

鳙小卫星 pBC174 的序列结构特性分析

简纪常¹, 夏德全²

(1. 湛江海洋大学水产学院, 广东 湛江 524025;

2. 中国水产科学研究院淡水渔业研究中心, 江苏 无锡 214081)

摘要:用双脱氧 DNA 测序法对鳙 (*Aristichthys nobilis*) 小卫星 pBC174 进行序列测定。其长度为 1 873 bp。在这段序列中, A-T 含量高达 64%, 从 1 511 碱基处开始, 含有多聚腺嘌呤-多聚胞嘧啶, 即 (AC)₁₃ 的短重复序列的隐蔽微卫星。序列中含有 20 种限制性内切酶的 57 个酶切位点, 97 个不同的正向重复序列, 其中 71 个 8 bp, 10 个 9 bp, 10 个 10 bp, 2 个 11 bp 和 4 个 17 bp。根据序列分析, 认为小卫星 pBC174 是由富含 A 的重复序列 TAAAGAAA 和富含 T 的重复序列 TTTTACA 等 2 个不同的核心序列组成。

关键词:鳙; 小卫星 pBC174; 酶切位点; 正向重复序列; 核心序列

中图分类号: Q959.4

文献标识码: A

文章编号: 1005-8737(2002)02-0186-04

在真核生物的基因组中, 含有大量的非编码重复 DNA 序列。这种重复序列一般位于染色体的中心粒和端粒^[1,2], 通常占基因组的 1%~7%, 在基因组中的拷贝数在 10 至 10⁶ 之间变化。这种非编码的重复序列或者是由简单重复序列如 (AC)_n 或 (CT)_n 组成, 或者具有单体串联重复特征, 串联单体的大小在几个 bp 到几千个 bp 之间变化。

非编码的重复 DNA 序列在生物进化或遗传过程中, 不受自然选择压力的影响, 故进化速度很快, 因此在生物种群结构分析、亲缘关系鉴定、监测遗传育种操作效应、寻找与控制主要性状的基因相连锁的分子遗传标记等研究方面具有重要的作用, 已引起了生物学家的极大关注。在各种重复序列中, 碱基的含量不尽相同, 如一些富含 G-C^[3~5], 一些富含 A-T^[6]。通过比较不同小卫星 DNA 的序列, 可以研究生物的进化, 特别是近缘物种的鉴定^[7]。本文着重分析克隆的鳙小卫星 pBC174 的序列组成及其结构特点, 旨在促进 pBC174 在鱼类分子遗传学和分子育种学上的广泛应用。

1 材料与方法

1.1 鳙小卫星 pBC174 的克隆

见文献[8]。

1.2 试剂与引物

荧光脱氧标记终止循环测序试剂盒购自美国 Perkin

Elmer 公司。第 1 次测序反应引物为 M₁₃ 反向测序引物, 序列为 5'-GCCGATACAATTTTCACACAGGA-3'。第 2 次测序引物是根据第 1 次测序结果设成的, 第 3、4 次引物以此类推。第 2 次测序引物为: 5'-TTCAAGCTCAAAC-CGGTGTG-3', 第 3 次测序引物为: 5'-CCTG-GTTCTTCTTCAAGGC-3', 第 4 次测序引物为: 5'-TGA-CAAGTTATCTGGATGG-3'。

引物合成在 Beckman 公司的 Oligo 1000M DNA 合成仪上进行, 序列测定在 ABI 373A 自动测序仪(美国生物系统公司产品)上进行。

1.3 测定与计算

双脱氧法测定 pBC174 序列的具体操作程序按测序试剂盒的说明书进行。用 Goldkey DNA 分析软件(ver. 1.0)对小卫星 pBC174 序列进行分析。

2 结果

电泳测得鳙小卫星 pBC174 的大小约为 1 900 bp, 故从小卫星 pBC174 的反向进行连续 4 次测序, pBC174 的序列(1 873 bp)结果见图 1。在 pBC174 的序列中, A-T 的含量约 64%。用 Goldkey DNA 分析软件对 pBC174 的序列进行分析, 发现从 1 511 位置开始, 含有多聚腺嘌呤-多聚胞嘧啶即 (AC)₁₃ 的简单重复序列, 可能是隐蔽微卫星。在 pBC174 的序列中, 有 97 个不同的正向重复序列, 结果见表 1。在整个序列中, 含有 20 种限制性内切酶的 57 个酶切位点, pBC174 的酶切位点见表 2。分析重复序列的数量和组成后, 认为小卫星 pBC174 是由富含 A 的重复序列 TAAAGAAA 和富含 T

