MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2012) **12**, 285–292 doi: 10.1111/j.1755-0998.2011.03082.x

Evaluation of two approaches to genotyping major histocompatibility complex class I in a passerine—CE-SSCP and 454 pyrosequencing

MARTA PROMEROVÁ,* WIESŁAW BABIK,† JOSEF BRYJA,* TOMÁŠ ALBRECHT,*‡ MICHAŁ STUGLIK† and JACEK RADWAN§

**Institute of Vertebrate Biology, Academy of Sciences of the Czech Republic, Květná 8, 60365 Brno, Czech Republic, †Institute of Environmental Sciences, Jagiellonian University, Gronostajowa 7, 30387 Kraków, Poland, ‡Department of Zoology, Faculty of Sciences, Charles University in Prague, Viničná 7, 12844, Prague, Czech Republic, §Institute of Nature Conservation, Polish Academy of Sciences, Al. Mickiewicza 33, 31-120 Kraków, Poland*

Abstract

Genes of the highly dynamic major histocompatibility complex (MHC) are directly linked to individual fitness and are of high interest in evolutionary ecology and conservation genetics. Gene duplication and positive selection usually lead to high levels of polymorphism in the MHC region, making genotyping of MHC a challenging task. Here, we compare the performance of two methods for MHC class I genotyping in a passerine with highly duplicated MHC class I genes: capillary electrophoresis-single-strand conformation polymorphism (CE-SSCP) analysis and 454 GS FLX Titanium pyrosequencing. According to our findings, the number of MHC variants (called alleles for simplicity) detected by CE-SSCP is significantly lower than detected by 454. To resolve discrepancies between the two methods, we cloned and Sanger sequenced a MHC class I amplicon for an individual with high number of alleles. We found a perfect congruence between cloning/Sanger sequencing results and 454. Thus, in case of multi-locus amplification, CE-SSCP considerably underestimates individual MHC diversity. However, numbers of alleles detected by both methods are significantly correlated, although the correlation is weak (r = 0.32). Thus, in systems with highly duplicated MHC, 454 provides more reliable information on individual diversity than CE-SSCP.

Keywords: avian, *Carpodacus erythrinus*, major histocompatibility complex, next-generation sequencing, scarlet rosefinch

Received 25 May 2011; revision received 30 August 2011; accepted 9 September 2011

Introduction

In all jawed vertebrates, classical major histocompatibility complex (MHC) genes encode trans-membrane proteins on cell surfaces, where these bind antigen peptides and present them to T cells for recognition and subsequent activation of the adaptive immune response (Klein 1986). A direct link has been found between MHC genes and health/disease in several model and nonmodel species (reviewed in Piertney & Oliver 2006; see also e.g. Evans & Neff 2009; Meyer-Lucht *et al.* 2010, Kloch *et al.* 2010). MHC genes have also been shown to play a role in mate choice and reproductive success (e.g. Yamazaki *et al.* 1976; Penn 2002; Schwensow *et al.* 2008; Eizaguirre *et al.* 2009; Agbali *et al.* 2010). Mainly for these reasons, they have been the subject of great interest in evolutionary and conservation biology over the last decades.

The MHC region contains highly dynamic and variable multi-gene families (Klein 1986). Frequent gene duplication, gene conversion and recombination, as well as the aforementioned pathogen-related positive selection often make MHC genotyping a tricky and time-consuming task. Especially, in some fish and passerine birds, genes of the MHC are highly duplicated, and in most cases, the different genes cannot be amplified separately (e.g. Westerdahl *et al.* 2004; Lenz *et al.* 2009; Promerová *et al.* 2009; Zagalska-Neubauer *et al.* 2010). In such cases, a special approach to genotyping is required. For a long time, the most popular methods for genotyping highly duplicated MHC genes were the denaturing gradient gel electrophoresis (DGGE), the single-strand conformation polymorphism (SSCP) and the reference strand-mediated conformational analysis (RSCA) (reviewed in Babik 2010).

Correspondence: Marta Promerová, Department of Population Biology, Institute of Vertebrate Biology AS CR, Studenec 122, 67502 Koněšín, Czech Republic, Fax: +420 568423121; E-mail: promerova@seznam.cz

However, none of these methods directly provide sequence information. If information on sequences is required to characterize particular variants (for simplicity called alleles hereafter, even if it is often not possible to assign them to a particular locus), these methods have to be combined with Sanger sequencing successive to either band extraction from gel or molecular cloning. Commonly observed preferential amplification of some alleles may further complicate genotyping, requiring very high numbers of clones to be sequenced to obtain complete information on sequence variation (e.g. Promerová *et al.* 2009; Čížková *et al.* 2010). With the development of next-generation sequencing, such as the 454 GS FLX Titanium pyrosequencing (hereafter referred to as 454), it is now possible to generate huge amounts of sequence data within a short time at competitive costs. Moreover, the currently available read length enables sequencing whole exons of MHC genes in a single pass (as a single sequencing read), thus providing directly phased sequence information (e.g. Galan *et al.* 2010; Zagalska-Neubauer *et al.* 2010).

