

鲤胰蛋白酶原基因的电子克隆及生物信息学分析

陈文波*, 李卫国, 赵亚军

(河南理工大学资源环境学院生物系, 河南 焦作 454000)

摘要: 为深入研究胰蛋白酶在鱼类中的生理功能和作用机制, 利用生物信息学的方法, 成功获得了鲤3种胰凝乳蛋白酶原 cDNA 序列(ccTRP1、ccTRP2 和 ccTRP3) 并对其进行了序列分析。结果表明, 三者均含有一个长度为 729 bp 的开放阅读框(open reading frame, ORF), 编码由 242 个氨基酸组成的胰蛋白酶原, 其中包括 15 个氨基酸组成的信号肽和 5 个氨基酸(LDDDK)组成的激活肽。总平均亲水性 GRAVY(grand average of hydropathicity) 分析表明, 三者均是亲水性蛋白。氨基酸比对结果显示, 三者具备胰蛋白酶原的保守结构特征, 如含有催化三联体氨基酸(His-57、Asp-102 和 Ser-195), 12 个半胱氨酸, 位于底物结合口袋底部 Asp-189 和口袋开口处的 Gly-216、Gly-226 等。同时, 鲤3种胰蛋白酶之间具有 90% 以上的同源性。进化树结果显示, 鲤3种胰蛋白酶均属于 Group I(阴离子型胰蛋白酶), 且三者的进化距离不一致。

关键词: 鲤; 胰蛋白酶原; 表达序列标签; 生物信息学

中图分类号: Q 785; S 917

文献标识码: A

胰蛋白酶 Trypsin(EC3. 4. 21. 4) 是由胰腺腺泡细胞分泌的一种重要蛋白酶, 属于丝氨酸蛋白酶类超家族^[1]。His-57, Asp-102 和 Ser-195 形成催化三联体, 特异性切割多肽链中赖氨酸和精氨酸残基中羧基端参与形成的肽键^[2]。在胰腺中, 胰蛋白酶以非活性酶原的形式被合成和分泌, 进入小肠后, 在有 Ca²⁺ 的环境中被肠激酶激活, 切割除去 N-端的酸性激活肽, 转变成有活性的胰蛋白酶。产生的胰蛋白酶不仅可以激活胰脏产生的其它蛋白酶如胰凝乳蛋白酶原、弹性蛋白酶原及羧肽酶原等, 同时还可以激活自身的胰蛋白酶原^[3-5]。

目前, 已在一些鱼类中克隆得到了胰蛋白酶原的 cDNA 序列, 如大西洋鳕(*Gadus morhua*)^[6]、牙鲆(*Paralichthys olivaceus*)^[7]、蓝鳕(*Paranotothenia magellanica*)^[8]、日本鳗鲡(*Anguilla japonica*)^[9]、冬鲽(*Pseudopleuronectes americanus*)^[10]、塞内加尔鲷(*Solea senegalensis*)^[11]、金头鲷(*Sparus aurata*)^[12]、草鱼(*Ctenopharyngodon idellus*)^[13]、翘嘴鲌(*Culter alburnus*)^[13] 等, 并且对其时空表达进行了研究^[11-14]。结果显示, 鱼类胰蛋白酶原

与哺乳类结构类似, 同样具有保守功能区域。但是也有差别, 如鱼类胰蛋白酶的催化效率远高于哺乳动物等。

随着生物信息学的发展, 越来越多物种的基因组文库或同一物种不同组织的 cDNA 文库已被构建, 同时部分物种的全基因组序列也已完成测序。这些庞大的序列信息储存使得电子克隆技术成为可能。所谓电子克隆(*In silico cloning*) 是一种基于表达序列标签(expressed sequence tags, ESTs) 和基因组数据库, 利用计算机来协助克隆基因的方法^[15]。其基本路线是以一条序列作为探针检索 EST 数据库, 然后将同源性较高的 ESTs 拼接成连接体(contig), 继而以新获得的连接体为新的查询探针, 继续搜索 EST 数据库, 直到没有新的 EST 可供拼接为止, 将拼接得到的序列在非冗余数据库中进行搜索, 判断是否为全新的序列^[16]。

鲤(*Cyprinus carpio*) 是我国淡水养殖鱼类中最常见的优良种类之一, 生长速度快, 肉质细嫩肥美, 营养丰富, 经济价值较高。尽管已分离纯化得到鲤胰蛋白酶, 但是其核苷酸和氨基酸全序列仍

收稿日期:2010-07-22 修回日期:2010-09-10

资助项目:河南理工大学博士基金项目资助(B2010-7)

通讯作者:陈文波, E-mail: chenwenbo@hpu.edu.cn

在研究中^[17]。为了深入研究胰蛋白酶在鲤消化系统中的作用,促进水产养殖业的发展,本文利用生物信息学的手段获得了鲤胰蛋白酶原 cDNA 序列及氨基酸序列,并对其蛋白结构、同源性和进化关系进行了分析,为进一步对其生理生化的研究及养殖生产提供理论基础。

