Next-generation Sequencing of MHC Class I Genes Reveals Trans-species Polymorphism in *Eutropis multifasciata* and Other Species of Scincidae

Shufang ZHANG1, Youfu LIN2, Yingzhi CHENG1, Haiyun YANG3, Xiaming ZHU1, Yu DU1,3, Longhui LIN4, Yanfu QU1, Lian CHEN2* and Hong LI1*

1 College of Life Sciences, Nanjing Normal University, Nanjing 210023, Jiangsu, China
2 College of Life Sciences, Nanjing Forestry University, Nanjing 210037, Jiangsu, China
3 Hainan Key Laboratory of Herpetological Research, College of Fisheries and Life Science, Hainan Tropical Ocean University, Sanya 572022, Hainan, China
4 Herpetological Research Center, College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou 311121, Zhejiang, China

**Abstract** The genes of the major histocompatibility complex (MHC) encode cell surface proteins that are essential for adaptive immunity. MHC genes show the most prominent genetic diversity in vertebrates, reflecting the adaptation of populations to their evolving environment, population survival and reproduction. In the present study, we used next-generation sequencing (NGS) to study the loci polymorphism of exon 3 of the MHC class I genes in an ovoviviparous skink, the many-lined sun skink, *Eutropis multifasciata* and five other species of Scincidae, to quantify genetic variation. In addition, we genotyped the same MHC class I genes of *E. multifasciata* using clone sequencing, to directly compare the effectiveness of both analytical techniques for MHC genotyping. NGS detected 20 MHC class I alleles in *E. multifasciata*, and 2 to 15 alleles in the other five Scincidae species. However, clone sequencing detected only 15 of those MHC class I alleles in *E. multifasciata*. In addition, trans-species polymorphism of MHC class I genes was studied by constructing a phylogenetic tree using the gene sequences obtained by NGS. Phylogenetic analysis revealed that MHC class I alleles were shared among different species of Scincidae with trans-species polymorphism, and did not exhibit specific genealogical inheritance. These results have important implications for understanding polymorphism interspecies diversity in the MHC genes of Scincidae, and the evolution of the MHC more broadly.

**Keywords** *Eutropis multifasciata*, major histocompatibility complex, next-generation sequencing, Scincidae, trans-species polymorphism

1. Introduction

The immune system is a complex network of organs, cells and proteins which play vital roles in protecting an organism from pathogens (Elbers and Taylor, 2016). The major histocompatibility complex (MHC) is part of the sophisticated “adaptive immune system”; it encodes the molecules that are responsible for antigen recognition, and subsequent binding (Sommer, 2005). The MHC is divided into three classes (MHC class I, II and III) based on the location and function of the genes within it. MHC class I genes primarily encode regions which identify and respond to polypeptides of endogenous antigens (i.e., antigens that have been newly synthesized in antigen-presenting cells) and initiate the subsequent immune response (Neefjes et al., 2011). In order to bind to different antigenic peptides, the antigen-binding regions (exons 2 and 3) of MHC class I must be highly polymorphic. Polymorphisms
in MHC class I genes have been studied across the Cercopithecinae (Caskey et al., 2019), Lacertidae (Yuan et al., 2014) and Muscicapidae (Biedrzycka et al., 2017), generally using exon 3 as the typing target. Perhaps not surprisingly, individuals with high polymorphisms in MHC class I genes have stronger immune function in response to a wide range of pathogens (Radwan et al., 2020).

Within the jawed vertebrate genome, the evolutionary history of the MHC has varied across clades, resulting in the MHC consisting of a highly polymorphic, multigene family. Polymorphisms of MHC genes are created by individual adaptation in a population, resulting in the occurrence of multiple alleles for a single gene locus. As such, MHC genes are important markers for species adaptations, like developing resistance to parasites (Hacking et al., 2018) or acclimating to different geographical locations (Sagonas et al., 2019).

Another form of polymorphism in MHC genes is trans-species polymorphism (TSP), which is the occurrence of identical or similar alleles among closely related species, that have not resulted from convergent evolution or gradual cross-fertilization (Klein et al., 2007). In general, MHC genes are homologous among species, have high sequence polymorphism, and linkage disequilibrium among different loci. For example, using exon 3 of the MHC class I genes researchers reconstructed the phylogenetic tree of three Eremias skink species (E. przewalskii, E. brenchleyi, and E. multisellata). The resulting tree showed that the MHC class I gene lineages were not clustered by species, rather, they exhibited TSP across the three species (Yuan et al., 2014). In such cases, MHC alleles could be older than the species and would thus have sufficient time to accumulate multiple interallelic variations, resulting in the development of the TSP (Klein et al., 2007).