Here, we compared the performance of two methods for genotyping of the third exon (encoding a part of the functionally important alpha 2 domain) of highly duplicated MHC class I in a passerine bird, the scarlet rosefinch (*Carpodacus erythrinus*). We evaluate the performance of capillary electrophoresis SSCP (CE-SSCP), which is commonly used for analysing MHC variability in nonmodel species (e.g. Bryja *et al.* 2005; de Bellocq *et al.* 2009; Baratti *et al.* 2010; Čížková *et al.* 2010; Promerová *et al.* 2011), and the recently developed, much debated massive parallel 454 pyrosequencing (Babik *et al.* 2009; Wegner 2009; Galan *et al.* 2010; Kloch *et al.* 2010; Zagalska-Neubauer *et al.* 2010). Applying both methods to the same data set facilitated the identification of their advantages and drawbacks, and hence their meaningful comparison.

Materials and methods

Samples

In this study, we used samples of 234 scarlet rosefinches from the Šumava Mountains National Park, Czech Republic (48°49′N, 13°56′E, ~750 m a.s.l.), from breeding seasons in 2000–2008, including 90 chicks from 19 nests (for more detailed information on the study site and population sampling, see Albrecht *et al.* 2009). From all birds, 20–30 μL of blood was taken by wing venipuncture and stored in 96% ethanol at −20 °C until DNA extraction. Genomic DNA was extracted using Dneasy Blood & Tissue Kit (Qiagen, Germany), according to the manufacturer's instructions.

CE-SSCP—laboratory procedures

After testing several primer combinations, to amplify the whole MHC class I exon 3, we used the primers HN34-forward (5′-CCATGGGTCTCTGTGGGTA-3′, Westerdahl *et al.* 2004) and CE-MHCREX3-reverse (5′-ACAGGAATTCTGCTCCCACC-3′, Promerová *et al.* 2009) located in conserved regions of the adjacent introns. Both primers were fluorescently labelled for CE-SSCP analysis (HN34 by 6-FAM and CE-MHCREX3 by NED); this way each allele was visualized by two peaks of different colours, facilitating reliable genotyping. PCR was performed in a final volume of 10 μL with 1.5 mM $MgCl_2$, 0.4 μM of each fluorescently labelled primer, 0.2 mM dNTPs, 1 U *Taq* polymerase (Fermentas) in 1× PCR buffer and 1 μL of extracted DNA. Amplification consisted of initial denaturation at 94 °C (3 min) followed by 30 cycles of denaturation at 94 °C (40 s), annealing at 59 °C (30 s) and extension at 72 °C (1 min), and a final extension step at 72 °C for 10 min.

We mixed 2 μL of the PCR product with 0.5 μL of GeneScan-500 LIZ™ size standard (Applied Biosystems) and 12 μL Hi-Di™ Formamide (Applied Biosystems), and denatured the mix at 95 °C for 3 min and then immediately chilled it in an ice-bath for 2 min. CE-SSCP electrophoresis was run at 22 °C in 5% nondenaturing conformation analysis polymer (CAP; prepared by dilution from 9% CAP polymer, Applied Biosystems, according to the manufacturer's instructions) on an ABI PRISM 3130 sequencer. Data were analysed using GeneMapper v3.7 (Applied Biosystems).

As repeating CE-SSCP analysis after a long period of time showed shifts in the peaks, probably due to changing polymer, buffer, capillaries, etc. (M. Promerová, unpublished data; D. Čížková, personal communication), all individuals were analysed within short time, and all laboratory work and subsequent peak determination was performed by the same person (MP). In cases with difficult peak determination, i.e. occurrence of unspecific peaks or peak shifting, samples were run once more and also dyads of parents and offspring within families were compared to ensure the most precise identification of alleles. For further details on CE-SSCP analysis, see also Promerová *et al.* (2009).