1 材料与方法

1.1 鲤胰蛋白酶原基因的电子克隆

以草鱼胰蛋白酶原的全长氨基酸序列 (GenBank 登录号: ACS71745) 为信息探针,对 GenBank 中鲤 EST 数据库 (common carp, taxid: 7962) 进行 tBlastn 同源性检索,将检索到的同源性较高的 EST 序列利用 DNASTar 软件包中的 SeqMan 子软件进行拼接,形成连接体 (contig)。然后用此连接体再次进行同源检索、拼接,重复以上过程直至没有更多的重叠 EST 检出,使得每个碱基至少经过 3 条以上 EST 序列验证以保证碱基的正确性,最终得到鲤胰蛋白酶原基因的 cDNA 序列片段。再以此片段在 GenBank 中进行 BLAST 同源搜索,与近缘物种的胰蛋白酶原进行序列比较,进而判断拼接得到的鲤胰蛋白酶原基因的 cDNA 序列的正确性以及其 ORF 序列完整性。

1.2 蛋白质结构预测与分析

用 DNATools 6.0 对开放阅读框进行分析并翻译成蛋白质;利用 SignalP 3.0 Server (<http://www.cbs.dtu.dk/service/SignalP>) 进行信号肽的预测;利用 ProtParam 软件 (<http://www.expasy.org/tools/protparam.html>) 进行编码蛋白的基本理化特性分析;利用 SPOMA (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) 对其蛋白质二级结构组成形式进行分析;利用 NetPhos 2.0 Server (<http://www.cbs.dtu.dk/services/NetPhos/>) 进行磷酸化预测;利用 NCBI 中 CDD 数据库 (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) 进行保守功能区域的预测。

1.3 同源性比较和进化分析

采用 Clustalx 1.83 软件对多个物种的胰蛋白酶原氨基酸序列进行比对,并用 DNASTar 软件包 MegAlign 子软件对各序列之间的同源性进行计算;运用 Mega 3.1 软件,采用 Neighbor-Joining 法

构建胰蛋白酶原的系统进化树。

所需各物种的胰蛋白酶原氨基酸序列 GenBank 登录号分别为人 (*Homo sapiens*): AAA61232; 大鼠 (*Rattus norvegicus*): AAA41985; 猪 (*Sus scrofa*): NP_001156363; 鸡 (*Gallus gallus*): AAA79912; 爪蟾 (*Xenopus laevis*): AAB17274; 巴丁鱼 (*Pangasianodon hypophthalmus*): AAP81159; 青鳉 (*Oryzias latipes*): NP_001098370; 斜带石斑鱼 (*Epinephelus coioides*): ADG29127; 鳚 (*Seriola quinqueradiata*): BAE16615; 鳊 (*Siniperca chuatsi*) 1: ACD70339, 2: ACD70340; 河豚 (*Takifugu rubripes*): AAU06120; 青鳉 (*Tautoglabrus adspersus*): AAR88364; 白鲷 (*Diplodus sargus*): ABX89622; 胡瓜鱼 (*Osmerus mordax*): ACO09805; 草鱼: ACS71745; 翘嘴鲌: ACS71744; 金头鲷: ABX89621; 日本鳗鲡: BAI59110; 冬鲈 1: AAC32751, 2: AAC32752; 牙鲆 1: BAA82362, 2: BAA82363; 塞内加尔鳎: BAF76141; 大西洋鳕: CAD30543; 蓝鳕鱼: CAA57701。

2 结果

2.1 鲤胰蛋白酶原基因的电子克隆结果

利用已知的草鱼胰蛋白酶原的氨基酸序列对 GenBank 中鲤 EST 数据库进行 tBlastn 同源性检索分析,发现 *E* 值 (E-value) 最高的 5 条序列均来自同一个鲤肠 cDNA 文库 (图 1)。对它们进行序列比对分析,结果显示这 5 条 EST 可拼接成 3 个独立的序列。随后用其它同源性较高的 EST 序列分别与这 3 个序列进行比对组装,结果得到 3 种鲤胰蛋白酶原 cDNA,分别命名为鲤胰蛋白酶原 1 (common carp trypsinogen 1, ccTRP1)、胰蛋白酶原 2 (common carp trypsinogen 2, ccTRP2) 和胰蛋白酶 3 (common carp trypsinogen 3, ccTRP3)。三

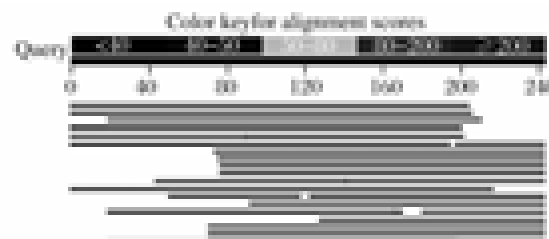


图 1 用草鱼胰蛋白酶原基因搜索 GenBank 中鲤 EST 数据库的结果

Fig. 1 Alignment result of grass carp trypsinogen searching in common carp EST database from GenBank

