

초어톨류사접수체 5a유전자(*tlr5a*)의 배열특성

손주성, 장성훈, 주창성

위대한 령도자 김정일동지께서는 다음과 같이 교시하시였다.

《양어과학과 기술에 대한 연구사업을 강화하고 선진적인 물고기기르기기술을 적극 받아들이어 우리 나라의 양어사업을 최신과학기술에 기초하여 발전시켜나가도록 하여야 합니다.》(《김정일선집》 증보판 제20권 178~179페이지)

포유동물에서 TLR(Toll-like receptor)족 유전자는 병원체결합분자형태(PAMPs)를 인식하고 염증성세포인자들과 I형인터페론을 발현시켜 병원성세균에 대한 유기체의 면역을 조절한다. 우리는 초어를 연구대상으로 하여 물고기의 선천성면역에서 중요한 역할을 하는 유전자인 *tlr5a*의 배열과 진화적특성을 밝혔다.

재료와 방법

실험에 리용한 프라이머 프라이머는 Primer premier 5.0프로그램으로 설계하였다.(표 1)

표 1. 초어*tlr5a* cDNA의 증폭에 리용한 프라이머

프라이머이름	배열(5'→3')	증폭위치
SMART II(TM)(정방향)	AAGCAGTGGTATCAACGCAGAGTACGCGGG	첫번째 사슬 cDNA의 증폭
CDSIII(반대방향)	AAGCAGTGGTATCAACGCAGAGTAC(T)30V N	첫번째 사슬 cDNA의 증폭
gcT5a-F(정방향)	AAGGGTCTTGAGGCGAGTG	열린읽기틀
gcT5a-R(반대방향)	CAAAGCAGCAGCGGAGTG	열린읽기틀
T5aG1(반대방향)	CACTCGCCTCAAGACCCT	5'-말단
T5aNG1(반대방향)	TGTTGTCACCCAAGTAGAGC	5'-말단
T5aG2(정방향)	AGGGTCTTGAGGCGAGTGGT	3'-말단
T5aNG2(정방향)	GGTCTTGAGGCGAGTGGTAT	3'-말단

cDNA의 전 배열의 증폭 SMART cDNA의 합성방법[1]을 리용하여 초어의 두신으로부터 분리한 mRNA를 역전사시켜 주형cDNA를 얻는다. 다음 Vector NTI프로그램으로 다중배열 상동성검색을 진행하고 진화적으로 보존된 배열을 찾아내어 축퇴프라이머를 합성하였다. 이 프라이머를 리용하여 PCR법으로 초어*tlr1*의 보존된 배열을 증폭한다. 다음 증폭된 토막을 pMD18-T운반체(《Promega》)에 재조합하고 배열분석을 진행한다. 배열분석자료에 근거하여 5'-와 3'-말단을 증폭하기 위한 프라이머를 설계한다. SMART II과 T5aG1(T5aNG1), T5aG2(T5aNG2)와 CDS III프라이머로 목적유전자의 양쪽 말단을 증폭한다. 증폭된 5'- 및 3'-말단의 PCR산물들을 pMD18-T운반체에 재조합하고 그 배열을 분석하였다.

생명정보학적분석 핵산배열의 분석은 Vector NTI프로그램으로 진행하였다. 단백질의 기능영역에 대한 분석은 SMART와 TMHMM프로그램으로 분석하였다.

결과 및 논의

1) 초어*tlr5a*의 핵산 및 아미노산배열에 대한 분석

SMART cDNA합성방법을 리용하여 얻어낸 초어*tlr5a*의 증폭산물들을 pMD18-T운반체에 클론화한 다음 배열분석을 진행하였다.

그 결과에 의하면 5'—말단 및 3'—말단 증폭산물로서 1 300, 798bp 토막이 얻어졌으며 중간배열로서 2 294bp의 배열이 분석되었는데 이 배열들은 호상 겹치므로 Vector NTI프로 그람으로 분석하면 3 054bp의 완성된 cDNA배열(폴리 A배열을 포함)이 얻어졌다.(그림 1)