454—laboratory procedures

For 454 pyrosequencing, the forward primer started with the Titanium Primer_A sequence (5′-CGTATCGCCTCCCTCGCGCCATCAG-3′) followed by a 6-bp tag identifying the amplicon, and the HN34 primer sequence, resulting in a forward fusion primer of 50 bp. The reverse primer started with the Titanium Primer_B sequence (5′-CTATGCGCCTTGCCAGCCCGCTCAG-3′) followed

by the sequence of primer CE-MHCREX3, resulting in a reverse fusion primer of 45 bp. The 234 individuals were sequenced using 39 unique tags. To assess PCR artefacts and sequencing errors, 43 of these individuals (18.4%) were amplified twice, i.e. each one independently with two different tags (these are further called replicates).

PCR was performed in a 20-μL reaction volume, with 0.4 μM of each fusion primer, 0.2 mM dNTPs, 1.5 mM $MgCl_2$ and 1 U *Taq* polymerase (Fermentas) in 1 × PCR buffer. Amplification consisted of an initial denaturation at 94 °C (3 min) and 30 cycles of denaturation at 94 °C (30 s), annealing at 59 °C (30 s) and extension at 72 °C (40 s), followed by a final extension step at 72 °C for 7 min. PCR products were visualized by agarose gel electrophoresis (5 μL of PCR product with 1 μL of loading dye on a 1.5% gel), and 10 μL of individual PCR products labelled by unique tags were pooled together, forming eight different pools.

As we only used 39 tagged primers, the picotiter plate was divided into eight sections and section information for each sequencing read was recorded during data extraction. Each tagged primer might have been used up to eight times, independently for each section, reducing the costs considerably. Only few PCR products displayed weaker bands on agarose gel, and the double volume (20 μL) of these was used for pooling. The final pools were purified using MinElute PCR Purification Kit (Qiagen) and eluted in 50 μL elution buffer. All pools were sequenced using a part of a single sequencing run on 454 GS FLX Titanium Roche instrument, on a plate containing also MHC amplicons of other species, at the Functional Genomics Center, Uni/ETH Zürich, Switzerland.

To verify the reliability of our 454 genotyping, we cloned (using CloneJet™ PCR cloning kit, Fermentas, according to the manufacturer's instructions) the PCR product of MHC class I exon 3 for one individual with 12 alleles identified by 454, and we sequenced 192 clones (using the vector primer pJET1.2 Forward and Big Dye Terminator Sequencing mix v3.1, Applied Biosystems, according to a standard BDT protocol).

Quality assessment and identification of true alleles —CE-SSCP

Importantly, neither the CE-SSCP peaks nor the sequences of MHC class I exon 3 can be assigned to specific loci in this study (Promerová *et al.* 2009). When studying MHC genes, most attention is usually paid to variation among alleles. However, some artificial variation might be created during PCR and subsequent cloning, by nucleotide substitutions and *in vitro* recombination (chimera formation) between true alleles (e.g. Keohavong & Thilly 1989; Zylstra *et al.* 1998; Cummings *et al.* 2010). Although there are ways to reduce them (Lenz & Becker 2008), such artefacts may still arise at substantial frequency. The usual approach to exclude artefacts is to only accept alleles as true if these occurred in at least two independent PCRs, because the probability of the same error occurring in more than one PCR is low (but see Lenz & Becker 2008). To minimize the frequency of artefacts, we performed only a modest number of 30 PCR cycles for both genotyping methods.

To validate alleles visualized by CE-SSCP and to obtain sequences of these alleles, we cloned the products of 19 individuals and ran CE-SSCP for these clones (see also Promerová *et al.* 2009). The CE-SSCP peaks of the clones were then compared with the individuals' CE-SSCP patterns. Those clones which fitted the CE-SSCP pattern of the individuals were sequenced (as described earlier). Sequences were analysed in BioEdit (Hall 1999). This way, we were able to assign a CE-SSCP peak pattern to a particular sequence.

Quality assessment and identification of true alleles—454

Quality of the 454 output was first assessed using standard settings of the 454 software, and multifasta files were produced. To proceed with the sequence analysis from this point, we used the jMHC software available at http://code.google.com/p/jmhc/ (Stuglik *et al.* 2011). From the multifasta file, only those sequences were extracted which contained the complete forward primer and complete 6-bp tag, at least the first 10 bases of the reverse primer and no ambiguous bases ('N' calls) within the region of interest. The sequences contained also small parts of introns, but as these were of no interest for genotyping, only exon 3 sequences were extracted. Reads were sorted according to tags, and an alignment of sequence variants was created for each amplicon separately. All singletons (sequences present only once in the data set) were removed. As functional MHC alleles do not contain frameshift mutations, we assumed that all variants containing 1- or 2-bp insertions or deletions (indels) were generated either as PCR or sequence reading errors. Moreover, indels are known to be common in 454 reads, mostly if homopolymers are present in the target sequences (Moore *et al.* 2006; Huse *et al.* 2007). Therefore, all variants containing other than 3-bp indels were excluded from further analyses.