者 cDNA 长度分别为 830 bp、840 bp 和 835 bp, 3' 端非编码区均含有一个加尾信号 (AATAAA)。经 NCBI ORF finder 和 DNAtools 6.0 分析, 三者

均包括一个长度为 729 bp 的开放阅读框 (open reading frame, ORF), 编码由 242 个氨基酸组成的胰蛋白酶原 (图 2)。

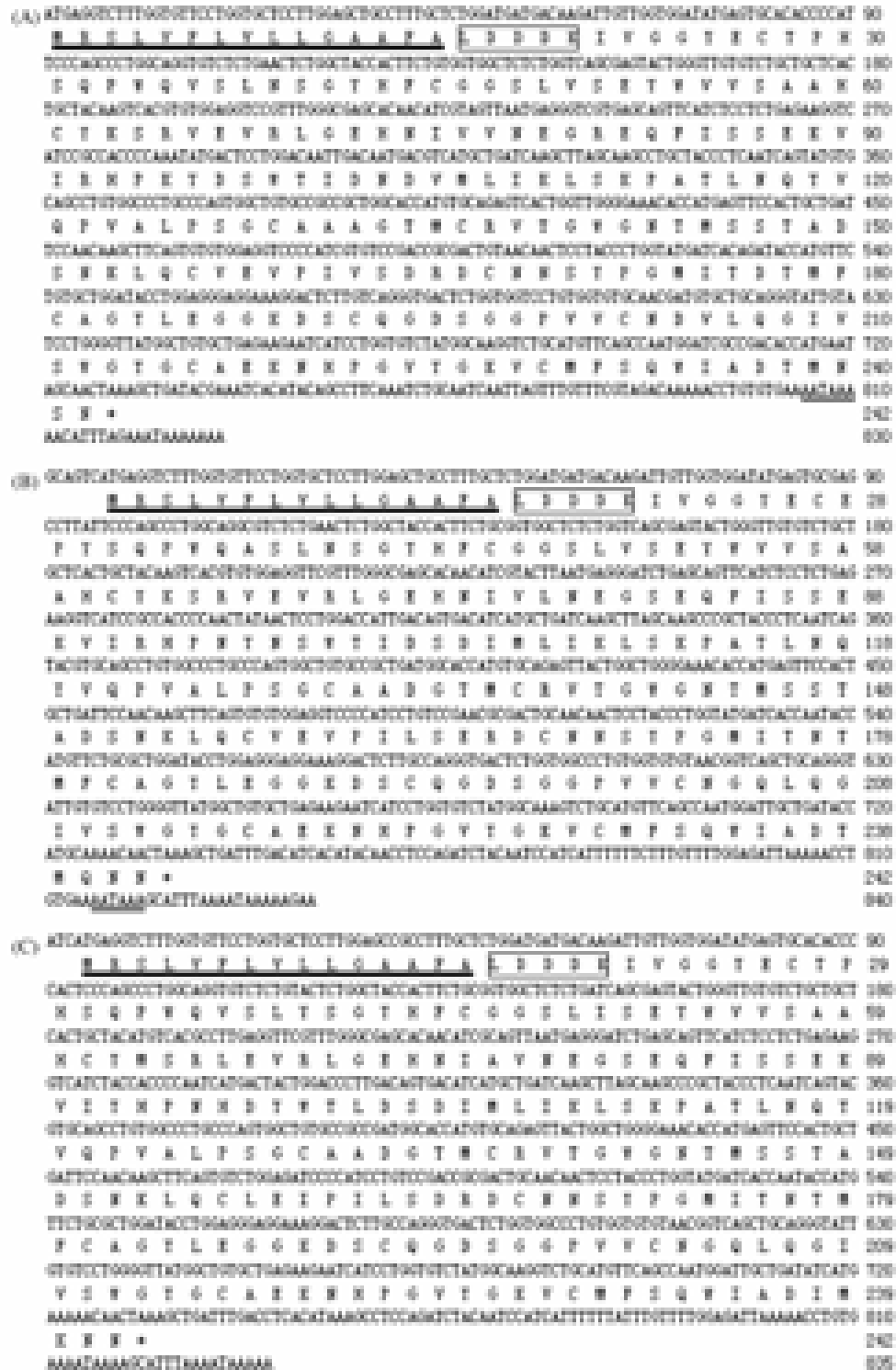


图 2 鲤胰蛋白酶原 1(A)、2(B)和 3(C) cDNA 序列及推导的氨基酸序列

信号肽部分用下划线表示, 激活肽用“□”方框表示, 加尾信号 (AATAAA) 用双下划线表示, 终止密码子用星号表示。

Fig. 2 Nucleotide sequences of common carp trypsinogens and the deduced amino acid sequences. Putative signal peptide was underlined. The activation peptide was indicated by “□”. The stop codon was denoted by asterisk.

