

ISSN 1682-8356
ansinet.org/ijps



INTERNATIONAL JOURNAL OF
POULTRY SCIENCE

ANSI*net*

308 Lasani Town, Sargodha Road, Faisalabad - Pakistan
Mob: +92 300 3008585, Fax: +92 41 8815544
E-mail: editorijps@gmail.com

Genetic Polymorphisms in a 1.2 kb Long Fragment within Intron 2 of Chicken *UBTD2* Gene

H.Y. Zhang^{1,2}, W.J. Yang^{1,2}, Y.Z. Luo¹ and J.L. Han^{2,3}

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 *UBTD2* 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 *UBTD2* gene except a few invalidated SNPs derived from ESTs. In this study, a 1.2 kb long fragment within intron 2 of chicken *UBTD2* 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* reference 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 *UBTD2* 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, *UBTD2* gene, intron 2, polymorphism

INTRODUCTION

Ubiquitin domain containing protein 2 (*UBTD2*) is an approved name (symbol) by the HUGO Gene Nomenclature Committee (HGNC) to replace the dendritic cell-derived ubiquitin (Ub)-like protein (DC-UbP). *UBTD2* 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 *UBTD2* 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 *UBTD2* 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 BU849695.1 (Strausberg *et al.*, 2002), AK022894.1 (Ota *et al.*, 2003) and AU117374.1 (Kimura *et al.*, 2006). The complete human *UBTD2* protein reference sequence carries 234 amino acids (NP_689490.2, released on June 27, 2006).

The *UBTD2* gene was found to be rather conserved and a homologous region to human *UBTD2* gene was accepted by the Chicken Gene Nomenclature Consortium (CGNC) as chicken *UBTD2* gene with a CGNC gene ID of 2027 based on following predictions. The predicated chicken *UBTD2* gene is mapped to chromosome 13: 8,075,587-8,122,847 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/genome_annotation.html) as a novel protein (first entry in November 2010 and last update in March 2013). Its gene ID was assigned to as ENSGALG00000002810, transcript ID as ENSGALT00000004443 and protein ID as ENSGALP00000004434 (updated in April 2013). The

annotated mRNA and protein were supported by two transcripts of European domestic ferret (JP020628.1) (León *et al.*, 2013) and zebrafish (NM_001002718.1) (Strausberg *et al.*, 2002; Guyon *et al.*, 2005), four mRNA sequences from African clawed frog (NM_001094450.1) (Klein *et al.*, 2002; Strausberg *et al.*, 2002), western clawed frog (NM_001101807.1) (Klein *et al.*, 2002; Hellsten *et al.*, 2010), Rhesus monkey (JU477829.1) (Gibbs *et al.*, 2007) and cattle (NM_001015679.1) (Harhay *et al.*, 2005) as well as four complete genomic DNA sequences from mouse (AL662780.20 and AL669844.20) (Gerhard *et al.*, 2004; Carninci *et al.*, 2005; 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_003763912) in *Gallus_gallus*-4.0 was annotated for the *UBTD2* 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 Chicken Genome Sequencing Consortium in November 2011 (see http://genome.wustl.edu/genomes/view/gallus_gallus) and it is a mixed assembly of HTG BAC clones (Hillier *et al.*, 2004) and contigs from the whole-genome shotgun project AADN00000000.3 based on a female red jungle fowl (*Gallus gallus*) of the inbred line UCD001.

The predicted chicken *UBTD2* gene also has three short exons but two very long introns of 26,286 bp and 19,814 bp each, however, the current version of chicken *UBTD2* protein in the GenBank seems to have 299 amino acids (XP_429403), of which 208 out of the first 221 residues are identical to its human counterpart. Nevertheless, the recently updated prediction of transcript sequence in the Ensembl (ENSGALT00000004443) shares an almost identical coding sequence in 705 bp (except five synonymous nucleotides) with the representative chicken mRNA sequence of BX930477 derived from a Leghorn chicken of Compton Line 15l (Hubbard *et al.*, 2005), of which 70 bp are of exon 1, 237 bp of exon 2 and the rest 398 bp of exon 3. Both chicken sequences are highly homologous to human reference mRNA sequence of NM_152277.2, encoding 234 amino acids (ENSGALP00000004434), of which 219 residues are identical to human *UBTD2* protein. It is surprised that the alignment of genomic re-sequencing data of a Chinese local Silkie chicken, a male White Cornish broiler and a female White Leghorn layer against the *Gallus gallus* reference genome sequence (Hillier *et al.*, 2004; Wong *et al.*, 2004) did not reveal any genetic variation within 47,261 bp long complete *UBTD2* genomic DNA