After these treatments, we performed a correlation test between coverage and number of alleles. As these were still correlated up to the minimal coverage of 140, we decided to remove all amplicons with coverage lower than 150 after these first treatments [70 of 277 amplicons (25.3%) were removed]. Subsequently, to distinguish true alleles from PCR or sequencing artefacts among the remaining sequences, we applied several cri-

teria first on the whole data set level and then on the per-individual level. These were adapted from Babik *et al.* (2009), Zagalska-Neubauer *et al.* (2010) and Galan *et al.* (2010). Following Galan *et al.* (2010), who showed that the probability of obtaining the same artefactual sequence three times is very low, we removed from our data set all sequences that did not reach three reads in any amplicon.

We calculated the maximum per amplicon frequency (MPAF) for each sequence variant in the whole data set, considering only amplicons with coverage at least 150. We sorted the sequence variants according to their MPAF and examined them from the bottom up. For each variant, we examined three (if available) amplicons, in which this sequence variant was most abundant. If the sequence variant could be explained as an artificial single-base substitution or recombinant of other alleles in all the examined amplicons, it was considered an artefact and removed. All variants with MPAF <1.5% were removed, because of the randomly chosen 20 sequences between 1% and 1.5% all could be classified as artefacts. The lowest identified MPAF for a true allele was 1.79% (corresponding to five reads in this case). The highest MPAF for an artefact was 6.3% (corresponding to 13 reads in this case). According to our findings, all variants with MPAF over 6.3% may safely be considered true alleles on the level of the whole data set.

It is still possible that some true alleles on the whole data set level occur as artefacts on per-individual level. This is mostly because of the fact that some true alleles might be recent recombinants of other alleles present in the population, and thus distinguishing them from PCR chimeras is difficult. Therefore, we applied a second filtering step on the per-individual level. For the MPAF value range in which both true alleles and artefacts may occur (i.e. up to 6.3%), we conservatively classified as artefacts all sequences that could be explained as a recombinant of two other alleles present in a given amplicon. Only alleles with more than two reads in an amplicon were considered.

Results and discussion

Variation detected by CE-SSCP

Using CE-SSCP, an approach routinely applied to genotype MHC genes, we obtained genotypes for all analysed individuals. We identified 83 alleles, of which we sequenced 32 by cloning/Sanger sequencing (these sequences translate to 31 unique amino acid sequences). To obtain these sequences, we cloned the PCR product of 19 individuals, and we screened over 600 clones by CE-SSCP (for more details see Promerová *et al.* 2009). Thus, obtaining sequences for all alleles detected by CE-SSCP is laborious and costly.

Variation detected by 454

All of the 32 sequences obtained by cloning and Sanger sequencing additional to CE-SSCP were also obtained by the 454. When applying 454 for genotyping, one of the first things to be taken into account is sufficient coverage, as low per-individual coverage may cause missing of some alleles. According to the software by Galan *et al.* (2010) that calculates the minimal required coverage for each amplicon, depending on the number of expected gene copies and the necessary number of sequences for each allele in an individual, we would need coverage over 100 reads when expecting four loci. The software only allows a maximum of eight alleles per individual, but we observed as much as 13 alleles in the most heterozygous scarlet rosefinches (Fig. 2).

In our final data set (coverage ≥150), the correlation between number of alleles and coverage was not significant ($P = 0.1$, $r^2 = 0.01$, $N = 207$), indicating that higher coverage does not imply higher number of alleles detected. The mean final coverage in our genotypes after the first filtering, i.e. removing all singletons and 1- or 2-bp indels, and removing amplicons with coverage lower than 150, was 301.38 ± 89.24 (SD) reads, with a minimum of 150 and a maximum of 683 reads. 454 pyrosequencing yielded genotypes for 77.8% of the individuals (182 of 234), with 189 different true nucleotide sequences (translating to 153 unique amino acid sequences), a much higher number of alleles than detected using CE-SSCP. This outcome is comparable to the results of Galan *et al.* (2010), who obtained 1407 DRB exon 2 genotypes of 1710 sequenced individuals of rodents (i.e. 82.3% genotyping success).

