



Single nucleotide polymorphism screening, molecular characterization, and evolutionary aspects of chicken *Piwi* genes

H.Z. Wang¹, T. Ma¹, G.B. Chang¹, F. Wan¹, X.P. Liu², L. Lu¹, L. Xu¹, J. Chen¹ and G.H. Chen¹

¹College of Animal Science and Technology, Yangzhou University, Yangzhou, Jiangsu, China

²Poultry Institute, Chinese Academy of Agricultural Sciences, Yangzhou, Jiangsu, China

Corresponding authors: G.B. Chang / G.H. Chen
E-mail: passioncgb@163.com and / ghchen@yzu.edu.cn

Genet. Mol. Res. 14 (4): 14802-14810 (2015)

Received April 4, 2015

Accepted August 1, 2015

Published November 18, 2015

DOI <http://dx.doi.org/10.4238/2015.November.18.45>

ABSTRACT. The P-element-induced wimpy testis (*Piwi*) gene is involved in germline stem cell self-renewal, meiosis, RNA silencing, and transcriptional regulation. *Piwi* genes are relatively well conserved in many species, but their function in poultry species is unclear. In this study, *Piwi* genes were sequenced using a target-sequence capture assay in quail and 28 breeds of chicken. Single nucleotide polymorphisms (SNPs) and evolutionary aspects of these chicken breeds were then analyzed. We found that SNP sites existed mainly in the introns of a few chicken breeds, and we selected an SNP on intron 4 for further verification by Sanger sequencing, the results of which were similar to those obtained by the target-capture sequencing assay. The evolutionary analysis revealed that there were more mutations in the Chahua and Leghorn breeds than in the other breeds, and that the phylogenetic tree was divided into four main categories that suggested that *Piwi* is evolutionarily conserved, and mutations in the introns might be associated with gametogenesis. The screened SNPs can be used as

candidate markers for *Piwi*, and our results provide basic information for the further study of *Piwi* function in poultry.

Key words: Chicken; Target-capture sequencing; *Piwi* genes; SNP

INTRODUCTION

The P-element-induced wimpy testis (*Piwi*) genes was first discovered in *Drosophila* germline stem cells (GSCs), and plays an important role in GSC self-renewal (Lin and Spradling, 1997). In the *Drosophila* ovary, *Piwi* is simultaneously expressed in stem and somatic cells, and has the dual function of endogenous and exogenous regulation. The results of subcellular localization analysis show that *Piwi* encodes nuclear-cytoplasmic proteins (Cox et al., 2000), suggesting that it may be involved in the modification process after nuclear mRNA transcription. Homology analysis has shown that PIWI protein congeners exist in many animals and plants. In mice, *Piwi* homologs, such as *Miwi*, encode cytoplasmic proteins that are specifically expressed in spermatocytes and spermatids (Deng and Lin, 2002). Lack of *Miwi* causes spermatogenic arrest at the round spermatid stage, similar to the deletion of *Crem*, which is thought to be the main regulator of spermatogenesis. In addition to what is required during spermatogenesis, *Piwi* genes are also associated with the formation and development of tumors in human and mice. Few studies have been conducted on the function of *Piwi* in poultry. In this study, we sequenced the entire *Piwi* of 28 chicken (*Gallus gallus*) breeds and quail (*Coturnix coturnix*, Aves: Galliformes: Phasianidae). The chicken breeds used included wild red jungle fowl, broilers, Leghorn (specific-pathogen free, SPF), and several Chinese breeds. SNPs in the chicken breeds were detected, and polymerase chain reaction-single-strand conformation polymorphism (PCR-SSCP) and Sanger sequencing were then used for verifying variant calls, in order to investigate *Piwi* evolution, distribution, and functional sites in poultry.

