

文章编号:1000-0615(2002)04-0295-06

编码鳗鲡生长激素基因的序列与结构

贡成良, 曹广力, 薛仁宇, 张传溪

(苏州大学生命科学学院, 江苏 苏州 215006)

摘要:序列分析表明:欧洲鳗鲡 GH 基因从 ATG 到 TAG 共计 2393 bp, 有 4 个外显子、3 个内含子, 推测编码 209 个氨基酸。第一内含子存在 3 个重复单体构成的微卫星序列。同源性分析表明:欧洲鳗鲡 GH 可分为 5 个结构域, GD1 - GD4 区域与 GH 的活性有关, 为 GH 与其受体专一性结合区域, GD5 区域与 GH 分子的结构、稳定性有关。进化分析表明, 鱼类 GH 分可成 3 个不同的组, 欧洲鳗鲡 GH 基因属 I 组, II 组可分含 4 个内含子基因及含 5 个内含子基因二个亚组, III 组 GH 基因有 5 个内含子。推测原始的 GH 基因在进化过程中, 通过丢失或增加内含子导致基因歧化。

关键词:鳗鲡生长激素基因; 微卫星序列; 同源比较; 进化

中图分类号: Q75 **文献标识码:** A

Sequence and structure of encoding eel growth hormone gene

GONG Cheng-liang, CAO Guang-li, XUE Ren-yu, ZHANG Chuan-xi

(Life Science College, Suzhou University, Suzhou 215006, China)

Abstract: The analysis of the sequence shows that the growth hormone gene (GH gene) of *Anguilla anguilla* contains 2393 bp from ATG to TAG, and consists of 4 exons and 3 introns, also it was deduced to encode 209 amino acids. The first intron consists of 3 repeated monoploid forming its minisatellite sequence. Alignment analysis indicated that the GH gene can be divided into 5 structure domains. GD1 - GD4 domains may be relative to the GH activity and the specially binding field of the receptor, and GD5 domain to molecular structure and stability of the GH respectively. Evolution study demonstrated that there were I, II and III groups existing among the fishy GH genes, and the *Anguilla anguilla* GH belongs to group I; group II can be redivided to 2 subgroups which were 4 introns or 5 introns, and group III has 5 introns only. It was presumed that the dismutation of the GH gene had happened through losing or adding introns during the evolution.

Key words: *Anguilla anguilla* GH gene; microsatellite; alignment; evolution

生长激素(GH)是一种单链多肽类激素,在脊椎动物的生长发育中起重要作用。由于其有潜在增加水生鱼类的生长速度,因此科学家已从多种鱼中获得了编码生长激素基因(GH)的克隆,希望通过基因工程生产大量的 GH,以配制促进鱼快速生长的饲料。在所调查的大部分真骨鱼中,仅发现一种 GH 基因,而在鲑鳟及罗非鱼中发现了两种 GH 基因。有报道认为鲑鳟中 GH 基因以两种形式存在,是在 25 ~ 100 百万年前,由原始祖先中发生四倍体化事件所导致的结果^[1]。在罗非鱼中, GH 基因存在两个

收稿日期:2001-10-16

资助项目:苏州大学“211”标志性成果研究项目(XB 113056)

作者简介:贡成良(1965-),男,江苏丹阳人,副教授,博士,主要从事基因调控、表达及基因工程制药研究。E-mail: chlgong@163.net

拷贝^[2]。草鱼^[3],斑点叉尾鲷^[4]中,GH基因组含5个外显子,4个内含子,类似于鸟类和哺乳类动物的GH基因;而虹鳟^[5],尼罗罗非鱼^[6],牙鲆^[7],黄尾鲈^[8]GH基因的最后—个外显子中有一个附加的内含子。对于不同的来源的GH基因序列分析,可以帮助我们认识GH基因的进化及内含子起源的踪迹,内含子的模型也许可以作为一个自然标记,有助于认识不同生物的进化关系。我们通过PCR克隆了欧洲鳗鲡GH基因,对其进行了序列分析,并将欧洲鳗鲡GH基因与其它GH基因进行了比较和进化分析,现将研究结果报道如下。