收稿日期: 2001-09-10。

作者简介: 简纪常(1964-), 男, 副研究员, 博士, 从事水产动物免疫学和分子生物学方面研究。E-mail: jianjichang@21cn.com

的重复序列 TTTTACA 等 2 个不同的核心序列组成。

```

1 CTACAAAACATATTTCTTGTGAAGAGTACTTGATTTTGAACCTTCTATACATACTCAT
61 AACTTCTTCTGAATTCACATAATCAAGGAAGAAATCCATATTTGACCTTTTGAACAAAT
121 TACGACCAAAAGAGTCTTCGATCTTGAACCTTCTTACAIACCTACTGACCTCATCTGAIG
181 TTTAATGAACCATAAATAGTCCATATTTGACCTTTTAAATGTATTTAAGGTCAAAAAGT
241 GCTCAGGGGGTAATTACTATATATACTCTTAGACICGTCTGTATGTATAACGAAGA
301 GAGTACTCAATTTTGAACATACTGAACACATTTGAGGCCAAAGAGCCCTTGATCATCAIG
361 TTTGAGCAAAACAATATATATCAGACCIAATGGAAATAATAGTATTTACCCCAAACTA
421 GTCTGCCTTGTCTTAAAAAATAAATGTTCAAGCATGTTCAAGCTCAAACCGGTGTGACG
481 CGTTTGTCTACGTTTCCAGGTGAATGGCATGTTCCTGGAAACGGTGTGCAACGGGIA
541 CGTTTCTCTTGTGTGGTGGGTGTGTTCAAAAATCAGCCCAATCAGCAGCAACATGTAT
601 ATAAACCAAGCGGTATTAAGAAACAGCTCATAAATGTAATTAACATCAAAAATAAATCA
661 CAAGAGCAAGTATATGTTTGCACATGTATATTTACATTAATGGCAAAAATAACTATAAT
721 AAGAGCACAAGTATAGATCATGTTCACAGTCCACATATCAACATATAATTTCTTGCAGA
781 CGTAAAGCAGAAATATTAAGAGTGAATATACTGTGTGTTAAAGAAAGTGAACTTAGAAC
841 AAAATGAGATATAAATCTAGATATTATGACGATCCAACATCTCTTTGTTTCTCTAG
901 TGTGTTTCAATCCTTCGCCCTGGTCTCTTCAAGGCTCCCATTTTCTTACAAGGTC
961 TTCAAATTTCTTCACAGCTTTTGTCTCTGACCTCCTTCATGGICTGCGCTGATGCTTCA
1021 TGCTTCATTATACCTTTTACACCCGGAGCAGTGGGCAGCCATATGCTTTGGCATTTG
1081 AGGGGTTAGGTGTCTTGTCTCAAGTGCACCTCAGTTATTTCTGCCAGTGTGGAATCGAA
1141 CCAGACTCCTTGCTTTCTATCTGGTTATGCAAAATGGCCATCTGATATCAGAAATGGTCA
1201 AAAGTTTATCAGTTAAAGAAAGCTTTTGCTAATATCACAATCAGATATCAAGTACCTGTTC
1261 ATTTTGTGATGCTTCTTGTAGCATACCTTAAATGACAAGTTATCTGGATGGAATTTCT
1321 GTATTTTCTATCCCTACACTAACCTACCCCTAAACCTACCCATCAGAAAATCTCTCT
1381 GCATTTTACATTTTCAAAAAAACTCATTTAGTATGATTTATAAGTTTGTTCCTCCAT
1441 AGGAACCTCAATTTAGGTCCCCAGGTGACATGAGTCTCATGAGTCTTTGTGATCCAG
1501 ATTGACATGTATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
1561 TTTTATATCTGTACAACTCTATATCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
1621 CAAGTCACTTTATTTATATAGTGTCTTTTACAAATCAGATTTTCAAAAGCAGCTTTTAC
1681 AGGGATAACAGGAAATAATTCACAACGTTTGTGTTGGGCTGTACAGCAGCTCTAGAAG
1741 ATAATAGTGGGATTGTCCAGCTTAAGAAAGTCCAGTGTGATTTGTCTCCATGTAAAT
1801 CATTAGTTATTAATTTAATTCATTTATCTATAGCAGCTCTGGAGAAAACGGTGTATGTCAT
1861 CGTCCAGTCTCAGT

