JAST (Journal of Animal Science and Technology) TITLE PAGE
Upload this completed form to website with submission

ARTICLE INFORMATION	Fill in information in each box below
Article Type	Research article
Article Title (within 20 words without abbreviations)	Investigation of MHC-B Linked LEI0258 Marker Diversity in Various Chicken Populations
Running Title (within 10 words)	Repeat Motif Diversity of MHC-B linked LEI0258 Marker
Author	Prabuddha Manjula[1], Denuwan Lasitha Perera[2], Roshani Fernando[3,4], Minjun Kim[3], Eunjin Cho[5], Jun Heon Lee[3,5*]
Affiliation	[1]Department of Animal Science, Uva Wellassa University, Badulla, Sri Lanka [2]Department of Animal Science, University of Ruhuna, Kamburupitiya, Sri Lanka [3]Department of Animal Science, Chungnam National University, Daejeon 34134, Korea [4]Department of Animal Science, Faculty of Agriculture, University of Peradeniya, Sri Lanka [5]Department of Bio-AI Convergence, Chungnam National University, Daejeon 34134, Korea
ORCID (for more information, please visit https://orcid.org)	Prabuddha Manjula: https://orcid.org/0000-0001-8074-8323 Denuwan Lasitha Perera: https://orcid.org/0009-0002-1516-8108 Roshani Fernando: https://orcid.org/0000-0001-6712-3302 Minjun Kim: https://orcid.org/0000-0002-8173-8431 Eunjin Cho: https://orcid.org/0000-0003-4800-1603 Jun Heon Lee: https://orcid.org/0000-0003-3996-9209
Competing interests	No potential conflict of interest relevant to this article was reported.
Funding sources State funding sources (grants, funding sources, equipment, and supplies). Include name and number of grant if available.	This study was supported by a grant from the National Research Foundation, Republic of Korea (grant number MHC-2022R1F1A1064025).
Acknowledgements	Not applicable.
Availability of data and material	Upon a reasonable request, the datasets of this study can be available from the corresponding author.
Authors' contributions Please specify the authors' role using this form.	Conceptualization: Manjula P, and Lee JH Data curation: Manjula P Formal analysis: Manjula P Methodology: Manjula P, and Lee JH Software: Manjula P Validation: Manjula P Investigation: Kim M, Cho E, and Lee JH Writing - original draft: Manjula P, Perera DL, and Fernando R Writing - review & editing: Manjula P, Perera DL, Fernando R, , Kim M, Cho E, and Lee JH
Ethics approval and consent to participate	All samples were obtained in accordance with the guidelines of the Institutional Animal Care and Use Committee of the National Institute of Animal Science (NIAS; 2012-C-037), and the guidelines of the Guide for the Care and Use of Laboratory Animals.

ACCEPTED

CORRESPONDING AUTHOR CONTACT INFORMATION

For the corresponding author (responsible for correspondence, proofreading, and reprints)	Fill in information in each box below
First name, middle initial, last name	Jun Heon Lee
Email address – this is where your proofs will be sent	junheon@cnu.ac.kr
Secondary Email address	prabuddha@uwu.ac.lk
Address	213, KTnG, Chungnam National University, 99, Daehakro, Yuseong-gu, Daejeon, Republic of Korea
Cell phone number	+82-10-5172-0816
Office phone number	+82-42-821-5779
Fax number	+82-42-825-9754

ACCEPTED

Abstract

The variable number tandem repeat (VNTR) LEI0258 is the most polymorphic marker in the chicken major histocompatibility complex (MHC) B region. Unlike other microsatellite markers, LEI0258 is characterized by R13 and R12 repeat motifs, resulting in greater allele variation across chicken breeds. The allele size of this marker can also vary due to deletions and insertions of one to several base pairs. Sanger sequencing of LEI0258 alleles allows for determining exact allele sizes (bp), elucidates its repeat motif combination patterns, reveals other polymorphisms in the non-repeating sequence region, and identifies allele size inconsistencies among populations. This study investigates LEI0258 diversity and its repeat motif and flanking sequence variation by Sanger sequencing of 621 LEI0258 alleles from Asian, African, and North American standard and commercial chickens. Eighty-eight different allele sizes (182–565 bp) were detected. Asian and African chickens exhibited more alleles than North American and commercial breeds. Eighteen shared alleles and numerous unique alleles were identified. There were 48 repeat motif combinations across the 88 allele sizes, including 16 novel combinations in Asian chickens and two in American and commercial chickens. In 26 alleles, the R13-R12 combinations consisted of a single copy of R13 with 2 to 28 R12 repeats; the remaining alleles contained various copy numbers for both R13 and R12 repeats. Moreover, the same allele size could occur with different motif combinations. Additional allele variation was observed due to single-nucleotide polymorphisms (SNPs) and insertions or deletions (indels) in the upstream or downstream of the LEI0258 marker. Collectively, the loss or gain of VNTRs and additional polymorphisms explain the broader allele variation in LEI0258; greater MHC diversity is observed in Asian and African chickens.

Keywords: chicken MHC-B, LEI0258 sequence, repeat motifs, VNTR marker

ACCEPTED

Introduction

Major histocompatibility complex (MHC) B diversity can be evaluated using microsatellite markers in MHC gene regions. The degrees of variation and polymorphism differ depending on the location of markers within the MHC region. Markers located near or within most polymorphic regions, such as Class I and II, provide more information regarding polymorphism content and allele diversity than those in the lower diversity MHC gene regions. The chicken LEI0258 marker exhibits high allele variation and is linked to the MHC-B locus; some of its alleles are correlated with MHC-B serological haplotypes [1]. Moreover, the relationship between the LEI0258 marker and previously defined MHC Class I (BF) haplotypes in the blue-egg Caipira chicken shows strong genetic disequilibrium. This marker has been associated with disease resistance, production, and reproduction traits in various chicken populations [2,3,4]. These findings suggest that LEI0258 can serve as a valuable and affordable genetic marker for investigating MHC genetic diversity in chickens. LEI0258 is a variable number tandem repeat (VNTR) marker characterized by two repeat sequences of 13 bp (ATGTCTTCTTTCT)n (R13) and 12 bp (TTCCTTCTTTCT)n (R12). The allele size of this locus is primarily determined by different combinations of the R13 and R12 repeat motifs. However, additional polymorphisms in the flanking region contribute to greater allele variation in this marker with similar or different repeat motif combinations. Recombination at the B-BNT gene also may contribute to this locus's evolution [5].

More than 50 different allele sizes (variations in length) have been reported for LEI0258 in diverse chicken populations, including breeds from Asia, Africa, Europe, and North and South America. A broader allelic range (182–552 bp) has been identified in many chicken breeds through either polymerase chain reaction (PCR)-capillary electrophoresis (CE) or PCR sequencing [1,5,6,7,8,9,10,11]. The number of alleles and the allele size range considerably vary among breeds; these differences may contribute to breed-specific characteristics due to unique allele counts and structural mechanisms.

The allele detection method (e.g., CE or Sanger sequencing), along with the genotyping instruments and size standards (500 Liz and 600 Liz), can result in different allele sizes. Unlike other microsatellite markers, LEI0258 fragment size primarily varies due to the gain or loss of 12 or 13 bp repeat motifs. However, alleles with 1–8 bp differences, caused by deletions upstream or downstream of the LEI0258 repeats, have also been identified. Due to the subjective nature of the microsatellite allele binning process, alleles identified in one population might be reported as different alleles in another population, misrepresenting the diversity of this marker.