MATERIAL AND METHODS

The chicken breeds included Xianju, Chahua, Luyuan, Baier, Tibetan, Gushi, Dagu, Gamecock, Langshan, Silky fowl, Xiaoshan, and Beijing fatty, which were obtained from the Chinese Research Institute of Agricultural Science; red jungle fowl was obtained from a wildlife rescue center in Yunnan Province, China; and Ross, AA, recessive white feather, Anka, Leghorn, Rucuo, Xueshan grass, Wenchang, Taihu, Liyang, green egg, normal Guangxi yellow, Guangxi yellow (necrospermia/aspermatism), partridge shank, and Gamecock x Ross hybrids were obtained from the Genetic Resources Laboratory of the College of Animal Science and Technology, Yangzhou University, China. We took brachial-vein blood samples from 10 birds per breed, and DNA was extracted by the phenol/chloroform method. The pooled DNA samples of 10 birds per breed were sent to BGI, Shenzhen, China for target-capture sequencing, with *G. gallus*-4.0 as a reference genome. We selected normal Guangxi yellow chickens, Guangxi yellow chickens (necrospermia/aspermatism), and Chahua chickens for the next verification experiment; there were 30 samples for each candidate breed.

Genomic DNA capture and sequencing

Genomic DNA was captured by hybridization in solution with custom-designed cRNA oligonucleotide baits, following the manufacturer protocol (NimbleGen Sequence Capture, Roche).

The raw data from the sequencing was processed in the following manner: firstly, removed the original data joint, pollution sequence and low quality reads, and then low-quality bases were removed [quality value ≤ 5 (E)] to ensure high-quality data. The *Piwi* SNP sites were detected by target-capture sequencing.

Discovery and validation of *Piwi* SNPs

The oligonucleotide pair sequences for each SNP were designed so that the capture probe contained the mutation site. *Piwi* (GenBank accession No. NM_001098852.1) was obtained from the National Center for Biotechnology Information (NCBI). For each potential SNP position, the reads that covered the *Piwi* region were partitioned into bins based on the position of the variant site in the read. The base quality score and the strand of the reference sequence that the reads aligned to during sequence alignment were obtained using the ClustalW function in MEGA software, in order to detect *Piwi* SNPs.

We filtered out some SNP sites in specific chicken breeds for verification by Sanger sequencing, based on the results of the SNP detection and by searching the dbSNP database (http://asia.ensembl.org/Gallus_gallus/Gene/Variation_Gene/Table). PCRs were performed in a 20- μ L volume containing 10X PCR buffer (TaKaRa, Dalian, China), 0.2 mM dNTP (TaKaRa, Dalian, China), 1 μ M of each primer (Sangon Biotech, Shanghai, China), 0.2 U *rTaq* DNA polymerase (TaKaRa, Dalian, China), and approximately 100 ng genomic DNA. Initial denaturation for 5 min at 98°C was followed by 35 cycles at 98°C for 30 s, 61.7°C for 30 s, and 72°C for 30 s, followed by a final 10-min extension at 72°C. The primers were designed for amplifying sequences containing SNPs using the Oligo7 software (<http://www.oligo.net/>). Subsequently, we obtained the specific bands of the different chicken breeds on polyacrylamide gel by SSCP experiments, and selected mutated sequences for re-sequencing to verify previous variant calls and detect SNP differences between the different chicken varieties.

Phylogenetic analysis

The complete sequences of 19 poultry breeds that contained SNP sites (including quail and reference breeds) were assembled in FASTA format and aligned using ClustalW. Based on these data, we constructed a phylogenetic tree using the maximum parsimony method in MEGA, which described the relationships between *Piwi* sequences that contained SNPs. The branch lengths of the tree represent the distance and the branch node numbers represent bootstrap values. The root of the tree assumed a degree of constancy in the molecular clock, and estimated the common ancestor sequence from the derived progeny sequences.

RESULTS

Distribution of SNPs on *Piwi*

Piwi is also known as *Piwi1* (Gene ID, 416804) in NCBI, its full name is *Piwi*-like 1, and it is located on chicken chromosome 15. The full-length gene sequence is 16,682 bp long, covers a 3,327,008-3,343,689-bp region, and contains 20 exons and 19 introns.

Compared to the reference *Piwi* sequences (ID, 416804), we found 36 mutations in 17 chicken breeds, excluding quail; these breeds were Ross, AA, recessive white feather broiler,

Leghorn (SPF), Langshan, Gamecock, Rucao, Xueshan, Chahua, Luyuan, Taihu, Liyang, green egg, Xianju, Guangxi yellow (necrospermia/aspermatis), Guangxi yellow (normal), and Gamecock x Ross hybrids. The SNP distribution is shown in Figure 1. All of the *Piwi* mutations were in the introns, which demonstrates that *Piwis* are relatively well conserved in different chicken breeds. As can be seen in Figure 1, the intron polymorphism of *Piwi* is rich. Introns 3, 4, 8, and 9 had relatively more mutations than the other introns, and were selected for further experimental validation and functional analyses.

