Gene 513 (2013) 40–52

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Genomic organization, polymorphisms and molecular evolution of the goose-type lysozyme gene from Zhikong scallop *Chlamys farreri*

Ling Li, Jianmin Zhao, Lingling Wang, Limei Qiu, Linsheng Song *

Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

ARTICLE INFO

Article history:
Accepted 11 October 2012
Available online 8 November 2012

Keywords:
Zhikong scallop *Chlamys farreri*
Gene structure
Transcription factor binding site
Repetitive sequence
Phylogenetic evolution

ABSTRACT

Lysozyme is a ubiquitous hydrolase that plays an important role in protecting the host against pathogenic infection. In the present study, the genomic DNA sequence of an invertebrate goose type (G type) lysozyme (designated CfLysG) was cloned from Zhikong scallop *Chlamys farreri* by genome walking technique. The full-length DNA of CfLysG gene was of 9455 bp, and the fragment from the transcription site to the polyadenylation site was of 8217 bp. A total of 104 SNPs and 29 ins–del polymorphisms were identified in the genomic sequence of CfLysG, and most of them were located in the promoter and intron regions, except for three SNPs located in the exon regions. Some putative transcription factor binding sites in the promoter region suggested the involvement of CfLysG in immune responses. There were six exons of 55, 60, 90, 113, 145 and 140 bp interrupted by five relatively large introns in the genomic DNA sequence. CfLysG exhibited a unique exon–intron organization which was different from both its vertebrate and invertebrate homologues. Though some introns were lost in the urochordate homologues, four of the five introns in CfLysG DNA shared homologous positions with vertebrate G type lysozyme genes, which indicated the existence of these introns in the ancestry of G type lysozyme. With respect to the number and size of both exons and introns, the gene organization of CfLysG was more similar to that of vertebrate G type lysozyme, but its amino acid sequence shared higher similarity with that of other invertebrate G type lysozymes. In the phylogenic tree, G type lysozymes from mollusk were clustered together and formed a sister clade to the urochordate and vertebrate G type lysozymes. G type lysozyme was separated from C and I type lysozymes and closely matched to the phage-type lysozyme. The results suggested that G type lysozyme might be evolutionarily closest to the lysozyme ancestor, and they would be helpful to understand the evolution of lysozyme genes.

1. Introduction

Lysozyme (muramidase, EC 3.2.1.17) is a ubiquitous hydrolase with lytic activity against bacterial peptidoglycan, which plays an important role in protecting the host against pathogenic infection. Since the discovery in 1922, lysozyme has long been extensively studied as a model for molecular evolution, protein structure and gene regulation. Several classes of the lysozyme have been described in highly diverged organisms from baceriophage to human (Labadie et al., 2007). Based on its differences in structural, catalytic and immunological criteria, lysozyme has been classified into the chicken (C), goose (G), phage, bacteria, plant, and invertebrate (I) types (Bachali et al., 2004; Beintema and Terwisscha van Scheltinga, 1996; Liu et al., 2006; Weaver et al., 1985).

G type lysozyme is identified from both vertebrates and invertebrates, including mollusks, oikopleuras, ascidians, fish, birds and mammals (He et al., 2012; Hikima et al., 2001; Irwin and Gong, 2003; Larsen et al., 2009; Nilsen et al., 2003; Savan et al., 2003; Simpson et al., 1980; Simpson and Morgan, 1983; Zhang et al., 2012; Zhao et al., 2007; Zhao et al., 2011; Zou et al., 2005). Recently, the genomic organizations of G type lysozyme have been characterized from various organisms, and they are different from urochordates to mammals. For example, the number of exons in G type lysozyme varies from a minimum of two in *Oikopleura dioica* to a maximum of seven in human g1 (Irwin and Gong, 2003; Nilsen et al., 2003). However, there is still no information about the genomic sequence and organization of G type lysozyme in protostomia.

