Genetic Polymorphisms in a 1.2 kb Long Fragment within Intron 2 of Chicken UBD2 Gene

H.Y. Zhang1,2, W.J. Yang1,2, Y.Z. Luo1 and J.L. Han1,3
1College of Animal Science and Technology, Gansu Agricultural University, Lanzhou 730070, China
2CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing 100193, China
3International Livestock Research Institute (ILRI), P.O. Box 30709, Nairobi, Kenya

Abstract: The widely expressed chicken UBD2 gene in different types of tissues from embryonic to adult developmental stages implies its important role in regulating protein ubiquitination and delivery of ubiquitinated substrates; however, there is no specific study on the genomic DNA structure and function of chicken UBD2 gene except a few validated SNPs derived from ESTs. In this study, a 1.2 kb long fragment within intron 2 of chicken UBD2 gene was amplified and directly sequenced from a flock of 11 White Leghorn chickens. A total 15 sequences were identified from these birds, of which seven were homozygous but the remaining four were heterozygous. These new sequences were analyzed together with the homologous region in Gallus gallus genome version 71.4 (Chromosome 13: 8104760–8105979 in Galgal4). There were 10 polymorphic sites including seven transitions and three transversions, which defined four haplotypes in these 16 sequences. The Gallus gallus reference sequence formed a specific and distinct haplotype while the White Leghorn chickens carried three haplotypes at very similar frequencies of four, five and six sequences each. These data were the first piece of evidence of genetic polymorphisms present in the chicken UBD2 gene which is therefore warranted for further investigation on the functional diversity in its complete genomic DNA sequence of different genetic backgrounds.

Key words: chicken, UBD2 gene, intron 2, polymorphism

INTRODUCTION
Ubiquitin domain containing protein 2 (UBD2) is an approved name (symbol) by the HUGO Gene Nomenclature Committee (HGNC) to replace the dendritic cell-derived ubiquitin (Ub)-like protein (DC-UbP). UBD2 protein was first identified and localized in cytoplasm, especially in mitochondrion. It was detected in human dendritic cells and found highly expressed in tumour cell lines but not detectable in most normal tissues (Liu et al., 2003). The nearly complete mRNA sequence of human UBD2 gene (cDNA clone MGC:30022) (Strausberg et al., 2002) or the C-terminal residues of this protein (Liu et al., 2003) was first isolated from mammary adenocarcinoma or dendritic cells. The former (GenBank accession no. BC019910) encoded 190 amino acids while the latter (AF251700) was initially identified as an Ub-like (UbL) protein with 106 amino acids at its C-terminal. It is mapped to 5q35.1 in human chromosome 5 (Liu et al., 2003; Schmutz et al., 2004). In fact both sequences were incomplete for the coding sequence but missing the first partial exon 1 and/or even partial exon 2 of the gene which was annotated as such following wrongly recognized positions of start codons. The latest version of human UBD2 gene reference mRNA sequence covering complete three exons, two introns and its 3'-untranslated region (NM_152277.2, last updated on April 17, 2013) was annotated with additional mRNA sequences of BU902027.1 and BU846695.1 (Strausberg et al., 2002), AK022894.1 (Ota et al., 2003) and AU117374.1 (Kimura et al., 2006). The complete human UBD2 protein reference sequence carries 234 amino acids (NP_689490.2, released on June 27, 2006).