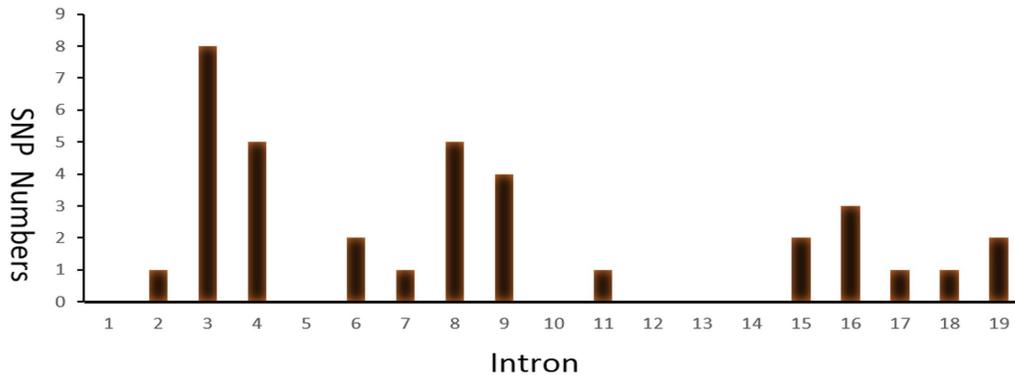


Figure 1. Distribution of SNP sites on the *Piwi* gene.

Comparative analysis of *Piwi* sequences in different chicken breeds

Complete *Piwi* sequences with SNP sites are shown in Figure 2.

[1	1111111111	1111111]
[12333345	688888990	0112222233	3334445]		
[3587446921	4013699073	7111238977	8884660]		
[6647462052	7473767861	3469904369	2330576]		
[2270953917	0657583156	6608457490	7463577]		
Gallus (reference)	AAACACCACT	GCTACAGTCA	TAAGTTGTAC	CAAGTTA		
Ross	A.....		
Langshan	A.....		
Gamecock	A.....		
Rucao	C.....		
Xueshan	...A.T.	...G...	A.A.C...		
Recessive white	A.....		
Chahua	.CGTGAA.	A.CAGGT.	CTT.CGTAGC.	...A..G		
LuyuanT.		
Gamecock × Ross	A.....		
TaihuG...	A.A.....		
Liyang	A.....		
Green-shelled-egg	A.....		
AA	A.....		
Guangxi (normal)	A.A.....		
Leghorn (SPF)	GCGT..A.AC	.A.GT....	CGTAGCTC.G		
Xianju	...A....	...GA...	AC.....		
Guangxi (necrospermia)	...TG.A.A.	CAGGT..CTT	CGTAGC.	...ACAG		

Figure 2. Sequence alignment in different chicken breeds. The left side of the table is the name of different chicken breeds with SNP sites. Points in the table represent the base that is consistent with the reference sequence, and different base indicates different SNP sites.

The SNP sites of the Chahua, Leghorn (SPF), and Guangxi yellow chickens (necrospemia/aspermatism) were significantly different from those of the other breeds. SNP sites in these three breeds were relatively richer and had similar distributions, while fewer SNP sites were found in the other chicken breeds. All coexisting SNP sites in these three breeds are presented in Table 1. As shown in Table 1, introns 4 and 9 had a large number of polymorphisms. We selected intron 4 for PCR amplification, SSCP, and Sanger sequencing, in order to verify previous variants found by the target-capture sequencing assay.

Table 1. Coexisting SNP Sites of Chahua Chicken, Leghorn (SPF) and Guangxi yellow chicken (necrospemia).

Polymorphic loci	Locus coordinates	Location on chromosome	Reference/mutation	dbSNPposition matches
SNP1	3342074	Intron 2	A/G	√
SNP2	3339312	Intron 4	T/C	√
SNP3	3339301	Intron 4	T/G	√
SNP4	3339205	Intron 4	G/A	√
SNP5	3338167	Intron 6	A/T	√
SNP6	3338153	Intron 6	A/G	√
SNP7	3337743	Intron 7	T/C	√
SNP8	3335682	Intron 9	C/T	√
SNP9	3335344	Intron 9	A/G	√
SNP10	3335053	Intron 9	C/A	√
SNP11	3331258	Intron 15	C/A	√
SNP12	3330630	Intron 16	C/A	√
SNP13	3329777	Intron 17	C/T	√

Verification and analysis of SNPs on intron 4

Intron 4 is 928 bp long and its SNPs are distributed within 300 bp. A PCR-SSCP assay and the re-sequencing of the intron 4 region were conducted in order to detect SNPs in the Chahua, Guangxi yellow (necrospemia/aspermatism), and Guangxi yellow (normal) chickens (Table 2). And the primer sequences are provided in Table 3.