The genes characteristics of the MHC are well described in mammals (especially humans and mice), birds, amphibians and fish, but little is known about their structural features in reptiles (Arnold and Hämmerling, 1991; Jongema et al., 2019), leaving a large phylogenetic gap. To understand the wider evolution of the MHC, we must study the sequence characteristics of reptilian MHC genes. The reptilian family Scincidae is large and diverse, both morphologically and ecologically. Because of this, the Scincidae provide an ideal model system to obtain fundamental information about the evolution of MHC diversity and TSP across the wider family. The many-lined sun skink (Eutropis multifasciata) is a medium-sized, tropical lizard, native to China, India, Bangladesh and the Philippines (Du et al., 2012). It is ovoviviparous and a suitable species to investigate the proximate and ultimate reasons for variation in the reproductive patterns of reptiles (Sun et al., 2012). Populations of this skink are expected to decline dramatically in the future, due to the medicinal and culinary values of this species (Dung et al., 2011). In 2023, the State Forestry Administration of the People’s Republic of China recognized this threat and identified that E. multifasciata was of important economic and scientific value and recommended state protection as within China. Given this context, the study of the MHC in E. multifasciata also provides information for the conservation of endangered species.

When sequencing genes, researchers generally employ a protocol using the polymerase chain reaction (PCR) to amplify the corresponding alleles. Initially, the cloning and then sequencing technique uses the plasmid incompatibility principle to isolate heterozygous alleles and obtain haplotypic gene information, commonly used for the confirmation of new alleles, obtaining gene sequences and polymorphism analyses (Christ et al., 1994). However, this may not be the most suitable method for sequencing MHC genes. Previous studies have found that in many species sequencing MHC genes with PCR can cause gene duplication, copy number variation and co-amplification of alleles from different loci, making the determination of individual genotypes impossible (Babik, 2010). Fortunately, next-generation sequencing (NGS) is a massively parallel sequencing technology with high throughput, scalability and speed. This technology facilitates the accuracy and reproducibility of genotyping polymorphic gene families, such as the MHC, in non-model organisms (Grogan et al., 2016; Shaheenuzzaman et al., 2020).

In the present study, we use NGS to investigate exon 3 of the MHC class I genes in six skink species (E. multifasciata, P. chinensis, S. modesta, S. reevesi, S. indicus and S. incognitus). MHC genotyping is difficult due to the high degree of polymorphism between MHC gene sequences and the variation in allele numbers between species (Rekdal et al., 2018). In order to compare the ability of different techniques to detect MHC genes, we also used clone sequencing to detect exon 3 of the MHC class I genes in E. multifasciata. The results of this study have important implications for understanding polymorphism and interspecific diversity in the MHC genes of the Scincidae.

2. Materials and Methods

2.1. Sample collection and DNA extraction  We collected 14 E. multifasciata and 8 other lizards in China (Table 1). The samples were kept in the Reptile Laboratory at Nanjing Normal University. DNA was extracted from the muscle tissues of 22 lizard samples using the DNeasy Tissue Kit (Qiagen, Germany) according to the manufacturer’s protocol.

2.2. MHC genotyping by an amplicon-based NGS approach  The universal primers (Murphy et al., 2009) 2MHCF (5’-CAG CAG ATG TAT GGC TGT GA-3’) and 2MCHR (5’-GCA
GAT CTC CTC CAG GTA G-3) were used to amplify exon 3 of the MHC class I genes (corresponding to the α-2 region) in six species (E. multifasciata, Plestiodon chinesis, Scincella modesta, S. reevesii, S. indicus, and S. incognitus) and a unique barcode was added to the 5’ end of the primers to identify different individuals. The PCR amplification reaction system was 50 μL, including 10 μL Premix Taq, 1 μL of each primer (10 μmol/L), 2 μL template DNA (100 ng) and ddH₂O. The following cycle conditions were used for PCR amplification: 94°C for 5 min; 35 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 1 min; 72°C for 10 min. PCR products were mixed and purified with the QIAquick Gel Extraction Kit (Qiagen, Germany). The size and quality of the pooled amplicons were verified using the TruSeq® DNA PCR-Free Sample Preparation Kit (Illumina, USA) on the 2100 Bioanalyzer (Agilent, USA). The library was then sequenced on the Illumina NovaSeq and 250 bp paired-end sequences were obtained (Novogene Bioinformatics Technology, China).

