

15种鱼转铁蛋白 cDNA 序列密码子使用比较分析

龙 华¹, 王祖荣²

(1. 中国水产科学研究院长江水产研究所, 农业部淡水鱼类种质资源与生物技术重点开放实验室, 湖北 荆州 434000; 2. 长江大学计算机科学学院, 湖北 荆州 434023)

摘要: 运用 DNAMAN v4.0 分析软件, 对克隆的鲫鱼血清转铁蛋白 cDNA 序列与文献中的 14 种鱼转铁蛋白 cDNA 序列的同源性、密码子使用频率和密码子使用偏性进行分析, 结果表明: 鲤形目鱼类 Tf 的 cDNA 序列同源性在 46% 以上, 而鲢形目鱼类 Tf 的 cDNA 序列同源性在 34%~99% 之间, 变化较大。密码子使用频率越高或使用偏性越小, 说明该密码子编码的氨基酸对蛋白质空间结构的形成和基本生理功能的影响也越重要; 在进化上分化时间越早的物种, 密码子使用频率和使用偏性的差异较小, 反之则较大; 在同源性方面, 在进化上比较接近的物种, 基因的密码子使用频率和使用偏性指标比较接近或基本相同。根据对 15 种鱼 Tf 的 cDNA 序列同源性和密码子分析, 推测 Tf 基因的进化顺序是: 鲤形目、鲢形目、鲢形目和鲮形目。

关键词: 鱼; 转铁蛋白; cDNA; 密码子; 使用

中图分类号: Q786

文献标识码: A

文章编号: 1673-2227-(2005)01-0006-05

Comparative analysis of codon usage of 15 fish transferrin cDNA sequences

LONG Hua¹, WANG Zu-rong²

(1. Key Opening Laboratory of Freshwater Fish Germplasm Resources & Biotechnology of Ministry of Agriculture, Yangtze River Fisheries Institute of Chinese Academy of Fishery Sciences, Jingzhou, Hubei 434000, China;

2. College of Computer Science, Yangtze University, Jingzhou, Hubei 434023, China)

Abstract: The analysis of the identity, the codon usage frequency and the codon usage preference were made to sequences of *Carassius auratus auratus* serum transferrin (Tf) cDNA cloned in our laboratory and 14 fish Tf cDNA from reference documentation by DNAMAN v4.0. The results show that the identity of Tf cDNA sequences among Cypriniformes fishes is more than 46% and the identity of Tf cDNA sequences among Oncorhynchus fishes is between 34%~99% with larger variation. The higher the codon usage frequency or the lower the codon usage preference, the more important the effect of amino acids coded by the codon to form protein space structure and perform basic physiological functions. Furthermore, the earlier the species differentiation time on evolution, the lower the difference of the codon usage frequency and preference and the higher by contraries. Moreover, species with near homology relationship on evolution share the near codon usage frequency and preference index or the similar index. According to the codon usage frequency and the codon usage preference analysis, the evolution sequence of Tf gene, viz. Cypriniformes, Melanogrammus, Oncorhynchus and Paralichthys, were presumed.

Key words: fish; transferrin; cDNA; codon; usage

收稿日期: 2005-03-17

资助项目: 国家“973”计划资助项目(2004XB117400)

作者简介: 龙 华(1964-), 男, 研究员, 硕士, 主要从事鱼类分子生物学研究。E-mail: longhua0000@yahoo.com.cn

鱼类转铁蛋白 (transferrin, Tf, 又称为铁传递蛋白) 是鱼类血清中一种非血红素结合铁的 β -球蛋白, 单一肽链, 分子量为 70 ~ 80 kD, 有 2 个三价铁离子结合位点, 是鱼体内铁的运输者。已经证实, 鱼类血清 Tf 的血清铁浓度、铁饱和度与鱼类的耐低氧性能和栖息水层有明显的关系, 并具有抗菌功能^[1]。20 世纪 80 年代后期, 人们开始对各类 Tf 基因的结构、序列以及表达以及 Tf 受体的功能等进行较为深入的研究^[2]。由于遗传密码的简并规则具有进化上的稳定性^[3], 即使在蛋白质的编码过程中, 同义密码子的使用频率和概率并不相同。密码子使用频率越高, 说明该密码子对该基因的结构和功能越重要, 在进化上越保守; 密码子使用偏性越小, 同样说明了上述问题。本项研究在克隆鲫鱼血清 Tf 的 cDNA 基础上, 应用 NCBI 数据库和计算机分析软件对 GenBank 序列数据库中的鱼类 Tf 的 cDNA 序列的密码子使用频率和密码子使用偏性进行分析, 进行基因的同源性和进化的分类, 还可以分析蛋白质的空间结构。