All of the SNP sites in the Sanger-sequenced regions were in agreement with the results of the target-capture sequencing assay. Guangxi yellow chickens (normal) had no mutations in intron 4 with respect to the reference breed; Chahua and Guangxi yellow (necrospemia/aspermatism) chickens had a few mutations.

Table 2. Sanger sequence validation for SNPs of intron 4.

Chicken breeds	3339205	3339301	3339312
<i>Gallus</i> (reference)	G	T	T
GH (normal)	G	T	T
GH (necrospemia)	A	G	C
CH	A	G	C

Table 3. Primers used for PCR amplification and Sanger sequencing.

Gene	Primer sequence (5' to 3')	Annealing temperature
<i>Piwi</i>	F: ATTCGGGCTCTAGAATACTT R: TGTGCTTTAATGTTTGTCTCTGC	61.7°C

Phylogenetic analysis

The phylogenetic tree is presented in Figure 3. The different chicken breed sequences were sequenced, assembled, and divided into four main clades. The sub-branches within each of these four groups represent the SNP clusters. Chahua, Leghorn (SPF), Guangxi yellow (necrospermia/aspermatism), and quail were classified together because they had more SNP sites than the reference chicken breeds. Other classifications were related to the birthplace of the chicken and spermatogenesis-associated traits. For example, Ross, recessive white feather broiler, etc., are non-Chinese species, and Gamecock, green egg, etc., are Chinese breeds. Guangxi yellow (necrospermia/aspermatism) and Guangxi yellow (normal) chickens could be classified in terms of sperm formation. However, these were rough classifications, and a more accurate phylogenetic tree requires further experimental validation.

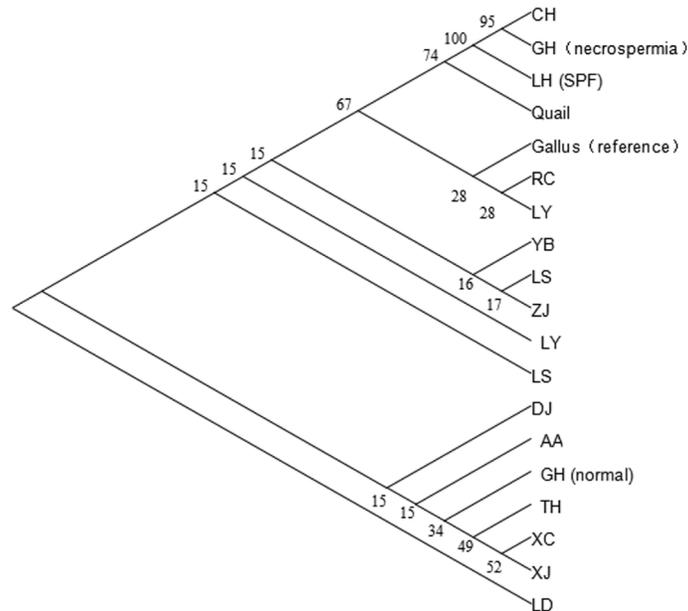


Figure 3. Phylogenetic tree of the *Piwi* gene in different chicken and quail. Mega-Clustal W analysis of sequences from 17 different chicken breeds. Neighbor joining trees were calculated by using phylogenetic reconstruction analysis using maximum parsimony distance computation. Bootstrap values (1000 replication) are given on the nodes.