454 reliability of genotyping

To assess 454 genotyping errors, Babik *et al.* (2009) suggested running a fraction of replicates in every experiment. Galan *et al.* (2010) replicated <6% of their samples, and this was enough to assess the reliability of genotyping in rodents. In our study, we replicated 43 of 234 (18.4%) individuals. We obtained a final coverage of over 150 reads for both replicates in only 25 cases (58.14%). Of the 25 replicates, 20 displayed identical genotypes. Four of the five nonidentical replicates differed in the final number of alleles by only one allele, one of the replicate pairs differed by three alleles missing in one of the replicates. Thus, from a total count of 197 alleles detected in the 25 replicated individuals, seven alleles were not recovered twice (3.55% genotyping error). In all cases, the replicates with fewer recovered alleles had a lower

coverage, although not significantly (paired *t*-test: $P = 0.09, N = 5$).

To verify the high number of alleles found by 454, we cloned the PCR product of one individual displaying 12 alleles (this individual displayed seven alleles on CE-SSCP). We sequenced 192 clones, of which 150 yielded high-quality sequences (coverage comparable to 454). This cloning confirmed all 12 alleles of the individual detected by 454 with a read number ranging from four to 25 clones per allele. Additionally to these, Sanger sequencing of clones yielded 33 variants present in only one read each (singletons) and one sequence that occurred in five copies. None of the sequenced clones contained 1- or 2- bp indels. As to the Sanger-singletons, they probably resulted from PCR and cloning artefacts, commonly observed in MHC studies (Longeri *et al.* 2002; Lenz & Becker 2008), as 28 of them represented 1-bp substitutions of sequences occurring at higher frequencies and five represented recombinant chimeras. Indeed, none of these sequences were found among 454 results that passed the basal criterion of minimum three reads. In contrast, the variant that arose in five copies was found among 454 sequences that passed the basal three reads criterion. However, this variant was later filtered out from 454 genotyping, as it only achieved a low MPAF, and it differed by 1 bp from a more common true allele with which it co-occurred. Thus, this variant is probably a true allele that was considered a 1-bp substitution artefact. This would mean that our method is rather conservative and may sometimes be too strict in filtering true alleles from artefacts. However, with the high amount of artefacts that can be obtained by 454 sequencing, it is important to set strict thresholds, as the risk of overestimating actual polymorphism by retaining artefacts appears more important than that of an underestimation. Also, it is very important to achieve sufficient coverage to apply strict thresholds, while still enough reads remain for a robust genotyping after filtering.

So far, four studies have been published on the application of next-generation sequencing for MHC genotyping (Babik *et al.* 2009; Galan *et al.* 2010; Kloch *et al.* 2010; Zagalska-Neubauer *et al.* 2010). Each of these studies had to face different problems, and the data analyses, such as the filtering of true alleles from artefacts, are study-specific. Unlike Babik *et al.* (2009) and Galan *et al.* (2010), who had knowledge on most of the expected sequences gained from previous studies (Axtner & Sommer 2007; Babik & Radwan 2007), we only knew the sequences for a small fraction (32 of 189) of the present alleles. All four aforementioned studies analysed a fragment much shorter than presented here. Galan *et al.* (2010) sequenced 222 bp of DRB exon 2, Babik *et al.* (2009) and Kloch *et al.* (2010) sequenced a 120–129 bp part of the same exon, and Zagalska-Neubauer *et al.* (2010) studied 188 bp of the passerine exon 2 of MHC class IIB. However, we analysed a whole exon of 273–276 bp, and the longer the analysed sequence, the higher the expected error rate for the 454 system. The most common sequencing mistakes occurring in our 454 data were indels. In most MHC studies, researchers are interested in functional genes, so sieving out sequences that would cause frameshifts, even if they represented true pseudogenes, is a reasonable approach. It is, therefore, enough to only extract the sequences of the right length. However, the biggest problem when dealing with MHC class I in the scarlet rosefinch was that many of the alleles were extremely rare. Approximately 20% of all alleles only occurred in one individual (but at high coverage) or one family. Moreover, several of the true alleles differ by only one or few substitutions (see also Promerová *et al.* 2009) and may be missed because of strict filtering thresholds. However, such cases should be relatively rare. Indeed, our cloning/Sanger sequencing experiment indicated that apart from 12 alleles confirmed by both methods, only one additional allele (7.6% of the individual's variability) might have been missed under our stringent criteria. Thus, it seems that the method of excluding artefacts, developed by Zagalska-Neubauer *et al.* (2010), which was adapted in the present study, seems to work satisfyingly, and we propose that it should be adopted as a standard method of analysis in systems without prior knowledge of MHC sequences.