AAAAATCTTTGCGTAAATGATAACATTTTCATTTGTGGGTAAACTGTTTTAACCTTTGTA	60
TTCTACAGGGATCTCACAA ATG GCAACAATACACACATTATCTCTGATCCTCCTTGGATTA	120
M A T I H T L S L I L L G L	14
TGCGTCAGCACTCAAATTTGTGAAATGCACCTCAGTGTGTTTCGGTTGGTGCTTCTGTCGCC	180
C V S T Q I V K C T S V C S V G A S V A	34
CTCTGCATCGATAAAGGTCTTCAAGATGTGCCAGAGCTTCCCTCCATATGTAAATAAAGTG	240
L C I D K G L Q D V P E L P P Y V N K V	54
GATTTGAGTAAGAACAATATTGCTGAACCTCAATGAAACATCCTTTTCTCATCTGGAAGGT	300
D L S K N N I A E L N E T S F S H L E G	74
CTACAGGTCCTTATACTCATGCACCAACAAGACTTGTGATCAGAAACAACACATTC	360
L Q V L I L M H Q T T R L V I R N N T F	94
AGAAGACTCTCTAACTTAACATCACTTCAGCTTGACTACAACAACCTTCCTGCGAATGGAT	420
R R <u>L S N L T S L Q L D Y N N F L R M D</u>	114
ACAGGAGCGTTTAAACGGATTATCCAACCTTAAAAATCTCACTCTTACTCAGTGCAGTTTA	480
<u>T G A F N G L S N L K N L T L T Q C S L</u>	134
GAGGATACCATTTTGTCTGGTGACTTCCTCAAACCTCTGGTGTCTCTTGAGATGCTTGTCT	540
E D T I L S G D F L K P <u>L V S L E M L V</u>	154
TTACGTGAAAACAACATTAAAAGAATCCAGCCAGCATTGCTCTTTTTTAAATATGAGGAGA	600
<u>L R E N N I K R I Q P A L L F L N M R R</u>	174
TTCCATGTGCTCGATCTCTCTCGCAACAAAGTGAAGAGCATCTGTGAAGAAGACCTCCTC	660
<u>F H V L D L S R N K V K S I C E E D L L</u>	194
AGCTTTTCAGGGTAAACATTTACGCTTCTGAAGCTGTCTCAGTGACACTGCAAGACATG	720
<u>S F Q G K H F T L L K L S S V T L Q D M</u>	214
AATGAGTACTGGTTAGGATGGGAAAAGTGTGGAAACCCATTTAAGAACATGTCCGTAAGT	780
N E Y W L G W E K C G N P F K N M S V S	234
GTATTGGACTTATCTGGAAATGGCTTTAATGATAACAATGCAAAGCTTTTCTTTGATGCA	840
V L D L S G N G F N D N N A K L F F D A	254
ATCACTGGTACCAAAAATACAAAGTCTCATTCTCAGTAACAGTCACAGCATGGGCAGTTCT	900
I T G T K I Q S L I L S N S H S M G S S	274
TCTGGTAATAATTCAAAAAGATCCAAACAAATTCACATTTAAGGGTCTTGAGGCGAGTGGT	960
S G N N S K D P N K F T F K G L E A S G	294
ATTAAGATTTTTCGATCTCTCCAATTCAAGCATTTTGTCTGTGCATATTCAGTATTTAGT	1020
I K I F D L S N S S I F A L S Y S V F S	314
TGTTTGTCTAGATCTAGAACAAATTACATTAGCAGAAAGTCGGATCAACAAGATTGAAAAA	1080
<u>C L S D L E Q I T L A E S R I N K I E K</u>	334
AGTGCATTTTTGGGTATGGCAAATTTGCTGAAGCTAAACCTGTCCAAAACTTCCTTGGT	1140
<u>S A F L G M A N L L K L N L S K N F L G</u>	354
AATATTGATTCTAATACATTTTCAGAATCTAGAGAAGCTTGAGGTGCTTGATTTGTCTTAT	1200
<u>N I D S N T F Q N L E K L E V L D L S Y</u>	374