The raw sequencing data contained a variety of error messages that interfered with the subsequent analysis, such as sequencing junctions, low-quality bases and undetected bases (expressed as N). Therefore, the data were filtered by four criteria: (1) filtering out splice sequences and barcode sequences from the reads; (2) splicing pairs of reads with overlaps using FLASH VI.2.7 (Magoč et al., 2011); (3) filtering the spliced data using QiIME (Caporaso et al., 2010) to remove chimeric sequences with high N or low quality; and (4) filtering out chimeric sequences from the spliced sequences. The Illumina data were then clustered using Vsearch (Rognes et al., 2018) to remove splice sequences and barcode sequences from the reads; (2) splicing pairs of reads with overlaps using FLASH VI.2.7 (Magoč et al., 2011); (3) filtering the spliced data using QiIME (Caporaso et al., 2010) to remove chimeric sequences with high N or low quality; and (4) filtering out chimeric sequences from the spliced sequences. The Illumina data were then clustered using Vsearch (Rognes et al., 2016) and amplicons with a frequency of less than 1% were discarded. We only considered sequences that were present in more than one individual to be alleles (Zhao et al., 2013).

2.3. MHC genotyping by clone sequencing Universal primers 2MHCF and 2MHCR were used to amplify exon 3 of MHC class I genes in E. multifasciata. PCR reaction procedures and systems were as described above. Amplified PCR products were then separated using agarose gel electrophoresis, and the amplified bands of 14 individuals of E. multifasciata were excised and cleaned (MiniElut RNA-Pure Kit, Genebase Biogene Technology). The purified PCR product was ligated to the pMD19-T vector (TaKaRa, Japan) and transformed into DH5α receptor cells. PCRs were then performed for positive clones. Each individual then had 10 colonies and the bacterial solution containing the target fragment was sent to Shanghai Sangon Biological Engineering Technology & Services Co., Ltd. for sequencing.

The initial quality and length filtering of raw data was performed with ChromasPro (v 2.1) to output FASTA files. The target sequences were obtained by removing redundant sequences using the MEGA X (Kumar et al., 2018). The sequences obtained from clone sequencing were imported into NCBI (https://www.ncbi.nlm.nih.gov/) for BLAST to identify MHC class I genes. The sequences were imported into MEGA X and aligned using Clustal W (Larkin et al., 2007). Haplotype information for MHC sequences was obtained by DnaSP 5.0. MHC class I alleles were named with reference to the naming principle: a new allele needed to be identified by its occurrence in at least two clones of the same individual or in different individuals (Kennedy et al., 2002) and named sequentially according to the frequency of allele occurrence.

2.4. Phylogenetic analysis The sequences from clone and next-generation sequencing were imported into DnaSP 5.0 to calculate the nucleotide diversity (π) and variable sites (S) of MHC class I genes. The TCS Network algorithm in Popart (Hayes et al., 2022) was used to construct separate network relationships between the diverse haplotypes of Scincidae. The exon 3 sequences of MHC class I derived from 6 species in the present study combined with other 4 species (Ctenophorus decresii, Eremias multiocellata, E. przewalskii, and E. brenchleyi)
were used for phylogenetic analysis (Supplementary Table S1 for GenBank accession numbers). The sequences were aligned using MAFFT implemented in PhyloSuite (Zhang et al., 2020). The phylogenetic tree was created by the maximum-likelihood (ML) method implemented in IQ-TREE with a bootstrap of 1,000 replicates. The phylogenetic tree was visualized with the iTOL online tool (https://itol.embl.de/).

3. Results

In this study, the exon 3 sequences of MHC class I genes were successfully amplified in 22 samples from *E. multifasciata*, *P. chinensis*, *S. modesta*, *S. reevesii*, *S. indicus*, and *S. incognitus*.