1 材料与方 法

鳗鲡:欧洲鳗鲡(*Anguilla anguilla*),购自苏州外贸冷冻厂。

工具酶:限制性内切酶、Tag酶、DNA连接酶购自Gibico BRL公司。

鳗鲡基因组DNA的抽提:取肌肉组织按常规方法进行。

克隆及测序:根据日本鳗鲡(*Anguilla japonica*)的序列,设计—对引物(P1:ATGGCATCAGGGTTCCTTC,P2:CTACAGGGTGCAGTTGCTTTC),引物由上海生物工程公司合成;PCR反应条件为93℃ 1 min,55℃ 1min,72℃ 2.5 min进行30个循环。PCR产物用低熔点胶回收,TA法克隆进pBluescript II SK+后,用引物追踪法进行序列分析。酶切、连接、转化等按常规法进行。

分子进化分析:应用PC/gene(Ver 6.8 Intekki Genekics) Clustal程序进行分析。

2 结果

2.1 欧洲鳗鲡GH基因组的结构

表1 欧洲鳗鲡GH基因第1内含子中的微卫星序列

Tab.1 Microsatellite sequence of the first intron of *Anguilla anguilla* GH gene

重复单体类型		长度(bp)	重复次数	AT含量(%)	重复单体类型		长度(bp)	重复次数	AT含量(%)
type	sequence	length	repeats	AT content	type	sequence	length	repeats	AT content
A1	tgtgAGTttt	10	1	70	A2	tgtgATTttt	10	9	80
A3	tgtgCTCttt	10	1	60	B1	gTAca	5	8	60
B2	gCAca	5	1	40	B3	gTGca	5	1	40
C1	ggtgaGataCaG	12	5	50	C2	ggtgaAataCaG	12	1	58.3
C3	ggtgaAata*G	10	1	60	C4	ggtgaAataCaA	12	1	66.7
C5	ggtgaAataGaA	12	1	66.7					

注:小写字母为微卫星序列重复单体的一致核苷酸,大写字母为非一致核苷酸,*表示重复单体中缺失核苷酸。

Notes: Conserved minisatellite sequence (lower case letter), misalignment nucleotide (capital letter), deletion nucleotide (asterisk).