```

图1 pBC174 的核苷酸序列

Fig.1 Nucleotide sequence of pBC174

3 讨论

小卫星 DNA 遍布于整个基因组中, 数量较多。不同的小卫星其碱基组成也不同, 有的 G-C 含量丰富^[4,5], 有的 A-T 含量丰富^[6]。在所克隆的鳊小卫星 pBC174 中, A-T 含量较高, 达 64%。这与鲤 (63%)^[9]、绿鲮 (65%)^[10]、鲢属鱼 (64%)^[11]、金头鲷 (67%)^[12] 等的小卫星 DNA 中 A-T 的含量很相似。说明鳊的 pBC174 小卫星可能与这些鱼的小卫星来源相似, 都是 A-T 含量高的小卫星 DNA。小卫星 pBC174 序列中内切酶的位点较多, 含有 20 种内切酶的 57 个酶切位点, 这些丰富的酶切位点对其进一步亚克隆或与其他外源 DNA 连接提供了更大的可能性。

对小卫星 pBC174 序列分析后, 没有发现类似于人源小卫星 33.15 的核心序列, 即 AGAGGTGTGGCAGGTGG^[4]。但在序列中含有 97 个不同的 8、9、10、11 和 17 bp 的正向重复序列。根据碱基组成的特点, 将其分成富含 A、T 和其他等 3 类。富含 A 的重复序列主要位于 850 bp 之前, 而富含 T 的重复序列主要位于 850 bp 之后。在这些富含 A 和 T 的

重复序列中, A-T 含量平均达 81%, 而其他重复序列的 A-T 含量只有 65%, 这与整个小卫星 pBC174 的 A-T 含量相近。

8 bp 的重复序列占总重复序列的 73%, 根据 8 bp 重复序列的重复次数和碱基组成, 认为富含 A 的核心序列可能是 TAAAGAAA, 这个核心序列与莫桑比克罗非鱼小卫星的核心序列 CTGAAAC^[12]、绿鲮的小卫星核心序列 TCT-GAAAAG^[10] 和金头鲷的小卫星核心序列 CTGAAAA^[2] 相似。正因为 pBC174 的核心序列与莫桑比克罗非鱼的核心序列相似, 所以 pBC174 与尼罗罗非鱼基因组杂交, 产生较多的 DNA 指纹图带 (另文发表)。富含 T 的核心序列可能是 TTTTACA, 这与美洲鲮的小卫星核心序列 GTTTT 和 TTTTG 碱基组成相似^[13]。其他富含 A 或 T 的 8 bp 重复序列可能是由这 2 个核心序列分别演变而来的。

值得注意的是, 小卫星 pBC174 在 1 511 处存在着连续重复的多聚腺嘌呤-多聚胞嘧啶即 (AC)₁₃ 的隐蔽微卫星, 该简单重复序列 (AC)₁₃ 亦称为微卫星, 类似于 Bentzen^[3] 克隆的鲑小卫星 Ssal 3' 的 (CA)_n 结构。由于其变异程度高, 数量多, 因而是确定动植物个体间相互关系以及进行群体遗传学

研究有效的新遗传标记^[14,15]。(AC)_n 微卫星在哺乳动物上也得到较好的研究^[16]。张亚平等^[17]用人工合成的微卫星 (AC)₁₅ 作探针,从大熊猫 DNA 文库中筛选到 10 个微卫星座位,设计了 10 个微卫星的特异性引物,对大熊猫的 DNA 样品进行 PCR 扩增,发现有 9 个呈现多态性,并且这 10 个微卫星 DNA 座位严格按照孟德尔方式遗传,能有效地鉴定大

熊猫的亲缘关系。(AC)_n 在分子遗传学的研究上具有广阔的应用前景,它分散在真核基因组中,如在蝇的 X-染色体上存在着 (AC)_n^[18]。鱧小卫星 pBC174 中的 (AC)₁₃ 微卫星在鱼类染色体上的定位、分布还有待进一步的研究,但它作为遗传标记在鱼类遗传学上具有潜在的应用价值。

