Identification and sequence characterization of LOC427400: a novel gene highly expressed in the Magnum of White Leghorn’s Oviduct

Ying Zhang, Bolin Ning, Jingejun Li, Qianyu He, Chunbo Wei and Shengjun Liu*

College of Animal Science and Veterinary Medicine, Heilongjiang BAYI Agricultural University, Daqing, Heilongjiang, P. R. China, 163319.

Received: 29-06-2017 Accepted: 24-02-2018 DOI: 10.18805/ijar.B-764

ABSTRACT
LOC427400 is a novel bird gene located on chromosome Z of the chicken genome. It was deciphered in our previous study by RNA sequencing of different parts of the chicken oviduct. Here we confirmed that LOC427400 expressed significantly high in the magnum comparing with the ovary, isthmus, uterus, liver, jejunum and breast muscle (P<0.05) by quantitative real-time PCR in 5 White Leghorn layers at the age of 34 wk. The full length of transcript sequence of LOC427400 was also obtained by 3' and 5' RACE. The new transcript was 2831 bp and the corresponding translated protein was 343 aa. The predicted protein show 2 conserved domains, RICIN and OST-beta, which indicated that LOC427400 protein might have functions like carbohydrate-binding and signal transduction. This study illustrated that LOC474 00 should be an important candidate gene related with the process of egg albumen formation and albumen biological functions.

Key words: Chicken, Egg albumen, Gene expression, LOC427400 gene, Rapid amplification of cDNA ends.

INTRODUCTION
Chicken eggs for the table should be fresh and good quality. Eggs provide many kinds of proteins for human health and these proteins are mainly from the egg white. Egg white (albumen) takes 60 ~ 63% of an egg weight (Mine, 2008) and it is widely preferred for its nutritive properties and low cholesterol content. Albumen with good quality is a high-viscosity structure. It makes the yolk in the center of the egg, protects the interior of eggs for its antimicrobial components (Raikos et al., 2006), and provides necessary nutrients and bioactive constituents for chicken embryo development (Benton et al., 2001; Silphaduang et al., 2006). The good quality of albumen is also emphasized because the increasing amount of eggs are used in the food industry such as bakery (Honkatukia et al., 2005).

As the storage time extends, the egg quality declines. Gel-like albumen becomes thin and running. Its antimicrobial capacity also becomes weaker. So the eggs are susceptible to bacterial contamination, such as salmonella (Omana et al., 2011). A potential reason for the albumen quality decline is that CO2 leak through the shell pores changes albumen pH towards alkaline values (Jin et al., 2011).

Most albumen proteins are thought to be secreted in the magnum of the laying hen’s oviduct. This process takes about 3 hours. Albumen formation involves complex processes, including expressions of tissue-specific genes, like ovalbumin (Ov) gene, interactions with steroid hormones, like estrogen, androgen etc., and other regulating factors (Dean et al., 2001; Dougherty and Sanders, 2005; Gupta et al., 2007; Park et al., 2006; Yildirim et al., 2017). Albumen quality is a typical quantitative trait of layers. Phenotypic variances would be partly interpreted by genetics (Hosseini and Tahmoorespur, 2013). But the researches on the genes or QTLs associated with albumen trait are relatively few comparing with other egg quality traits. As of March 2016, 19 QTLs are associated with HU, 34 QTLs for AH, and 11 QTLs for AW in the ChickenQTLdb (Chicken QTL database: http://cn.animalgenome.org/cgi-bin/ QTLdb/GG/index). SNPs associating with early or late AW significantly were reported on chromosomes 1, 3, 5, 18, 19, 23 and Z by genome-wide association studies (Abasht et al., 2009).

Albumen proteomic researches have greatly expanded the identified albumen proteins in various layer breeds (Guerin-Dubiard et al., 2006; Mann, 2007; D’Ambrosio et al., 2008; Mann and Mann, 2011; Wang et al., 2012; He et al., 2014). The changes of albumen protein components during the degeneration process was also reported (Omana et al., 2011). However, there were few published studies about the functions of the genes nor the SNPs associating the albumen traits. The genetic mechanisms influencing albumen quality is still not clear.

LOC427400 is a novel gene (ENSGALG0000-0000581), located on Chromosome Z of chicken genome and was between nucleotide positions 7480622 to 7495980.
totally 15.3 kb long. The transcript of the gene (ENSGALT0000009333) is 2284 bp, composing 6 exons that range from 69 bp (exon 4) to 1377 bp (exon 6) and the corresponding protein was 194 aa. The gene was deciphered by RNA sequencing and transcriptome analysis of different tissues of the White Leghorn layers (ovary, magnum, isthmus, and uterus). It was found that the expression level of LOC427400 was distinctively higher in the magnum than other parts of the layer’s oviduct (the RNA sequencing data have not been published).