欧洲鳗鲡GH基因的序列如图1所示,全基因从ATG到TAG共计2393 bp,分4个外显子(第1外显子150bp,第2外显子117bp,第3外显子162bp,第4外显子201bp),3个内含子(第1内含子434bp,第2内含子852bp,第3内含子477bp)。欧洲鳗鲡GH基因的3个内含子都起始于GT,终止于AG,符合内含子共同剪接位点序列。4个外显子共计630 bp,编码209个氨基酸,分子量为23.6 kDa。推测N端20个氨基酸为欧洲鳗鲡GH的信号肽序列。欧洲鳗鲡GH基因在外显子、内含子组织结构上不同于以往报道的GH基因,比骨鲮类中的草鱼少1个内含子,比鲱形总目中的红大马哈鱼、大西洋鲑以及鲈形总目中的罗非鱼、石首鱼少2个内含子。如图1、表1所示,在鳗鲡GH基因的第一内含子中出现微卫星序列,根据重复单体序列特征,可分为A、B、C三种基本类型,三者串连排列。其中A(tgtga/cg/tt/cttt)重复11次,B(gt/ca/gca)重复10次,C[ggtgag/aatac/g(-)a/-g/a]重复9次。类似的重复序列在尖吻鲈GH的第3内含子中也曾发现^[9]。欧洲鳗鲡的第3内含子在1875~1893位点出现aaacacttattgagtgttt序列,该序列可以形成茎环结构,其中1875~1881为茎,1882~1886为环。

```

ATGCCATCAGGGTTCCTTCTGTGGCCGGTCTTACTGGTCAGCTTCTCTGTGAACCGGGTGGAGCCCAATTCCCTCTACAACCTCTTACC 90
M A S G F L L W P V L L V S F S V N A V E P I S L Y N L F T
AGCGCTGTTAACCGAGCAGCACACCTGCACACACTGGCTGCCGAATATACAAGGAATTGtaagctcagcgtcacatccatgcatgagtggt 180
S A V N R R A Q H L H T L A A E I Y K E F
ggaataagcagatcgttttacacatttcaggtctaggggtgttcaaaaaaaagtgggaaatcaccttctgtgacgttaataataaaagga 270
tttataaagaagtcctcatataaaaaatgcacccagttctctgtgagttttgtacaggtgagatcacagtgattttttgtacaggtga
gatacagtgatgtttttgtacaggtgaaatacagtgatgtttttgtacaggtgaaatagtgatgtttttgtacaggtgagatcacagtggt 450
atttttgtacaggtgagatcacagtgatgtttttgtacaggtgagatcacagtgatgtttttgtacaggtgagatcacagtgatgtttttgt
caggtgagatcacagtgatgtttttgtgctctttgtgcaacacagGAGCGAAGCATCCACCCGAGGCCACAGACAGCTCAGCAAGACCT 630
E R S I P P E A H R Q L S K T S
CCCCATGGCCGGCTGTACTCCGACTCTATCCCTACCCCCACAGGCAAGATGAAACGCAGGAAATCGgtgagtggtcagtgccaa 720
P L A G C Y S D S I P T P T G K D E T Q E K S
aggaagagggaaattctatcaggcgtgtcctctgtctcggttacagttcaatgctttcagcaatcaagctgagcagaaaaattggtgtaat 810
ccaaaggttaagttatttcagtgtaaatgcatttaaccacaagatctcaacacaatattcaaatcaagaggttaaaaaatgagaacaa 900
aggaagcagaaaaatggccatagtcattaagatgaagtcacatagcaggaatattgtaaatagattttttataaatgaaaaactgg 990
tgcatagctatttaaaacatcaaacattgcttggtagagtgtagaaatttgcataagacaaattataaacaacaaatttatttggt 1080
tgactttctcaagctctcttgatgaaattcttatttttaatttccttcagtagagctgtagagaaactcctaaacatataccaggatct 1170
aatcaacatttgggtgtgcttctgtaaaaaatgaaagacatgcatgtgtaaatgaaagacagattttatcctgggggtatttcaactat 1260
ttaaccttctactgtgattgtgtgtaaaaaatgaaagacatgcatgtgtaaatgaaagacagattttatcctgggggtatttcaactat 1350
gctataattgcgggttaaaaccgcaacagctcttttaatttcagttcatgtttcagattggggaggttctggtttatacatcagtgag 1440
ttctgcagcctaactcagtaactctgctttgcaagcagggcccgaggttaaggggctgatgtatccagtttaaacggactgctcataa 1530
gtttatcgccctctatttgcagGATGGTACTTGTGCGCATCTCCTCAGCCCTGATCCAGTCATGGGTGTATCCTCTGAAGACCCCTGA 1620
D G Y L L R I S S A L I Q S W V Y P L K T L S
GCGATGCTTCTCAAAACAGCCTGATGTTGGGACCTCTGATGGGATCTTTGATAAGCTGGAGGACCTGAACAAGGCATCAATGAATTA 1710
D A F S N S L M F G T S D G I F D K L E D L N K G I N E L M
TGAAAGtataaagggttaaaagggtgaaagtaatgccaggtagtaaaaatgtgtgagggattggccttgaacttaagggtccaggcaac 1800
K
tggattcaaatgtgggttatattccccagaagtgaacaactgattgttttcttaatttctttcaagcttttaaacacttattgagtg 1890
tttaaacagttatgaccttattttgcaataaacacagattagattatgagtcataatgtgtgattatgaggaacatacagattttattg 1980
ttaaatggaggagcctccaaaaataatgctacaaatataacatttcaggcaaaaaaacaaaaaacaaaaaagtc 2070
caagataaaatcggaggttaaacacatattcttaaaatctcaaatgtgttggtcataatcaatcccaaacatgcatggttgaataaccg 2160
ctgtactaagccctgacctatctcctcacagGATGTAGGTGACGGTGGTATTACATTGAGGATGTGAGAAATCTCCGGTACGAGA 2250
D V G D G G I Y I E D V R N L R Y E N F
TCGACGTACACCTTAGGAACGATGCCGGCCTGATGAAGAACTATGGCCTGCTGGCCTGCTTTAAGAAAGACATGCACAAAGTGGAGACCT 2340
D V H L R N D A G L M N Y G L L A C F K K D M H K V E T Y
ACCTGGAAGTCACAAAGTGCAGGCGCTTTGTAGAAAGCAACTGCACCCCTGTAG 2393
L E V T K C R R F V E S N C T L *

```

图 1 欧洲鳊 GH 基因的序列

Fig.1 Sequence of *Anguilla anguilla* GH gene

注:大写字母为外显子核苷酸序列;小写字母为内含子序列;大写斜体字母为据外显子推测的氨基酸序列;小写黑体为推测的形成茎环结构序列;下划线、框盒、阴影分别示 A、B、C 微卫星序列重复单体,大写黑体字母为与日本鳊 GH cDNA 不同的核苷酸。