Though G type lysozyme is a functional conserved molecule, its sequence varies significantly among diverse species, and even differs in the same species. In our previous study, eight sites of single nucleotide polymorphisms (SNPs) and two sites of insert–deletion (ins–del) polymorphisms were identified in the promoter region of G type lysozyme gene from Zhikong scallop *Chlamys farreri*, and the −391 A/G polymorphism was proved to be associated with the resistance of *C. farreri* to *Listonella anguillarum* (Li et al., 2009). Similarly, nine sites of SNPs and three sites of ins–del polymorphisms were also reported in the promoter

Abbreviations: (prime), denotes a truncated gene at the indicated side; bp, base pair(s); C type, chicken type; CDS, coding sequence; dNTP, deoxyribonucleoside triphosphate; EDTA, ethylenediamine tetra-acetic acid; EST, expressed sequence tag; G type, goose type; I type, invertebrate type; ins-del, insert-deletion; kb, kilobase(s) or 1000 bp; LB, Luria-Bertani (medium); RACE, rapid amplification of cDNA ends; RNase, ribonuclease; SDS, sodium dodecyl sulfate; SNP, single nucleotide polymorphism.

* Corresponding author. Tel.: +86 532 82898552; fax: +86 532 82898578.
E-mail address: lshsong@ms.qdio.ac.cn (L. Song).

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.10.080

Table 1
The major primers (from 5′ to 3′) used for identification of polymorphisms.

Primer name	Sequence (5′–3′)	Position
LysF1	GCTGAGTAGACTGGAACAAGCGGAATGA	−795 to −768
LysR1	AAGACGAGTGAAGGCCGGGGAGTAGGT	−60 to −34
LysF2	ATGAACCCACTGGCAGTACTCACACTTCT	1–29
LysR2	GTGTTGTCCATGGGGATGAAGTCGTG	1209–1234
LysF3	TACATCATCTCAAATCGGTGTTGCTTCAG	1041–1069
LysR3	TTTGGACGTGTACGGTATAATACTACCCT	3510–3538
LysF4	AGGGTAGTATTATACCGTACACGTCCAAA	3510–3538
LysR4	ATTATGCTGTGTTGGCGTTATGCAGTAC	5559–5586
LysF5	GTACTGCATAACGCCAACACAGCATAAT	5559–5586
LysR5	CTCGAACTCGACAGGACCACAGAAAAC	6939–6965
LysF6	CTCGATGTACTTTGCCGGTTCCTGTAAG	6860–6887
LysR6	AATGCCAGCCGTGGCTTGAAATAAGGTG	7947–7974

region of G type lysozyme from Japanese scallop *Mizuhopecten yessoensis*, and these mutations organized into two haplotypes which were associated with different transcription factor binding sites (He et al., 2012).

As increasing number of lysozymes have been identified from different species, the diversity raises the scientific questions on their evolution. Efforts have been devoted to understanding the relationships between lysozymes of different origins by comparison of their sequences and tertiary structures. However, no statistically significant sequence similarities have been detected between members of different classes but interesting structural similarities have been discovered (Matthews et al., 1981; Weaver et al., 1985). These comparisons have led to the assumption that the genes encoding these lysozymes have evolved from a remote common ancestor and these lysozymes have become a classical example of divergent evolution. Grütter et al. (1983) suggested that G type lysozyme represented an evolutionary link between C type and bacteriophage lysozymes, and all these three lysozymes evolved from the same predecessor. However, Thunnissen et al. (1995) considered that G type lysozyme took up a central position in the lysozyme superfamily and it was structurally closest to the lysozyme ancestor.

In our previous studies, the first invertebrate G type lysozyme was identified from *C. farreri* by EST approach and RACE technique. In the present study, the genomic DNA sequence of CfLysG was cloned, and its organization and architecture were analyzed comprehensively to better understand the phylogenetic evolution of lysozymes. Meanwhile, the polymorphisms in the DNA sequence of CfLysG were investigated to provide candidate markers for the selection of disease resistant scallops.

2. Materials and methods

2.1. DNA isolation and construction of genome walking library

The genomic DNA was extracted from adductor muscle of *C. farreri*, using the traditional phenol–chloroform method (Li et al., 2009). The genome walking library was constructed following the instruction of Universal GenomeWalker™ Kit (Clontech, USA). Five restriction endoribonucleases, *Sca* I, *Sma* I, *Stu* I, *Dra* I and *Ssp* I were selected to digest the genomic DNA. The adaptor 5′-GTAATACGACTCACTATAGGGCA CGCGTGGTCGACGGCCCGGGCTGGT-3′ 3′-H_2N-CCCGACCA-PO_4-5′ was ligated to both ends of the genomic DNA fragments to create GenomeWalker libraries.