ACCEPTED

49 Sequencing of the alleles provides their exact size (bp) and reveals additional polymorphisms, such as single-
50 nucleotide polymorphisms (SNPs) and insertions or deletions (indels). Thus, comparing alleles identified by CE
51 with those determined by PCR sequencing validates allele sizes across global chicken populations. Despite these
52 efforts, this marker remains uncharacterized in numerous breeds, suggesting greater allele diversity in LEI0258.
53 Therefore, this study investigates reported Sanger sequences of LEI0258 alleles from Asian, African, and North
54 American chickens, along with those from our study populations, to summarize the global allele distribution and
55 diversity of LEI0258 marker. Allele variation was examined in terms of fragment size (bp) and variants (repeat copy
56 numbers, unique combinations, SNPs, and indels), and the underlying evolutionary mechanisms. This information
57 will assist researchers in studying this marker in new chicken populations by providing clarity with respect to correct
58 allele sizes.

59 # Materials and Methods

60 **Ethics approval**

61 All samples were obtained following the guidelines of the Institutional Animal Care and Use Committee of the
62 National Institute of Animal Science (NIAS; 2012-C-037), and the guidelines of the Guide for the Care and Use of
63 Laboratory Animals.

64 **Sequencing data for LEI0258 marker**

65 The sequence data for this study consisted of two sources; the first data set was obtained by re-sequencing 59
66 homozygous samples corresponding to 21 fragment sizes from South Korean, Bangladesh, and Sri Lankan local
67 chicken populations [12]. Sequences of each sample were obtained using the primer pair (CAJF01F) 5'-
68 TCGGGAAAAGATCTGAGTCATTG, and (CAJF01R) 5'-TGATTTTCAGATCGCGTTCCTC [6]. These primers
69 cover the region that flanks the primers generally used for PCR-CE of the LIE0258 locus [6,7]. Therefore, the exact
70 allele size (bp) can be obtained by Sanger sequencing its forward and reverse strands.
71 The second sequence data set included 562 sequences retrieved from the NCBI database, with GenBank accession
72 numbers. These include commercial and standard chicken breeds in North America [6], Asian chicken breeds
73 [5,7,13,14], and African chicken breeds [15,16].

74 **Data analysis**

75 **Allele diversity**

76 The number of total, common, and population unique alleles (private alleles) was analyzed and visualized using
77 GenAlEx and R ggplot2 package [17].

ACCEPTED

78 **Sequence data analysis**

79 The sequences were edited using the BioEdit sequence alignment editor [18] and aligned using the Clustal multiple
80 aligned option in MEGA7 software [19]. The copy number of repeats of R13 (“ATGTCTTCTTTCT”) and R12
81 (“TTCCTTCTTTCT”) were counted. As LEI0258 is a compound of R13 and R12 repeats, respectively, all possible
82 repeat copy combinations were summarized. To identify polymorphisms in the regions surrounding the repeats, the
83 flanking sequences were defined based on the coverage provided by the LEI258 sequencing primers (not the
84 conventional genotyping primers) [6], which span the whole LEI0258 marker region. Accordingly, the upstream
85 flanking region was defined from position -1 to -64, and the downstream region was defined from position +1 to +88,
86 including the last R12 repeat. SNPs and indels within these defined flanking regions were detected using DNASP6
87 software [20].

88 Repeat combinations and flanking sequence polymorphisms (SNP and indels) were identified using sequences
89 corresponding to 90 fragment sizes (182 bp to 565 bp) from African, Asian, North American, and commercial
90 chicken populations. Within each population, 100% identical sequence of each allele was removed to avoid
91 redundancy. Therefore, at least one sequence for each fragment size was kept. Occasionally, fragments of similar
92 sizes with a different flanking sequence identity and repeat copy numbers were included for the analysis.

93 Evolution relationships between alleles were analyzed based on the indel variations and SNP polymorphisms in the
94 upstream and downstream of the repeat region. Flanking sequence polymorphisms, both indels and SNPs were
95 converted to binary values. Alleles were grouped based on the sequence identity and segregation points.
96 Relationships were visualized using a Neighbour-joining network tree analysis in the SplitTree program [21].

97 **Correlation between allele size and R13-R12 repeat combinations.**

98 To establish a relationship between allele size and repeat motif combination, exploratory analysis focusing on the
99 copy numbers of the R13 (“ATGTCTTCTTTCT”) and R12 (“TTCCTTCTTTCT”) motifs in relation to allele sizes
100 (measured in base pairs). The copy numbers of each motif were manually determined from aligned sequences, while
101 allele sizes were derived from sequencing data. Two primary analyses were conducted: 1. The correlation between
102 R13 and R12 repeat numbers, aimed at evaluating whether the two motifs exhibit co-variation; and 2. The
103 relationship between total repeat copy number and allele size, intended to assess whether repeat expansion
104 contributes to increased fragment length. Scatter plots were generated to visualize these relationships both across the
105 entire dataset and within individual populations. While no formal hypothesis was conducted, the Pearson correlation
106 coefficient was calculated and used as a descriptive measure to indicate the relative strength of the association

ACCEPTED

107 between the variables. These analyses were intended to provide an initial, qualitative understanding of the potential
108 correlations rather than definitive statistical conclusions.

109
Results

110 **Global variation and evolution of different fragment sizes**

111 The global diversity of LEI0258 allele sizes and repeat combinations was analyzed via sequencing. Eighty-eight
112 allele sizes, ranging from 182 to 565 bp, were identified. Asian chicken populations exhibited the highest allele
113 diversity, with 63 of the 88 alleles. African chickens had 56 alleles, whereas North American standard and
114 commercial breeds had only 28 alleles. All three populations shared 19 allele sizes; in terms of private alleles, 23
115 were detected in Asian breeds, 20 were observed in African breeds, and only five were found in North American
116 standard breeds (Figure 1). The wide range of LEI0258 allele variation has important applications in population
117 genetics, breed differentiation, and the preliminary characterization of MHC diversity, offering a valuable tool for
118 selective breeding, conservation programs, and potential association studies targeting disease resistance and
119 production traits.

120 American and European chicken populations display lower diversity because they have been selectively bred for
121 economically important traits or to maintain specific MHC alleles. These experimental populations, selected for
122 serologically known MHC alleles, and the purebred standard populations [6], exhibit limited genetic variability
123 compared with local chickens due to their small population sizes.

124 **Diversity in repeat motif (R13 and R12) sequences**

125 PCR fragment size significantly varied due to combinations of the R13 and R12 repeat elements. The numbers of
126 R13 and R12 repeat copies varied among allele sizes and populations (Table 2). In the first Asian chicken dataset, 21
127 LEI0258 allele sizes were observed, including four different R13 repeats (1, 12, 15, and 22 copies) and 15 different
128 R12 repeats (3, 4, 5, 7, 8, 9, 11, 12, 15, 16, 18, 20, 21, 25, and 27 copies). Almost all repeat copies were consistent
129 with previous reports [6,7,9,13,22].

130 Seventeen different combinations of R13 and R12 were identified in this dataset. Fourteen contained a single copy
131 of R13 but exhibited varying R12 copy numbers according to allele size. One new combination (R13 = 1, R12 = 25)
132 was identified in the Korean native chicken with a 465 bp allele.

133 Considering the entire dataset, 18 different variants in R13 copies and 26 different variants in R12 copies were
134 identified. The range of R13 copy numbers (1-29) was consistent with previous reports, although 11 of these copy

ACCEPTED

135 numbers were not found in any of the populations studied. A single copy of R13 was the most common; 55% of all
136 combinations contained one copy of R13. The number of R12 repeats ranged from 2 to 28 copies, which exceeds the
137 range of 2-20 copies [6]. All possible R12 copy numbers were observed, except for 23 copies.

138 Higher combinations of R13 and R12 repeats (48 combinations) were identified across all three populations.
139 Notably, each population displayed unique distributions of repeat copy numbers (Figure 2a-d). Some of these
140 combinations were common across all three populations or shared between Asian and African, or Asian and
141 American chickens.