Notes: Extron (capital letter), intron (lowercase letter), deduced amino acid sequence (capital and italic letter), deduced structure of stem loop; A, B, C repeated minisatellite sequence (underline, box and shadow respectively); different nucleoid from cDNA of *anguilla japonica* (black capital letter)

2.2 鱼类 GH 氨基酸序列的同源性比较

鱼类 GH 氨基酸序列与 GenBank 中鱼类 GH 氨基酸序列比较结果如图 3 所示。欧洲鳊 GH 与日本鳊 GH 氨基酸一致序列达 99%,与猪 GH 氨基酸一致序列达 52%,高于本文中所收集的其它鱼类 GH。根据 Masanori 等^[10]的描述, GH 可分为 5 个不同的结构域(GD1 - GD5),它们对 GH 基因家族共同结构的形成具有重要作用。用 Chou - Fasman 对 GH 二级结构分析表明, GD1、GD2、GD3、GD4 和 GD5 分别含有 β-折叠 + α-螺旋、反转、β-折叠 + α-螺旋、α-螺旋和 α-螺旋 + β-折叠结构,这些结构特征同人类 GH 的一致。

2.3 鱼类生长激素基因进化分析

应用鱼类 GH 氨基酸序列,对 GH 基因进行进化分析,结果如图 4 所示。可将鱼类 GH 分成 3 个不同的组, I 组中欧洲鳊 GH 基因有 3 个内含子, II 组可分二个亚组,其中 II-1 亚组中的草鱼、斑点叉尾鲷及鲢 GH 基因有 4 个内含子, II-2 中虹鳟、红大马哈鱼 2、红大马哈鱼 1(红大马哈鱼 1 和 2 来自不同研究结果)的 GH 基因有 5 个内含子; III 组中尖吻鲈、牙鲆、尼罗罗非鱼的 GH 基因有 5 个内含子。

□:推测的信号肽切割位点;.:相对保守氨基酸残基;* :完全一致氨基酸残基;GD1 - GD5:GH 高度保守的 5 个区域;黑体字:预测的決定 GH 活性的重要氨基酸残基。

AA:欧洲鳗鲡;AP:日本鳗鲡,M24066;SC:野猪,X53325;PP:鲢鲢,M63713;HM:鲢,M94348;OS:大马哈鱼(未定种类),E01334;AC:秋白鲑,X77245;SS:鲑(未定种类)E00964;SS2:大西洋鲑,M21573;TT :金枪鱼,E02688;BP:尖吻鲈,X59378;AB:布氏棘鲷,X59377;SA:金头鲷,S54890;ES:舌齿鲈,X65716;OA:美洲棘拟银鱼(作者译),AF236091;ON:尼罗罗非鱼,M26916;SS3:塞内加尔鲷,U01143;CM:真鲷,E02560;EP:飞鲮,E02013;CD:条纹鲮,X55176;PO:牙鲆,X12887(未定种类指原文献中未确定分类地位)

□:deduced cutting cite of signal peptide;.:relative conserved amino acid residues;* :perfect match amino acid residues;GD1 - GD5:much conserved 5 domains;black body:predicted important active amino acid residues.

AA: *Anguilla anguilla* ; AP: *Anguilla japonica* , M24066; SC: *Sus scrofa* , X53325; PP: *Pangasius pangasius* , M63713; HM: *Hypophthalmichthys molitrix* , M94348; OS: *Oncorhynchus* sp. , E01334; AC: *Coregonus autumnalis* , X77245; SS: *Salmon* , E00964; SS2: *S. salar* , M21573; TT : *Thunnus thynnus* , E02688; BP: *Barramundi perch* , X59378; AB: *Acanthopagrus butcheri* , X59377; SA: *Sparus aurata* , S54890; ES: *Dicentrarchus labrax* , X65716; OA: *Odontesthes argentinensis* , AF236091; ON: *O. niloticus* , M26916; SS3: *Solea senegalensis* , U01143; CM: *Chrysophrys major* , E02560; EP: *Euthynnus pelamis* , E02013; CD: *Caranx delicatissimus* , X55176; PO: *Paralichthys olivaceus* , X12887