Many studies supported that the genes highly expressed in specific tissues played important roles in species-specific traits (Hincke et al., 1999; Gautron et al., 2007). The birth of new genes contributes a lot to the evolutionary research and drew much attention of biological scientists (Long et al., 2003; Kaessmann et al., 2009). Based on the hypothesis above, this study investigated the expression patterns and sequence structures of the LOC427400 gene which might play important roles in the process of egg albumen formation.

**MATERIALS AND METHODS**

**Quantitative real-time PCR:** The LOC427400 gene expression levels in various tissues of layers were examined using quantitative real-time PCR. The relative expression levels of LOC427400 were normalized to GAPDH gene to control possible differences in RNA isolation and cDNA reverse transcription efficiencies between different samples.

4 White Leghorn layers (34 week old) were sacrificed when they were observed that the egg had already moved into the magnum part of the oviduct. 7 Tissue samples (ovary, magnum, isthmus, uterus, liver, jejunum and breast muscle) of each layer were taken and frozen immediately in liquid nitrogen, and stored at -80°C. Total RNA was isolated by the Total RNA Extraction Kit (Omega Bio-Tek Inc., Beijing, China). Specific primers used in the real-time PCR were designed with PrimerExpress 3.0 software and the sequences were listed in Table 1.

**Table 1:** Primer sequences used in quantitative real-time PCR and RACE.

<table>
<thead>
<tr>
<th>Primer1</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH-F</td>
<td>CTTCCGTGTCGCCAACC</td>
</tr>
<tr>
<td>GAPDH-R</td>
<td>CATCAGCAGGACCTTCAC</td>
</tr>
<tr>
<td>qLOC-F</td>
<td>TCTGAAGCCGTGGATAAT</td>
</tr>
<tr>
<td>qLOC-R</td>
<td>GGTCACTAGACCCGAGA</td>
</tr>
<tr>
<td>3GP1</td>
<td>TGGTGCTTCCTGGGATTT</td>
</tr>
<tr>
<td>3nGP3</td>
<td>GCTTTCTGGCTCCTTGCAAA</td>
</tr>
<tr>
<td>5GP1</td>
<td>CCGCGACACAGACAGAGT</td>
</tr>
<tr>
<td>5GP2</td>
<td>CCGCGACACAGACAGAGT</td>
</tr>
<tr>
<td>5nGP3</td>
<td>GGGCTGACAGGCACTGGTGTC</td>
</tr>
</tbody>
</table>

1 F = forward primer; R = reverse primer.
2 Primers used in quantitative real-time PCR.
3 Primers used in 3’ and 5’ RACE. GSP = Gene Specific Primer; nGSP = nested Gene Specific Primer. 3GP and 3nGP were primers used in 3’ RACE; 5GP1, 5GP2 and 5nGP were primers used in 5’ RACE.

Relative gene expression was conducted using the standard curves of GAPDH (as reference) and LOC427400 genes. The PCR products of the 2 genes were cloned into pMD19-T vector (Takara Bio Inc., Otsu, Shiga, Japan) after multiplication in DH5α *Escherichia coli* competent cells (Tiangen Biotech, Beijing, China). Ten-fold serial dilutions of the relevant plasmid DNA samples were made duplicate and served as the calibrator. Three biological replicates were used for each sample and a no-template control was included in each run. PCR was performed in a total volume of 15 μl with 1 μl template cDNA, 100 nM primers and 1xPCR mix (Power SYBR® Green PCR Master Mix, Applied Biosystems). The thermal cycling procedure was as 95°C for 10 min, 40 cycles of 95°C for 15 s, 60°C for 1 min, 95°C for 15 s, 60°C for 30 s and 95°C for 15 s.

The LOC427400 expression levels of all tissues were illustrated relative to GAPDH gene. The ANOVA procedure in SAS 8.0 (SAS Institute Inc., Cary, NC) was used to identify the expression differences of the LOC427400 gene in different chicken tissues.

**Rapid amplification of cdNA ends (RACE):** We cloned the full length sequence of LOC427400 transcript with 5’ and 3’ RACE System for Rapid Amplification of cDNA Ends (Invitrogen, USA) following the manufacturer’s protocols. The cDNA used for the RACE PCR was reverse transcribed using the mRNA extracted from the magnum tissues (the same with those used in the above experiment). The sequences of the gene specific primers (GSP) used in 3’ and 5’ RACE were shown in Table 1. The GSP primers were designed with Oligo 6.0 software. Purified PCR products were cloned into pMD 19T vector (TaKaRa Bio Inc., Otsu, Shiga, Japan) and reproduced in DH5α *Escherichia coli* competent cells (Tiangen Biotech, Beijing, China) for plasmid propagation. The recombinant plasmids with target fragments were sequenced subsequently.