142 Missing alleles in North American chickens, particularly those with the 1-3, 1-9, 1-10, and 1-24 combinations, might
143 be present in other populations [6]. All of these alleles and repeat possibilities were found in Asian and African
144 chickens.

145 Sixteen new repeat combinations (1-25, 1-26, 1-27, 1-28, 8-3, 9-4, 10-3, 11-3, 12-3, 12-4, 17-4, 17-5, 17-6, 18-6,
146 19-7, and 20-6) were exclusively present in Asian chickens (Figure 2, Table 2); the 22-3, 25-3, 27-3, and 28-3
147 combinations were not observed in Asian chickens. However, these combinations were found in American and
148 African chickens. The African chicken population examined in this study did not exhibit any unique combinations of
149 R13 and R12; one unique combination (27-3) was found in North American chickens, and another unique
150 combination (29-3) was present in commercial chickens.

151 **Correlation between fragment size and repeat motifs**

152 Although many mathematically possible combinations of R13 and R12 motifs exist, only 48 combinations were
153 identified; this number is higher than the 37 that was reported by a previous study [9]. Twenty-six allele sizes had
154 only one R13 repeat with 2-28 R12 repeat motifs (Figure 3a-d), comparable to the previous findings [9]. In the
155 Asian, African, and North American populations, this combination displayed a linear relationship with increasing
156 allele size up to certain sizes. However, for at least 23 allele sizes (i.e., 292, 305, 317, 318, 331, 344, 356, 419, 420,
157 421, 431, 433, 443, 445, 458, 474, 483, 484, 487, 512, 513, 539, 552, and 565 bp), the linear trend shifted, revealing
158 more than eight R13 repeats and various copy numbers for the R12 repeat (2-8). Mutation mechanisms led to allele
159 sizes that differed by only one base but had completely different VNTR motifs. Some of these alleles were private,
160 being unique to specific populations. For example, the 317 bp and 318 bp fragments differ by a single base but have
161 completely different R13 and R12 repeat motifs: 9-4 and 10-3, respectively. Additionally, the 331 bp allele had two
162 different motif combinations in different populations: in the Asian wild red jungle fowl (GenBank accession

ACCEPTED

163 number: KF535097), this allele had a 1-14 R13/R12 combination, whereas in domestic chicken (GenBank accession
164 number: KX365371), it had an 11-3 combination for the same repeat units. Similarly, the 356 bp allele found in an
165 African breed (GenBank accession number: MG518246) had a 1-16 combination; the same allele in a Chinese
166 chicken breed (GenBank accession number: KX365375) had a 12-4 combination.

167 **Allele variation due to SNP and indel polymorphisms in the flanking region**

168 A higher rate of polymorphism in the flanking region is also associated with differences in allele size. Evolutionary
169 mechanisms maintaining a fixed number of alleles with greater variation in LEI0258 have been discussed [5].
170 LEI0258 alleles can have the same fragment size and repeat motif combinations but differ in their flanking
171 sequences; additionally, the same fragment size with different motif combinations is possible.

172 SNPs and indels were observed in the flanking region. Downstream of R12, an 8 bp deletion ("ATTTTGAG") from
173 +21 to +28, and a 2 bp deletion (TT/Δ) at –31 and –32 in the upstream region, are the primary causes of changes in
174 allele size by 2 or 8 bp. For example, a 2 bp deletion in the 309 bp allele creates a 307 bp allele. Similarly, 2 bp
175 deletions in the 381 bp and 285 bp alleles result in 379 bp and 283 bp alleles, respectively. The 8 bp deletion was
176 mostly observed in alleles ranging from 182 to 217 bp and in the 241 bp allele. However, these two mutations were
177 never found together. Moreover, single nucleotide deletions were reported; a few of these were monotonous,
178 whereas the one-base deletion at +31 was more frequent. Five probable SNPs were identified at +3, +13, +31, +43,
179 and +46, along with three SNPs at –12, –30, and –45 (Table 1). Additional allele variation beyond 170 bp, based on
180 the most common flanking sequence variants (SNPs and indels), is summarized in a network tree, which shows that
181 these alleles are divided according to the presence of the main indel variants, whereas the subclusters are due to the
182 SNPs.

ACCEPTED

183 Discussion

184 VNTR, in which non-coding DNA sequences are organized as tandem repeats, is among the most polymorphic
185 genomic loci in higher vertebrates, invertebrates, bacteria, and plants. A locus with a 2-6 bp sequence repeat is
186 known as a microsatellite, whereas sequence repeat loci ranging from 10-100 bp are categorized as minisatellites.
187 Although the LEI0258 marker is located in a non-coding region, the interplay between its repeat motifs (R12 and
188 R13) and flanking sequence polymorphisms (such as SNPs and indels) may influence the structural evolution of the
189 locus. This could potentially impact regulatory elements or create a linkage with nearby functional MHC genes,
190 recommending further investigation into its functional significance [23]. While many VNTR markers are considered

selectively neutral with no known functional effects, some can influence biological functions directly or are closely linked to important gene regions, undergoing selection through the "hitch-hiking effect" [24,25]. In chickens, LEI0258 is a compound minisatellite marker situated within the B-BTN gene and is linked to the MHC-B complex on microchromosome 16.

Due to the high mutation rates in the repeating and flanking regions of LEI0258 and its significant associations with disease and economically important traits, this VNTR marker is considered highly valuable for the molecular identification of MHC alleles across chicken populations. To date, more than 50 distinct allele sizes have been reported for LEI0258 using CE. Substantially different amplicon sizes at this locus can be separated by gel or CE. Moreover, many of these alleles have been sequenced in at least one population. Comparisons of allele sizes estimated by electrophoresis with sequences of the same alleles have shown small base pair disparities. This size inconsistency ranges from 1 to 20 bp. In addition to indel polymorphisms and the loss or gain of repeat copy numbers that contribute to size differences, external factors such as PCR amplification, genotyping instruments, internal size standards, and allele-calling software may explain these small base pair differences. Therefore, sequencing is necessary to determine whether these allele differences are due to polymorphisms (indels and SNPs) or extrinsic factors.

ACCEPTED

Sequencing each allele is not always practical or cost-effective for genotyping large sample sizes. LEI0258 exhibits high heterozygosity. If new genotypes or alleles are heterozygous, additional steps, such as cloning, are required before sequencing. When sequencing all alleles is limited, the use of available sequence information and control or known allele sizes is more appropriate in identifying new LEI0258 alleles in any population. Nonetheless, CE is effective when the same DNA analyzer and internal size standards are consistently used, which reduces allele size inconsistencies.

This subtyping system enables the unprecedented differentiation of LEI0258 alleles among numerous chicken breeds worldwide. Asian and African chicken populations include many indigenous breeds. Compared with commercial breeds, Asian, African, and European local breeds show considerable diversity in allele size and a high polymorphism rate in the repeating regions (R12 and R13) [5,9,26]. This reflects the unique genetic variation preserved in Asian and African indigenous chickens.

In contrast, commercial breeds have fewer alleles. Highly selected layer lines have fewer alleles than broiler breeds [12,27]. The North American chicken populations exhibit low diversity because they have been selectively bred for

219 a few MHC alleles and improved for economically important traits. Specifically, the experimental population
220 (selected for serologically known MHC alleles) and purebred [6] displayed limited genetic variability in the MHC.

221 Measures of allele number or allele richness are more sensitive than heterozygosity to founder events followed by
222 population expansion [28]. Loss of alleles during a founder event reduces allele richness but not heterozygosity [29].
223 This partially explains the observed high heterozygosity despite the reduction of allele number, which could also
224 indicate a heterozygote advantage, as MHC heterozygotes may recognize a broader range of antigens.

225 Furthermore, the presence of numerous alleles and their frequencies hold the potential for a response to selection
226 [30,31,32]. We speculate that the diversity of LEI0258 alleles reflects the immense allele diversity of the MHC
227 region that remains to be discovered in indigenous chickens.