AATCATATATGGATGCTTGGATATGAATCATTTTCGAGGGCTTCCAAATCTACTCAACCTA	1260
N H I W M L G Y E S F R G L P N L L N L	394
AATTTAACAGGAAATGCTCTCAAACATTTACATGCATTCGCAACCTTACCAAGACTGGAG	1320
N L T G N A L K H L H A F A T L P R L E	414
AAGCTCTACTTGGGTGACAACAAAATTTTATCTGTCTTTTATTTAATCAAAAATTTCCAAA	1380
K L Y L G D N K I L S V F Y L I K I S K	434
TATCTTACAACCCTTTACCTGGAACATAACATACTTTCTTCCTTGTCTGAGATCTCTTCACA	1440
Y L T T L Y L E H N I L S S L S D L F T	454
ATACTAGAGGAATTTTCCTCAAATTGAGGAAATTGTTTTTCGAGGTAATGAGCTTCTTTAT	1500
I L E E F P Q I E E I V F R G N E L L Y	474
TGCCCTAATGAAAGACACAAAGTGCTTTCACAAAAAATACAAATCCTTGATCTTGCATTT	1560
C P N E R H K V L S Q K I Q I L D L A F	494
GCAGGTTTGGAGTTATCTGGTCAGAAGGAAAATGTTTAAACGTGTTTAAACAATCTTCAC	1620
A G L E V I W S E G K C L N V F N N L H	514
CAGTTAAAACAGCTTTCTCTGAGTCACAACCTTGCTACAGTCTCTTCCCAAAGACATTTTT	1680
Q L K Q L S L S H N L L Q S L P K D I F	534
AAAGACCTTACCTCTTTGTACTTTTTGGATTTGTCCTTCAACTCTTTGAAGTACCTTCCA	1740
K D L T S L Y F L D L S F N S L K Y L P	554
AACGGTATATTCCCTGAAAGTCTTCAAATTCTTAATCTTGAATATAATTCTATTTATTCA	1800
N G I F P E S L Q I L N L E Y N S I Y S	574
GTAGATCCAAATCTCTTTAGCACCTCAGCTACCTCAGCCTGATAAAAAACGATTTCCGT	1860
V D P N L F S T L S Y L S L I K N D F R	594
TGTGATTGCAATTAAGGGATTTCCAACTTGGCTAAACCAAACCAATGTAATCATTCT	1920
C D C K L R D F Q T W L N Q T N V I I S	614
CACTCCATTGAGGATGTGATATGTGCCAGTCCTGAGGATCAGTACATGGTTCCGGTTGTG	1980
H S I E D V I C A S P E D Q Y M V P V V	634
AGATCCAGCATACAATGTGAGGATGAAGAGGACGAGAGAAACGCTGAAAAACTGAGGCTT	2040
R S S I Q C E D E E D E R N A E K L R L	654
GTGCTTTTTATTTTCTGTACCGCACTTATCACGTTACTCACTGCTAGCGCCATCATTTAT	2100
V L F I F C T A L I T L L T A S A I I Y	674
GTCCGTCGACGTGGCTACATCTTCAAGCTTTACAAAAAATCATTGGCAAACCTTGTGGAT	2160
V R R R G Y I F K L Y K K L I G K L V D	694
GGAAAGCGAGAGGAGCCTGATCCTGATCAATTCTTGTATGATGTGTTTCTTTGTTTGT	2220
G K R E E P D P D Q F L Y D V F L C F S	714
TCCAATGATATTAAGTGGGTAGAAAGAGCACTGCTGAACAGGCTAGACTCTCAGTTCTCA	2280
S N D I K W V E R A L L N R L D S Q F S	734
GAGCAGAACACACTCCGCTGCTGCTTTGAGGAGCGAGACTTCATACCCGGGGAGGACAAT	2340
E Q N T L R C C F E E R D F I P G E D N	754
CTTACCAACATGCGAAATGCTATCCAGAATAGTCATAAAACCCTTTGTGTGGTGTCTGAG	2400
L T N M R N A I Q N S H K T L C V V S E	774
CATTTCTGAAGGACGGCTGGTGCCTAGAAACCTTCACCCTGGCACAGTGCAGGATGCTA	2460
H F L K D G W C L E T F T L A Q C R M L	794
GTGGAGCTAAAGGACATTCTGGTGGTGTGTTGTAGGGAACATAACCGCAGTACAGGCTA	2520
V E L K D I L V V L V V G N I P Q Y R L	814
CTGAAGTACGAACAACTGAGATCCTACATTGAGAACAGAAGATACCTTCTGTGGCCCGAT	2580
L K Y E Q L R S Y I E N R R Y L L W P D	834
GACAGCCAGGACTTGGAGTGGTTTTATGACCAACTTCTGCATAAAATTAGAAAAACACC	2640
D S Q D L E W F Y D Q L L H K I R K N T	854
AAGGTTAAACAAACAAATACTAAAGTCAATGAGGGTGAAAAAATCTTGACGCAGCAAAT	2700
K V K Q T N T K V N E G E K N L D A A N	874

초어톨류사접수체 5a유전자(*tlr5a*)의 배열 특성

GTTCAAGCAGACACTGCAGTG TAA CTATAAAGGACTTTGTATAATTCCTTGCTTTGCTCC	2760
V Q A D T A V *	881
CAGGTTCTCATCCGTCCTACTTTGTTGAAATTCTCTCCATTGGACAATGACATCAACAAA	2820
TATTAAGCATATAAAAGCTGGAAACCATACTGTATTGTATAGTTATAAATGTTATTTTGT	2880
ATTATCCATAATGTGTATGTACTTCTCAGAGTCAGTTGGGGTAGTTAAATGTCAGTGGAA	2940
CGGTGACATTTTTTATCTATTATGTAAATACTCC ATTTA TATCACCTTTTTTTCATCATAG	3000
TTTTCAGGATTTGGCA AATAAA GTCTTAGAATAAAAAAAAAAAAAAAAAAAAAA	3054