1 材料与方法

1.1 鱼类种类与序列来源

本文所采用的 Tf 的 cDNA 序列来源见表 1。

1.2 数据分析

将克隆的 cDNA 序列提交 GenBank 序列数据库中, 以 BLAST 程序进行核苷酸序列同源性比较。同源性和密码子分析采用 DNAMAN v4.0 程序。

2 结果与讨论

2.1 15 种鱼 Tf 的 cDNA 序列同源性比较

对 15 种鱼 Tf 的 cDNA 序列进行碱基比较, 发现鱼类 Tf 的 cDNA 长度约为 2 600 bp 左右。在序列同源性方面, 鲤形目鱼类 Tf 的 cDNA 序列同源性在 46% 以上, 而鲱形目鱼类 Tf 的 cDNA 序列同源性在 34% ~ 99% 之间, 变化较大 (表 2)。相比之下, 克隆的鲫鱼 Tf 的 cDNA 序列与其它鱼 Tf 的 cDNA 序列同源性均在 44% 以上, 较其它鱼类之间有更高的同源性。

表 1 15 种鱼 Tf 的 cDNA 序列来源

Tab. 1 Resources of cDNA sequences of 15 species transferrins

编号 no.	鱼 名 species name	拉丁名 Latin name	Tf-cDNA 序列来源 resources of Tf- cDNA sequences	Tf-cDNA 序列号 accession number Tf- cDNA sequence	cDNA 长度 (bp) cDNA length
1	鲫鱼	<i>C. auratus</i>	本实验室 ^[4]		2 325
2	银鲫	<i>C. gibelio</i>	GenBank	AF518745. 1 GI:22087582	2 372
3	鲤鱼	<i>C. carpio</i>	GenBank	AF457152. 1 GI:18034629	2 206
4	草鱼	<i>C. idella</i>	GenBank	AY383546. 2 GI:46277824	2 387
5	青鳉	<i>O. latipes</i>	文献 ^[5]		2 353
6	黑线鳉	<i>M. aeglefinus</i>	GenBank	AJ300650. 1 GI:11877337	1 837
7	银大麻哈鱼	<i>O. kisutch</i>	GenBank	D89084. 1 GI:1694683	2 504
8	马苏大麻哈鱼	<i>O. masou</i>	GenBank	D89087. 2 GI:42662205	2 153
9	红大麻哈鱼	<i>O. nerka</i>	GenBank	D89085. 1 GI:5837766	2 429
10	虹鳟	<i>O. mykiss</i>	GenBank	D89083. 1 GI:5837758	2 634
11	褐鳟	<i>S. trutta</i>	GenBank	D89091. 1 GI:5837808	2 437
12	美洲红点鲑	<i>S. fontinalis</i>	GenBank	D89089. 1 GI:5837795	2 431
13	粗壮红点鲑	<i>S. namaycush</i>	GenBank	D89090. 1 GI:5837801	2 421
14	日本鲑	<i>S. pluvius</i>	GenBank	D89088. 1 GI:5837787	2 437
15	牙鲈	<i>P. olivaceus</i>	GenBank	D88801. 1 GI:3242406	2 296

2.2 15 种鱼 Tf 的 cDNA 序列密码子使用比较分析

通过对 15 种鱼 Tf 的 cDNA 中密码子使用频率和使用偏性最大和最小的前 3 个列表(见表 3),可以看出:使用频率最大的有 CUG(亮氨酸密码子)和 AAG(赖氨酸密码子),其次是 GAG、GAA 和 CAG(谷氨酸和谷氨酰胺密码子),还有 GCC(丙氨酸密

码子)和 UGG(色氨酸密码子)。使用频率最小的有 UAG 和 UAA(终止密码子),其次为 CGA(精氨酸密码子)、UCG(脯氨酸密码子)和 UCG(丝氨酸密码子);使用偏性最小的为 AUG(启动密码子)和 UGG(色氨酸密码子),使用偏性最大的有 UCG(丝氨酸密码子)及 CGC、CGG 和 CGA(精氨酸的密码子)。