表 1 pBC174 中的正向重复序列

Table 1 Direct repetitive units within pBC174

位置 Location	富含 A 的序列 A-rich sequences	大小/bp Size	位置 Location	富含 T 的序列 T-rich sequences	大小/bp Size
1	CTACAAAA	8	35	TTTGAAC	8
645	CTACAAAA		108	TTTGAAC	
127	CAAAGAGT	8	668	TTTATTTA	8
233	CAAAGAGT		1630	TTTATTTA	
339	CAAAGAGT		692	TTTACATT	8
398	ATAATAGT	8	1386	TTTACATT	
1741	ATAATAGT		882	TTCTTCIT	8
441	ATAAATGT	8	925	TTCTTCIT	
631	ATAAATGT		355	ATCATGTT	8
822	AAGAAAGT	8	737	ATCATGTT	
1764	AAGAAAGT		888	TITGTTTT	8
648	CAAAATAA	8	1710	TTTGTTTT	
705	CAAAATAA		891	TTTTTCTA	8
434	TAAAAAAA	8	946	TTTTTCTA	
1398	TAAAAAAA		1036	TTTTTACA	8
617	TAAAGAAA	8	1384	TTTTTACA	
820	TAAAGAAA		1646	TTTTTACA	
1214	TAAAGAAA		359	TGTTTGCA	8
217	TAAATGTA	8	676	TGTTTGCA	
632	TAAATGTA		306	CTCAATTT	8
798	AAAGTGAA	8	1447	CTCAATTT	
825	AAAGTGAA		1427	GTTTGTTT	8
437	AAAATAAAT	9	1709	GTTTGTTT	
649	AAAATAAAT		768	AATTTCTT	8
819	GTAAAGAAAG	10	965	AATTTCTT	
1213	GTAAAGAAAG		1035	CTTTTACA	9
1258	TTCATAAAAA	10	1645	CTTTTACA	
1394	TTCATAAAAA		1478	TTTGTTC	9
			1782	TTTGTTC	
其他序列 Other sequences			其他序列 Other sequences		
521	AAACGGTG	8	12	ATTTCAAG	8
1864	AAACGGTG		769	ATTTCAAG	
667	CAAGTATA	8	115	CAAATTC	8
728	CAAGTATA		963	CAAATTC	
808	TATACTGT	8	471	CGGTGTGCA	9
1565	TATACTGT		525	CGGTGTGCA	
1470	CATGAGTC	8	1727	AGCAGCTCT	9
1480	CATGAGTC		1832	AGCAGCTCT	
1755	GTCCAGCT	8	49	TACATACTCA	10
1862	GTCCAGCT		155	TACATACTCA	
24	AGAGTACT	8	126	CCAAAGAGTC	10
300	AGAGTACT		341	CCAAAGAGTC	
288	TATATCAG	8	445	ATGTTCAAGA	10
377	TATATCAG		455	ATGTTCAAGA	
289	ATATCAGC	8	37	TTGAACCTTCT	11
572	ATATCAGC		143	TTGAACCTTCT	
378	ATATCAGA	8	95	TCCATATTGACCTTTT	17
1185	ATATCAGA		201	TCCATATTGACCTTTT	
71	AATTCACA	8	1353	TAAACCTACCCATCACA	17
655	AATTCACA		1601	TAAACCTACCCATCACA	

表 2 pBC174 中的内切酶及其位点

Table 2 Endonucleases and endonuclease sites of pBC174

内切酶 Edonuclease	酶切位点位置 Location of endonuclease sites
<i>Aat</i> I	935
<i>Dpn</i> I	139, 351, 736, 873
<i>Dpn</i> II	139, 351, 736, 873
<i>Eco</i> R I	71
<i>Eco</i> R II	497, 517, 920, 1160, 1461
<i>Hae</i> I	335, 935, 1175
<i>Hae</i> II	336, 936, 1176
<i>Hinc</i> II	742
<i>Hind</i> II	742
<i>Hind</i> III	1220
<i>Hinf</i> I	131, 167, 273, 343, 1133, 1144, 1473, 1483
<i>Hinf</i> II	1220
<i>Hpa</i> II	470, 1045
<i>Mae</i> I	418, 858, 897, 1734
<i>Mae</i> II	540, 780, 1707
<i>Mae</i> III	1466, 1624
<i>Oxa</i> I	462, 626, 976, 1221, 1673, 1730, 1759, 1835, 1866
<i>Rhc</i> I	1479
<i>Sst</i> IV	350
<i>Xba</i> I	857, 1733