3.1. Comparison of clone sequencing and NGS

The nucleotide sequences of all alleles identified by sequencing did not reveal a stop codon, indicating that these alleles all encode functional proteins. A total of 20 alleles were identified in 14 individuals. According to the naming principle of MHC class I functional proteins. A total of 20 alleles were identified in 14 individuals. According to the naming principle of MHC class I functional proteins, 20 MHC class I sequences were named *Eumu-UA*01 to *Eumu-UA*20 (Table 2). Fifteen alleles were identified by clone sequencing and all of these alleles were also detected by NGS. Furthermore, clone sequencing identified two alleles that were not detected by NGS. Table 2 shows the allele frequency of MHC I gene exon 3 of *E. multifasciata* based on clone sequencing and NGS.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Clone sequencing</th>
<th>NGS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of observed individuals</td>
<td>Allele frequency</td>
</tr>
<tr>
<td><em>Eumu-UA</em>01</td>
<td>6</td>
<td>0.125</td>
</tr>
<tr>
<td><em>Eumu-UA</em>02</td>
<td>3</td>
<td>0.063</td>
</tr>
<tr>
<td><em>Eumu-UA</em>03</td>
<td>5</td>
<td>0.104</td>
</tr>
<tr>
<td><em>Eumu-UA</em>04</td>
<td>4</td>
<td>0.083</td>
</tr>
<tr>
<td><em>Eumu-UA</em>05</td>
<td>2</td>
<td>0.042</td>
</tr>
<tr>
<td><em>Eumu-UA</em>06</td>
<td>4</td>
<td>0.083</td>
</tr>
<tr>
<td><em>Eumu-UA</em>07</td>
<td>6</td>
<td>0.125</td>
</tr>
<tr>
<td><em>Eumu-UA</em>08</td>
<td>1</td>
<td>0.021</td>
</tr>
<tr>
<td><em>Eumu-UA</em>09</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><em>Eumu-UA</em>10</td>
<td>1</td>
<td>0.021</td>
</tr>
<tr>
<td><em>Eumu-UA</em>11</td>
<td>4</td>
<td>0.083</td>
</tr>
<tr>
<td><em>Eumu-UA</em>12</td>
<td>3</td>
<td>0.063</td>
</tr>
<tr>
<td><em>Eumu-UA</em>13</td>
<td>2</td>
<td>0.042</td>
</tr>
<tr>
<td><em>Eumu-UA</em>14</td>
<td>3</td>
<td>0.063</td>
</tr>
<tr>
<td><em>Eumu-UA</em>15</td>
<td>2</td>
<td>0.042</td>
</tr>
<tr>
<td><em>Eumu-UA</em>16</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><em>Eumu-UA</em>17</td>
<td>2</td>
<td>0.042</td>
</tr>
<tr>
<td><em>Eumu-UA</em>18</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><em>Eumu-UA</em>19</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><em>Eumu-UA</em>20</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

samples with 3 to 7 alleles, with half of the samples having 3 alleles and only one sample having 7 alleles. The number of alleles detected by NGS ranged from 4 to 14, with four samples containing nine alleles, five samples containing more than ten alleles, and five samples containing fewer than nine alleles (Supplementary Table S2).

3.2. Class MHC class I genes sequence polymorphism

There were some differences in nucleotide diversity among different species. The number of variable sites detected for *E. multifasciata* was 90, with a nucleotide diversity of 0.101. No singleton variable sites were presented, but 90 parsimony informative sites were detected (Table 3).

In the other five Scincidae species, the number of variable sites ranged from 39 to 97, and the nucleotide diversity was 0.103–0.232. There were no singleton variable sites in exon 3 of the MHC class I genes in *S. modesta*, and the value of the parsimony informative sites was 54. No parsimony informative sites were present in *S. reevesii*, and the value of the singleton variable sites was 39. Three skinks (*P. chinensis*, *S. indicus*, and *S. incognitus*) had both singleton variable sites and parsimony informative sites. A total of 15 alleles were present in *S. indicus*, 13 alleles were present in *S. indicus*, and 10 alleles were present in *P. chinensis*. *S. modesta* and *S. reevesii* each contained at least two allelic loci (Table 3).

3.3. MHC class I genes trans-species polymorphism

There were eight alleles (*Eumu-UA*01, *Eumu-UA*02, *Eumu-UA*05, *Eumu-UA*06, *Eumu-UA*07, *Eumu-UA*08, *Eumu-UA*11, and *Eumu-UA*14) shared with four species (*E. multifasciata*, *S. indicus*, *S. incognitus*, and *P. chinensis*). *E. multifasciata* and *S. indicus* shared 13 alleles. *E. multifasciata* and *P. chinensis* shared 10 alleles. No shared alleles were found among *E. multifasciata*, *S. modesta*, and *S. reevesii*. The *E. multifasciata* had five unique alleles (*Eumu-UA*09, *Eumu-UA*10, *Eumu-UA*15, *Eumu-UA*16, and *Eumu-UA*18), and the *S. indicus* only had one allele *Spinc-UA*01 (Figure 1).