The UBD2 gene was found to be rather conserved and a homologous region to human UBD2 gene was accepted by the Chicken Gene Nomenclature Consortium (CGNC) as chicken UBD2 gene with a CGNC gene ID of 2027 based on following predictions. The predicted chicken UBD2 gene is mapped to chromosome 13: 8,075,687–8,122,847 forward strand of Ensembl Gallus gallus version 71.4 (Galgal4, released in April 2013). In the Ensembl database, the annotation was produced by the Ensembl gene build method (http://www.ensembl.org/info/docs/genebuild/gencode_annotation.html) as a novel protein (first entry in November 2010 and last update in March 2013). Its gene ID was assigned to as ENSGALG000000022810, transcript ID as ENSGALT00000044443 and protein ID as ENSGALP0000004434 (updated in April 2013). The
annotated mRNA and protein were supported by two transcripts of European domestic ferret (JP020658.1) (León et al., 2013) and zebrafish (NM_001002718.1) (Strauberg et al., 2002; Gwyn et al., 2005), four mRNA sequences from African clawed frog (NM_001094450.1) (Klein et al., 2002; Strauberg et al., 2002), western clawed frog (NM_001108701.1) (Klein et al., 2002; Hellsten et al., 2010), Rhesus monkey (JU477829.1) (Gibbs et al., 2007) and cattle (NM_001056791.1) (Hartley et al., 2006) as well as four complete genomic DNA sequences from mouse (AL62780.20 and AL89844.20) (Gerhard et al., 2004; Carninci et al., 2006; Church et al., 2009) and human (CH471062.2 and AC024561.5) (Venter et al., 2001). In the GenBank database, the reference genomic sequence of chromosome 13 genomic scaffold (NW_003783912) in Gallus_gallus-4.0 was annotated for the UBDT2 gene using gene prediction method of GnomON process (http://www.ncbi.nlm.nih.gov/genome/annotation_euk/process/) with supporting evidence of similarities to 10 expressed sequence tags (ESTs). Its Entrez gene ID was assigned to as 416202 (updated on April 21, 2013), reference protein as XP_429403 and mRNA as XM_429403 (both released on December 16, 2011). The Gallus_gallus-4.0 is from the whole-genome assembly released by the International Hicken Genome Sequencing Consortium in November 2011 (see http://genome.wustl.edu/genesview/gallus_gallus) and it is a mixed assembly of HTG BAC clones (Hillier et al., 2004) and contigs from the whole-genome shotgun project AADN3006712.1 which is assembled into the chromosome 13 of Gallus_gallus-4.0. Similar result was also obtained through a further blast search against the Ensembl Gallus gallus genome version 71.4 (Galgal4, released in April 2013). This fragment is located within the intron 2 of chicken UBDT2 gene which motivated our effort to re-design of pair of specific primers UBDT2f and UBDT2r for this study (Fig. 1). All primers were synthesized by Shanghai Sangon Biological Engineering Technology & Services Co., Ltd. (Shanghai, China).

8104723 8104752
UBDT2 CACAACTGCATCTGAAACAGCATCCTCAG
C348 GC . . GAGT . CATCC . G . . . . . . . .
UBDT2f . . . . . . . . . . . . . . . . . . . . . .

81056020 8105992
UBDT2 TGATGTGAGTTGACATGGTGTTAC
C350 A . . GA . T . . . . . . . .
UBDT2r . . . . . . . . . . . . . . . . . . . . . .

Fig. 1: The primer sequences and comparison with homologous regions in Gallus gallus UBDT2 gene (all sequences are in 5' - to 3'- directions). Dots indicate identical nucleotides to UBDT2 gene sequences. Numbers on top of the sequences are starting and ending nucleotide positions scored against the chromosome 13 of Ensembl Gallus gallus genome version 71.4 (Galgal4, released in April 2013).
The 10 polymorphic sites (left) and median joining network relationship (right) between the four haplotypes identified from 11 White Leghorn chickens and the single red jungle fowl. Circles filled with black colour are three haplotypes identified in the White Leghorn chickens while the single circle with grey colour is the *Gallus gallus* reference sequence. The sizes of the circles are in proportion to the number of shared sequences. Empty circle represents a median vectors (mv1). Numbers in bold font are 10 specific polymorphic sites linking every two haplotypes.

**PCR conditions and direct sequencing:** All PCR reactions were performed with 100-200 ng of genomic DNAs in a 50 µL volume using LAX-Taq polymerase (Beijing Huitian Dongfang Sci. & Tech. Co., Ltd., Beijing, China) following the same procedure of Li et al. (2010). All PCR products were purified and directly sequenced by the Beijing Sunbiotech Co., Ltd. (Beijing, China) using the same PCR primers of UBDT2f and UBDT2r.