2.2 鲤 3 种胰蛋白酶原的结构特征分析

将推导出的鲤 3 种胰蛋白酶原氨基酸序列利用 SignalP3.0 Server 分别进行信号肽分析,发现 3 种蛋白氨基端均含有序列一致的长度为 15 个氨基酸(MRSLVFLVLLGAFA)组成的信号肽(图 2)。

利用 NCBI 中 CDD 数据库预测保守功能区域结果显示,鲤 3 种胰蛋白酶原皆含有类胰蛋白酶丝氨酸蛋白酶超家族所共有的结构域(Tryp_SPC),这个家族大部分成员都具备一个共同的特征就是以非活性的酶原形式被合成和分泌。同样,在 3 种胰蛋白酶原成熟肽的第 5 和 6 个氨基酸(K-I)之间存在一个肠激酶的切割位点。

磷酸化分析显示,在鲤 ccTRP1 中存在潜在的 6 个丝氨酸磷酸化位点、1 个苏氨酸磷酸化位点和 2 个酪氨酸磷酸化位点;ccTRP2 和 ccTRP3 中都存在 5 个丝氨酸和 2 个酪氨酸磷酸化位点,但没有苏氨酸磷酸化位点。

2.3 鲤胰蛋白酶原的基本特性分析

利用 ProtParam 软件对 3 种鲤胰蛋白酶原成熟蛋白进行基本特性分析,结果显示 ccTRP1 含有原子总数为 3375,结构式为 $C_{1074}H_{1647}N_{295}O_{340}S_{19}$,其分子量为 24.7 ku,等电点为 5.14;在氨基酸组成上,甘氨酸和缬氨酸含量相同各有 23 个,均占 10.1%;其次是丝氨酸 22 个占 9.7%(表

1)。总的带负电荷和总的带正电荷的氨基酸残基数分别为 24 和 16。预测该蛋白不稳定指数为 27.96,分类为稳定蛋白,在哺乳动物体外网状细胞中的半衰期为 5.5 h。

ccTRP2 含有原子总数为 3373,结构式为 $C_{1075}H_{1641}N_{293}O_{343}S_{19}$,其分子量为 24.8 ku,等电点为 4.92;在氨基酸组成上,甘氨酸含量最高有 24 个,占 10.6%;其次是丝氨酸 23 个占 10.1%(表 1)。总的带负电荷和总的带正电荷的氨基酸残基数分别为 23 和 14。预测该蛋白不稳定指数为 28.17,分类为稳定蛋白,在哺乳动物体外网状细胞中的半衰期为 5.5 h。

ccTRP3 含有原子总数为 3386,结构式为 $C_{1086}H_{1649}N_{291}O_{340}S_{20}$,其分子量为 24.9 ku,等电点为 4.98;在氨基酸组成上,甘氨酸含量最高有 24 个,占 10.6%;其次是丝氨酸 22 个占 9.7%(表 1)。总的带负电荷和总的带正电荷的氨基酸残基数分别为 23 和 13。预测该蛋白不稳定指数为 32.78,分类为稳定蛋白,在哺乳动物体外网状细胞中的半衰期为 5.5 h。

总平均亲水性 GRAVY (grand average of hydropathicity) 分析表明,三者 GRAVY 值分别为 -0.250、-0.292 和 -0.230,这表明三者均为亲水性蛋白。

表 1 鲤胰蛋白酶原成熟蛋白的氨基酸组成

Tab. 1 Amino acid composition of common carp matured trypsinogens

氨基酸组成 amino acid	数目 no.			氨基酸组成 amino acid	数目 no.		
	ccTRP1	ccTRP2	ccTRP3		ccTRP3	ccTRP1	ccTRP2
Ala	11	11	11	Leu	11	13	15
Arg	6	5	4	Lys	10	9	9
Asn	13	15	13	Met	7	7	8
Asp	14	11	13	Phe	4	4	4
Cys	12	12	12	Pro	10	10	10
Gln	9	11	10	Ser	22	23	22
Glu	10	12	10	Thr	10	9	9
Gly	23	24	24	Trp	6	6	6
His	6	5	7	Tyr	10	11	12
Ile	10	11	13	Val	23	18	15

2.4 鲤胰蛋白酶原的二级结构分析

用 SOPMA 软件分析 3 种蛋白的二级结构,结果表明三种胰蛋白酶原成熟蛋白主要形成 4 种二级结构形式 α 螺旋、延伸直链、无规则卷曲和 β -折叠片(表 2)。其中,都以无规则卷曲的区间

居多。同时,三者羧基末端都具有一段较长的 α 螺旋区域。

2.5 氨基酸序列比对和同源性分析

鲤 3 种胰蛋白酶原氨基酸序列与部分其它脊椎动物进行比对(图 3),结果显示在胰蛋白酶原

中一些保守的氨基酸在鲤中同样存在,例如:①胰蛋白酶催化活性必需的高度保守的催化三联体氨基酸(His-57, Asp-102 和 Ser-195)在鲤中同样保守,位置分别是 Hs-45, Asp-89 和 Ser-181);②位于底物结合口袋底部的 Asp-189 和口袋开口处的 Gly-216、Gly-226 在鲤中的位置是 Asp-175、Gly-198、Gly-208;③具有构成6个二硫键所需的12个半胱氨酸残基(Cys-12、30、46、114、121、141、152、166、177、187、201、214);④丝氨酸蛋白酶的典型特征序列即位于催化活性部位丝氨酸附近的氨基酸(Gly-Asp-Ser-Gly-Gly-Pro)