Comparison of performance

We compared the individual number of alleles obtained by CE-SSCP analysis with that of 454. They are significantly correlated (Spearman, $P < 0.01$, $r = 0.32$, $N = 182$; Fig. 1). As the maximal number of alleles detected by CE-SSCP was eight, we also performed the correlation for individuals displaying eight or less alleles on 454, and they are significantly correlated (Spearman, $P < 0.01$, $r = 0.264$, $N = 124$). However, the low r^2 suggests that these results are not in full congruence, and thus analyses using outcomes of CE-SSCP might be less powerful. On the per-individual level, the 454 results indicate a significantly higher number of alleles than CE-SSCP (Fig. 2; CE-SSCP: mean = 5.07 ± 1.04 (SD); 454: mean = 7.99 ± 1.63 (SD); Wilcoxon paired test: $P < 0.01$, $N = 182$). As we used the same primer combination for both analyses, it is unlikely that this is because of preferential amplification of more alleles during the 454 sample preparation.

The fact that CE-SSCP underestimates the actual MHC variability might be due to often indistinguishable patterns of peaks, i.e. different alleles displaying identical peaks, the same problem was already reported,

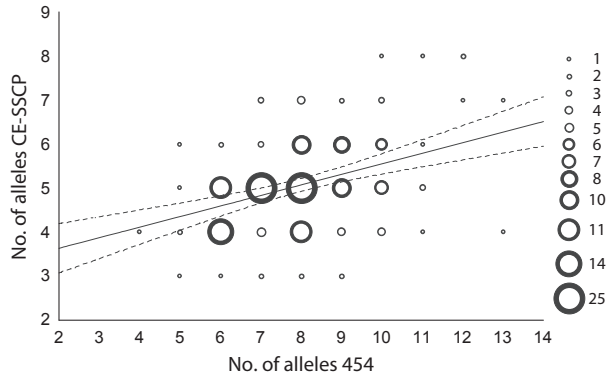

Fig. 1 Relationship between the number of alleles detected by CE-SSCP and 454 in the same individuals. Linear regression curve and 95% confidence interval shown (N = 182). The size of the circles corresponds to number of cases, as indicated by the legend on the right.

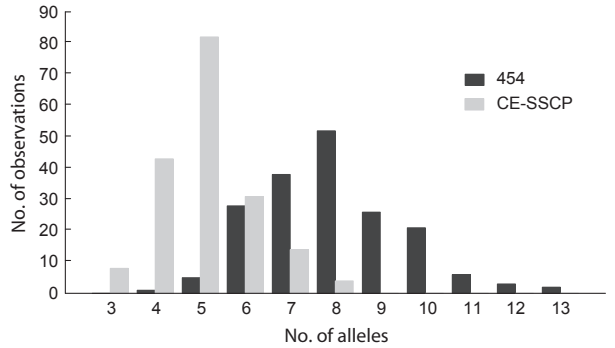

Fig. 2 Distribution of number of alleles per individual according to CE-SSCP and 454 in the scarlet rosefinch (N = 182).

e.g. by Bryja *et al.* (2005). In our data set, there was at least one allele which was sequenced several times, but was not detectable on the CE-SSCP electrophoretogram (Promerová *et al.* 2009). Alcaide *et al.* (2010) suggested solving these problems by running the same fragment amplified by several primer pairs simultaneously on SSCP, thus gaining more information for each allele. However, this is only feasible for less complicated patterns than passerine MHC, as in cases of locus-specific amplification.

Because using CE-SSCP might cause missing considerable parts of the allelic variability in populations with complex MHC genotypes, 454 seems to be a better choice for high-throughput studies. An advantage of CE-SSCP and similar classical methods over 454, for instance, is the possibility to re-run only a small fraction of samples, in case these did not yield satisfactory results, and also the possibility to add some samples to the analyses at any time. However, currently available desktop versions of next-generation sequencers from major manufacturers and emerging new technologies should enable satisfactory scalability. These new platforms (Ion Torrent and Illumina MiSeq appear most promising at the time of writing) may also decrease the overall cost of sequencing considerably, as the per-base cost of sequencing is orders of magnitude lower than for 454, and read lengths are increasing rapidly. Eventually, the sequencing itself will constitute only a minor fraction of the total cost of laboratory analyses. Still, the use of 454 and other next-generation sequencing platforms for MHC typing is not devoid of specific problems. Although it is rapidly improving, the error rate of 454 pyrosequencing is not negligible, especially if the target sequences contain homopolymers (Moore *et al.* 2006; Huse *et al.* 2007), a feature apparently shared by Ion Torrent. Finally, additional work may be needed to establish universally applicable criteria for genotyping with next-generation sequencing (see, for example, the different approaches in analysing the same gene in Babik *et al.* 2009 and Galan *et al.* 2010). However, the procedures developed so far for genotyping multilocus MHC systems with 454 technology are easily transferable to other emerging technologies and would probably require only minor adjustments to accommodate platform-specific features.