3 讨论

有关鳗鲡 GH 基因的 cDNA 序列(E01353,M24066)业已报道,但至今未见鳗鲡 GH 基因组的有关研究。研究表明:欧洲鳗鲡 GH 基因组从 ATG 到 TAG 共计 2393 bp,有 4 个外显子、3 个内含子,在组织结构上不同于以往报道的任何 GH 基因。在尖吻鲈 GH 基因的第 3 内含子中曾发现微卫星序列,且在不同来源尖吻鲈 GH 基因中表现出多态性^[9]。在欧洲鳗鲡 GH 基因的第 I 内含子中,也发现微卫星序列,但组织方式不同于尖吻鲈 GH 基因,由 3 个重复单体 A、B、C 串连构成。尖吻鲈 GH 基因的第 3 内含子的多态性是由于微卫星序列的重复次数不同所引起,欧洲鳗鲡 GH 基因的第 I 内含子是否也呈现多态性,这需要对不同来源欧洲鳗鲡 GH 基因进一步分析才能下结论。我们有理由相信对于不同的来源的 GH 基因序列分析,可以帮助我们认识 GH 基因的进化及内含子起源的踪迹,内含子的模型也许可以作为一个自然标记,有助于认识不同生物的进化关系。

除去信号肽序列部分,在所比较的 GH 中,有 37 个氨基酸残基完全保守,这 37 个氨基酸残基在人、鼠、马、山羊、鸭的 GH 中也呈现较好的保守性,可以认为这 37 个氨基酸残基对 GH 专一性生物活性有重要作用。由疏水性分析及同源性比较结果推测,鳗鲡 GH 的 5 个不

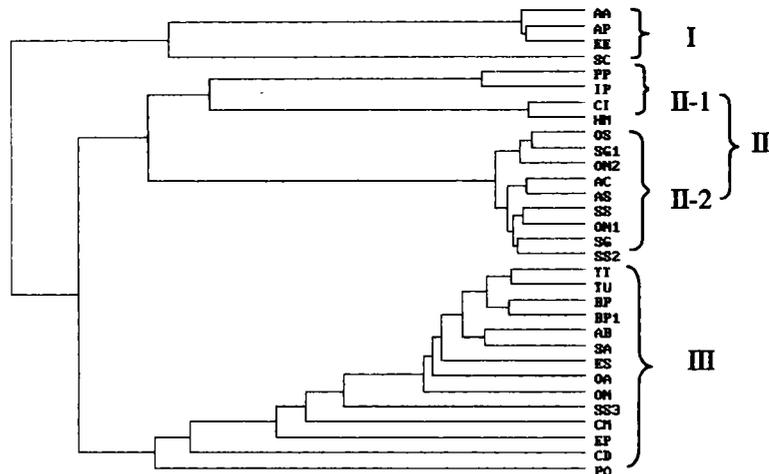


图 3 生长激素基因进化分析

Fig.3 Evolution analysis of the GH genes

AA:欧洲鳗鲡;AP:日本鳗鲡,M24066;SC:野猪,X53325;PP:鲢鲢,M63713;HM:鲢,M94348;OS:大马哈鱼(未定种类),E01334;AC:秋白鲑,X77245;SS:鲑(未定种类)E00964;SS2:大西洋鲑,M21573;TT :金枪鱼,E02688;BP:尖吻鲈,X59378;AB:布氏棘鲷,X59377;SA:金头鲷,S54890;ES:舌齿鲈,X65716;OA:美洲棘拟银鱼(作者译),AF236091;ON:尼罗罗非鱼,M26916;SS3:塞内加尔鲷,U01143;CM:真鲷,E02560;EP:飞鲮,E02013;CD:条纹鲮,X55176;PO:牙鲆,X12887;EE:鳗鲡(未定种类),E01353;IP:斑点叉尾鲷,S69215;CI:草鱼,X60988;SG1:虹鲟,M22731;ON2:红大马哈鱼2,U14551;AS:大西洋鲑,X14305;ON1:红大马哈鱼1(红大马哈鱼1和2来自不同研究结果),U14535;SG:虹鲟,J03797;TU:金枪鱼(未定种类),E01953;BP1:尖吻鲈,U16816