在鲤3种胰蛋白酶原中同样保守;⑤决定结合底物特异性的一个关键氨基酸 Tyr-172 所在基序在鲤(¹⁵⁶YPG)中同样具备。利用 MegAlign 进行同源性分析显示,鲤3种胰蛋白酶原之间的同源性均大于90%,且 ccTRP1 与 ccTRP2 的同源性最高为93%。与其它物种相比较,三者均与草鱼的同源性最高,在88.8%~90.5%;其次是与青鳉和金头鲷,同源性在75%~80%。尽管与翘嘴鲌同属鲤科,但是与其同源性却较低,在66%左右。与哺乳动物及两栖类的同源性也在60%~68%。

表2 鲤胰蛋白酶原成熟蛋白的二级结构组成

Tab.2 The secondary structure composition of common carp matured trypsinogens

	α 螺旋 α helix	延伸直链 extended strand	无规则卷曲 random coil	β 折叠片 β turn
ccTRP1	14.54%	25.99%	46.70%	12.78%
ccTRP2	15.86%	24.67%	48.9%	10.57%
ccTRP3	15.86%	27.75%	44.49%	11.89%

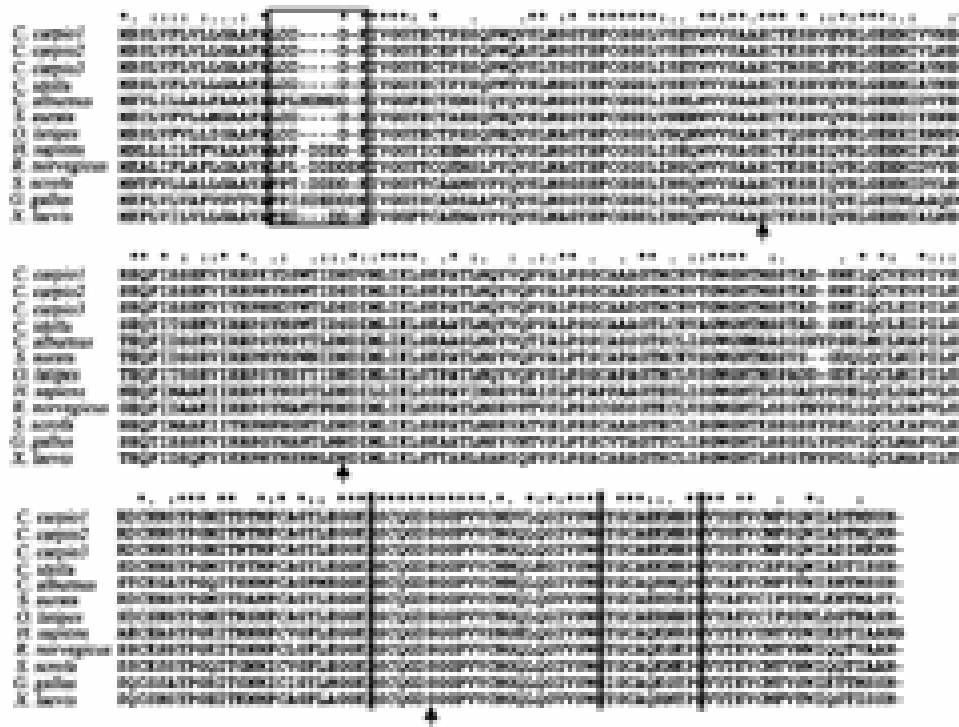


图3 鲤与其它部分物种的胰蛋白酶原的氨基酸序列比对

一致的、高度保守度的和低保守度的氨基酸分别用(*), (:), (.)标出。激活肽用“□”标出,催化三联体氨基酸用上箭头表示,底物结合口袋底部天冬氨酸和开口处的两个甘氨酸用阴影表示。

Fig.3 Alignment of the deduced amino acid sequences of common carp trypsinogens and some other vertebrate trypsinogens deposited in GenBank

The identical, highly conserved and less conserved amino acid residues are indicated by (*), (:), (.) respectively. The activation peptide was marked by“□”. The catalytic triad amino acid residues were indicated by upward arrows. The Asp-189 at the bottom of the substrate-binding pocket and Gly-216, -226 lining the sides of the binding pocket were shaded.