Conclusion

CE-SSCP provides information for a relative comparison of individual MHC variability, as the results of CE-SSCP and 454 are significantly, albeit weakly, correlated. However, with complex MHC systems, CE-SSCP may serve only as a rough estimate of individual MHC diversity and will, thus, have very limited application, as the results may grossly underestimate the actual variability. Furthermore, CE-SSCP indicates differences in allelic composition, but does not provide particular nucleotide sequences. To obtain these, usually it has to be combined with cloning and Sanger sequencing (Bryja *et al.* 2005). In such complicated multi-locus systems, as the passerine MHC, obtaining sequences using these standard methods might be extremely costly and time-consuming. Thus, CE-SSCP is rather applicable on species, in which MHC genotyping is less complicated than in passerine birds (e.g. Galliform birds or mammals). Using 454 GS FLX Titanium pyrosequencing, on the other hand, it is possible to obtain individual sequence genotypes for large data sets in a short time. So far, this method has been relatively expensive, but there is a trend of rapidly decreasing costs and increasing efficiency. According to our findings, adequate coverage coupled with careful genotyping gives rewarding results, as shown by the perfect congruence between 454 typing and cloning/Sanger sequencing of an individual with high number of alleles.

Acknowledgements

We are grateful to M. Kuenzli and W. Qui for running 454 sequencing, to M. Vinkler, J. Schnitzer, P. Munclinger, R. Poláková, L. Červa, J. Jelínek, F. Zicha, M. Lundák, P. Jaška and H. Mrkvičková for help in the field. We also thank D. Čížková for advice concerning data analysis and R. Burri for valuable comments on the manuscript. This project was financially supported by the Grant Agency of the Academy of Sciences of the Czech Republic (project no. IAA600930608), by the Czech Science Foundation (projects 206/06/0851 and P505/10/1871) and the Jagiellonian University (DS/WBINOZ/INOŚ/762/10).

References

Agbali M, Reichard M, Bryjová A, Bryja J, Smith C (2010) Mate choice for nonadditive genetic benefits correlate with MHC dissimilarity in the rose bitterling (*Rhodeus ocellatus*). *Evolution*, **64**, 1683–1696.

Albrecht T, Vinkler M, Schnitzer J, Poláková R, Munclinger P, Bryja J (2009) Extra-pair fertilizations contribute to selection on secondary male ornamentation in a socially monogamous passerine. *Journal of Evolutionary Biology*, **22**, 2020–2030.

Alcaide M, López L, Tanferna A, Blas J, Sergio F, Hiraldo F (2010) Simultaneous analysis of multiple PCR amplicons enhances capillary SSCP discrimination of MHC alleles. *Electrophoresis*, **31**, 1353–1356.

Axtner J, Sommer S (2007) Gene duplication, allelic diversity, selection processes and adaptive value of MHC class II DRB genes of bank vole, *Clethrionomys glareolus*. *Immunogenetics*, **59**, 417–426.

Babik W (2010) Methods for MHC genotyping in non-model vertebrates. *Molecular Ecology Resources*, **10**, 237–251.

Babik W, Radwan J (2007) Sequence diversity of MHC class II DRB genes in the bank vole *Myodes glareolus*. *Acta Theriologica*, **52**, 227–235.

Babik W, Taberlet P, Ejsmond MJ, Radwan J (2009) New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Molecular Ecology Resources*, **9**, 713–719.

Baratti M, Ammannati M, Magnelli C, Massolo A, Dessì-Fulgheri F (2010) Are large wattles related to particular MHC genotypes in the male pheasant? *Genetica*, **138**, 657–665.

de Bellocq JG, Suchentrunk F, Baird SJE, Schaschl H (2009) Evolutionary history of an MHC gene in two leporid species: characterisation of MHC-DQA in the European brown hare and comparison with the European rabbit. *Immunogenetics*, **61**, 131–144.

Bryja J, Galan M, Charbonnel N, Cosson JF (2005) Analysis of major histocompatibility complex class II gene in water voles using capillary electrophoresis-single stranded conformation polymorphism. *Molecular Ecology Notes*, **5**, 173–176.

Čížková D, de Bellocq JG, Baird SJE, Piálek J, Bryja J (2010) Genetic structure and contrasting selection pattern at two major histocompatibility complex genes in wild house mouse populations. *Heredity*, **106**, 727–740.