The evolutionary relationship of MHC class I was explored by constructing a phylogenetic tree of the sequences of exon 3 from ten lizard species (Figure 2). It was found that the MHC class I exon 3 sequences of the *Eremias* lizards (*E. multiiocellata*, *E. przewalskii*, and *E. brevichelya*) clustered into one large branch, while those of *C. decresii* clustered into a single branch. However, the MHC class I exon 3 sequences of the six lizards in the family Scincidae were not clustered into one branch according to the family. Although the phylogenetic tree formed sub-branches of Scincidae, Lacertidae and Agamidae, it did not show obvious phylogenetic characteristics, and there was TSP in the exon 3 sequence of MHC class I genes.
Table 3 Genetic diversity of MHC class I genes in 6 species of Scincidae.

<table>
<thead>
<tr>
<th>Species</th>
<th>Allele numbers</th>
<th>Nucleotide diversity</th>
<th>Variable sites</th>
<th>Singleton variable sites</th>
<th>Parsimony informative sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. multifasciata</td>
<td>20</td>
<td>0.101</td>
<td>90</td>
<td>0</td>
<td>90</td>
</tr>
<tr>
<td>P. chinensis</td>
<td>10</td>
<td>0.103</td>
<td>46</td>
<td>20</td>
<td>26</td>
</tr>
<tr>
<td>S. modesta</td>
<td>2</td>
<td>0.214</td>
<td>54</td>
<td>0</td>
<td>54</td>
</tr>
<tr>
<td>S. indicus</td>
<td>13</td>
<td>0.206</td>
<td>97</td>
<td>25</td>
<td>72</td>
</tr>
<tr>
<td>S. incognitus</td>
<td>15</td>
<td>0.169</td>
<td>93</td>
<td>31</td>
<td>62</td>
</tr>
<tr>
<td>S. reevesii</td>
<td>2</td>
<td>0.232</td>
<td>39</td>
<td>39</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 1 Median-joining network of MHC class I exon 3 alleles found in Scincidae. Each circle represents a haplotype, and the size of the circle indicates the number of samples contained in that haplotype. The line between two circles indicates that the two haplotypes are related, and the short line on the line indicates the number of nucleobase substitutions required to change from one haplotype to another.
4. Discussion

4.1. Characteristics of MHC class I genes in *E. multifasciata* The MHC plays an important role in the immune response of vertebrates. The inherent level of genetic variation and function within a species' MHC underlies disease resistance, directly affecting the survival of individuals and even the fitness of entire populations (Minias et al., 2016). Additionally, the level of genetic variation in MHC genes can be an important indicator for evaluating the evolutionary potential of adaptation in endangered wildlife (Akiyama et al., 2017). Genetic diversity within the MHC is reflected in allelic and nucleotide diversity, as well as in the number of MHC genes. In the present study, amplicon-based NGS yielded a total of 20 MHC class I alleles in *E. multifasciata*. In comparison with other reptile families, the number of MHC class I alleles in the ornate dragon lizard (*C. ornatus*) was close to that of *E. multifasciata*, ranging from 3–14 (Radwan et al., 2014). In contrast, the Tuatara (*Sphenodon spp.*) possessed 49 MHC class I alleles (Miller et al., 2010), which was significantly higher than the *E. multifasciata*. However, only six MHC class I alleles were found in a more closely related skink species *C. decresii* (Yuan et al., 2014), which was lower than that of *E. multifasciata* in the present study. The main driver of genetic diversity within the MHC is pathogen-mediated selection, but other evolutionary processes (e.g., genetic drift and gradual infiltration) may also increase genetic diversity within the MHC (Sagonas et al., 2019).

