DNA Sequence Variations of Alleles at Two Chicken Microsatellite Loci of MCW0216 and LEI0234

1,2 Mei-Qing Wei, 2,3 Sheng-Cheng Zeng, 1 Yan-Ming Wei and 3,4 Jian-Lin Han
1 College of Life Science, Gansu Agricultural University, Lanzhou 730070, China
2 Key Laboratory of Farm Animal Genetic Resources and Utilization of Ministry of Agriculture, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (CAAS), Beijing 100193, China
3 CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing 100193, China
4 International Livestock Research Institute (ILRI), 00100 Nairobi, Kenya

Corresponding Author: Jian-Lin Han, CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (CAAS), Beijing 100193, China

ABSTRACT
To verify the quality of genotyping data and understand the property of microsatellite markers to elucidate chicken genetic diversity, the DNA sequence variations of major alleles at two chicken microsatellites of MCW0216 and LEI0234 were assessed through direct sequencing of the PCR products of 26 selected chicken blood samples. The results of MCW0216 locus were interesting and useful because they showed very complicated DNA sequence structures within both flanking and repeat regions which validated alleles that were different from each other by single nucleotides. The three SNPs present within both primer regions for genotyping the MCW0216 locus called for an action to re-design new genotyping primers to avoid allelic drop-out during genotyping exercises. The LEI0234 locus was a complex of either perfect or compound tetra-nucleotide microsatellite carrying small-sized alleles with 7 to 18 perfect CTTT repeat units while all big alleles longer than 275 bp consisted of more than 20 basic CTTT repeats being interrupted by a few CTTT motifs as a stabilizing effector contributing to the expansion of microsatellite length and allelic polymorphism. The presence of many cryptic alleles that were identical in fragment sizes but different in either flanking sequences or repeat unit compositions at both loci further warranted an intensive DNA sequence characterization of major alleles for a correct interpretation of microsatellite data.

Key words: Chicken, microsatellite, allele, sequence, genetic variability

INTRODUCTION
Microsatellite DNA markers are characterized by Short Tandem Repeats (STRs) or Simple Sequence Repeats (SSRs). They are among the most variable types of DNA sequences in the genome and therefore have been developed into most popular genetic markers since their discovery in the early 1980s (Ellegren, 2004; Gholizadeh and Mianji, 2007). They consisted of STRs of core sequences of mono- to hexa-nucleotides in length. They are overrepresented in the genome and occupy between 0.5 and 3% of the total genome size (Mayer et al., 2010). Their polymorphisms are derived mainly from the variability in number of STRs rather than in primary sequence. Based on the type of variation in DNA sequence within STRs region, Weber (1990) divided microsatellites into three categories: prefect, imperfect and compound STRs. Perfect STRs are one type of repeat
units without interruption and separated from any other type of repeat units. Imperfect STRs are one type of two or more repeat units separated by no more than three consecutive non-repeat nucleotides; each of their terminal uninterrupted repeat units (outside of non-repeat nucleotides) must be at least three full repeat units in length. Compound STRs are one type of repeat units separated by no more than three consecutive non-repeat nucleotides from another type of repeat units. Microsatellites are widely used as tools for molecular forensics, parentage testing, analysis of genetic diversity and structure within populations and assessment of phylogenetic relationships among populations.

Based on the experience and knowledge generated from studies involved microsatellites in genetic characterization of chicken genetic resources (Van Marle-Koster and Nel, 2000; Wimmers et al., 2000; Rosenberg et al., 2001; Hillel et al., 2003; Ya-Bo et al., 2006), 30 autosomal microsatellites were recommended by the joint ISAG (International Society for Animal Genetics) and FAO (Food and Agriculture Organization of the United Nations) Standing Committee as an international panel of molecular markers to be used for measuring genetic diversity within and genetic relationships among domestic chicken breeds (Hoffmann et al., 2004). However, all these microsatellites were only isolated and defined by single reference DNA sequences, of which 29 were reported to carry STRs of di-nucleotide repeat units, the remaining one locus carried tetra-nucleotide STRs (Crooijmans et al., 1993, 1994, 1996, 1997; Cheng et al., 1995; Ruyter-Spira et al., 1996, 1998; Gibbs et al., 1997; McConnell et al., 1999). The chicken genome sequence published later added the second sequences to all of these microsatellites (Hillier et al., 2004).

We recently selected and applied 20 out of the 30 microsatellites for genetic characterization of Papua New Guinean indigenous chickens and found that the allelic number per marker ranged from 4 to 14 and in some loci alleles were either different by a single nucleotide or by a big gap of more than 10 base pairs (bp), or showing irregular peak patterns (Sheng-Cheng et al., 2010). To understand some of these unexpected phenomena, we conducted this study from September 2009 to January 2011 to verify the quality of genotyping data and subsequently to shed more light on the complete understanding of the microsatellite data to elucidate chicken genetic diversity.