2.6 系统进化树分析

从 GenBank 上获得已知的脊椎动物胰蛋白酶原的氨基酸序列经 Clustal W 方法比对后,利用 Mega 3.1 以邻位相接(NJ)构建系统进化树(图 4)。由图可以得出鱼类胰蛋白酶原分为 3 支

(Group I、II、III),按照 Roach 等^[18]有关脊椎动物胰蛋白酶原的聚类方法,3 种鲤胰蛋白酶原型均位于 Group I,属于阴离子胰蛋白酶原。同时,三者的进化距离不一样,ccTRP1 较之 ccTRP2 和 ccTRP3 分化较早。

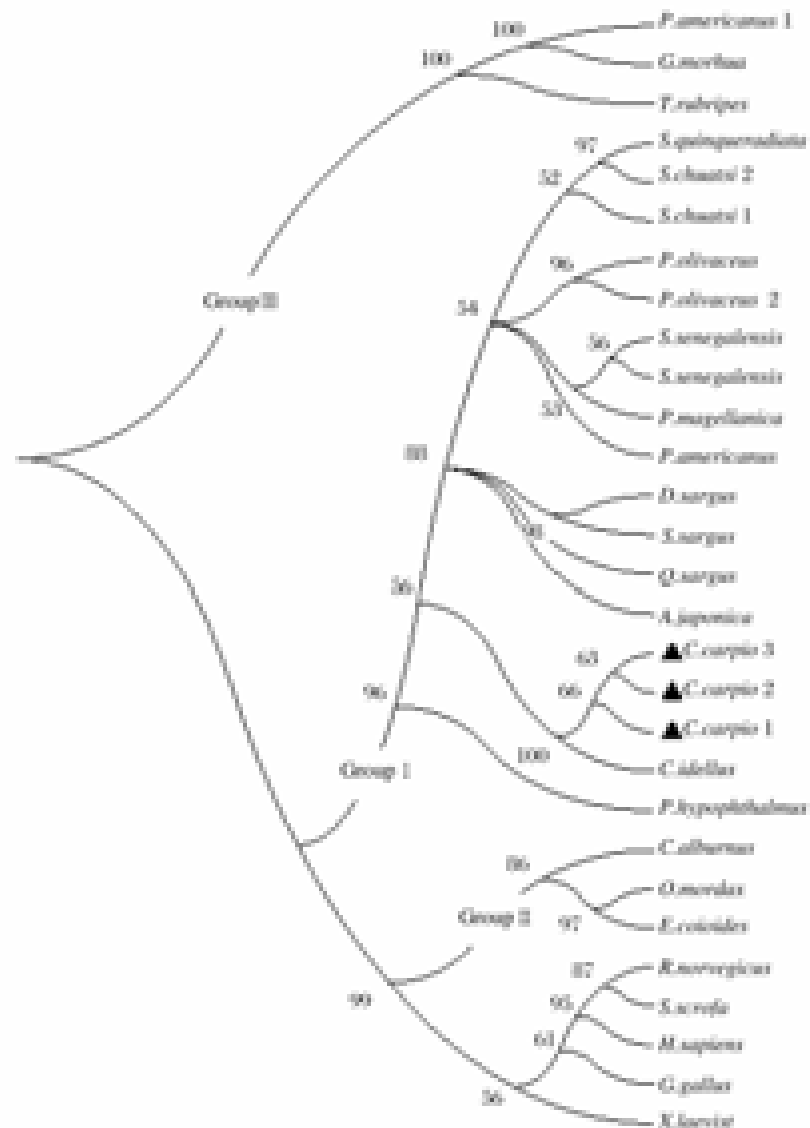


图 4 胰蛋白酶原的进化树分析

各物种胰蛋白酶原序列均来自 GenBank 数据库,采用 Mega 3.1 软件(Neighbor-Joining 法)构建。图中各分支数值代表置信度(%)。鲤 3 种胰蛋白酶原用▲标出。

Fig. 4 Phylogenetic tree based upon the alignment of amino acid sequences from a wide range of vertebrates

The full-length amino acid sequences of various trypsinogens were extracted from GenBank and analyzed using the Neighbor-Joining bootstrap method by Mega 3.1 with 1000 bootstrap replicates. The number shown at each branch indicates the bootstrap values(%). The common carp trypsinogens were noted by ▲.

3 讨论

实验通过电子克隆的手段从鲤 EST 数据库

成功拼接出 3 种胰蛋白酶原,并利用生物信息学的手段对编码的蛋白质基本性质、蛋白结构、同源性和进化关系进行了研究。鲤 3 种胰蛋白酶原都

具备胰蛋白酶的保守结构特征,如催化三联体、12个半胱氨酸形成6个二硫键以及催化活性部位丝氨酸附近的氨基酸基序(Gly-Asp-Ser-Gly-Gly-Pro)等。这表明鲤3种胰蛋白酶在功能上是保守的,都具备切割多肽链中赖氨酸和精氨酸残基中的羧基端的能力。以往研究发现,丝氨酸蛋白酶家族表面的两个环对酶与底物的结合非常重要^[19]。在鲤3种胰蛋白酶原中,构成两个环的氨基酸是一致的。另外,和底物特异性有关的 Tyr-172 及其所在基序¹⁵⁶YPG也完全一致。这些结果表明鲤3种胰蛋白酶对底物的选择基本相同。尽管功能区域是保守的,但是鲤3种胰蛋白酶原在氨基酸组成上也有一定的区别,如缬氨酸、亮氨酸和异亮氨酸的数目差异较显著(表1)。这些氨基酸的差异对其结构有无影响以及是否造成其它功能的差异还需要进一步深入研究。