그림 1. 초어*tlr5a*의 핵산과 아미노산배열

테두리를 친 부분은 보존된 TIR구조배열, 1개의 밀선을 그은것은 로이신평반복모티프배열(LRR), 2개의 밀선을 그은것은 로이신평반복모티프(LRR)의 C-말단배열, 회색으로 표시한 부분은 세포막투과배열, 사선강조체로 표시한것은 가능한 폴리아데닐산신호배열, 강조체에 굵은 밀선으로 표시한것은 전사배열의 불안정성을 보여주는 모티프배열, 번역개시코돈과 중지코돈은 강조체로 표시

초어*tlr5a*에서 5'— 및 3'—비번역배열의 길이는 각각 78, 330bp이고 열린읽기틀의 길이는 2 646bp이며 3'—비번역배열에 배열의 불안정성을 보여주는 1개의 모티프(ATTTA)가 있다. 이 배열에 의하여 암호화되는 단백질의 크기는 881aa이다.

그림 1에서 보면 이 유전자에는 번역시작코돈(ATG)과 중지코돈(TAA)이 있으며 3'—비번역배열에는 전사중지를 알리는 폴리아데닐신호(AATAAA)와 전사를 종결짓는 폴리 A배열이 있다.

2) 초어TLR5a접수체단백질의 기능영역분석

SMART와 TMHMM프로그램으로 분석하면 초어TLR5a접수체유전자는 1개의 신호펩티드배열(1~23aa)과 12개의 로이신평반복배열(97~579aa), 1개의 C—말단로이신평반복배열(591~641aa), 1개의 세포막투과배열(654~676aa)과 세포질안에 위치하는 1개의 TIR도메인(706~853aa)을 가지고있다.(표 2, 그림 2)

표 2. 초어TLR5a접수체에서 아미노산배열의 기능영역배열

배열이름	부터(aa)	까지(aa)	E-값
신호펩티드(signal peptide)	1	23	1 690
로이신평반복모티프(LRR)	97	120	0.498
로이신평반복모티프(LRR)	147	170	6.4
로이신평반복모티프(LRR)	172	195	203
로이신평반복모티프(LRR)	316	339	92.4
로이신평반복모티프(LRR)	340	363	4.32
로이신평반복모티프(LRR_TYP)	364	387	0.011
로이신평반복모티프(LRR)	388	409	19.7
로이신평반복모티프(LRR)	410	431	90.9
로이신평반복모티프(LRR)	433	452	6.18
로이신평반복모티프(LRR_TYP)	513	536	$2.61 \cdot 10^{-4}$
로이신평반복모티프(LRR_TYP)	537	560	0.022
로이신평반복모티프(LRR)	561	579	75.5
C—말단로이신평반복모티프(LRRCT)	591	641	$1.21 \cdot 10^{-9}$
세포막투과영역	654	676	1 770
톨인터로이킨접수체도메인(TIR)	706	853	$6.34 \cdot 10^{-16}$



그림 2. 초어TLR5a접수체의 기능영역

포유동물의 TLR5접수체는 세포질의 바깥에 세균의 편모를 특이적으로 인식하는데 적응된 로이신 풍부반복배열들을 여러개 가지고있고 세포질의 안쪽에는 진화상 보존되었으며 신호전달에 관여하는 TIR도메인을 가지고있다.[2]

맺 는 말

1) 초어*tlr5a*의 cDNA전배열의 크기는 3 054bp이다. 5'-비번역배열의 크기는 78bp이고 열린읽기틀(ORF)의 크기는 2 646bp이며 3'-비번역배열의 크기는 폴리 A를 포함하여 330bp이다. 열린읽기틀은 881개의 아미노산을 암호화한다.

2) 초어TLR5a접수체단백질은 12개의 로이신 풍부반복배열과 1개의 막투과배열, 보존된 TIR도메인배열을 가지고있다.

참 고 문 헌

[1] S. Giovanna et al.; Genomics, **64**, 230, 2000.

[2] S. I. Yoon et al.; Science, **335**, 859, 2012.

주체104(2015)년 12월 5일 원고접수

Sequence Characteristics of the *tlr5a* Gene from Grass Carp, *Ctenopharyngodon idella*

Son Ju Song, Jang Song Hun and Ju Chang Song

The full length of the grass carp *tlr5a* is 3 054 nucleotides (nt). The 5'-untranslated region(UTR) is 78nt and open reading frame is 2 646nt. The 3'-UTR is 330nt, including a poly A tail. The ORF of *tlr5a* encodes 881 amino acids(aa). It has 12 LRR motifs, a C-terminal LRR motif, one transmembrane zone and a TIR domains.

Key words: grass carp, *Ctenopharyngodon idella*, *tlr5a*, gene sequence, LRR motif