2.2. Molecular cloning of the genomic DNA of CfLysG gene

The gene-specific primers, designed based on the cDNA sequence of CfLysG (GenBank ID. DQ227696), and the adaptor primers were used to clone the genomic DNA of CfLysG. The primary genome walking PCR was carried out in a PTC-100 Peltier Thermal Controller Cycler (MJ Research, USA) in a 25 μl reaction volume containing 2.5 μl of 10× PCR buffer, 1.5 μl of $MgCl_2$ (25 mM), 2 μl of dNTP mix (2.5 mM), 1 μl of adaptor primer AP1 and gene-specific primer (10 μM), 15.8 μl of PCR grade water and 1 μl of genome walking library. After hot start, 0.2 μl (1 U) of *Taq* polymerase was added to the mixture. The PCR temperature profile was two-step cycle PCR: 7 cycles of 94 °C for 25 s and 72 °C for 3 min, then another 32 cycles of 94 °C for 25 s and 67 °C for 3 min. The samples were finally incubated for further 7 min at 67 °C. The PCR products were analyzed on a 1% agarose/EB gel, along with DL2000 DNA size markers (TaKaRa, Japan).

For the secondary genomic-walking PCR, 1 μl of each 100× diluted primary genomic-walking PCR (including negative controls) products was used as template, and the primers were AP2 and corresponding nested primers. The PCR temperature profile was almost the same as the primary genomic-walking PCR reaction except for the slightly fewer cycles.

The PCR products were detected by agarose gel electrophoresis, purified and ligated into pMD18-T simple vector, and then transformed into *Escherichia coli* JM109 cells. The recombinants were identified through blue–white color selection in ampicillin-containing LB plates, and the positive clones were sequenced on an ABI 3730 Automated Sequencer (Applied Biosystem, USA). The generated sequences were screened for vector contamination using VecScreen software (http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html), and the adaptor sequences were then removed. The full-length sequence was generated by overlapping the sequence fragments resulted from the PCR with the 17 gene-specific primers (Supplementary data, Table S1).

2.3. Identification of polymorphisms in CfLysG

Six pairs of gene specific primers (Table 1), designed based on the sequence of CfLysG, were used to amplify the genomic DNA of CfLysG. PCR reaction was performed in 25 μl reaction volume containing 50 ng of DNA template, 1 μl of each primer (10 μM), 2.5 μl of 10× PCR buffer, 1.5 μl of $MgCl_2$ (25 mM), 2 μl of dNTP mix (2.5 mM), 15.8 μl of PCR grade water and 0.2 μl (1U) of *Taq* polymerase (TaKaRa, Japan). The PCR temperature profile was as follows: 94 °C for 5 min; 35 cycles of 94 °C for 40 s, 64 °C for 40 s and 72 °C for 2 min; and a further 10 min elongation at 72 °C.

The PCR products from six scallops were detected by electrophoresis and the fragments were purified. The objective fragments were then cloned into pMD18-T vector, transformed into *E. coli* Top10 cells, and at least two positive clones were sequenced for each fragment. The alignments of nucleotide sequences of CfLysG gene were performed using Vector NTI Suite 9 and the polymorphisms were detected from the sequence alignments of different scallops. Moreover, nucleotide sequences of exons were translated by Primer 5.0 to identify synonymous SNP and non-synonymous SNP.

Table 2
Species, genomic sequence, accession number and exon lengths of G type lysozyme gene used to compare the gene structures.

Code name	Organism	Accession no.	Exon 1	Exon 2	Exon 3	Exon 4	Exon 5	Exon 6	CDS
Human	*Homo sapiens*	NC_000002	88	68	141	197	139	237	639
Chicken	*Gallus gallus*	X61001	118	74	102	197	139	711	636
Japanese flounder	*Paralichthys olivaceus*	AB050591	79	78	215	151	231		588
Ascidian	*Ciona intestinalis*	NW_001955274	185	80	407				564
Oikopleura	*Oikopleura dioica*	AJ564629	>382	320					627
Zhikong-scallop	*Chlamys farreri*	JX233811	70	60	90	113	145	312	603