228 Common alleles are segregated among populations that separated genetically long ago. This phenomenon partially
229 explains the evolutionary stability of MHC allele sizes over generations, given their biological importance in disease
230 resistance. Sequence analysis of these common allele sizes indicates that alleles can differ due to SNP
231 polymorphisms in the non-repeating sequences, despite the presence of identical repeat variations. These
232 observations reflect the parallel evolutionary mechanisms underlying the substantial MHC diversity.

233 The higher number of alleles (n=14) shared between African and Asian chickens suggests a historical genetic
234 relationship. However, inferring such genomic relationships using a single gene region is inappropriate. According
235 to the literature, most populations share several common alleles. Theoretically, closely related breeds share many
236 alleles, whereas more unrelated populations likely share fewer alleles but have distinct allele distribution. One
237 population might possess several alleles with high frequencies, while similar alleles may have low frequencies in
238 other populations. This is because the number of alleles, or allele richness, and allele frequencies depend on
239 population size, breeding history, founder effect, and genetic drift. Since MHC allele diversity significantly
240 correlates with the pathogenic dynamics of living environments, selecting alleles and their effects align with the
241 rare-allele hypothesis.

242 In this study, many rare alleles were found in Asian and African chickens, based on both allele size and
243 polymorphisms in the flanking sequences. Significant negative Tajima's D values (< 0) supported the observation of
244 an excess of rare alleles. This may be due to the linkage of the target locus with a gene in the MHC-B region or
245 population expansion after recent bottleneck events.

ACCEPTED

246 **Diversity in repeat motif R13 and R12 sequences**

247 We observed that alleles of similar size but with different repeat motif copy number combinations occurred within
248 the same populations and among distant populations. This could be misleading when discriminating populations
249 based on shared allele-based genetic distance estimation. Similar observations have been discussed [9], suggesting
250 that this is due to homoplasy. The high mutation rate in microsatellite markers is likely responsible for homoplasy,
251 so it can no longer be assumed that alleles identical by state (IBS) are identical by descent (IBD). Therefore,
252 phylogenetic relationships estimated based on allele size similarity or shared alleles at a single locus are not reliable
253 indicators of an IBD relationship.

254 Several factors affect the mutation rate of VNTR loci. The size of the repeat motif and repeat copy numbers are
255 central because microsatellite mutation is primarily due to intramolecular slipped-strand mispairing (SSM) during
256 replication. Recombination might also contribute to minisatellite mutations during or after replication, in addition to
257 SSM, which is generally responsible for indel mutations. The evolution of the LEI0258 locus aligns with these
258 mechanisms and can be explained by the classical stepwise mutational model (SMM) [33]. Considering that R13
259 and R12 repeat copy number variation determines the allele size range, the mutation rate of these two repeats is
260 crucial. According to the SMM, the addition or deletion of one unit of R13 or R12 results in an increase or decrease
261 in allele size. It is evident that in all populations, R13 exhibited a lower mutation probability (17 of 29 possibilities)
262 than the R12 repeat unit (25 of 26 possibilities). For example, in the 182-417 bp allele range, only one R13 repeat
263 copy number was observed, whereas the R12 copy number linearly increased. However, this mutation pattern
264 shifted in larger alleles, where the R13 copy number tended to increase while the R12 copy number decreased after
265 the 417 bp allele. Therefore, it remains difficult to explain the mutation rate of compound markers such as LEI0258
266 solely based on repeat copy numbers and linear or geometric relationships between mutations and allele size. This
267 presents a limitation for automated allele size calling programs (e.g., TANDEM), because the power function
268 relationship is typically applied for microsatellite repeats (di-, tri-, and tetra-nucleotides) to transform allele size into
269 integers. Compound markers with different mutation rates may not perform well under this algorithm.

270 The evidence for the role of recombination in the evolution of non-coding VNTR loci has been reviewed [34].
271 Homologous recombination and localized recombination between non-identical homologous alleles of VNTR both
272 occur, resulting in novel allele variation and sequence conversion. Evidence of recombination at the LEI0258 locus
273 using flanking sequences has been studied [5]. Moreover, the butyrophilin-like (*BTN*) gene cluster in the chicken,

ACCEPTED

274 where LEI0258 is located, undergoes duplication and gene shuffling events between the *BG-like* and *B32.2-like*
275 genes. These events may lead to localized recombination between existing alleles, resulting in novel alleles with
276 new repeat combinations that deviate from the typical pattern of repeat copy number combinations.

277 **Allele variation in the flanking region**

278 SNPs and indel variation in the flanking regions also drive LEI0258 allele variation. We observed alleles with
279 different sizes that had identical tandem repeat copy numbers; the size difference was caused by small deletions in
280 the flanking regions. For example, the 307 bp and 309 bp alleles have identical LEI0258 repeat units but differ by a
281 2 bp deletion in the upstream region. Similarly, some alleles possessing identical or different LEI0258 repeats have
282 an 8 bp deletion. SNPs in the flanking region generated new alleles, regardless of allele size or copy number
283 variation. However, we identified more alleles based on this flanking variation than previously reported. Previous
284 studies assessing the correlation between the serological B haplotype [6,9] and MHC-B SNP panel-based haplotypes
285 [26,35,36] showed that different B or SNP haplotypes are associated with the same LEI0258 allele (e.g., the 193 bp
286 allele is found in B15.1, B11, and B27). The use of these flanking sequence polymorphism differences may help to
287 determine whether distinct B haplotypes are associated with different alleles at the same LEI0258 loci.

288 # Conclusion

289 Sequence information for the LEI0258 marker from diverse breeds around the world reveals LEI0258 allele
290 diversity and population-specific MHC diversity. Moreover, various mutation mechanisms (SSM, SNPs, indels, and
291 recombination) contribute to this allele variation. Sequence data helps overcome the limitations associated with
292 STR-based subtyping. However, novel LEI0258 alleles might still exist in understudied chicken populations. The
293 comparison of newly detected LEI0258 alleles in local chickens with those in standard chicken populations and
294 reference cell lines exhibiting known MHC-B serological haplotypes could facilitate inferences regarding the MHC
295 serological haplotypes in local chickens as a preliminary step in MHC characterization.

296

ACCEPTED

297

298 # Recommendations

299 When typing new populations for LEI0258 diversity, comparisons with existing reference alleles of the same size
300 should be performed to avoid allele inconsistency. Additionally, sequence information for novel alleles should be
301 compared with common and shared alleles to understand their evolutionary patterns.

302 # Funding Source

303 This study was supported by a grant from the National Research Foundation, Republic of Korea (grant number
304 MHC- 2022R1F1A1064025).

305

ACCEPTED

306 # References

307 1. Chazara O, Juul-Madsen HR, Chang CS, Tixier-Boichard M, Bed'hom B. Correlation in chicken between the
308 marker LEI0258 alleles and Major Histocompatibility Complex sequences. BMC Proc. 2011;4:3-5.
309 https://doi.org/10.1186/1753-6561-5-S4-S29.

310 2. Nikbakht G, Esmailnejad A. Chicken major histocompatibility complex polymorphism and its association with
311 production traits. Immunogenetics. 2015;67:247-52. https://doi.org/10.1007/s00251-015-0832-7.

312 3. Banat GR, Tkalcic S, Dzielawa JA, Jackwood MW, Saggese MD, Yates L, et al. Association of the chicken
313 MHC B haplotypes with resistance to avian coronavirus. Dev Comp Immunol. 2013;39:430.
314 https://doi.org/10.1016/j.dci.2012.10.006.

315 4. Lwelamira J, Kifaro GC, Gwakisa PS, Msoffe PLM. Association of LEI0258 microsatellite alleles with
316 antibody response against Newcastle disease virus vaccine and body weight in two Tanzania chicken ecotypes.
317 Afr. J. Biotechnol. 2008;7:714-20.