4.2. Comparison of MHC typing using clone sequencing and next-generation sequencing To understand the functional significance of allelic diversity in the MHC within the context of evolutionary ecology, pathogen resistance and conservation, it is imperative that the MHC composition of individuals is correctly assessed. Several methods have been used to type MHC genes, such as
denaturing gradient gel electrophoresis (DGGE); single-strand conformation polymorphism (SSCP); reference strand-mediated conformational analysis (RSCA); and direct sequencing after PCR purification (Lenz et al., 2008). The most widely used of these methods are clone sequencing and next-generation sequencing, which have their own advantages and limitations. The number of classical MHC genes may vary intra- or inter-species (de Groot et al., 2015; Plasil et al., 2022). The traditional method is to amplify individual samples using primers and pool the amplification products from the individual samples for sequencing. However, the MHC genes of many species exhibit gene duplications, copy number variations, and common alleles from different loci, making it impossible to determine individual genotypes using traditional methods (Babik, 2010). In addition, clone sequencing can only be performed on small batches of samples, which is cumbersome and less sensitive, i.e., susceptible to allelic deletions. Clone sequencing is also limited by the size of the cloning vector, the larger the amplified fragment, the more difficult it is to clone. NGS technology is easy to use and powerful, it detects large samples of MHC alleles concurrently during typing. However, it is difficult to obtain complete sequences and usually targets one or more exons (Cheng et al., 2022).

In the present study, the sequences of exon 3 of the MHC class I genes of E. multiformis were sequenced using conventional clone and next-generation sequencing techniques, respectively. A total of 15 MHC alleles were detected by both clone sequencing, and NGS techniques, however NGS also detected an additional five new MHC alleles. This demonstrates that NGS is superior for sequencing highly polymorphic MHC genes, and can thus reduce the risk of allelic loss caused by preferential amplification of fewer alleles. For example, Eumaleo-UA*01 was detected in only 6 individuals by clone sequencing, whereas it was detected in all 14 individuals by NGS. Therefore, NGS technology has clear advantages over traditional clone sequencing for genotyping of the MHC. This builds on previous studies utilizing NGS for MHC in other taxa. Genotyping of the MHC in Korean native chickens using long-range PCR in combination with NGS confirmed that NGS technology could be used effectively to detect MHC variants in avian fauna (Edirweera et al., 2022). Another study utilizing NGS to genotype the MHC of the Stokes skink Egeria stokesii revealed that 43 MHC I alleles were present in the offspring and two MHC I alleles were missing in its parents. This provides a useful reference for MHC genotyping in other taxa (Pearson et al., 2017). In addition, clone sequencing and next-generation sequencing were used to type the MHC genes of 28 goat antelope, Rupicapra rupicapra, showing differences in MHC typing between the two methods in 25% of the individuals and again confirming the superior ability of next-generation sequencing to detect highly polymorphic genes (Stipoljev et al., 2020).

However, NGS methods can be prone to sequencing errors and false positives, i.e., false “alleles”. To minimize the occurrence of false positives, sequencing depth and coverage can be increased, thus avoiding errors generated by sequencing, and additional bioinformatics methods can be developed to address these issues (Cheng et al., 2022; Sebastian et al., 2018). Therefore, the use of NGS technology to obtain MHC gene sequences should take the corresponding sequencing depth into account. In the present study, NGS was used based on the amplicon principle, i.e., using a high sequencing depth to exclude sequencing errors (the average amplicon reads were 62,533 base pairs) to ensure genotype accuracy. In addition, the deeper the sequencing depth, the higher the probability of detecting alleles, which could compensate for the limitations of difficult allele amplification in clone sequencing (differences in amplification efficiency between alleles) (Biedrzycka et al., 2017).

4.3. Intraspecific and trans-species polymorphisms of MHC class I genes MHC genes form a highly polymorphic gene family with a higher rate of variable sites than other genes, and most animal MHC genes have genetic variability (Kohn et al., 2006). However, reptiles have low levels of genetic diversity in their MHC genes. The nucleotide diversity of different populations of Podarcis erhardii ranged from 0 to 0.00564 (Hurston et al., 2009). The nucleotide diversity of three populations in the southern Flinders Ranges of E. stokesii ranged from 0.151 to 0.163 (Pearson et al., 2017). The π value of the MHC allele for 208 Sphenodon punctatus from 12 island populations was 0.152 ± 0.013. Compared to the above lizards, the π value of E. multiformis was lower (0.101). The level of genetic diversity within a species can predict its evolutionary potential and viability (Guo et al., 2019). The higher the level of genetic diversity of the species, the smaller the impact of external environmental change, and the stronger the ability to survive (Pauls et al., 2013). Multiple factors can influence the level of genetic diversity within a species, such as environmental variables, genetic drift within populations, bottleneck effects, inbreeding and life history traits, and diversity in a population. (Raposo Do Amaral et al., 2013).