表 2 15 种鱼 Tf 的 cDNA 序列同源性比较
Tab.2 Identity comparison of cDNA sequences of 15 fish transferrins

编号 no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1															
2	70.5														
3	58.6	75.6													
4	67.8	84.7	65												
5	46.6	58.4	65	67											
6	44.5	47.1	66	66	96										
7	46.6	52	66	63	62	63									
8	51.9	51	67	63	62	62	99								
9	47.1	35	33	36	32	33	34	38							
10	46.4	42	48	50	52	53	47	48	36						
11	47.9	40	49	52	51	49	49	47	37	80					
12	47.8	40	45	45	45	45	46	44	39	80	79				
13	48.3	44	52	54	50	49	51	50	38	71	72	67			
14	47.2	43	47	52	48	48	50	49	36	71	70	68	89		
15	49.9	36	34	33	35	32	33	34	35	33	34	35	32	33	

2.3 讨论

(1) 鲤形目鱼类之间 Tf 的 cDNA 序列同源性比较一致,而鲑形目鱼类之间 Tf 的 cDNA 序列变异较大,推测与鲑形目鱼类较鲤形目鱼类进化高等有关;鲤形目鱼类碱基 A 和 T 的含量高,而鲑形目鱼类碱基 G 和 C 的含量高,这与鲑形目鱼类进化更高等,其 cDNA 序列更稳定有关。

(2) 使用频率分析结果显示:使用频率最大的氨基酸(如亮氨酸、赖氨酸、谷氨酸、丙氨酸、色氨酸等)在蛋白质二级结构中形成 α -螺旋或转角的几率相对较大。使用频率最小的终止密码子可能并不是必要的,而精氨酸、脯氨酸和丝氨酸或极性较强或带正电荷较多,可能与电荷的平衡有关。

(3) 使用偏性分析结果显示:使用偏性最小的 AUG 启动密码子(甲硫氨酸)是不可缺少的,全部都使用(100%)。色氨酸为非极性并具有苯环结构,对维持空间结构极为重要;使用偏性最大的氨基酸(如丝氨酸和精氨酸),与密码子使用频率对比来看,丝氨酸和精氨酸对 Tf 的结构和功能可能并不十分重要。

(4) 应用密码子使用频率来分析基因的同源性和物种进化顺序,也是一种值得肯定的研究方法。分析结果表明:密码子使用频率越高或使用偏性越小,这些密码子在进化上被认为是最保守的,对基因的同源性影响极大。如在进化上分化时间越早的物种之间,其基因的密码子使用频率和/或使用偏性的

差异较小,反之则较大;在源性方面,在进化上比较接近的物种之间,其基因的密码子使用频率和/或使用偏性指标比较接近或基本相同。

(5)根据对 15 种鱼 Tf 的 cDNA 序列同源性和密码子分析,推测 Tf 基因的进化顺序是:鲤形目、鲱形目、鲑形目和鲾形目。用密码子使用频率和使用偏性分析得出的同源性和进化分类的结论,与 cDNA 序列源性分析得出的结论是吻合的。

(6)应用计算机技术研究不同物种同源基因的密码子信息来分析蛋白质的空间结构是可行性的。分析结果表明:密码子使用频率越高或使用偏性越小,说明该密码子对基因的结构和功能越重要;其编码的氨基酸对蛋白质空间结构(如 α -螺旋、电荷平衡、疏水性能等)的形成和基本生理功能(如受体结合、金属离子结合)的影响也越重要。

表 3 15 种鱼 Tf 的 cDNA 序列密码子使用频率和使用偏性比较分析
Tab.3 Comparative analysis of codon usage frequency and preference of 15 fish transferrin cDNA sequences