目前,在冬鲮^[10]、塞内加尔鲷^[11]等鱼类中已克隆得到多个胰蛋白酶原 cDNA 序列,其中塞内加尔鲷胰蛋白酶原1有3个亚型,尽管同源性高达96%,但是它们在肠、肝脏、头肾和胃中的表达量各不相同。鲤3种胰蛋白酶原同属于阴离子胰蛋白酶原,同源性也高达90%以上。同时,我们利用草鱼胰蛋白酶原进行 tBlastn 同源检索时,在鲤的 EST 数据库中得到了数目不同且高度同源的胰蛋白酶 EST 序列,其中以 ccTRP2 为最高。这预示着在鲤中尽管存在功能相对保守的3种胰蛋白酶原,但是其表达分布或者在同一组织中的表达量可能各不相同,提示它们可能在不同或相同的组织中发挥着特异的作用。

胰蛋白酶在胰脏以酶原的形式被合成和分泌,在小肠中被肠激酶切去^[6,11]。氨基酸序列比对显示,鱼类和哺乳类的激活肽长度不尽相同,哺乳类比鱼类的激活肽要多3~4个氨基酸。典型的激活肽包含一簇阴离子氨基酸残基,能够为肠激酶所识别^[20]。在大部分鱼类中激活肽包含3个酸性氨基酸(D和E),而在哺乳动物中则是连续4个天冬氨酸^[13]。3种鲤胰蛋白酶原的激活肽长度均由5个氨基酸组成(LDDDK),与同属于阴离子胰蛋白酶的草鱼、金头鲷和青鲈一致,与属于阳离子胰蛋白酶的翘嘴鲃变化较大。目前认为,不同物种的激活肽序列及其组成的差别可能是由进化过程中的选择性压力所造成^[3,13,21]。

参考文献:

- [1] Halfon S, Baird T T, Craik C S. Trypsin, handbook of proteolytic enzymes [M]. 2nd. // Barrett A J, Rawlings N D, Woessner J F, Eds. Cysteine, Serine and Threonine Peptidases, vol. 2. Amsterdam: Elsevier, 2004: 1483 - 1488.
- [2] Rypniewski W, Perrakis A, Vorgias C E, et al. Evolutionary divergence and conservation of trypsin [J]. Protein Eng, 1994, 7(1): 57 - 64.
- [3] Chen J M, Kukor Z, LeMarechal C, et al. Evolution of trypsinogen activation peptides [J]. Mol Biol Evol, 2003, 20(11): 1767 - 1777.
- [4] Rust M B. Nutritional physiology [M] // 3rd ed. Halver J E, Hardy R W. Fish Nutrition [M]. Amsterdam: Elsevier, 2002: 367 - 452.
- [5] 王镜岩, 朱圣庚, 徐长发. 生物化学 [M]. 3版. 北京: 高等教育出版社, 2002: 405.
- [6] Gudmundsdottir A, Gudmundsdottir E, Oskarsson S, et al. Isolation and characterization of cDNAs from Atlantic cod encoding two different forms of trypsinogen [J]. Eur J Biochem, 1993, 217(3): 1091 - 1097.
- [7] Suzuki T, Srivastava A S, Kurokawa T. cDNA cloning and phylogenetic analysis of pancreatic serine proteases from Japanese flounder, *Paralichthys olivaceus* [J]. Comp Biochem Physiol B Biochem Mol Bio, 2002, 131(1): 63 - 70.
- [8] Genicot S, Rentier-Delrue F, Edwards D, et al. Trypsin and trypsinogen from an Antarctic fish: molecular basis of cold adaptation [J]. Biochem Biophys Acta, 1996, 1298(1): 45 - 57.
- [9] Kurokawa T, Suzuki T, Ohta H, et al. Expression of pancreatic enzyme genes during the early larval stage of Japanese eel *Anguilla japonica* [J]. Fish Sci, 2002, 68(4): 736 - 744.
- [10] Douglas S E, Gallant J W. Isolation of cDNAs for trypsinogen from the winter flounder, *Pleuronectes americanus* [J]. J Mar Biotechnol, 1998, 6(4): 214 - 219.
- [11] Manchado M, Infante C, Asensio E, et al. Molecular characterization and gene expression of six trypsinogens in the flatfish Senegalese sole (*Solea senegalensis* Kaup) during larval development and in tissues [J]. Comp Biochem Physiol B Biochem Mol Biol, 2008, 149(2): 334 - 344.
- [12] Psochiou E, Sarropoulou E, Mamuris Z, et al. Sequence analysis and tissue expression pattern of *Sparus aurata* chymotrypsinogens and trypsinogen