Nucleotide sequences of LOC427400 transcript were translated into amino acid residues and open reading frame (ORF) was predicted by DNAMAN software package. Protein sequence alignment was done with Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) and shaded by Boxshade program.

**RESULTS AND DISCUSSION**

The normalized LOC427400 expressions in the magnum were significantly higher than those in the other tissues examined (*P < 0.05, Fig 1*). And the gene expression levels in the ovary, isthmus, uterus, liver, jejunum, and breast muscle were not significantly different. Real-time-PCR is commonly used as a validation tool for verify the gene expression results of microarray and RNA-Seq analysis (Seal et al., 2012). The results in this part were consistent with the previous RNA-Seq analysis. Magnum is the main tissue for albumen formation. It is reasonable to infer that
Fig 1: Relative expression of LOC427400 to GAPDH gene in 7 tissues of 4 White Leghorn layers. There was an egg in the magnum when all the 4 layers sacrificed. Error bars are the standard errors of the means. The expression level of LOC427400 gene in the magnum was significantly higher than the other 6 tissues ($P < 0.05$).

Fig 2: Nucleotide and translated amino acid sequences for LOC427400 gene. Nucleotides are numbered on the left side of the sequences. Amino acids are in single letters under the cDNA sequences. The underlined bases denote the location of the primer sequences for rapid amplification of cDNA ends (RACE). Polyadenylation signal (AATAAA) and stop codon are indicated by wave underline and asterisk respectively. 5GSP1, 5GSP2 and 3nGSP were primers used in 5’ RACE and 3GSP and 3nGSP were primers used in 3’ RACE.
the LOC427400 gene might play important roles in the albumen formatting process.

By 3' and 5' RACE, 386 bp 3'-end and 161 bp 5'-end cDNA fragments of LOC427400 were obtained individually (Fig 2). The full-length cDNA sequence of LOC427400 was 2831 bp, 544 bp longer than that in Ensembl genome browser database (release 84). This cDNA sequence was then blasted to the Chicken Genome using UCSC Blat web server (http://genome.ucsc.edu/) and the identification was 99.8%. The blasting results also show that there were no homologous sequences of LOC427400 among chicken, human, rat, mouse, zebrafish, opossum, and x_tropicalis, which supported LOC427400 gene was bird specific new gene.

A maximum 1032 bp long open-reading frame was predicted by DNAMAN software using the new LOC427400 cDNA sequence. The translated protein is 343 amino acids, 49 aa longer than the original predicted protein (Fig 2). The predicted molecular mass is 38.5 kDa and its theoretical pI is 9.33 (http://web.expasy.org/cgi-bin/protparam/protparam).

ENSGALP00000009319 was the corresponding protein of the LOC427400 gene (ENSGALG00000005812) and it belonged to an Ensembl protein family (family ID: ENSFMOO71001445971). Besides ENSGALP00000009319, there were totally four proteins in this family. The other three proteins were ENSAPLP00000007096 (duck), ENSFALP00000010297 (Flycatcher), and ENSTGUP000-

Fig 2: The amino acid sequence alignment of the translated LOC427400 protein and the 4 proteins in the family ENSFMOO71001445971.

The sequences were aligned using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) with default parameters. The identical amino acid residues are shown in black, and similar acid residues are shown in gray; the hyphens denote gaps.
LOC427400 located on chicken chromosome Z, highly expressed in the hen’s oviduct. These illustrate that LOC427400 was a special and important gene.

Laying eggs is one special character of birds. In a short period of time, chicken transported massive proteins, water and mineral substance into the magnum of the oviduct and form the albumen of an egg. This study show a new bird gene, LOC427400, highly expressed in the magnum during the albumen formation. The predicted peptide sequence analysis of LOC427400 illustrated its function may be related with carbohydrate-binding, enzymatic activity and signal transduction. From the researches above, we can infer that the LOC427400 is an important gene associated with the process of the albumen formation and it is worthy of further study.

ACKNOWLEDGEMENT
This work was funded by the Startup Foundation for the Doctors in Heilongjiang BAY1Agricultural University (Grant No. XYB2013-08) and the Natural Science Foundation of Heilongjiang Province (Grant No. C2016044).

REFERENCES