AA: *Anguilla anguilla* ; AP: *Anguilla japonica* , M24066; SC: *Sus scrofa* , X53325; PP: *Pangasius pangasius* , M63713; HM: *Hypophthalmichthys molitrix* , M94348; OS: *Oncorhynchus* sp. , E01334; AC: *Coregonus autumnalis* , X77245; SS: *Salmon* , E00964; SS2: *S. salar* , M21573; TT : *Thunnus thynnus* , E02688; BP: *Barramundi perch* , X59378; AB: *Acanthopagrus butcheri* , X59377; SA: *Sparus aurata* , S54890; ES: *Dicentrarchus labrax* , X65716; OA: *Odontes arg* , AF236091; ON: *O. niloticus* , M26916; SS3: *Solea senegalensis* , U01143; CM: *Chrysophrys major* , E02560; EP: *Euthynnus pelamis* , E02013; CD: *Caranx delicatissimus* , X55176; PO: *Paralichthys olivaceus* X12887; EE: *Eel* , E01353; IP: *Ictalurus punctatus* , S69215; CI: *Ctenopharyngo idella* , X60988; SG1: *S. gairdner* , M22731; ON2: *Oncorhynchus nerka* , U14551; AS: *Salmo salar* , X14305; ON1: *Oncorhynchus nerka* , U14535; SG: *S. gairdneri* , J03797; TU: *Tuna* , E01953; BP1: *Barramundi perch* , U16816

同的结构域(GD1 - GD5)中, GD1 - GD4 区域位于分子的外侧, 为 GH 的结合决定区域, 与 GH 的活性有关, 为 GH 与其受体专一性结合区域, 其中 R35、L39、H40、A43、S73、S97、L131、L141、D147 氨基酸残基与 GH 专一性结合的决定关系非常密切(图 2 中黑体字母), 而 GD5 则位于分子的内部, 该区域不决定 GH 的专一性, 与 GH 分子的结构、稳定性有关。

应用鱼类 GH 氨基酸序列, 对 GH 基因进行进化分析, 可以将鱼类 GH 分成 3 个不同的组, II 组中的二个亚组的成员拥有共同的祖先, 有可能原始的 GH 基因在进化过程中, 通过丢失或增加 1 个内含子形成二个不同的亚组; 含有 5 个内含子的 GH 基因可分为两个明显不同的进化分枝(II-2 亚组和 III 组), 从而分属于二个不同的组。从系统进化分析来看, 鳊 GH 更多地与哺乳类动物猪 GH 相似。

参考文献:

- [1] Allendorf F W, Thorgaard G H. Tetraploidy and the evolution of salmonid fishes [A]. The evolutionary genetics of fishes [C]. Plenum, New York, NY, 1984. 1 - 53.
- [2] Ber R, Daniel V. Sequence analysis suggests a recent duplication of the growth hormone encoding gene in *Tilapia nilotica* [J]. Gene, 1993, 125:143 - 150.
- [3] Ho W K K, Wong M W, Chan A P Y. Clone and sequencing of the grass carp (*Ctenopharyngodon idellus*) growth hormone gene [J]. Biochim Biophys Acta, 1991, 1090:245 - 248.
- [4] Tang Y, Lin C M, Chen T T, et al. Structure of the channel catfish (*Ictalurus punctatus*) growth hormone gene and its evolutionary implication [J]. Mol Mar Bio Biotechnol, 1993, 2:198 - 206.
- [5] Agellon L B, Davies S L, Chen T T, et al. Structure of a fish (rainbow trout) growth hormone gene and its evolutionary implications [J]. Proc Natl Acad Sci USA, 1988, 85:5136 - 5140.
- [6] Ber R, Daniel V. Structure and sequence of the growth hormone encoding gene from *Tilapia nilotica* [J]. Gene, 1992, 113:245 - 250.
- [7] Tanaka M, Toma Y, Ohkubo T, et al. Sequence of the flounder (*Paralichthys olivaceus*) growth hormone-encoding gene and its promoter region [J]. Gene, 1995, 165:321 - 322.
- [8] Ohkubo T, Araki M, Tanaka M, et al. Molecular cloning and characterization of yellowtail GH gene and its promoter: a consensus sequence for telost and avian pit-1/ghf-1 binding sites [J]. J Mol Endocrinol, 1996, 16:63 - 72.
- [9] David L Y, Ronald J. A minisatellite polymorphism in intron III of the barramundi (*Lates calcarifer*) growth hormone gene [J]. Genome, 1996, 39:934 - 940.
- [10] Masanori W, Masanori Y, Minoru Y, et al. Conserved and unique amino acid residues in the domains of the growth hormones [J]. J Biological Chemistry, 1989, 264(1):312 - 316.