DISCUSSION

Target-sequence capture technology is widely used in the study of genetic mutations that cause disease and are involved in evolution, and has many advantages: firstly, it can be customized to the genomic sequence in the region of interest, and genomic DNA fragments in the sequence capture chip (or solution) are then enriched. Subsequently, these DNA fragments can be sequenced by second-generation sequencing technologies. In addition, it can be used to study genome copy-number variation. Lastly, more samples can be studied. Useful mutations have been found in mice by target-capture sequencing technology (Mark and Jacob, 2009). Vallender (2011) found mutations in the coding regions of chimpanzees (*Pan troglodytes*) and rhesus

macaques (*Macaca mulatta*) using whole-exon capture sequencing, and compared them with the whole human genome. Cosart et al. (2011) used exon-capture sequencing and high-throughput sequencing to filter out SNP sites on target candidate genes in domesticated and wild species. In this study, *Piwi* mutations were all in introns, indicating that it is relatively well conserved. Chen et al. (2011) predicted the structure and function of the PIWI protein in quail, and suggested that the protein is also highly conserved.

Our results suggest the following: 1) A few *Piwi* mutations occur in most of the chicken varieties, and the mutational location is relatively concentrated, particularly in the 3,340,834-bp region, which indirectly proved the conserved nature of *Piwi*. 2) Most mutations have occurred in Chahua, Leghorn (SPF), and Guangxi yellow (necrospermia/aspermatism) chickens. The distribution of SNP sites in these three breeds was similar. Normal Guangxi yellow and Guangxi yellow (necrospermia/aspermatism) chickens exhibit significant differences in sperm development, and these phenotypic differences may be related to genetic mutations. This was confirmed by a verification experiment. 3) All of the *Piwi* mutations were in the introns. Introns were once considered sequences of indefinite nature, or genome junk, but more and more researchers have realized that introns have important biological functions (Gazave et al., 2007). Intron RNA contains a variety of non-coding RNA (ncRNA), including micro-RNA, small nucleolus RNA, and guidance RNA for RNA editing (Mattick et al., 2001). Intron RNA is an important component of ncRNA through the process of intron gain and loss, affects ncRNA variation, and is the driving force of eukaryotic speciation (Yoshihama et al., 2007). Introns are involved in gene expression and regulation, and play an important role in the processing and transportation of mRNA (Nott et al., 2003). Intron mutations can cause changes in the expression of insulin-like growth factor 2, estrogen receptors, FSH β , Nodal, CPT-I, α , β -actin, β -keratin, and immunoglobulin (Comeron and Kreitman, 2000). All of the mutations detected in this study have occurred in introns, and it is possible that *Piwi* mutations affect spermatogenesis or other traits in different varieties of chicken; these inferences need further research.

There are many ways to construct a phylogenetic tree, and *Piwi* mutations are caused by single nucleotide variations, so using SNPs to analyze phylogenetics is reliable. Pavy et al. (2008) conducted an SNP microarray analysis of black spruce (*Picea mariana*) and white spruce (*Picea glauca*), and compared their results with those obtained using simple sequence repeats, amplified fragment length polymorphism, and other molecular markers. Their results suggest that black and white spruce began to differentiate 10 million years ago (Pavy et al., 2008). This et al. (2007) studied variation in the *VvmybA1* genes, which are transcription factors for grape anthocyanin synthesis, and showed that grape color evolution was related to artificial selection during domestication. Ince et al. (2005) studied the *matK* genes (that are related to chloroplasts) in 142 species of plant, including the genus *Malus*, and found that different families, genera, and species could be distinguished by SNP analysis. Moreover, the reliability of this method was between 80 and 100% (Ince et al., 2005).

The phylogenetic tree that was constructed in the present study exhibited several noteworthy characteristics: 1) Chinese and non-Chinese species were separated by cluster analysis, indicating that there may be differences in sperm development and other breeding traits between different breeds; 2) Guangxi yellow chickens (normal) and Guangxi yellow chickens (necrospermia) were separated by the clustering, which confirmed the function of *Piwi* in sperm formation. Mutations in *Piwi* introns may affect gene expression, which affects different phenotypic traits; 3) molecular techniques are widely used to study the genetic structures and relationships of different chicken breeds. Zhou et al. (1997) analyzed genetic structure using random amplified polymorphic DNA (RAPD) in four Chinese chicken breeds and one introduced breed. Zhang et al. (1996) analyzed

the genetic structure of sex-linked color in Chinese chicken breeds using RAPD, and studied the genetic structure and relationships of chicken breeds using five microsatellite markers. Zhou and Lamont (1999) analyzed 42 microsatellite loci in 23 highly inbred chickens, and reported that microsatellite analysis is an effective method for studying genetic diversity. Yang et al. (2000) analyzed genetic variation between chicken groups using microsatellite markers. In the present study, a phylogenetic tree was constructed using *Piwi* SNP sites that were measured by target-capture sequencing in different chicken breeds, in order to investigate the genetic relationships and evolution between them, which is another method for studying genetic differences between chicken breeds at the molecular level.