sequences (http://www.ensembl.org/Gallus_gallus/Location/SequenceAlignment?db=core;ftype=Xref_Uni_prot_genename;g=ENSGALG00000002810;id=UBTD2-201; r = 13: 8075587 - 8122847; t = ENSGALT00000004443). However, there is no specific study on chicken *UBTD2* gene structure or protein function so far. In this study, we characterized the nucleotide polymorphisms in a 1.2 kb fragment within intron 2 of chicken *UBTD2* gene.

MATERIALS AND METHODS

Samples: Blood samples were collected from 11 birds of a commercial White Leghorn chicken flock.

Primers: During our separate study on chicken *BL2* gene using the C348 and C350 primers (Fig. 1) from Shaw *et al.* (2007), a clean but non-specific PCR product in around 1300 bp was detected in one of the samples. This product was directly sequenced using the same PCR primers in sequencing reactions and the sequence was clearly not our interest of targeted *BF2* gene. A blast search of this new sequence against the NCBI database found a single match by 100% coverage and 99% identity to a fragment in a whole genome shotgun sequence (AADN03006712.1) 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 *UBTD2* gene which motivated our effort to re-design of pair of specific primers *UBTD2f* and *UBTD2r* for this study (Fig. 1). All primers were synthesized by Shanghai Sangon Biological Engineering Technology & Services Co., Ltd. (Shanghai, China).

| | | |
|---------------|--|---------|
| | 8104723 | 8104752 |
| UBTD2 | CACAAGTGCACATGTGAACAGCACTTCCAG | |
| C348 | GC . . GAGT . CATCC . . G | |
| UBTD2f | | |
| | 8106020 | 8105992 |
| UBTD2 | TGGATGTGAGAGGTGACATGGGTGACACC | |
| C350 | A . . GA . . T TG | |
| UBTD2r | | |

Fig. 1: The primer sequences and comparison with homologous regions in *Gellus gallus UBTD2* gene (all sequences are in 5'- to 3'- directions). Dots indicate identical nucleotides to *UBTD2* 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)

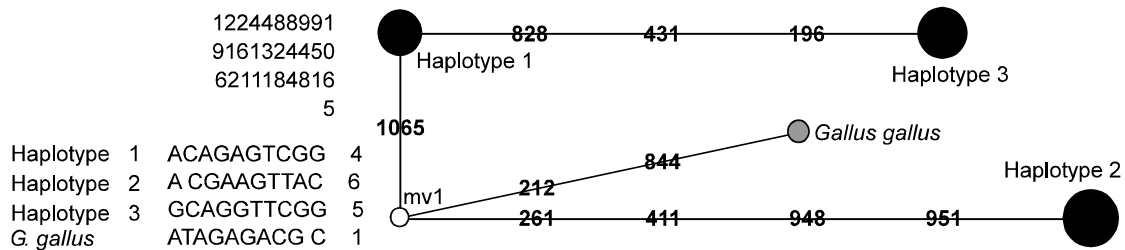


Fig. 2: The 10 polymorphic sites (left) and median joining network relationship (right) between the four haplotypes identified from 11 White Leghorn chickens and the single red jungle fowl. Circles filled with black colour are three haplotypes identified in the White Leghorn chickens while the single circle with gray 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 μ L 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 UBTD2f and UBTD2r.