318 5. Guangxin E, Sha R, Zeng S, Wang C, Pan J, Han J. Genetic variability, evidence of potential recombinational
319 event and selection of LEI0258 in chicken. Gene. 2014;537:126-31.
320 https://doi.org/10.1016/j.gene.2013.12.040.

321 6. Fulton JE, Juul-Madsen HR, Ashwell CM, McCarron AM, Arthur JA, O'Sullivan NP, et al. Molecular genotype
322 identification of the Gallus gallus major histocompatibility complex. Immunogenetics. 2006; 58:407-21.
323 https://doi.org/10.1007/s00251-006-0119-0.

324 7. Han B, Lian L, Qu L, Zheng J, Yang N. Abundant polymorphisms at the microsatellite locus LEI0258 in
325 indigenous chickens. Poult Sci. 2013;92:3113-9. https://doi.org/10.3382/ps.2013-03416.

326 8. Izadi F, Ritland C, Cheng KM. Genetic diversity of the major histocompatibility complex region in commercial
327 and noncommercial chicken flocks using the LEI0258 microsatellite marker. Poult Sci. 2011;90:2711-7.
328 https://doi.org/10.3382/ps.2011-01721.

329 9. Chazara O, Chang CS, Bruneau N, Benabdeljelil K, Fotsa JC, Kayang BB, et al. Diversity and evolution of the
330 highly polymorphic tandem repeat LEI0258 in the chicken MHC-B region. Immunogenetics. 2013;65:447-59.
331 https://doi.org/10.1007/s00251-013-0697-6.

332 10. Touko BAH, Keambou CT, Han JM, Bembidé C, Skilton RA, Ogugo M, et al. Molecular typing of the major
333 histocompatibility complex B microsatellite haplotypes in Cameroon chicken. Anim. Genet. 2015; 56:47-54.
334 https://doi.org/10.1017/s2078633614000538.

335 11. Da Veiga Lima-Rosa CA, Canal CW, Fallavena PRV, De Freitas LB, Salzano FM. LEI0258 microsatellite
336 variability and its relationship to B–F haplotypes in Brazilian (blue-egg Caipira) chickens. Genet. Mol. Biol.
337 2005;28:386-9. https://doi.org/10.1590/s1415-47572005000300008.

338 12. Manjula P, Kim M, Cho S, Seo D, Lee JH. High Levels of Genetic Variation in MHC-Linked Microsatellite

ACCEPTED

339 Markers from Native Chicken Breeds. Genes (Basel). 2021;12:240. https://doi.org/10.3390/genes12020240.

340 13. Huang X, Li L, Zhang JF, He DL, Zhang XQ, Chen JB, et al. Evaluation of Diversity and Evolution of the
341 Microsatellite LIE0258 in chicken MHC–B from South China. Acta Veterinaria et Zootechnica Sinica (in
342 Chinese). 2016;47:2175-83. https://doi.org/10.11843/j.issn.0366–6964.2016.11.04.

343 14. Haunshi S, Devara D, Ramasamy K, Ullengala R, Chatterjee RN. Genetic diversity at major histocompatibility
344 complex and its effect on production and immune traits in indigenous chicken breeds of India. Arch Anim
345 Breed. 2020;63:173-82. https://doi.org/10.5194/aab-63-173-2020.

346 15. Mwambene PL, Kyallo M, Machuka E, Githae D, Pelle R. Genetic diversity of 10 indigenous chicken ecotypes
347 from Southern Highlands of Tanzania based on Major Histocompatibility Complex-linked microsatellite
348 LEI0258 marker typing. Poult Sci. 2019;98:2734-46. https://doi.org/10.3382/ps/pez076.

349 16. Hailu A, Kyallo M, Yohannes T, Sendeku W, Getu A, Dagnachew S, et al. Genetic Diversity and Population
350 Structure of Indigenous Chicken Ecotypes (Gallus gallus domesticus) in Ethiopia using LEI0258 Microsatellite.
351 Int J Poult Sci. 2020;19:102-10. https://doi.org/10.3923/ijps.2020.102.110.

352 17. Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and
353 reasearch – an update. Bioinformatics. 2012;28:2537-9 https://doi.org/10.1093/bioinformatics/bts460

354 18. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows
355 95/98/ nt. Nucleic Acids Symposium Series. 1999 [cited 2025 April 1]. https://ci.nii.ac.jp/naid/10030689140.

356 19. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across
357 Computing Platforms. Mol Biol Evol. 2018;35:1547-9. https://doi.org/10.1093/molbev/msy096.

358 20. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6:
359 DNA Sequence Polymorphism Analysis of Large Data Sets. Mol Biol Evol. 2017;34:3299-302.
360 https://doi.org/10.1093/molbev/msx248.

361 21. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol.
362 2006;23:254-67. https://doi.org/10.1093/molbev/msj030.

363 22. Ngeno K, van der Waaij EH, Megens HJ, Kahi AK, van Arendonk JAM, Crooijmans RPMA. Genetic diversity
364 of different indigenous chicken ecotypes using highly polymorphic MHC-linked and non-MHC microsatellite
365 markers. Animal Genetic Resources. 2015;56:1-7. https://doi.org/10.1017/S2078633614000484.

366 23. Ellegren, H. Microsatellites: simple sequences with complex evolution. Nat. Rev. Genet. 2004;5:435-445.
367 https://doi.org/10.1038/nrg1348.

368 24. Rockman MV, Wray GA. Abundant raw material for CIS-Regulatory evolution in humans. Mol. Biol. Evol.
369 2002;19:1991-2004. https://doi.org/10.1093/oxfordjournals.molbev.a004023.

ACCEPTED

370 25. Gymrek M, Golan D, Rosset S, Erlich Y. LobSTR: A short tandem repeat profiler for personal genomes.
371 Genome Res. 2012;22:1154-1162. https://doi.org/10.1101/gr.135780.111.

372 26. Manjula P, Fulton JE, Seo D, Lee JH. Major histocompatibility complex B variability in Korean native chicken
373 breeds. Poult Sci. 2020;99:4704-13. https://doi.org/10.1016/j.psj.2020.05.049.

374 27. Esmailnejad A, Nikbakht Brujeni G, Badavam M. LEI0258 microsatellite variability and its association with
375 humoral and cell mediated immune responses in broiler chickens. Mol Immunol. 2017;90:22-6.
376 https://doi.org/10.1016/j.molimm.2017.06.027.

377 28. Greenbaum G, Templeton AR, Zarmi Y, Bar-David S. Allelic richness following population founding events--a
378 stochastic modeling framework incorporating gene flow and genetic drift. PLoS One. 2014;9:e115203.
379 https://doi.org/10.1371/journal.pone.0115203.

380 29. Nei M, Maruyama T, Chakraborty R. The bottleneck effect and genetic variability in populations. Evolution.
381 1975;29:1-10. https://doi.org/10.1111/j.1558-5646.1975.tb00807.x.

382 30. Petit RJ, El Mousadik A, Pons O. Identifying populations for conservation on the basis of genetic markers.
383 Conserv Biol. 1998;12:844-55. https://doi.org/10.1046/j.1523-1739.1998.96489.x.

384 31. Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics. Nat Rev Genet.
385 2010;11:697-709. https://doi.org/10.1038/nrg2844.

386 32. Caballero A, García-Dorado A. Allelic diversity and its implications for the rate of adaptation. Genetics.
387 2013;195:1373-84. https://doi.org/10.1534/genetics.113.158410.

388 33. Ohta T, Kimura M. A model of mutation appropriate to estimate the number of electrophoretically detectable
389 alleles in a finite population. Genet Res. 1973;22:201-4. https://doi.org/10.1017/s0016672300012994.