This study focused on *Piwi* sequences in different varieties of chicken and quail, which were sequenced by target-capture sequencing. Based on the SNP sites detected, *Piwi* is relatively well conserved. *Piwi* is influential in sperm formation, and several SNPs on intron 4 could be screened and used as candidate marker sites for related traits (e.g., spermatogenesis). Our phylogenetic tree was constructed based on sequences that contained SNP sites in different chicken breeds. Using this molecular-level approach, genetic differences between chicken breeds can be studied, and relationships between Chinese and non-Chinese breeds can be explored.

ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (#31172199).

REFERENCES

- Chen R, Chang GB, Zhang Y, Chen GH, et al. (2011). Cloning and expression analysis of piRNAs and *Piwi1* gene in quail. *Sci. Agric. Sin.* 44: 1727-1735.
- Cameron JM and Kreitman M (2000). The correlation between intron length and recombination in *Drosophila*: dynamic equilibrium between mutational and selective forces. *Genetics* 156: 1175-1190.
- Cosart T, Beja-Pereira A, Chen S, Ng SB, et al. (2011). Exome-wide DNA capture and next generation sequencing in domestic and wild species. *BMC Genomics* 12: 347.
- Cox DN, Chao A and Lin H (2000). *Piwi* encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* 127: 503-514.
- Deng W and Lin H (2002). *miwi*, a murine homolog of *piwi*, encodes a cytoplasmic protein essential for spermatogenesis. *Dev. Cell* 2: 819-830.
- Gazave E, Marqués-Bonet T, Fernando O, Charlesworth B, et al. (2007). Patterns and rates of intron divergence between humans and chimpanzees. *Genome Biol.* 8: R21.
- Ince AQ, Karaca M, Onus AN and Bilgen M (2005). Chloroplast *matK* gene phylogeny of some important species of plants. *Akdeniz Universites Iziraatfakultesi Dergisi* 18: 157-162.
- Lin H and Spradling AC (1997). A novel group of pumilio mutations affects the asymmetric division of germ line stem cells in the *Drosophila* ovary. *Development* 124: 2463-2476.
- Mark DA and Jacob K (2009). Mutation discovery in the mouse using genetically guided array capture and re-sequencing. *Mamm. Genome* 20: 424-436.
- Mattick JS (2001). Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.* 2: 986-991.
- Nott A, Meislin SH and Moore MJ (2003). A quantitative analysis of intron effects on mammalian gene expression. *RNA*: 607-617.
- Pavy N, Pelgas B, Beauseigle S, Blais S, et al. (2008). Enhancing genetic mapping of complex genomes through the design of highly multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. *BMC Genomics* 9: 1471-2164.
- This P, Lacombe T and Cadle-Davidson M (2007). Wine grape (*Vitis vinifera* L.) color associates with allelic variation in the domestication gene *VvmybA1*. *Theor. Appl. Genet.* 14: 723-730.
- Vallender EJ (2011). Expanding whole exome resequencing into non-human primates. *Genome Biol.* 12: R87.
- Yang Y, Zhu Q and Hu GA (2000). The use of microsatellite markers for detection of population genetic variation in chicken. *J. Sichuan Univ. (Nat. Sci. Edn.)*. 37: 148-152.

- Yoshihama M, Nguyen HD and Kenmochi N (2007). Intron dynamics in ribosomal protein genes. *PLoS One* 2: e141.
- Zhang XQ, Yang GF and Shi ZD (1996). The genetic structure and relationship analysis of *Gallus domesticus* by means of microsatellite polymorphisms. *Anim. Biotechnol. Bull.* 5: 54-59.
- Zhou HJ and Lamont SJ (1999). Genetic characterization of biodiversity in highly inbred chicken lines by microsatellite markers. *Anim. Genet.* 30: 256-264.
- Zhou HJ, Chen KW and Zhang XY (1997). Analysis on genetic structure of population in local chicken breeds by RAPD technique. *China Poultry*. 8: 7-10.