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 *UBTD2* 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 in 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(π)] 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 *UBTD2* gene, we searched the current versions of both GenBank (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=416202) and Ensembl (http://asia.ensembl.org/Gallus_gallus/Transcript/Prot_Variations?db=core;g=ENSGALG00000002810;r=13:8075587-8122847;t=ENSGALT00000004443) databases and found that there were six synonymous single nucleotide polymorphisms (SNPs) located within the coding sequences of chicken *UBTD2* 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.*, 2006). None of them has been subjected to validation, therefore there may be false positives resulted from simultaneous mutations during cloning sequencing of various cDNA libraries. Nonetheless, similar to mammals, the widely expressed chicken *UBTD2* 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 *UBTD2* 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

- Bandelt, H.-J., P. Forster and A. Röhl, 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16: 37-48.
- Boardman, P.E., J. Sanz-Ezquerro, I.M. Overton, D.W. Burt, E. Bosch, W.T. Fong, C. Tickle, W.R. Brown, S.A. Wilson and S.J. Hubbard, 2002. A comprehensive collection of chicken cDNAs. *Current Bio.*, 12: 1965.
- Carninci, P., T. Kasukawa, S. Katayama, J. Gough, M. Frith, N. Maeda, R. Oyama, T. Ravasi, B. Lenhard and C. Wells, 2005. The transcriptional landscape of the mammalian genome. *Science.*, 309: 1559-1563.
- Carre, W., X. Wang, T.E. Porter, Y. Nys, J. Tang, E. Bernberg, R. Morgan, J. Burnside, S.E. Aggrey, J. Simon and L.A. Cogburn, 2006. Chicken genomics resource: sequencing and annotation of 35,407 ESTs from single and multiple tissue cDNA libraries and CAP3 assembly of a chicken gene index. *Physiological Genomics*, 25: 514-524.
- Church, D.M., L. Goodstadt, L.W. Hillier, M.C. Zody, S. Goldstein, X. She, C.J. Bult, R. Agarwala, J.L. Cherry and M. DiCuccio, 2009. Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biology*, 7: e1000112.
- Gerhard, D.S., L. Wagner, E. Feingold, C. Shenmen, L. Grouse, G. Schuler, S. Klein, S. Old, R. Rasooly and P. Good, 2004. The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res.*, 14: 2121.
- Gibbs, R.A., J. Rogers, M.G. Katze, R. Bumgarner, G.M. Weinstock, E.R. Mardis, K.A. Remington, R.L. Strausberg, J.C. Venter and R.K. Wilson, 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science*, 316: 222-234.
- Guyon, J.R., A.N. Mosley, S.J. Jun, F. Montanaro, L.S. Steffen, Y. Zhou, V. Nigro, L.I. Zon and L.M. Kunkel, 2005. α -sarcoglycan is required for early zebrafish muscle organization. *Experimental Cell Res.*, 304: 105-115.
- Harhay, G., T. Sonstegard, J. Keele, M. Heaton, M. Clawson, W. Snelling, R. Wiedmann, C. Van Tassell and T. Smith, 2005. Characterization of 954 bovine full-CDS cDNA sequences. *BMC Genomics*, 6: 166.
- Hellsten, U., R.M. Harland, M.J. Gilchrist, D. Hendrix, J. Jurka, V. Kapitonov, I. Ovcharenko, N.H. Putnam, S. Shu, L. Taher, I.L. Blitz, B. Blumberg, D.S. Dichmann, I. Dubchak, E. Amaya, J.C. Detter, R. Fletcher, D.S. Gerhard, D. Goodstein, T. Graves, I.V. Grigoriev, J. Grimwood, T. Kawashima, E. Lindquist, S.M. Lucas, P.E. Mead, T. Mitros, H. Ogino, Y. Ohta, A.V. Poliakov, N. Pollet, J. Robert, A. Salamov, A.K. Sater, J. Schmutz, A. Terry, P.D. Vize, W.C. Warren, D. Wells, A. Wills, R.K. Wilson, L.B. Zimmerman, A.M. Zorn, R. Grainger, T. Grammer, M.K. Khokha, P.M. Richardson and D.S. Rokhsar, 2010. The genome of the Western clawed frog *Xenopus tropicalis*. *Science*, 328: 633-636.
- Hillier, L.W., W. Miller, E. Birney, W. Warren, R.C. Hardison, C.P. Ponting, P. Bork, D.W. Burt, M.A. Groenen and M.E. Delany, 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, 432: 695-716.
- Hubbard, S.J., D.V. Grafham, K.J. Beattie, I.M. Overton, S.R. McLaren, M.D. Croning, P.E. Boardman, J.K. Bonfield, J. Burnside and R.M. Davies, 2005. Transcriptome analysis for the chicken based on 19,626 finished cDNA sequences and 485,337 expressed sequence tags. *Genome Res.*, 15: 174-183.
- Kimura, K., A. Wakamatsu, Y. Suzuki, T. Ota, T. Nishikawa, R. Yamashita, J.-I. Yamamoto, M. Sekine, K. Tsuritani and H. Wakaguri, 2006. Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res.*, 16: 55-65.
- Klein, S.L., R.L. Strausberg, L. Wagner, J. Pontius, S.W. Clifton and P. Richardson, 2002. Genetic and genomic tools for *Xenopus* research: The NIH *Xenopus* initiative. *Developmental Dynamics*, 225: 384-391.
- León, A.J., D. Banner, L. Xu, L. Ran, Z. Peng, K. Yi, C. Chen, F. Xu, J. Huang and Z. Zhao, 2013. Sequencing, annotation, and characterization of the influenza Ferret infectome. *J. Virol.*, 87: 1957-1966.
- 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.