MATERIALS AND METHODS

Samples: A total of 23 chicken samples (Table 2, 3) were selected for this study according to the genotyping results of Sheng-Cheng et al. (2010). All samples were in blood preserved on Whatman FTA® filter papers (Whatman BioScience, Maidstone, UK) and total genomic DNAs were extracted following the procedure of Smith and Burgoyne (2004).

Primers: Both forward (F) and reverse (R) PCR primers of either MCW0216 or LEI0234 loci that were mapped on chicken chromosome (GGA) 13 or GGA 2 were designed according to extended homologous sequences of UCD001 inbred red jungle fowl (G. gallus; Hillier et al., 2004) retrieved from the Ensembl database (Flicek et al., 2010) while the sequencing primer (F1) for MCW0216 was obtained from the recommendations (Hoffmann et al., 2004) (Table 1). The expected sizes of PCR products were around 1023 bp for MCW0216 and 954 bp for LEI0234. It was noticed that the directions of both PCR primers included in the recommendations that were taken directly from their original designs (Ruyter-Spira et al., 1998) for genotyping MCW0216 locus were upside down when they were aligned against the chicken genome reference sequence, therefore we correctly re-assigned the original ‘reverse’ PCR primer as our forward sequencing primer (F1).
PCR conditions and procedures: This study performed two separate PCR protocols. All PCRs were performed with 3 μL (70-100 ng μL⁻¹) template genomic DNAs in a 60 μL final reaction volume. The reactions of both MCW0216 and LEI0234 contained 0.5 μL of each primer (10 pmol), 4 μL (2.5 mM) dNTPs (Tiangen Biotech (Beijing) Co., Ltd., Beijing, China), 6 μL 10× buffer and 0.8 μL (2.5 units μL⁻¹) Taq DNA polymerase (Tiangen). The thermocycling procedure of PCR included an initial denaturation at 94°C for 5 min, followed by 30 cycles of denaturation at 94°C for 30 sec, primer annealing at 64°C for MCW0216 as well as 55.8°C for LEI0234 for 30 sec each and 72°C for 30 sec and finished by a final extension at 72°C for 10 min.

Direct sequencing and data analysis: The PCR products were purified using the gel purification kit following instructions of the manufacturer (Beijing Sunbiotech Co., Ltd., Beijing, China) and then directly sequenced using both PCR primers from two directions for both MCW0216 and LEI0234 loci while the sequencing primer (F1) was only used to sequence L66 sample for a verification of the sequencing data due to its complex structure. The BigDye® Terminator v3.1 Cycle Sequencing Kit was applied for sequencing reactions that were visualized on an ABI 3730 Sequence Analyzer (Applied Biosystems, Foster City, CA, USA) by the Beijing Sunbiotech Co., Ltd. (Beijing, China). The raw data was manually edited using Chromas version 1.45 (http://www.technelysium.com.au/chromas.html) and aligned with MEGA4 software (Tamura et al., 2007). In order to explain the structural changes of DNA sequences within the repeat units and flanking regions, we only restricted our sequences that were masked by the forward and reverse primers included in the recommendations (Hoffmann et al., 2004) for data analysis in this study.

RESULTS AND DISCUSSION

Our genotyping data showed that MCW0216 locus had six alleles ranging from 136 to 150 bp and with three of them different from each other by a single nucleotide while LEI0234 was the only marker carrying the tetra-nucleotide STRs and also one of the most polymorphic loci, carrying 16 alleles varying significantly in sizes ranging from 216 to 323 bp with big gaps larger than 10 bp (Sheng-Cheng et al., 2010).

DNA sequence variation of alleles at MCW0216 locus: According to our early genotyping results (Sheng-Cheng et al., 2010), six major alleles and allelic sizes in unexpected distribution of single nucleotide differences observed at MCW0216 locus (Fig. 1) in 136, 141, 143, 144, 145 and 150 bp from 11 samples were chosen for direct sequencing and analyzing their nucleotide variations. The original single reference sequence (AF030586) isolated from a chicken Uni-ZAPXR cDNA library that was screened by hybridization to a radioactively end-labeled (TG)₁₅ probe (Ruyter-Spira et al., 1998) retrieved from the GenBank database and the homologous
sequence of UCD001 inbred red jungle fowl (NW_001471449, G. gallus; Hillier et al., 2004) retrieved from the Ensembl database (Flicek et al., 2010) were included for further alignment and comparison.