390 34. Harding RM, Boyce AJ, Clegg JB. The evolution of tandemly repetitive DNA: recombination rules. Genetics.
391 1992;132:847-59. https://doi.org/10.1093/genetics/132.3.847.

392 35. Fulton JE, Lund AR, McCarron AM, Pinegar KN, Korver DR, Classen HL, et al. MHC variability in heritage
393 breeds of chickens. Poult Sci. 2016;95:393-9. https://doi.org/10.3382/ps/pev363.

394 36. Fulton JE, Berres ME, Kantanen J, Honkatukia M. MHC-B variability within the Finnish Landrace chicken
395 conservation program. Poult Sci. 2017;96:3026-30. https://doi.org/10.3382/ps/pex102.

396 37. Khobondo J, Machuka, E, Bwire CW, Kyalo M, Githae D, Mwakubambanya, R, et al. The LEI0258
397 microsatellite revealed the presence of at least four gene pools in the Kenyan indigenous chicken populations.
398 Unpublished. 2017. LEI sequence submitted to NCBI.

ACCEPTED

399 **Table 1**. Sequence variation in LEI0258 alleles in Asian chicken breeds.

Consensus size (bp)	Upstream (-1 to -62)				[1]R13	[2]R12	Downstream (1 to 52 bp)						[3]Origin	[4]Reference Gene bank Accession
	-45	(-32 to 31)	-30	-12			+3	+13	+21-28	+31	+36	43		
	T/C	TT/Δ	G/A	G/A	n	n	C/T	T/C	ATTTTGAG/Δ	Δ/A	A/T	T/A		
193					1	3	T		Δ				JF, NG	MN936024
205					1	4			Δ				JF	MN936026
217					1	5			Δ				CD	MN936027
217					1	5			Δ				CD	MN936028
217					1	5			Δ				CD	MN936029
217					1	5			Δ				KNC	MN936030
217					1	5			Δ				YO	MN936031
247		Δ			1	7		C				A	BS	MN936032
247		Δ			1	7		C				A	YO	MN936033
249					1	7						A	AS	MN936034
249					1	7						A	NO	MN936035
249					1	7						A	YO	MN936036
249					1	7							YO	MN936037
249					1	7						A	CD	MN936038
249					1	7						A	HI	MN936039
249					1	7						A	NN	MN936040
249	C				1	7		C				A	OG	MN936041
249	C				1	7						A	HI	MN936042
271		Δ			1	9		C				A	HF	MN936043
273					1	9						A	HF	MN936044
273					1	9						A	KNC	MN936045
295		Δ			1	11							CO	MN936046
307		Δ	A		1	12							NY	MN936047
307		Δ	A		1	12							HY	MN936048
309					1	12					T		SL	MN936049
309					1	12					T		KNC	MN936050
309					1	12					T		KNC	MN936051
309					1	12					T		KNC	MN936052
344					12	3							JF	MN936053

400
401
402
403
404

ACCEPTED

Continue table 1.

Consensus size (bp)	Upstream (-1 to -62)				[1]R13	[2]R12	Downstream (1 to 52 bp)						[3]Origin	[4]Reference GenBank Accession number
	-45	(-32 to 31)	-30	-12			+3	+13	+21-28	+31	+36	43		
	T/C	TT/Δ	G/A	G/A	n	n	C/T	T/C	ATTTTGAG/Δ	Δ/A	A/T	T/A		
344					12	3							JF	MN936054
344					12	3							CD	MN936055
345[a]					1	15					T		NN	MN936056
357					1	16					T		NN	MN936057
357					1	16					T		SL	MN936058
357					1	16					T		CO	MN936059
357					1	16					T		KNC	MN936060
357					1	16					T		KNC	MN936061
357					1	16					T		KNC	MN936062
357					1	16					T		KNC	MN936063
379		Δ	A		1	18							HI	MN936064
379		Δ	A		1	18							HI	MN936065
381					1	18					T		HF	MN936066
381					1	18					T		CO	MN936067
381					1	18					T		CO	MN936068
381					1	18					T		CO	MN936069
381					1	18					T		JF	MN936070
381					1	18					T		CO	MN936071
405					1	20					T		SL	MN936072
405					1	20					T		SL	MN936073
417[b]					1	21					T		YO	MN936074
443					15	8							CO	MN936075
443					15	8							KNC	MN936076
443					15	8							KNC	MN936077
465[b]					1	25						A	YO	MN936078
474					22	3							NW	MN936079
474					22	3							NW	MN936080
474					22	3							NW	MN936081
489[a]					1	27						A	HI	MN936082

[1]R13 repeat motifs: "ATGTCTTCTTTCT"; [2] number of R12 repeat motifs "TTCCTTCTTTCT", [3] population used for sequencing, KNC = Korean native chicken, JF = Bangladesh red jungle fowl; NN = Naked neck chicken, CO = commercial broiler, SL= Sri Lankan native chicken, NW= Korean native white line, HI= Bangladesh Hilly chicken, CD= Bangladesh common chicken, AS = Aseel chicken, YO= Yeonsan Ogye, HF= Hanhyup F line. [a]Private allele: the allele only appeared in one breed; [b]New allele detected in Korean native chicken. [4]GenBank accession number for the current study submitted to GenBank (accession number from MN936024 to MN936082).

ACCEPTED

411 **Table 2**. Unique LEI0258 fragment sizes and their respective R13-R12 repeat combinations in Asian, African, and
412 North American chicken breeds.

413

[1]Asian chicken Allele size (bp) (R13-R12)	[2]African chicken Allele size (bp) (R13-R12)	[3]North American & commercial chicken Allele size (bp) (R13-R12)
206 (1-4)	192 (1-3)	403 (1-20)
231 (1-7)	200 (1-4)	487 (23-3)
235 (1-6)	204 (1-4)	513 (25-3)
241 (1-7)	234 (1-6)	539 (27-3)
292 (8-3)	250 (1-7)	565 (29-3)
317 (9-4)	256 (1-8)	
318 (10-3)	284 (1-10)	
320 (1-13)	294 (1-11)	
331 (1-14)	303 (1-12)	
343 (1-15)	304 (1-12)	
344 (12-3)	332 (1-14)	
355 (1-16)	391 (1-19)	
421 (17-4)	392 (1-19)	
432 (16-6)	416 (1-21)	
433 (17-5)	428 (1-22)	
445 (17-6)	429 (1-22)	
458 (18-6)	431(16-6)	
465 (1-25)	452 (1-24)	
477 (1-26)	512 (25-3)	
483 (19-7)		
484 (20-6)		
489 (1-27)		
501 (1-28)		

414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440

441 [1] allele size and repeat motifs identified from the sequence that has been submitted to the NCBI database [7,13] and
442 current study (GenBank accession, MN936024 - MN936082)
443 [2,3] allele size and repeat motifs identified from the sequence that has been submitted to the NCBI database
444 [6,15,16,37]