**Data analysis:** DNA sequences were manually edited, aligned, and analyzed using the MEGA software version 5.0 (Tamura et al., 2011) and DnaSP software version 5.10.1 (Librado and Rozas, 2009). A median joining network was constructed using NETWORK software version 4.6.1.1 (Bandelt et al., 1999) to illustrate the phylogenetic relationship of the haplotypes.

**RESULTS AND DISCUSSION**

As what is shown in Fig. 1, the last partial sequences of both C348 and C350 primers had relatively high matches to homologous regions in chicken UBDT2 gene, which explained the ‘non-specific’ amplification of this fragment when they missed the first target in chicken BF2 gene. The sequences of newly designed primers were therefore directly taken from the same homologous regions of this gene. The quality of sequencing data of all 11 samples were sufficient for analysis with 1220 readable nucleotides, of which seven birds were homozygous but other four heterozygous genotypes. A total of 15 sequences were obtained from these genotypes (GenBank accession no. KF145057-KF145071) and they were aligned and analyzed together with homologous region in the *Gallus gallus* reference genome version 71.4 (Chromosome 13: 8104760 - 8105979 on the Galgal4).

There were 10 polymorphic sites, of which two were singleton variable sites and the remaining eight were parsimony informative sites while seven were transitions and other three were transversions (left in Fig. 2). These sites defined four haplotypes, three from the White Leghorn flock and the other of the *Gallus gallus* reference sequence. Nei’s haplotype diversity (Hd) was 0.742 ± 0.057 and nucleotide diversity (Pi) 0.0034 ± 0.0003 (Nei, 1987). Test of pair-wise number of nucleotide differences (Slatkin and Hudson, 1991; Rogers and Harpending, 1992) indicated no population size change in the White Leghorn chicken flock assuming no recombination among the haplotypes.

As shown in the right of Fig. 2, the frequencies of haplotypes 1, 2 and 3 are very similar to each other in this White Leghorn chicken flock while the single *Gallus gallus* reference sequence is an independent haplotype which is relatively far away from and linked through the only median vector (mv1) to all three chicken haplotypes. Haplotypes 1 and 3 are separated from each other by only three polymorphic sites while haplotype 2 is the most distant to other three haplotypes.

Because there is no single specific study on the chicken UBDT2 gene, we searched the current versions of both GenBank (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?db=bigsnp&locusid=416202) and Ensembl (http://asia.ensembl.org/Gallus_gallus/Transcript/PutVariation?db=core;g=ENSGALG00000002810;r=13:8075587-8122847; t= ENSGALT0000004443) databases and found that there were six synonymous single nucleotide polymorphisms (SNPs) located within the coding sequences of chicken UBDT2 gene. These SNPs, mostly in substitutions, were simply derived from the alignment of more than 10 chicken ESTs of cDNA libraries constructed from various tissues of Leghorn chickens among other commercial breeds or strains, often in pooled tissues of mixed strains (Boardman et al., 2002; Hubbard et al., 2005; Carre et al., 2008). None of them has been subjected to validation, therefore they may be false positives resulted from simultaneous mutations during cloning sequencing of various cDNA libraries. Nonetheless, similar to mammals, the widely expressed chicken UBDT2 gene in different types of tissues from embryonic to adult developmental stages implies its important role in regulating protein...
ubiquitination and delivery of ubiquitinated substrates (Song et al., 2010). The 10 polymorphic sites and three additional haplotypes identified in a single White Leghorn chicken flock involved in this study are certainly the first piece of evidence of genetic polymorphisms present within the introns of chicken UBD2 gene which is warranted for further investigation on its complete genomic DNA sequence of different genetic backgrounds (Ommeh et al., 2010).

ACKNOWLEDGEMENTS

The work was supported by funding to CAAS-ILRI joint laboratory from CAAS and the National Key Project of Scientific and Technical Supporting Programs Funded by Ministry of Science & Technology of China during the 11th 'Five-Year' Plan (No. 2008BADB2B01).

REFERENCES


Li, X.X., L.X. Han and J.L. Han, 2010. No specific primer can independently amplify the complete exon 2 of chicken BLB1 or BLB2 genes. Int. J. Poult. Sci., 9: 192-197.