The results showed that there were three insertions/deletions (indels) in the upstream flanking sequences involving two single nucleotides of a cytosine insertion in five sequences at position -1 and an adenine deletion in the reference sequence (AF030586) at -59 as well as one octa-nucleotide segment ranging from -39 to -46 in a single allele of 136 bp. Two single nucleotide polymorphisms (SNPs) of G/A (at position 1) and T/C (at position 19) transitions were detected in a single and all the downstream flanking sequences, respectively. The repeat units were composed of the expected, simple AC di-nucleotide that was screened using the (TC)_{13} probe (Ruyter-Spira et al., 1998); however, six sequences (including the reference sequence derived from the red jungle fowl) of an allele of 143 bp seemed to carry interrupted repeat units of (AC)_{5}AT(AC)_{4}AA that were probably fixed into a non-microsatellite fragment. The indel at -1 position was attributed to defining the allele of 144 bp that was different by a single nucleotide from two neighboring alleles of 143 and 145 bp and also the alleles of 136 and 150 bp that were scored in even numbers different from the remaining alleles in odd numbers. There were 'cryptic' alleles of the same fragment sizes of either 143 or 145 bp but different either in basic repeat unit structures or in downstream flanking sequences. The indel at position -1 in upstream flanking region and the SNP at position 19 in the downstream flanking region did not link with any of the particular repeat sequence structures (Table 2).
On the other hand, the extended sequences revealed one SNP of C/T at the last second nucleotide of the forward primer and two SNPs of C/T and A/C in the first and the last fourth nucleotides of the reverse primer included in the recommendations (Hoffmann et al., 2004) for genotyping MCW0216 locus.

**DNA sequence variation of alleles at LEI0234 locus:** Following our genotyping results (Sheng-Cheng et al., 2010), 16 major alleles of 216, 219, 263, 275, 279, 283, 287, 291, 295, 299, 303, 307, 311, 315, 319 and 323 bp carried by 17 heterozygous samples were selected for direct sequencing of their PCR products and examining of their DNA sequence variations. The original single reference sequence (Z94837) isolated from a pooled chicken genomic DNA library enriched for clones containing the tetra-nucleotide motif of (TTTC.AAAG)n (McConnell et al., 1999) retrieved from the GenBank database and the homologous sequence of UCD001 inbred red jungle fowl (NW_001471633, G. gallus; Hillier et al., 2004) retrieved from the Ensembl database (Flicek et al., 2010) were included for further alignment and analysis.

There were three mutations in the downstream flanking region. One deletion of ATT ranging from the 38th to the 40th nucleotides was attributed to defining a single allele of 216 bp that was present only once in H25 sample and different by three nucleotides from the allele of 219 bp carrying the same repeat sequence in S13 sample. One A/G SNP with the guanine was present only once in a single sequence of P05 sample at the 57th nucleotide. One T/C SNP with the thymine was present in three sequences of H25 and S13 samples as well as the reference red jungle fowl at the 112th nucleotide. The repeat sequences had either perfect tetra-nucleotide CTCTT STRs or very complicated compound STRs that in principle consisted of the expected CTCTT STRs ranging from 7 to 31 times in all alleles. The two small-sized alleles of 216 bp and 219 bp carried seven perfect STRs which jumped up to 18 perfect STRs of allele of 263 bp, leading to the first big gap in allelic size distribution. All the remaining alleles that were longer than 275 bp carried compound STRs that were irregularly broken down in their upstream sequences by one to four CTCTT interrupting motifs, which subsequently created two separated repeat segments with a fixed three CTCTT units at the beginning and other 16 to 28 CTCTT perfect repeats at the end of the repeat sequences (except for one allele of 295 bp with its first CTCTT being replaced by CTTC in two sequences). These irregular structures present in the repeat sequences, on one hand contributed to form another big gap between alleles of 263 and 275 bp and on the other hand increased significantly the allelic fragment sizes of LEI0234 locus (Table 3).

Fig. 2: The peak pattern and allelic distribution of at LEI0234 locus
Table 3: Allelic polymorphism in DNA sequences at LEI0234 locus