ACCEPTED

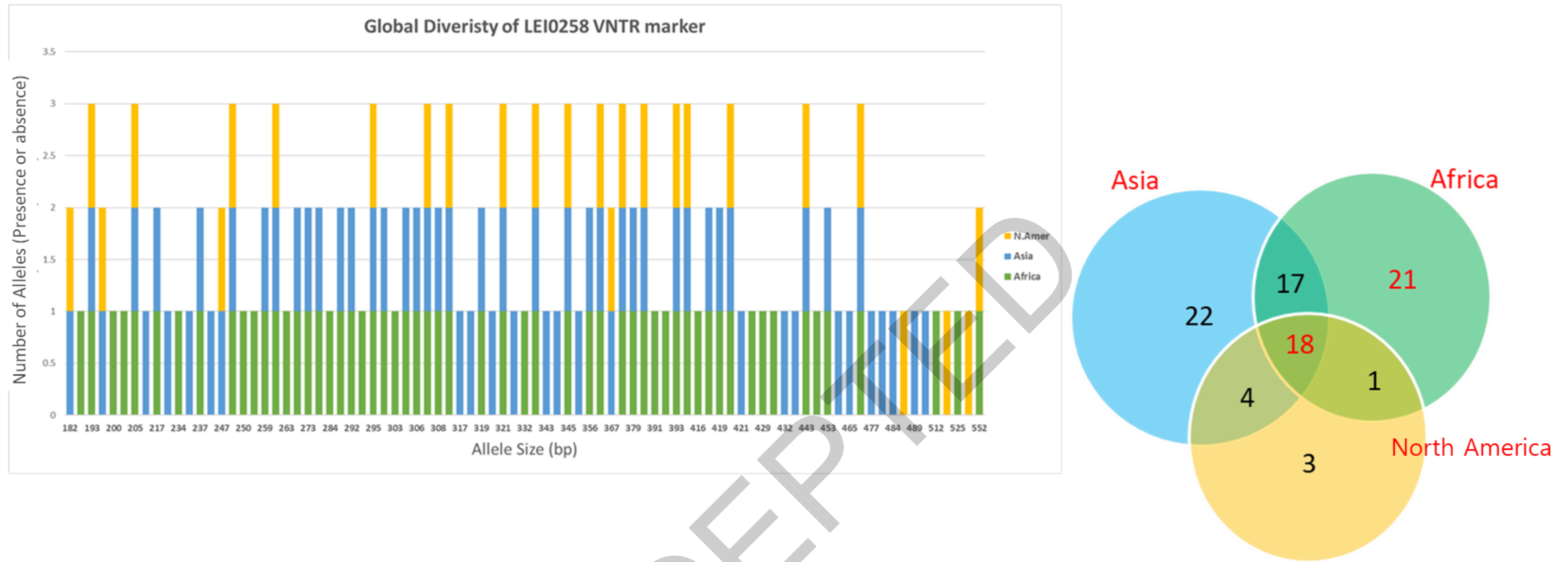
445 **Table 3**. LEI0258 allele range and VNTR (R13-R12) repeat motif combinations in Asian, African, and North American chicken populations.

Chicken population	Allele range (bp)	[1]R13 repeat motif	No. of different R13 Repeat motif	[1]R12 repeat motif	No. of different R12 Repeat motif
Red Jungle Fowl	193-489	1, 8-10, 12, 17-20	09	3 -10, 13-15, 18, 26, 27	14
Asian chicken	182-552	1, 8-12, 15-20, 22	13	2-21, 24-28	25
North American	182-565	1, 15, 16, 22, 23, 25, 27, 28	08	2-5, 7-8, 11-20	16
African chicken	192-552	1, 9, 15, 16, 22, 25, 28	07	3-22, 24	21
All population	182-565	1- 29	29	2- 27	25

446 [1]All possible R13 repeat motif $(ATGTCTTCTTTCT)_n$ and R12 repeat motif $(TTCCTTCTTTCT)_n$

447 Detail information of VNTR (R13, R12) was obtained based on the 621 LEI0258 sequences from the NCBI database [5,6,7,15,16,26,27]

ACCEPTED



ACCEPTED

Figure 1. LEI0258 allele size distribution in Asian, African, and North American chicken populations (based on LEI0258 sequence information).

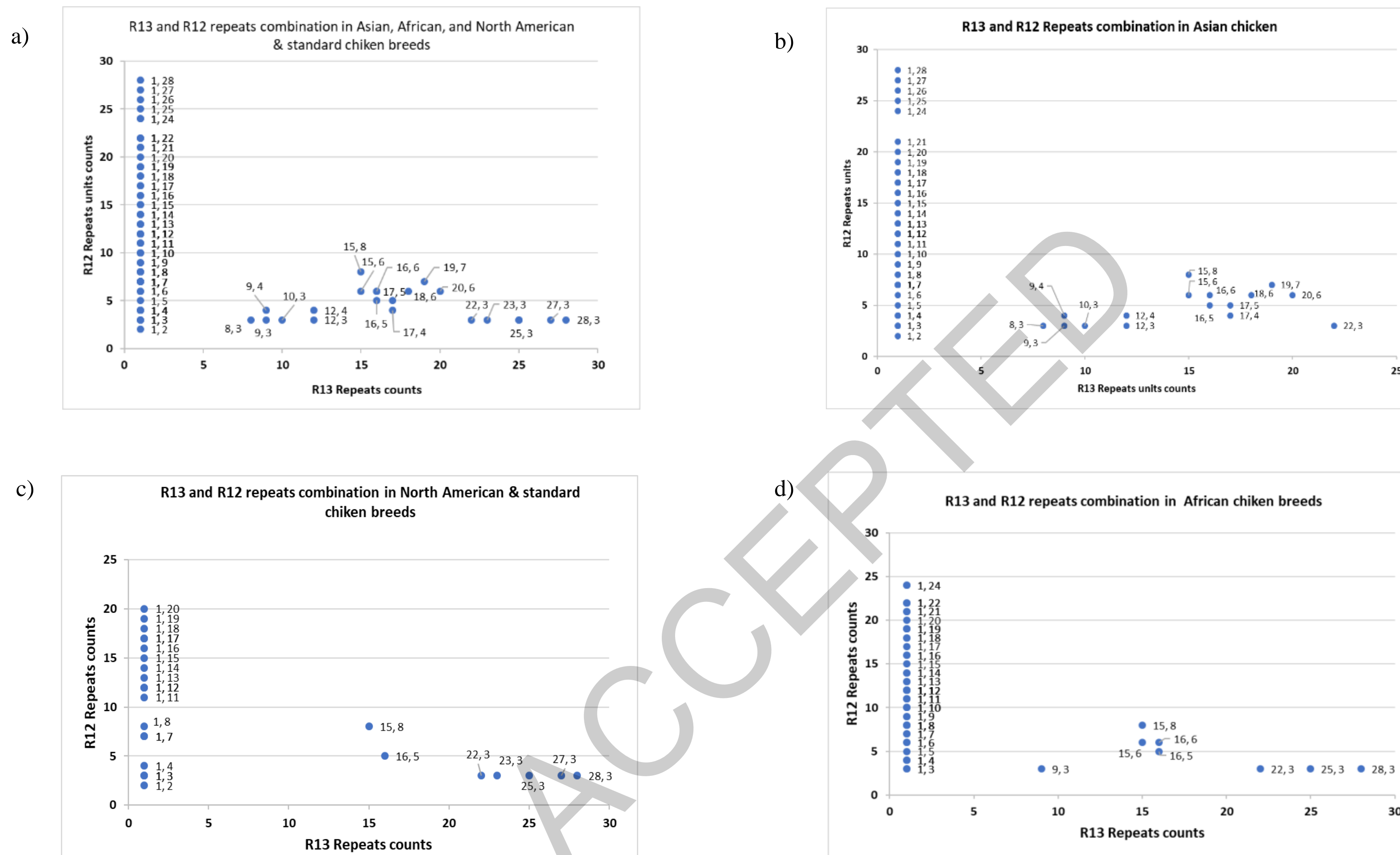
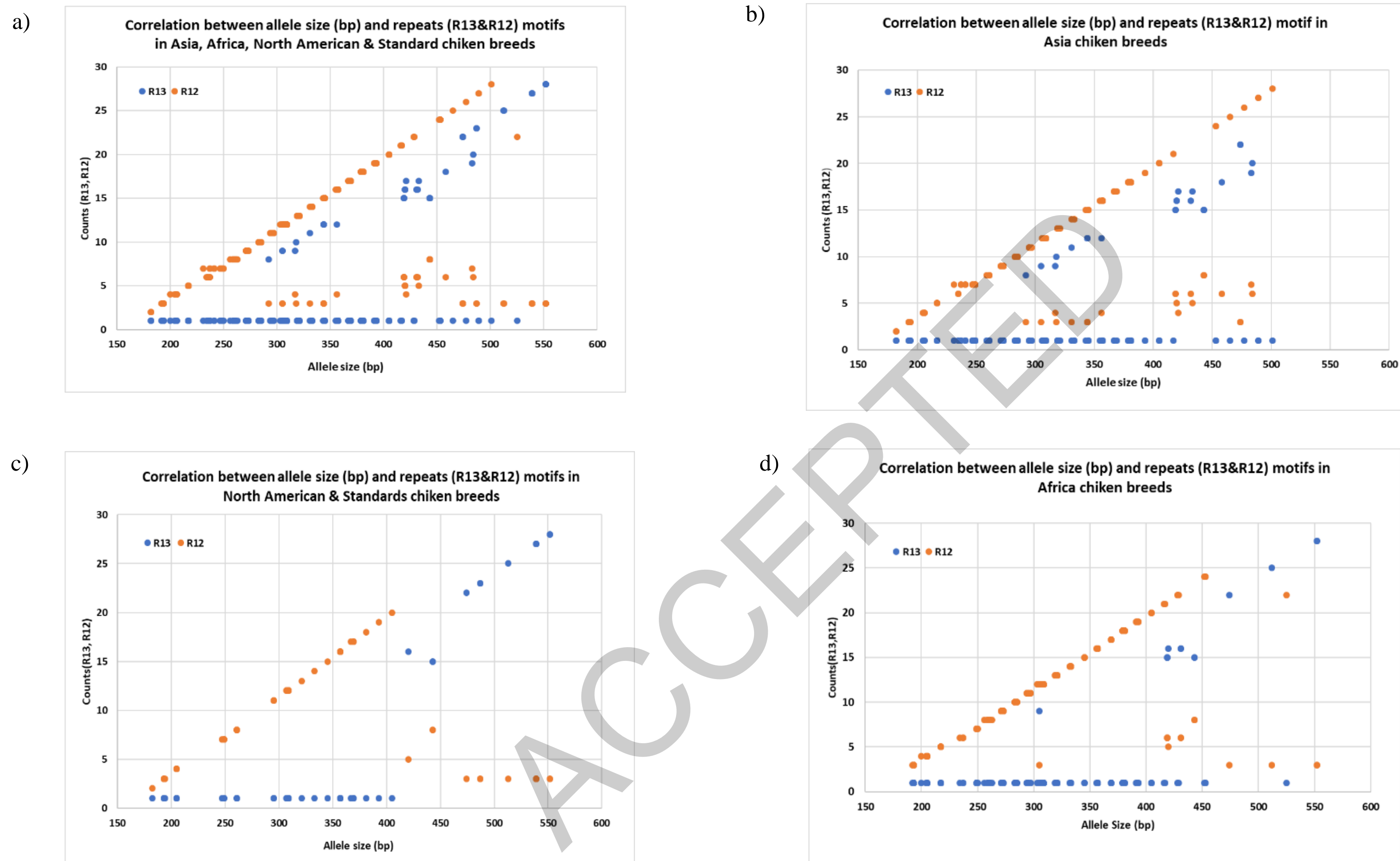