- Librado, P. and J. Rozas, 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25: 1451-1452.
- Liu, S., H. An, N. Li, Y. Yu, N. Lin, T. Wan, M. Zhang, W. Wang and X. Cao, 2003. Cloning and identification of a novel human ubiquitin-like protein, DC-UbP, from dendritic cells. *Biochemical and Biophysical Research Communications*, 300: 800-805.
- Nei, M., 1987. *Molecular Evolutionary Genetics*. Columbia University Press.
- Ommeh, S., L. Jin, H. Eding, F. Muchadeyi, S. Sulandari, M. Zein, G. Danbaro, C. Wani, S. Zhao and Q. Nie, 2010. Geographic and breed distribution patterns of an A/G polymorphism present in the Mx gene suggests balanced selection in village chickens. *Int. J. Poult. Sci.*, 9: 32-38.
- Ota, T., Y. Suzuki, T. Nishikawa, T. Otsuki, T. Sugiyama, R. Irie, A. Wakamatsu, K. Hayashi, H. Sato and K. Nagai, 2003. Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nature Genetics*, 36: 40-45.
- Rogers, A.R. and H. Harpending, 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, 9: 552-569.
- Schmutz, J., J. Martin, A. Terry, O. Couronne, J. Grimwood, S. Lowry, L.A. Gordon, D. Scott, G. Xie and W. Huang, 2004. The DNA sequence and comparative analysis of human chromosome 5. *Nature*, 431: 268-274.
- Shaw, I., T.J. Powell, D.A. Marston, K. Baker, A. van Hateren, P. Riegert, M.V. Wiles, S. Milne, S. Beck and J. Kaufman, 2007. Different evolutionary histories of the two classical class I genes *BF1* and *BF2* illustrate drift and selection within the stable MHC haplotypes of chickens. *The J. Immunol.*, 178: 5744-5752.
- Slatkin, M. and R.R. Hudson, 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, 129: 555-562.
- Song, A.X., C.J. Zhou, X. Guan, K.H. Sze and H.Y. Hu, 2010. Solution structure of the N-terminal domain of D-UbP/*UBTD2* and its interaction with ubiquitin. *Protein Sci.*, 19: 1104-1109.
- Strausberg, R.L., E.A. Feingold, L.H. Grouse, J.G. Derge, R.D. Klausner, F.S. Collins, L. Wagner, C.M. Shenmen, G.D. Schuler and S.F. Altschul, 2002. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proceedings of the National Academy of Sciences of the United States of America*, 99: 16899-16903.
- Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar, 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, 28: 2731-2739.
- Venter, J.C., M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans and R.A. Holt, 2001. The sequence of the human genome. *Science*, 291: 1304-1351.
- Wong, G.K.-S., B. Liu, J. Wang, Y. Zhang, X. Yang, Z. Zhang, Q. Meng, J. Zhou, D. Li and J. Zhang, 2004. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature*, 432: 717-722.