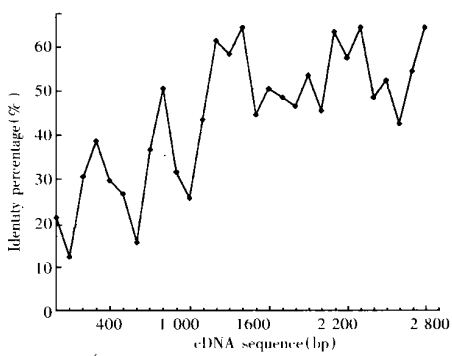


图 2 水渠鲇鱼、斑马鱼、虹鳟鱼及牛鲨 *rag*1 基因 cDNA 序列每 100 bp 一致位点百分率的折线图

Fig 2 A line diagram showing the percentage identity of *rag*1 cDNA among channel catfish, zebrafish, rainbow trout and bull shark

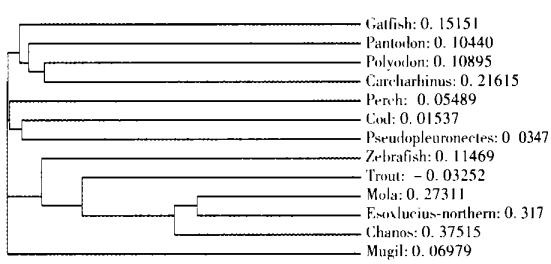


图 3 13 种硬骨鱼的系统进化树

Fig 3 The unrooted phylogenetic tree of *rag*1 sequences in the bony fishes

将所获 channel catfish recombination activating gene1 DNA sequence 提交 GenBank, 申请到 3 个序列号: AY423858, AY423859, AY423860

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

3　讨论

RAGs, 尤其是 *rag*1 基因进化缓慢, 高度保守, 因此可以通过已知物种 *rag*1 基因及 *rag*1 蛋白的序列资料和比对结果, 设计简并引物(degenerate primer)来扩增和克隆其它物种的 *rag*1 基因片段。我们就是根据多序列比较的结果, 找出 *rag*1 基因的保守区域, 参照斑马鱼等硬骨鱼类 *Rag*1 蛋白氨基酸密码子的偏好性, 按照简并引物设计的一般原则, 如尽量避开多个密码子编码的氨基酸, 避免引物 3′末端的简并, 在高简并位点用 I(脱氧次黄嘌呤)替代等, 设计出用来扩增 channel catfish *rag*1 基因片段的一组简并引物。

在用简并引物作 PCR 扩增时, PCR 的热力学框架能显著地改变扩增的成功率。要先在非严谨退火温度下(35～45 ℃)下进行 2～5 个循环, 然后再在较严谨退火温度下进行 25～40 个循环[8]。这是因为低退火温度可使引物的短互补区域与靶序列杂交, 扩增几个循环后 5′端延伸部分就可掺入扩增产物内, 可作为后续扩增的模板, 再换到高退火温度则可提高其特异性。我们参照文献[8]的方法, 分析了每条简并引物简并位点的多少、位置以及它们对扩增效率的影响, 采取了适当降低退火温度、提高反应体系中引物的浓度(相应提高 Mg^{2+} 浓度)和延长变性反应时间、增加延伸时间等措施, 成功扩增出了 channel catfish genomic DNA 中 *rag*1 基因的片段。