<table>
<thead>
<tr>
<th>Allele size (bp)</th>
<th>Samples</th>
<th>Repeat region</th>
<th>Downstream flanking region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NW_001471633</td>
<td>(CTTT)₄</td>
<td>ATT A C</td>
</tr>
<tr>
<td></td>
<td>Z24837</td>
<td>(CTTT)₄ (CTTTCTTT)₆ (CTTT)₁₂</td>
<td>ATT A T</td>
</tr>
<tr>
<td>216</td>
<td>H25</td>
<td>(CTT)₈</td>
<td></td>
</tr>
<tr>
<td>219</td>
<td>F06</td>
<td>(CTTT)₄</td>
<td></td>
</tr>
<tr>
<td>219</td>
<td>S13</td>
<td>(CTTT)₄</td>
<td></td>
</tr>
<tr>
<td>265</td>
<td>E81, K21</td>
<td>(CTTT)₈</td>
<td></td>
</tr>
<tr>
<td>275, 273, 283, 287,</td>
<td>C123, E48, F16, F24,</td>
<td>(CTTT)₄ CTTT (CTTT)₁₇,₁₉,₂₁</td>
<td></td>
</tr>
<tr>
<td>291, 307, 319</td>
<td>K21, F37, T20, V07, V09</td>
<td>(CTTT)₄ (CTTTCTTTCTTT)₆ (CTTT)₁₉,₂₁</td>
<td></td>
</tr>
<tr>
<td>295</td>
<td>H25, L15</td>
<td>CTTG (CTTT)₄ CTTT (CTTT)₁₉</td>
<td></td>
</tr>
<tr>
<td>299, 307, 315, 319, 323</td>
<td>F16, F24, S13, U14, U38</td>
<td>(CTTT)₄ CTTTCTTTCTTT (CTTT)₄ CTTT (CTTT)₂₁,₂₃,₂₅</td>
<td></td>
</tr>
<tr>
<td>287, 303, 311, 316</td>
<td>C97, E81, F05, T20, V07</td>
<td>(CTTT)₄ (CTTTCTTT)₆ (CTTT)₁₇,₂₁,₂₃,₂₅</td>
<td></td>
</tr>
<tr>
<td>323</td>
<td>V09</td>
<td>(CTTTCCTTT)₄ (CTTT)₈</td>
<td></td>
</tr>
</tbody>
</table>

Note that dots indicate identity with the reference sequences and ‘-’ for indel.

Fig. 3: Plot comparing the lengths of longest stretch of perfect repeats and total repeats at LEI0234 locus

Although the number of interrupting CCTT motifs among the upstream repeat sequences increased along with allelic size expansions, supporting the assumption of Anmarkrud et al. (2008) that the interrupting motifs act as a stabilizing effecter to increase microsatellite length and hence allelic size polymorphism, the plot of the longest stretch which consisted of perfect CTTT repeat units against the total lengths of either complete microsatellite allelic fragments or only repeat sequences (Fig. 3), still had a significant and positive correlation (r = 0.94). This implied that the introduction of such single interrupting motifs with a few representations in one particular region within the repeat sequences may have very limited stabilizing effect. This finding ruled out any possibly biased impact of either higher or lower mutation rate of the tetra-nucleotide microsatellites compared to that of the di-nucleotide counterparts (Weber and Wong, 1993; Armour et al., 1994; Chakraborty et al., 1997; Schug et al., 1998).

Similar to the sequence variations detected at the MCW0216 locus, there were also a number of 'cryptic' alleles of 219, 287, 291, 295, 307, 315, 319 and 323 bp which were identical in fragment
sizes but different in either downstream flanking sequences or repeat unit compositions. The extended sequences also revealed one SNP of C/T at the last third nucleotide of the forward primer included in the recommendations (Hoffmann et al., 2004) for genotyping LEI0234 locus.

CONCLUSION

This study provided data of the first comprehensive DNA sequence characterization of major alleles at two chicken microsatellites. The results of MCW0216 locus were not only interesting as its alleles carried very complicated DNA sequences with a number of mutations present in both flanking region and repeat regions, but also very useful because the identification of the single cytosine in/del explained and confirmed our early genotyping data with various alleles that were different from each other by single nucleotides. The three SNPs present within both primer regions for genotyping MCW0216 locus called for newly designed genotyping primers located in completely conserved flanking regions so as to avoid possible allelic drop-out due to reduced PCR amplification efficiency resulted from the mismatch of the current genotyping primers and some alleles. The LEI0234 locus turned out to be a complex of either perfect or compound microsatellite that carried most alleles with rather different DNA sequences within the repeat regions. The STRs of all alleles longer than 275 bp consisted of more than 20 basic CTTT repeats which were irregularly interrupted by 1-4 CCTT motifs, acting as a stabilizing effecter that may contribute to the expansion of microsatellite length and allelic polymorphism that validated our early genotyping data. The identification of many ‘cryptic’ alleles that were identical in fragment sizes but different in either flanking sequences or repeat unit compositions at both loci demanded an intensive and complete DNA sequence analysis of major microsatellite alleles for correct data interpretation of genetic diversity studies (Lopez-Giraledz et al., 2007; Yin et al., 2011).

ACKNOWLEDGMENTS

This study was supported by the National Key Project of Scientific and Technical Supporting Programs funded by Ministry of Science and Technology of China during the 11th ‘Five-Year’ Plan (No. 2008BADB2B01), the Sino-German bilateral cooperative project (14/10-11CHN11(2010-3-23)) and the CAAS-ILRI joint research program.

REFERENCES