最大频率%		最小频率%		最小偏性%		最大偏性%			
编号	max usage	min usage	min usage	max usage	编号	max usage	min usage	min usage	max usage
no.	frequency	frequency	preference	preference	no.	frequency	frequency	preference	preference
	(%)	(%)	(%)	(%)		(%)	(%)	(%)	(%)
1	UGG(48)	CGA(1)	UGG(100)	CGA(2)	9	GAG(42)	CGA(0)	UGG(100)	CGA(0)
	AGA(36)	UAC(1)	AUG(100)	UCG(4)		AAG(37)	CCG(0)	AUG(100)	CCG(0)
	CAG(32)	UAG(4)	UAU(90)	UAG(5)		GCC(37)	UCG(4)	GAG(76)	UCG(5)
2	AAA(46)	UCG(1)	UGG(100)	UCG(1)	10	AAG(42)	CGG(0)	UGG(100)	CGG(0)
	AUG(43)	UAG(1)	AUG(100)	CGC(2)		GCC(39)	CGA(1)	AUG(100)	CGA(3)
	CAG(42)	CCG(1)	UGA(69)	CGG(2)		GAG(36)	CCG(1)	CAG(76)	CCG(3)
3	AUG(46)	CGC(0)	UGG(100)	CGC(0)	11	AAG(58)	CCG(0)	UGG(100)	CCG(0)
	AAG(45)	CCG(1)	AUG(100)	UCG(3)		GAG(39)	UAA(1)	AUG(100)	CUG(3)
	CUG(39)	UCG(3)	UAU(73)	CCG(3)		CUG(38)	CGA(1)	UUC(74)	CGA(4)
4	AAA(64)	UAG(0)	UGG(100)	UAG(0)	12	AGG(51)	CGA(0)	UGG(100)	CGA(0)
	GAU(42)	CGA(0)	AUG(100)	CGA(0)		CUG(44)	GAC(1)	AUG(100)	CUG(4)
	GAA(30)	UCG(3)	AAA(69)	UCG(3)		CCA(41)	UAG(2)	UCA(80)	CGC(4)
5	AAG(37)	CGA(1)	UGG(100)	CGA(2)	13	AAG(51)	UGA(0)	UGG(100)	UGA(0)
	CAG(34)	UAA(3)	AUG(100)	UCG(4)		CUG(42)	CGA(0)	AUG(100)	CGA(0)
	CUG(34)	UAG(3)	UUC(72)	CGU(4)		GCC(38)	UAG(1)	UAA(88)	ACG(2)
6	GCC(41)	UAA(0)	UGG(100)	UAA(0)	14	AAG(52)	CCG(0)	UGG(100)	CCG(0)
	GCU(31)	UAG(0)	AUG(100)	UAG(0)		GCC(39)	CGA(0)	AUG(100)	CGA(0)
	GAC(31)	UGA(0)	UAA(82)	UGA(0)		GAG(38)	CGG(0)	GAG(76)	CGG(0)
7	CAA(40)	UGG(1)	UGG(100)	UCG(2)	15	AAG(35)	GGC(1)	UGG(100)	GGC(4)
	GAA(36)	GCG(1)	AUG(100)	GCG(3)		CUG(34)	CGG(4)	AUG(100)	CGC(6)
	GGA(34)	ACG(2)	GAA(77)	CCG(7)		UGG(34)	CGC(4)	CAG68)	CGG(6)
8	AAG(47)	UAA(0)	UGG(100)	UAA(0)					
	GAG(42)	UGA(0)	AUG(100)	UGA(0)					
	GCC(39)	CCG(0)	AAC(81)	CCG(0)					

注:使用频率为 1 000 个密码子中的频率,使用偏性为 100 个密码子中的偏性。
Note: The codon usage frequency is frequency in 1000 codons and the codon usage preference is preference in 100 codons.

参考文献:

- [1] 龙 华,曾 勇,李 谷. 鱼类血清转铁蛋白的研究现状与应用前景[J]. 水产学报,2001,25(2):181-186.
- [2] 龙 华,郑 英. 转铁蛋白生理功能及其动态模型[J]. 湖北农学院学报,2004,24(1):30-35.
- [3] 特木尔巴. 同义密码子非均匀使用频率的研究[J]. 内蒙古民族师院学报(自然科学版),2000,15(2):152-156.
- [4] 龙 华. 鲫血清转铁蛋白 cDNA 克隆及系统发育进化序列分析[J]. 水产学报,2004,28(3):250-254.
- [5] Hirono I, Uchiyama T, Aoki T. Cloning, nucleotide sequence analysis, and characterization of cDNA for medaka (*Oryzias latipes*) transferrin[J]. J Mar Biotech, 1995,2:193-198.