Figure 2. R13 and R12 repeat motif variations in chickens based on 621 sequences. The variation in a) all populations, b) Asian chicken breeds, c) North American and European Standard chicken populations, and d) African chickens. Unique combinations are observed in each population, although most variants are common to all three populations. The Asian population is distinct in exhibiting 16 unique combinations.

ACCEPTED



ACCEPTED

Figure 3. Correlations between R13 and R12 repeat motifs (y-axis) and allele size (x-axis). a) The 86 size fragments and their R13 and R12 repeat size variation in Asian, African, North American, and Standard chicken populations. b–d) Unique population patterns.

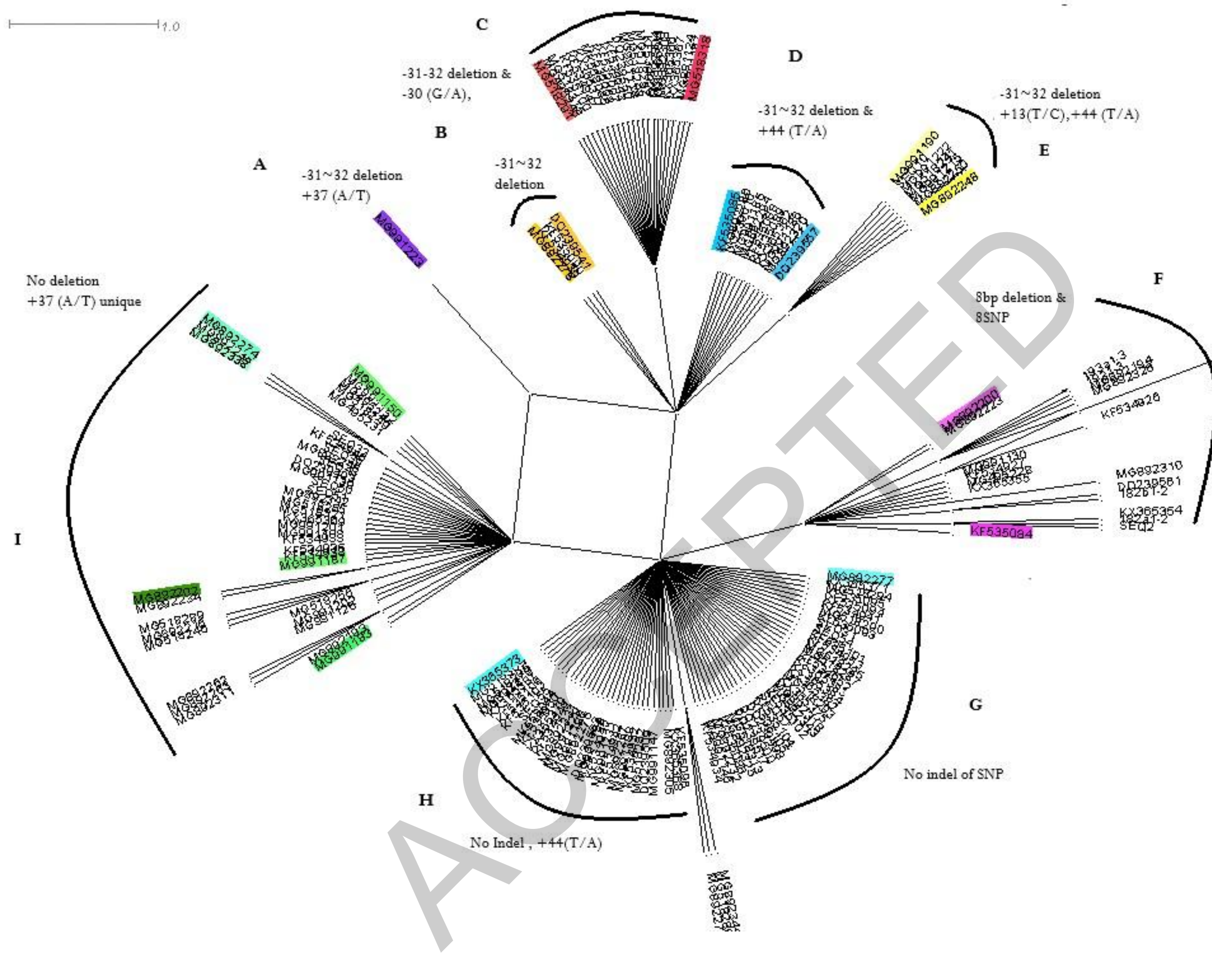


Figure 4. Neighbor-joining network for 175 allele sizes using single-nucleotide polymorphisms (SNPs) and insertions or deletions (indels) variation in the LEI0258 flanking region.

ACCEPTED