参考文献:

- [1] Garrido-Ramos M A, Jamilena M, Lozano R. Cloning and characterization of a fish centromeric satellite DNA[J]. Cytogenet Cell Genet, 1994, 65: 233 - 237.
- [2] Willard H F. Centromeres of mammalian chromosomes[J]. Trends Genet, 1990, 6: 410 - 4.
- [3] Bentzen P, Wright J M. Nucleotide sequence and evolutionary conservation of a minisatellite variable number tandem repeat cloned from Atlantic salmon, *Salmo salar* [J]. Genome, 1993, 36: 177 - 271.
- [4] Jeffreys A J, Wilson V, Thein S L. Hypervariable minisatellite regions in human DNA[J]. Nature, 1985, 314: 67 - 73.
- [5] Nakamura Y. Variable number of tandem repeat (VNTR) marker for human gene mapping[J]. Science, 1987, 235: 1 616 - 1 622.
- [6] 孙乃恩, 孙东旭, 朱德煦. 分子遗传学[M]. 南京: 南京大学出版社, 1990.
- [7] Franck J P, Wright J K. Conservation of a satellite DNA sequence (SATB) in the tilapia and haplochromine genome (Pisces: Cichlidae) [J]. Genome, 1993, 36: 187 - 194.
- [8] 简纪常, 夏德全. 鱮小卫星 DNA 的克隆[J]. 中国水产科学, 1999, 6(4): 18 - 20.
- [9] Datta U, Dutta P, Mandal P K, et al. Cloning and characterization of a highly repetitive fish nucleotide sequence[J]. Gene, 1988, 62: 331 - 336.
- [10] Denovan E M, Wright J M. A satellite DNA family from pollock (*Pollachius virens*) [J]. Gene, 1990, 87: 279 - 283.
- [11] Goodier J L, Davidson W S. Characterization of a repetitive element detected by Nhe I in the genomes of salmo species [J]. Genome, 1994, 37: 639 - 645.
- [12] Wright J M. Nucleotide sequence, genomic organization and evolution of a major repetitive DNA family in tilapia (*Oreochromis mossambicus*) [J]. Nucleic Acids Research, 1989, 17 (13): 5 071 - 5 079.
- [13] Moyer S P, Ma D P, Thomas T L. Characterization of a highly repeated satellite DNA from the Cyprinid fish *Notropis lutrensis* [J]. Comp Biochem Physiol, 1988, 91(4): 639 - 646.
- [14] Bowcock A M, Weber J S, Ford O R. High resolution of human evolutionary trees with polymorphic microsatellites [J]. Nature, 1994, 368: 455 - 457.
- [15] Weber Y P, May P E. Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction [J]. Am J Hum Genet, 1989, 44: 388 - 396.
- [16] Echt C S, May-Junquardt P. Survey of microsatellite DNA in pine [J]. Genome, 1997, 40: 9 - 17.
- [17] 张亚平, 王文, 宿兵, 等. 大熊猫微卫星 DNA 的筛选及其应用[J]. 动物学研究, 1995, 16(4): 301 - 306.
- [18] 刘树俊, 程光潮. DNA 指纹技术中所用的探针及其发展[J]. 遗传, 1994, 16(2): 40 - 43.

Analysis on sequence characters of pBC174 of bighead carp *Aristichthys nobilis*

JIAN Ji-chang¹, XIA De-quan²

(1. Fisheries College of Zhanjiang Ocean University, Zhanjiang 524025, China;

2. Freshwater Fisheries Research Center, Chinese Academy of Fishery Sciences, Wuxi 214081, China)

Abstract: The minisatellite pBC174 of bighead carp (*Aristichthys nobilis*) was sequenced using Sanger method. Its size is 1 877 bp in length. The A-T content is up to 64%. Within the pBC174, there is a cryptic microsatellite (AC)₁₃ at the 1 511 position, and 57 restriction endonuclease sites of 20 endonucleases exist in this sequence. The pBC174 contains 97 different direct repetitive sequences, among which there are 71 direct repeats of 8 bp, 10 of 9 bp, 10 of 10 bp and 4 of 17 bp. According to sequence analysis, A-rich repeat TAAAGAAA and T-rich TTTTACA are two different core sequences which constitute minisatellite pBC174.

Key words: *Aristichthys nobilis*; minisatellite pBC174; endonuclease site; direct repetitive sequence; core sequence