Cummings SM, McMullan M, Joyce DA, van Oosterhout C (2010) Solutions for PCR, cloning and sequencing errors in population genetic analysis. *Conservation Genetics.*, **11**, 1095–1097.

Eizaguirre C, Yeates SE, Lenz TL, Kalbe M, Milinski M (2009) MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Molecular Ecology*, **18**, 3316–3329.

Evans ML, Neff BD (2009) Major histocompatibility complex heterozygote advantage and widespread bacterial infections in populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology*, **18**, 4716–4729.

Galan M, Guivier E, Caraux G, Charbonnel N, Cosson J-F (2010) A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics*, **11**, 296.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology.*, **8**, R143.

Keohavong P, Thilly WG (1989) Fidelity of DNA-polymerases in DNA amplification. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 9253–9257.

Klein J (1986) *Natural History of the Major Histocompatibility Complex*. Wiley, New York.

Kloch A, Babik W, Bajer A, Sinski E, Radwan J (2010) Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole Myodes glareolus. *Molecular Ecology*, **19**, 255–265.

Lenz TL, 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.

Lenz TL, Eizaguirre C, Becker S, Reusch TBH (2009) RSCA genotyping of MHC for high-throughput evolutionary studies in the model organism three-spined stickleback *Gasterosteus aculeatus*. *BMC Evolutionary Biology*, **9**, 57.

Longeri M, Zanotti M, Damiani G (2002) Recombinant DRB sequences produced by mismatch repair of heteroduplexes during cloning in *Escherichia coli*. *European Journal of Immunogenetics.*, **29**, 517–523.

Meyer-Lucht Y, Otten C, Püttker T, Pardini R, Metzger JP, Sommer S (2010) Variety matters: adaptive genetic diversity and parasite load in two mouse opossums from the Brazilian Atlantic forest. *Conservation Genetics.*, **11**, 2001–2013.

Moore MJ, Dhingra A, Soltis PS *et al.* (2006) Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biology*, **6**, 17.

Penn DJ (2002) The scent of genetic compatibility: sexual selection and the major histocompatibility complex. *Ethology*, **108**, 1–21.

Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity*, **96**, 7–21.

Promerová M, Albrecht T, Bryja J (2009) Extremely high MHC class I variation in a population of a long-distance migrant, the Scarlet Rosefinch (*Carpodacus erythrinus*). *Immunogenetics*, **61**, 451–461.

Promerová M, Vinkler M, Bryja J *et al.* (2011) Occurrence of extra-pair paternity is connected to social male's MHC-variability in the scarlet rosefinch *Carpodacus erythrinus*. *Journal of Avian Biology*, **42**, 5–10.

Schwensow N, Eberle M, Sommer S (2008) Compatibility counts: MHC-associated mate choice in a wild promiscuous primate. *Proceedings of the Royal Society Series B—Biological Sciences*, **275**, 555–564.

Stuglik MT, Radwan J, Babik W (2011) jMHC: software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing. *Molecular Ecology Resources*, **11**, 739–742.

Wegner KM (2009) Massive parallel MHC genotyping: titanium that shines. *Molecular Ecology*, **18**, 1818–1820.

Westerdahl H, Wittzell H, von Schantz T, Bensch S (2004) MHC class I typing in a songbird with numerous loci and high polymorphism using motif-specific PCR and DGGE. *Heredity*, **92**, 534–542.

Yamazaki K, Boyse EA, Mike V *et al.* (1976) Control of mating preferences in mice by genes of major histocompatibility complex. *Journal of Experimental Medicine*, **144**, 1324–1335.

Zagalska-Neubauer M, Babik W, Stuglik M, Gustafsson L, Cichoń M, Radwan J (2010) 454 sequencing reveals extreme complexity of the class II major histocompatibility complex in the collared flycatcher. *BMC Evolutionary Biology*, **10**, 395.

Zylstra P, Rothenfluh HS, Weiller GF, Blanden RV, Steele EJ (1998) PCR amplification of urine immunoglobulin germline V genes: strategies for minimization of recombination artefacts. *Immunology and Cell Biology*, **76**, 395–405.

Data Accessibility

Final 454 genotypes, replicate 454 genotypes and CE-SSCP number of alleles uploaded as Supporting Information.

Original CE-SSCP electrophoretograms are deposited in the Dryad repository: doi:10.5061/dryad.bf20k4q4.

Sequences of alleles obtained by 454 are accessible at Genbank under accession numbers: JN712931–JN713119.

Supporting Information

Additional supporting information may be found in the online version of this article.

Appendix S1 Individual 454 genotypes with final coverage after filtering, 454 replicates, and a table with 454 and CE-SSCP number of alleles.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.