- [J]. *Comp Biochem Physiol B Biochem Mol Biol*, 2007, 147(3):367-377.
- [13] Ruan G L, Li Y, Gao Z X, *et al.* Molecular characterization of trypsinogens and development of trypsinogen gene expression and tryptic activities in grass carp (*Ctenopharyngodon idellus*) and topmouth culter (*Culter alburnus*) [J]. *Comp Biochem Physiol B Biochem Mol Biol*, 2010, 155(1):77-85.
- [14] Murray H M, Perez-Casanova J C, Gallant J W, *et al.* Trypsinogen expression during the development of the exocrine pancreas in winter flounder (*Pleuronectes americanus*) [J]. *Comp Biochem Physiol A Mol Integr Physiol*, 2004, 138(1):53-59.
- [15] 胡骛, 萧浪涛. 生物信息学在新基因全长 cDNA 电子克隆中的应用 [J]. *生物技术通报*, 2007, 4:93-96.
- [16] 刘卉玲, 崔群维, 吴国良, 等. 牛 *Pbx-1* 基因的电子克隆及生物信息学分析 [J]. *激光生物学报*, 2009, 18(1):17-23.
- [17] Cohena T, Gertlera A, Birka Y. Pancreatic proteolytic enzymes from carp (*Cyprinus carpio*) - I purification and physical properties of trypsin, chymotrypsin, elastase and carboxypeptidase B [J]. *Comp Biochem Physiol B Biochem Mol Biol*, 1981, 69(3):639-646.
- [18] Roach J C. A clade of trypsins found in cold-adapted fish [J]. *Proteins*, 2002, 47(1):31-44.
- [19] Hedstrom L, Szilagyi L, Rutter W J. Converting trypsin to chymotrypsin: the role of surface loops [J]. *Science*, 1992, 255(5049):1249-1253.
- [20] Louvard M N, Puigserver A. On bovine and porcine anionic trypsinogens [J]. *Biochim Biophys Acta*, 1974, 371(1):177-185.
- [21] Roach J C, Wang K, Gan L, *et al.* The molecular evolution of the vertebrate trypsinogens [J]. *J Mol Evol*, 1997, 45(6):640-652.

***In silico* cloning and bioinformatics analysis of trypsinogen genes in common carp (*Cyprinus carpio*)**

CHEN Wen-bo^{*}, LI Wei-guo, ZHAO Ya-jun

(Department of Biology, School of Resource and Environment, Henan Polytechnic University, Jiaozuo 454000, China)

Abstract: Trypsin is one member of the serine protease family, which is synthesized as proenzyme trypsinogen by pancreas. After being secreted into the intestine, enterokinase removes the N-terminal activation peptide converting it into its active form, trypsin which can specifically cleave at the peptide bond on the carboxyl side of basic L-amino acids such as arginine or lysine residues. Meanwhile, trypsin is one of the most abundant proteases in the digestive system of teleost fish, and plays important roles in the protein absorption from the diet. For further studying the physiological functions and mechanism of trypsinogen in common carp (*Cyprinus carpio*), one of the most economically important fish in China, we got three distinct trypsinogen cDNAs (ccTRP1, ccTRP2 and ccTRP3) successfully using *in silico* cloning technique, and then they were analysed by the method of bioinformatics. The results showed that three of them had an open reading frame of 729bp in length encoding 242 amino acids, the same sequence of a signal peptide consisted of 15 amino acids and an activation peptide of 5 amino acids, LDDDK. It also showed that they possessed the conserved domains of trypsin-like serine protease superfamily, and had the conserved structural characteristics, such as the catalytic triad (His-57, Asp-102, and Ser-195), 12 cysteines forming 6 disulfide bonds, Asp-189 at the bottom of the substrate-binding pocket and Gly-216 and Gly-226 lining the sides of the binding pocket, the motif of Gly-Asp-Ser-Gly-Gly-Pro in serine protease family, and so on. But, three of them also had some different characteristics with amino acid composition and phosphorylation sites. These results suggested that common carp trypsinogens we got in the present study may be functional, however, their functions may be different to a certain degree. The analysis of GRAVY (grand average of hydropathicity) and the instability index indicated that they belonged to the hydrophilic and stable proteins. In the aspect of advanced structure, they mainly contained four secondary structural forms (alpha helix, extended strand, random coil and beta turn) by SOPMA software analysis, and the most abundant was random coil. The alignment and homology analysis based on the amino acid sequences revealed that it had the identity of beyond 90% among the three common carp trypsinogens, and ccTRP1 and ccTRP2 had the highest identity of 93%. Furthermore, compared to other species they had the highest identity with grass carp (*Ctenopharyngodon idellus*). Although both common carp and topmouth culter (*Culter alburnus*) belonged to Cyprinidae, they only had the identity of 66%, indicating that they may possess distinct property. The phylogenetic tree, based upon the alignment of amino acid sequences of trypsinogens, showed that three of them clustered with teleost anionic trypsinogen Group I, but the evolutionary distances were not the same. These preliminary results from common carp trypsinogens by the method of bioinformatics can provide the foundations for further study of its expression profiles, the molecular characteristics and evolution of fish trypsinogen.

Key words: common carp (*Cyprinus carpio*); trypsinogen; expressed sequence tag (EST); bioinformatics

Corresponding author: CHEN Wen-bo. E-mail: chenwenbo@hpu.edu.cn