生物信息学(bioinformatics)[9]是运用数学的原理、方法以及 IT 的工具技术, 综合基因组计划已获得的海量数据, 对生物大分子信息的获取、处理、储存、分发、分析以及解释等方面进行全方位综合阐释, 以了解其生物学意义, 破解"生命之码"的一门科学。我们用多种生物信息学的方法对所得 channel catfish *rag*1 基因的 DNA 序列进行分析。在该段序列中找到一 2 235 bp 的开放阅读框架和一 293 bp 的内含子, 在这一内含子的两侧有正确的 RNA 剪接位点: 一个 GT 供体位点和一个 AG 受体位点。内含子序列的相似性比较可以显示物种之间在进化上的关系, 3 个内含子序列的多序列比较显示相似率为 22.9%, 但 channel catfish *rag*1 的内含子与 zebrafish 的相似率为 49.1%, 与 rainbow trout 的相似率为 44.3%, 而 zebrafish 与 rainbow trout 的相似率为 50.1%。硬骨鱼中只有斑马鱼的 *rag*1 有两个内含子, 它的下游内含子虽然稍大, 但与我们找到的 channel catfish *rag*1 基因的内含子以及 rainbow trout *rag*1 基因的内含子位置相当, 序列相似, 提示在进化上三者分开之前有相同的起源。所得序列的 BLAST-N 和 BLAST-X (去除内含子后)分析结果显示, channel catfish 的 *rag*1 基因与其它物种的相似程度高(> 70%), 而 *Rag*1 蛋白核心区域的相似率达 80%以上, 又一次证明了 *rag*1 基因进化缓慢、高度保守。

多序列的比较比两两比较更能找出序列的保守区域和评估物种之间的进化关系。所得 channel catfish *rag*1 基因的 DNA 序列去除内含子后与 zebrafish、rainbow trout、bull shark *rag*1 基因 cDNA 全序列做多序列比较(CLUSTAL W multiple sequence alignment), 结果四者一致的碱基位点占 43.9%, 而自第 1 352 位碱基至 2 925 位碱基的相似率为 53.1%, 自第 1 位碱基至 1 351 位碱基的相似率为 33.2%($P < 0.001$), 有非常显著的差异, 提示不同物种 *rag*1 基因的变异相对集中在氨基末端的区域。进化树又叫种系发生树[(unrooted) phylogenetic tree], 利用不同物种同一基因序列的异同来构建进化树, 可用于研究生物的进化, 确定各物种亲缘关系的亲疏。它的算法主要有两类, CLUSTAL W 采用是独立元素法(discrete character methods)中的"最大简约法", 在多序列比较结果的基础上绘出种系发生树。根据 13 种硬骨鱼的 *rag*1 基因序列资料比较结果绘出的种系发生树, 反映了这 13 种硬骨鱼在进化上关系的亲疏。

参考文献:

[1] 金伯泉. 细胞和分子免疫学[M]. 第 2 版. 北京: 科学出版社, 2001. 98—101.

[2] Gartner F, Alt F W, Monroe R J, *et al*. Antigen-independent appearance of recombination activating gene (RAG)-positive bone marrow B cells in the spleens of immunized mice[J]. J Exp Med, 2000, 192(12): 1745—1754.

[3] Hayman J R, Lobb C J. Heavy chain diversity region segments of the channel catfish: structure, organization, expression and phylogenetic implications[J]. J Immunol, 2000, 164(4): 1916—1924.

[4] Ghaffari S H, Lobb C J. Structure and genomic organization of a second cluster of immunoglobulin heavy chain gene segments in the channel catfish [J]. J Immunol, 1999, 162(3): 1519—1529.

[5] Petrie H L, Ainsworth A J. Ontogeny of channel catfish lymphoid organs [J]. Vet Immunol Immunopathol, 2001, 81(1—2): b 113—127.

[6] Wilson M R, Zhou H, Bengten E, *et al*. T-cell receptors in channel catfish: structure and expression of TCR alpha and beta genes[J]. Mol Immunol, 1998, 35(9): 545—557.

[7] Willett C E, Cherry J J, Steiner L A. Characterization and expression of the recombination activating genes (*rag*1 and rag2) of zebrafish[J]. Immunogeneties, 1997, 45(6): 394—404.

[8] 迪芬巴赫 C W, 德维克斯勒 G S. PCR 技术试验指南[M]. 北京: 科学出版社, 1998. 88—101.

[9] Valencia A. Bioinformatics: biology by other means[J]. Bioinformatics, 2002, 18(12): 1551—1552.

(编辑　张　宁)

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net