The six species of lizards in this study are listed as threatened under three kinds of animal protection laws for the state, and S. indicus is assessed as a near-threatened species. Because of these classifications, and the vulnerability of these species, our sample sizes were small. The sample size may have affected the nucleotide diversity of the alleles to some extent. However, results generated from this study could help us gain
a simple understanding of the level of genetic diversity in a given population, with implications for the conservation and management of particular populations.

Both the allele network and the phylogenetic tree showed that alleles were shared among the six lizard species in the family Scincidae, and they were not clustered by family in the phylogenetic tree. This indicated the existence of TSP in the MHC class I genes. This TSP could have arisen because the allelic lineage emerged before the species was formed and was retained after the species became established (Klein et al., 1998). It could also be possible that the six species have overlapping ranges, with hybrid gradual infiltration between them (Guo et al., 2011), and that the evolutionary relationships exhibited by their MHC class I genes may be the result of hybridization. The phenomenon of TSP in MHC locus allele lineages has been reported in many vertebrates. For example, a phylogenetic tree constructed from exon 3 of the MHC class I genes of the oviparous E. brenchleyi, ovoviviparous E. multiocellata and E. przewalskii showed that the MHC genes of the three species did not form their own genetic lineages according to species and did not show specific evolutionary lineage inheritance (Yuan et al., 2014). The phylogenetic tree showed that there was no interspecific mixing between the MHC class I alleles of E. multiocellata and the three species of the genera Eremias. It is therefore hypothesized that the divergence between the family Scincidae and Lacertidae occurred before the allelic lineage was formed. In addition, the phylogenetic relationships between the MHC I genes of E. multiocellata and other lizards that are closely related were also studied, and alleles from each class did not fit neatly into several branches. This suggested that either the shared MHC class I genes of E. multiocellata had recently undergone too much gene duplication to exhibit homologous clustering of alleles, or that the alleles had not evolved independently after gene duplication (Winternitz et al., 2023). Some MHC class I alleles exhibit a greater degree of phylogenetic similarity between species than within species. This can be common for immune genes in general, which are subject to parasitemediated balancing selection and maintain allelic diversity over time (Těšický and Vinkler, 2015). Convergent evolution, resulting from similar selective pressures, can also cause shared polymorphisms among independent evolutionary lineages (Kriener et al., 2000; Srithayakumar et al., 2012).

5. Conclusions

In most animals, the major histocompatibility complexes are polymorphic and polygenic, which makes genotyping problematic. This study demonstrated that when sequencing highly variable regions, such as the MHC, NGS technology performs better than clone sequencing. The level of genetic diversity of the MHC genes in E. multiocellata is low. In addition, the exon 3 sequences of the MHC gene of six species of lizards in the family Scincidae display TSP. This study contributes to our understanding of the MHC structure in lizards of the family Scincidae, and explores methods suitable for MHC genotyping in reptiles, while providing a reference for future studies on reptile genetic diversity.

Data availability The raw data presented in the study have been uploaded on the SRA database from NCBI under the BioProject PRJNA962605.

Acknowledgements We would like to thank Wenjia SHEN for her lab assistance during the research. We thank Georgia WARD-FEAR and Zheng LI for their constructive comments on the revised version. This work was supported by grants from the National Natural Science Foundation of China (32171495 and 31971414), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

References

rhesus macaques: Ancestry and plasticity. Immunol Rev, 267: 228–245
Hacking J. D., Stuart-Fox D., Godfrey S. S., Gardner M. G. 2018. Specific MHC class I supertype associated with parasite infection and color morph in a wild lizard population. Ecol Evol, 8: 9920–9933
Lenz T. L., Becker S. 2008. Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci—Implications for evolutionary analysis. Gene, 427: 117–123
Radosavljevic M., Kodak S., Cvetkovic D., Nikolai S., Milić B., Demir H. 2016. Improvement of the MHC-I and MHC-II typing methods in Bos taurus. BMC Genomics, 17: 204
Sommer S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool, 2: 16
evolutionary convergence at MHC in two broadly distributed mesocarnivores. Immunogenetics, 64: 289–301
Winternitz J., Chakarov N., Rinaud T., Ottensmann M., Krüger O. 2023. High functional allelic diversity and copy number in both MHC classes in the common buzzard. BMC Ecol Evol, 23: 24

How to cite this article: