Behavioral Ecology and Sociobiology (2020) 74:38
https://doi.org/10.1007/s00265-020-2822-5

ORIGINAL ARTICLE

Similarity at the major histocompatibility complex class II does not influence mating patterns in bluegill (*Lepomis macrochirus*)

Lucas Silveira[1] · Shawn R. Garner[1] · Bryan D. Neff[1]

Received: 19 September 2019 / Revised: 20 February 2020 / Accepted: 27 February 2020

Abstract

In many species, individuals prefer mates that are genetically dissimilar at the major histocompatibility complex (MHC), likely because it improves offspring resistance to pathogens. Here, we provide the first characterization of the MHC class II peptide binding region in bluegill (*Lepomis macrochirus*) and examine its effect on mating patterns. We captured female and male bluegill during spawning and sequenced these fish at the MHC. We found strong evidence that positive selection promotes genetic diversity at the MHC in bluegill, with a 5:2 ratio of non-synonymous to synonymous mutations. However, we found no evidence that the MHC led to disassortative mating between females and parental males. Extra-pair mating and the presence of specialized cuckolder males may have an important, albeit still unresolved, role in shaping mating patterns at the MHC in bluegill.

Significance statement

The immune genes of the MHC allow individuals to recognize and respond to pathogens. High sequence diversity at the MHC is thought to enable individuals to recognize a wider range of pathogens. In many animals, individuals select mates that have dissimilar MHC sequences, which can improve the immune function of their offspring. We provide the first characterization of the MHC class II gene in bluegill sunfish and show that selection appears to be increasing sequence diversity at this gene. However, we found no evidence that the MHC shapes mating patterns in this species.

Keywords Sunfish · Major histocompatibility complex · Mate choice · Genetic quality

Introduction

Biologists have long been interested in understanding mate choice and the nature of the benefits that it provides. When selecting a mate, individuals may choose for direct benefits (e.g. brood defence, territory quality, fecundity) or indirect benefits (genetic quality). In mating systems in which individuals select for indirect benefits, individuals may be congruent in their mating preferences and favour individuals that provide additive genetic benefits, termed 'good genes' (Kirkpatrick 1996). The outcome of this choice however, should lead to fixation or near-fixation of the targeted allele by directional selection and consequently the loss of any benefit to mate choice. This issue has become known as the lek paradox (reviewed by Kotiaho et al. 2008). Alternatively, individuals may choose mates for non-additive genetic benefits, which are often referred to as 'compatible genes' (Neff and Pitcher 2005). Compatible genes (alleles) increase fitness through either overdominance or through epistasis (Neff and Pitcher 2005). In this case, individuals choose mates based on a set of genes that will most complement their own; often because those genes are dissimilar and will increase the heterozygosity of their offspring, or conversely, because those genes are similar and will reduce outbreeding depression and the disruption of co-adapted gene complexes (Neff 2004; Lenz et al. 2009). Growing evidence suggests that mate choice for non-additive genetic benefits is common (Neff and Pitcher 2005).

The genes of the major histocompatibility complex (MHC) may be important determinants of genetic quality. MHC genes encode cell-surface proteins that bind short peptides (antigens)

Communicated by I. Hamilton

✉ Bryan D. Neff
bneff@uwo.ca

[1] Department of Biology, University of Western Ontario, 1151 Richmond Street, London, ON N6A 5B7, Canada

Published online: 10 March 2020

from foreign pathogens and present those peptides to other immune cells, which initiate an immune response (Murphy et al. 2012). There are two classes of MHC molecules: class I appears on all nucleated cells and primarily functions to present intracellular pathogens to cytotoxic T cells; class II appears on antigen-presenting cells such as B cells and primarily functions to present extracellular pathogens to helper T cells (Murphy et al. 2012). The peptide binding region of the MHC protein acts as a pocket with high specificity for antigens (Murphy et al. 2012). Brown et al. (1993) used X-ray crystallography to visualise the peptide-binding region of human MHC and identify key amino acid positions that determine the specificity of the peptides that are bound. Differences in those key residues lead to different folding arrangements of the protein which ultimately lead to differences in the ability of the MHC molecule to bind particular pathogens (Murphy et al. 2012). Langefors et al. (2001), for example, showed that a specific allele in Atlantic salmon (*Salmo salar*) was linked to higher resistance to infection from the bacterium *Aeromonas salmonicida*. Individuals that possess different MHC alleles, particularly at the amino acid level of the peptide biding region, are expected to recognize a greater range of pathogens. Penn et al. (2002) showed that mice (*Mus domesticus*) that were MHC heterozygous were more likely to survive bacterial infection than mice that were MHC homozygous. Worley et al. (2010) similarly showed that red junglefowl (*Gallus gallus*) that were MHC heterozygous survived infection longer than MHC homozygous individuals, and that the survival difference was independent of genome-wide heterozygosity. In Chinook salmon (*Oncorhynchus tshawytscha*), a positive effect of MHC heterozygosity on survival was found in populations with high rates of bacterial infection (Evans and Neff 2009). Genetic variation at the MHC may thus be associated with genetic quality, particularly through non-additive effects on immunity and survival.

Another approach to examine selection on the MHC is to examine sequence variation within a population. Models have been developed that examine the rates of synonymous to non-synonymous mutations to infer selection. One such model is PAML developed by Yang (2007). Using this model, signals of positive Darwinian selection acting on MHC genes have been found in Grey partridge (*Perdix perdix*), frog (*Rhacophorus omeimontis* and *Polypedates megacephalus*), mummichog (*Fundulus heteroclitus*) and guppies (*Poecilia reticulata*) (Cohen 2002; Fraser et al. 2010b; Promerová et al. 2013; Zhao et al. 2013). Population-level sequence analysis thus provides additional evidence that selection promotes increased genetic variation at the MHC.

The potential survival benefits through enhanced immunity provide an opportunity for mate choice for indirect benefits from the MHC. Individuals might be predicted to choose MHC dissimilar mates in order to increase the number of unique alleles in their offspring. Early studies on mammals showed that mating preferences led to a reduction in the number of MHC homozygous individuals relative to expectations under random mating (e.g. Potts et al. 1991; reviewed by Tregenza and Wedell 2000). More recently, effects of MHC on mating patterns have been observed in a number of fishes. Landry et al. (2001) found that Atlantic salmon mated with individuals that increased the number of amino acid differences between MHC class II alleles in their offspring relative to expectations under random mating. In Chinook salmon, Neff et al. (2008) similarly showed that the observed mating patterns would lead to offspring with a greater number of amino acid differences at the MHC class II than expected under random mating. In brown trout (*Salmo trutta* L.), mating patterns instead appear to favour an intermediate level of amino acid dissimilarity between MHC class II alleles in the offspring (Forsberg et al. 2007). In three-spined stickleback (*Gasterosteus aculeatus*), females prefer the odour of males with more unique MHC class II alleles relative to the odour of males with fewer unique alleles (Reusch et al. 2001). In contrast, several studies in fishes have supported the presence of a post-copulatory fertilization advantage for MHC-similar mates, possibly as a mechanism to reduce outbreeding depression associated with interspecies hybridization (Yeates et al. 2009; Gasparini et al. 2015).

Despite many reports of MHC-based mating patterns, a recent meta-analysis suggests that in many cases the effect of MHC genotypes on mating patterns is small or absent (Kamiya et al. 2014). This weak effect might occur because other factors interfere with the expression of MHC-based mating preferences and lead to an overall weak effect of MHC on mating patterns. For example, in Chinook salmon, even though females may prefer MHC-dissimilar mates, male aggression towards unreceptive females can override their choice, especially when sex ratios are male-biased (Garner et al. 2010). If potential mates provide important direct benefits, it might also lessen the role of MHC dissimilarity in mating preferences, although in mice it appears that these competing demands are resolved by female MHC preferences primarily influencing the choice of extra-pair mates (Potts et al. 1991).

Here, we use bluegill (*Lepomis macrochirus*) to investigate mating patterns based on MHC similarity. Bluegill are a freshwater fish native to North America, and the mating system in Lake Opinicon was first described by Gross (1979). It typically lasts from late May to early July wherein multiple spawning bouts occur throughout the lake. In the days leading up to spawning, parental males congregate in the shallows of the lake and form colonies that can range in size from 5 to 150 nests (Gross 1982). Each parental male builds a bowl-shaped nest in the sediment through sweeping motions of his caudal fin. On the day of spawning, females swim around a colony and enter males' nests to initiate spawning. Spawning is synchronous and rarely lasts more than 1 day at any particular

colony. Parental care, which is provided solely by the nest-tending parental male, involves fanning of the eggs and defense from nest predators, and it is essential for offspring survival (Gross 1982). Bluegill are also characterized by an alternative reproductive tactic, wherein smaller precocious males called 'cuckolders' steal fertilizations from parental males. Cuckolders provide no care for their offspring but leave that to the nest-defending parental males (Neff and Gross 2001). Small cuckolders use a sneaking tactic to ambush a spawning pair and release sperm near the eggs before darting out of the nest. Larger (older) cuckolders switch tactics and instead mimic females in coloration and behavior (Gross 1982). These female mimics are able to remain in the nest, often immediately between a spawning pair. Cuckolders, particularly sneakers, are more opportunistic (Stoltz and Neff 2006) and might circumvent, to some degree, potential choice for MHC-dissimilar mates. Bluegill frequently hybridize with pumpkinseed (*Lepomis gibbosus*) in Lake Opinicon, resulting in hybrid individuals that typically have low fertility (Konkle and Philipp 1992; Immler et al. 2011). Although bluegill and pumpkinseed nests co-occur in many areas of Lake Opinicon, hybridization between these species occurs not because females spawn in the nests of parental males of the other species, but instead because bluegill cuckolders frequently intrude into spawning events in pumpkinseed nests (Garner and Neff 2013).

Here, we present the first characterization of the MHC class II locus in bluegill. We first used the random sites codon model-based approach in PAML ver. 4 (Yang 2007) to investigate the patterns of selection at this locus. Next, to determine if mating patterns are disassortative with respect to MHC genotypes, we caught males and females in the act of spawning, and then analysed their genotypes at the MHC class II. If MHC heterozygosity increases immunity, we predict that females should mate most frequently with MHC-dissimilar parental males. If females instead use MHC dissimilarity to distinguish between bluegill and pumpkinseed males, we predict that female bluegill should most frequently mate with MHC-similar parental males.

Methods

Study species and sample collection

Sample collection occurred at the Queen's University Biological Station (QUBS) on Lake Opinicon (44.5° N, 76.3° W). In the summer of 2015, a 6-km transect along the shoreline of Lake Opinicon was swam daily by a group of divers to identify the formation of colonies and the arrival of breeding females, which marked the beginning of a spawning bout at that colony. On the day that spawning activity was first observed at a colony, divers floated motionless over the colony to observe the activity and watch for females entering nests. Spawning pairs were observed until a female visibly dipped, a movement where the female tips onto her side and releases a small batch of her eggs into a parental male's nest. Once dipping was observed at least five consecutive times between a pair, both parental male and female were caught with a dip net. This dipping threshold was used to ensure that the female had mated with the parental male and was not in his nest solely to evaluate him as a potential mate. A female bluegill may mate with multiple males during one or more breeding bouts, and a parental male's nest often contains eggs from multiple females (Neff 2001). A mesh cover was then placed over the male's nest to protect the eggs from predation by other fish. The nest was also marked with a uniquely numbered ceramic tile to allow for identification of the nest. The mating pair was then brought to a boat where total body length was recorded and a small fin clip was removed from each fish's caudal fin and stored in 95% ethanol for later genetic analysis. Both fish were then returned to the water and the cover was removed from the parental male's nest. Parental males typically returned to their nests immediately and commenced courting other females. Occasionally, a previously caught parental male (identified by nest tile and fin clip) was recaptured with a new female (identified by the absence of a fin clip), or two females were captured while simultaneously mating with a parental male. In these cases, each parental male-female pairing was treated as a novel pair (5 of 35 pairs in our sample were comprised of previously caught males). In addition to spawning pairs that contained a parental male and female, we also captured four sneakers in the act of spawning and collected a fin clip from these fish. A total of 69 bluegill (30 parental males, 35 females, 4 cuckolders) were thus included in our study. To minimize observer bias, blinded methods were used such that the identities of mating pairs were not known at the time of the genetic analysis.

MHC primer design and sequencing

Primers for amplification of MHC II in bluegill were designed based on sequences from a bluegill brain transcriptome (Partridge et al. 2016). Briefly, Partridge et al. (2016) used high-throughput sequencing to characterize the sequences of expressed transcripts from the brains of 20 bluegill collected in Lake Opinicon. The resulting transcriptome consisted of 235,547 transcripts. Using the transcriptome as a local database, NCBI BLAST was used to search for potential MHC class II putative peptide binding region sequences by using known MHC sequences from striped sea bass (*Morone saxatilis*, Genbank id: L33967) and three-spined stickleback (*Gasterosteus aculeatus*, Genbank id: DQ016429). Probing the transcriptome for exon 2 of MHC class II with these sequences yielded a single transcript in bluegill. Using the bluegill MHC II transcript sequence, Primer-Blast was used to

develop a primer pair (forward: GCATTCCTCAGTGG TCCGC and reverse: TGTACCAGTTCCCAATGTTG) that spanned a 239 base pair region of the putative MHC II locus.

To test the MHC II primers, DNA was first extracted from bluegill fin clips via Proteinase K digestion and ethanol precipitation (Neff et al. 2000). Next, DNA from three parental males was amplified at the MHC using polymerase chain reaction (PCR). The PCR amplicon was cloned using a pGEM T-easy vector kit following manufacturer's instructions (Promega Corp, Madison, Wisconsin) and used to transform *Escherichia coli*, which were then grown on lysogeny broth agar plates. Bacterial colonies containing the insert were collected and re-amplified using the sequencing primers SP6 and T7. Eight insert-containing colonies from three individuals were sequenced by the London Regional Genomics Centre (London, Ontario). The resulting sequences were analysed with NCBI BLAST, which confirmed that the bluegill MHC amplicon had high similarity to the putative peptide binding region of MHC class II in other teleost fishes, including 88% identity with striped sea bass and 87% identity with orange-spotted grouper (*Epinephelus coioides*). Bluegill MHC sequences were then aligned with the human MHC class II peptide binding region to identify the specific amino acid positions likely to comprise the key residues of the pathogen peptide binding region following the X-ray crystallography determinations of Brown et al. (1993).

Next-generation sequencing was used to sequence 50 of the 69 bluegill collected from Lake Opinicon. First, samples were PCR-amplified with modified versions of the MHC primers that included a UniA tail on the forward primer and a UniB tail on the reverse primer. After the PCR, the product was visualized on a gel to ensure amplification occurred and then cleaned using ethanol precipitation. A second PCR with primers specific to the UniA and UniB tails was then used to attach an Ion Torrent adaptor and sample-specific barcode in the forward direction, and an Ion Torrent adaptor sequence in the reverse direction. The sample-specific barcode (a unique 10–11 bp sequence) allowed multiple individuals to be pooled in a single sequencing run, with the resulting sequences assigned to individuals based on these unique barcodes. A QIAquick PCR Purification kit was used to purify the products after the second PCR. DNA concentrations in each sample were measured using a Nanodrop (ND-3300, NanoDrop Technologies) and pooled in equal concentrations. The resulting library was sequenced on an Ion Torrent Personal Genome Machine (Life Technologies) at the University of Windsor Environmental Genomics Facility.

AmpliSAT software was used to sort and clean the resulting Ion Torrent sequence data (Sebastian et al. 2016). Briefly, AmpliSAT de-multiplexes, clusters and filters the raw sequencing data allowing for the removal of artefacts and the assignment of alleles to the amplicon. Sequence variants that appeared as less than 1% of an individual's total reads were discarded as sequencing errors following protocols established by Galan et al. (2010). Chimeric sequences were identified within an individual as low-frequency sequence variants that were a combination of two common alleles possessed by that individual and were removed. After these clean-up steps, all individuals possessed either one or two unique alleles (consistent with a single unduplicated MHC II locus in bluegill).

In addition to the individuals sequenced using the Ion Torrent method, 19 of the 69 individuals were sequenced using a Sanger method. The Sanger method was also used for nine individuals that were already sequenced by Ion Torrent to confirm that the results were comparable. For the Sanger-sequencing method, the MHC amplicons were PCR-amplified as described above, and then sequenced in both directions at the London Regional Genomics Centre. The resulting chromatograms were manually examined in BioEdit (Hall 1999) to determine the sequences and identify heterozygous positions, which were characterized by two peaks of similar intensity at a variable site. Based on the alleles identified using the Ion Torrent sequencing, each combination of alleles would result in a unique pattern of variation on the chromatograms. We then assigned each individual an MHC genotype based on their chromatogram sequence. The Ion Torrent and Sanger methods produced identical genotypes for the nine individuals that were sequenced using both methods.

MHC characterization

To investigate how selection is acting on the MHC in bluegill at the codon level, the random sites codon model-based approach in PAML ver. 4 (Yang 2007) was used as described in Fraser et al. (2010a). Briefly, PAML uses the estimated rates of synonymous to non-synonymous substitutions (dN/dS, ω) to examine how codons are varying according to several models. Four potential models were assessed for their likelihood using a nested log-likelihood ratio test. Model M1a is the 'nearly neutral' model which creates an estimate for the proportion of codons undergoing purifying selection (p_0, $0 < \omega_0 < 1$) and the remaining proportion of codons that are neutrally evolving ($p_1 = 1 - p_0$, $\omega_1 = 1$). Model M2a is the 'positive selection' model which includes model M1a with the addition of a third class of codons where positive selection is occurring ($\omega_2 > 1$) defined by the proportion p_2 ($= 1 - p_0 - p_1$). Models M7 and M8 apply a less restrictive definition for ω between 0 and 1 with the use of a β distribution. The β distribution is a flexible probability density function used to capture further ω variation across codons and is estimated from the data (see Yang et al. 2000). Model M7 is analogous to M1a and serves as the null β model ($0 < \omega_0 < 1$). Model M8 is the 'positive selection plus β' model which is analogous to M2a which again is equivalent to M7 but allows for a proportion of

codons that are undergoing positive selection (p_1, $\omega_1 > 1$). Using a Bayes Empirical Bayes approach (Yang et al. 2005), codons were allotted to different selection classes under a 95% posterior probability cut-off. Likelihood ratios were calculated in PAML in order to compare between models.

ML-NullFreq was used to test for an excess in observed homozygosity which may indicate the presence of a null allele that was undetectable through sequencing (e.g. because of a mutation in the primer-binding sites; Kalinowski and Taper 2006). ML-NullFreq was further used to predict the frequency of a potential null allele and recalculate the frequencies of other alleles based on this prediction.

Mating pair analysis

We used two measures to compare the MHC similarity of mating pairs. First, expected heterozygosity of a pair's offspring was calculated. Second, following Landry et al. (2001), for each mating pair, we calculated the number of amino acid differences between the MHC alleles of the male and female. An average of the four values was then calculated, which represents the expected number of amino acid differences between MHC alleles in the pair's offspring. These two measures were calculated using the entire length of the MHC amplicon, and again using only the key peptide binding residues identified by Brown et al. (1993) as being most important for determining the binding properties of the peptide binding region.

Following Neff et al. (2008), Monte Carlo simulations were then used to create expected distributions for offspring heterozygosity and amino acid differences between pairs under a model of random mating. The simulation randomly paired females and males from the entire population and maintained the observed number of mates for each fish. Each simulation was repeated 10,000 times to generate a distribution under random mating. The observed values for offspring heterozygosity and amino acid differences between pairs were then compared to these random distributions to determine if the observed values differed significantly from the expectations under random mating. Specifically, two-tailed p values were calculated as the proportion of the simulated values that were at least as far from the median simulated value as the observed value.

We used a power analysis to quantify our power to detect non-random mating at the MHC if it were actually occurring, given a particular effect size. For this analysis, we first calculated the maximum offspring heterozygosity and amino acid differences between pairs for each female based on the observed male genotypes. We then assigned each female a mate at random, varying the rate at which she would choose her optimal mate in 10% increments ranging from 0 to 100%. This process was repeated 10,000 times. Power at each increment was then calculated as the proportion of the replicates that were significant based on a two-tailed test when mates were assigned at random (i.e. the proportion of the trials in which a significant effect would be detected given the null distribution we generated). Because similar results were obtained when analyses were based on the full MHC amplicon sequence or only key residues, for this and subsequent analyses we present only the results based on the full MHC amplicon sequence.

We also used a Monte Carlo simulation to examine the potential effects of a null allele on our inferences about mating patterns. First, individuals with a single MHC allele were randomly assigned as homozygous for the observed allele or heterozygous for the observed allele and the null allele. The probability of being assigned the null allele was proportional to the population-wide frequency of the null allele relative to the frequency of the observed allele (e.g. individuals with a common allele were more likely to be true homozygotes than individuals with a rare allele). Next, in order to calculate offspring heterozygosity and amino acid differences between pairs, we assigned the null allele a sequence equivalent to one of the observed alleles at frequencies proportional to the observed allele frequencies. For each generation of the simulation, the null allele had a single identity. We repeated this process 10,000 times, and for each generation calculated offspring heterozygosity and average amino acid differences between pairs for the observed mating pairs and for randomized mating pairs. We calculated a p value as the proportion of the replicates for which the observed value was greater than the random value.

Results

Collection summary

Parental males were collected from eight colonies and had an average length of 189 ± 12 mm (mean ± SD; range 165–209 mm; $n = 30$). The females captured in the nests of the parental males had an average length of 143 ± 26 mm (mean ± SD; range 80–198 mm; $n = 35$). The cuckolders had an average length of 80 ± 14 mm (mean ± SD; range 67–100 mm; $n = 4$). No significant correlation was observed between the length of parental males and females from observed mating pairs ($p = 0.11$, Rho = 0.28, $n = 35$).

MHC characterization

Ion Torrent sequencing resulted in 143,122 useable sequence reads across 50 bluegill samples. In total, ten different putative MHC class IIB alleles were found with five variable amino acid sites (Table 1). No insertions/deletions or stop codons were observed in these sequences, which is consistent with each sequence representing an expressed protein-coding gene. No apparent gene duplications were identified due to the

Table 1 Unique alleles observed at the MHC class II putative peptide binding region in bluegill (*Lepomis macrochirus*). Amino acid sequences are shown only for polymorphic positions. Asterisks denote analogs of sites found to be key to peptide binding in human MHC class II peptide binding region through three-dimensional imaging (Brown et al. 1993). Adjusted frequencies with the inclusion of a null allele were calculated using ML-NullFreq (Kalinowski 2006). Genbank accession numbers are included for each allele

Allele	Frequency	Adjusted frequency	Genbank accession #	Polymorphic amino acid position				
				15	23*	39	55*	62*
BG1	0.31	0.24	MK620857	N	E	V	R	Q
BG2	0.14	0.1	MK620858	H	E	V	R	Q
BG3	0.17	0.1	MK620859	H	E	L	R	Q
BG4	0.05	0.03	MK620860	N	Q	V	R	I
BG5	0.09	0.07	MK620861	H	E	V	H	Q
BG6	0.05	0.05	MK620862	H	Q	L	R	Q
BG7	0.09	0.07	MK620863	N	Q	V	R	I
BG8	0.06	0.05	MK620864	N	E	V	H	I
BG9	0.01	0.01	MK620865	N	Q	L	R	I
BG10	0.02	0.01	MK620866	H	E	V	R	I
Null		0.24						

absence of more than two alleles in any individual. Across all individuals, the observed heterozygosity was 53%, whereas the expected heterozygosity assuming Hardy-Weinberg equilibrium was 83%. MLNullFreq detected a significant excess of homozygosity ($p < 0.001$) and predicted that there was a null allele with a frequency of 0.24 (Table 1).

When comparing selection models on the MHC sequences, the M2a-positive selection model was significantly more likely than the M1a nearly neutral model (LnLRT = 64.55; $p < 0.0001$). The M8-positive selection model was also significantly more likely than the M7 null model (LnLRT = 64.78; $p < 0.0001$). These models indicate that the peptide binding region in bluegill is under positive selection and that 6.5% of codons showed signs of positive selection while the remaining 93.5% were under purifying or neutral selection. In both positive selection models, the Bayes Empirical Bayes method showed 5 codons in the peptide binding region of MHC class II were under positive selection (15, 23, 39, 55, and 62, with posterior probabilities > 0.97 for each).

Mating pair analyses

For the observed mating pairs, 84% of the offspring would be expected to be heterozygous at the MHC II locus. There was no significant difference between the observed heterozygosity and the expectations under random mating when we analysed the complete MHC sequence ($p = 0.84$, Fig. 1a). When we instead calculated heterozygosity based only on the key residues of the peptide binding region, there was again no significant difference between the observed heterozygosity and the expectations under random mating ($p = 0.98$, Fig. 1b).

For the observed mating pairs, the mean number of amino acid differences between pairs was 1.86 (range = 0 to 3). There was no significant difference between the observed amino acid differences between pairs and the expectations under random mating ($p = 0.98$, Fig. 1c). When we instead calculated amino acid differences between pairs based only on the key residues of the peptide binding region, the mean amino acid differences between pairs was 0.99 (range = 0 to 2). There was no significant difference between the observed amino acid differences between pairs at the key peptide binding region residues and the expectations under random mating ($p = 0.92$, Fig. 1d).

The power of our study to detect non-random mating as a function of the frequency with which individuals select an MHC-optimal mate is summarized in Fig. 2. The power to detect non-random mating with respect to offspring heterozygosity was 95% when individuals selected an optimal mate 72% of the time, and 80% when individuals selected an optimal mate 62% of the time. The power to detect non-random mating with respect to amino acid differences between pairs was 95% when individuals selected an optimal mate 35% of the time, and 80% when individuals selected an optimal mate 25% of the time.

When the presence of a null allele was included in simulations, offspring heterozygosity did not differ significantly between the observed mating pairs and random mating pairs ($p = 0.52$, Fig. 3a). Amino acid differences between pairs also did not differ significantly from the expectations under random mating when a null allele was included in the simulations ($p = 0.51$, Fig. 3b).

Discussion

MHC characterization

Here, we present the first characterization of the MHC class II in bluegill. Using primers designed from an existing bluegill transcriptome (Partridge et al. 2016), we sequenced a 239 base pair section of the putative peptide binding region. Multiple indicators are consistent with these primers capturing a

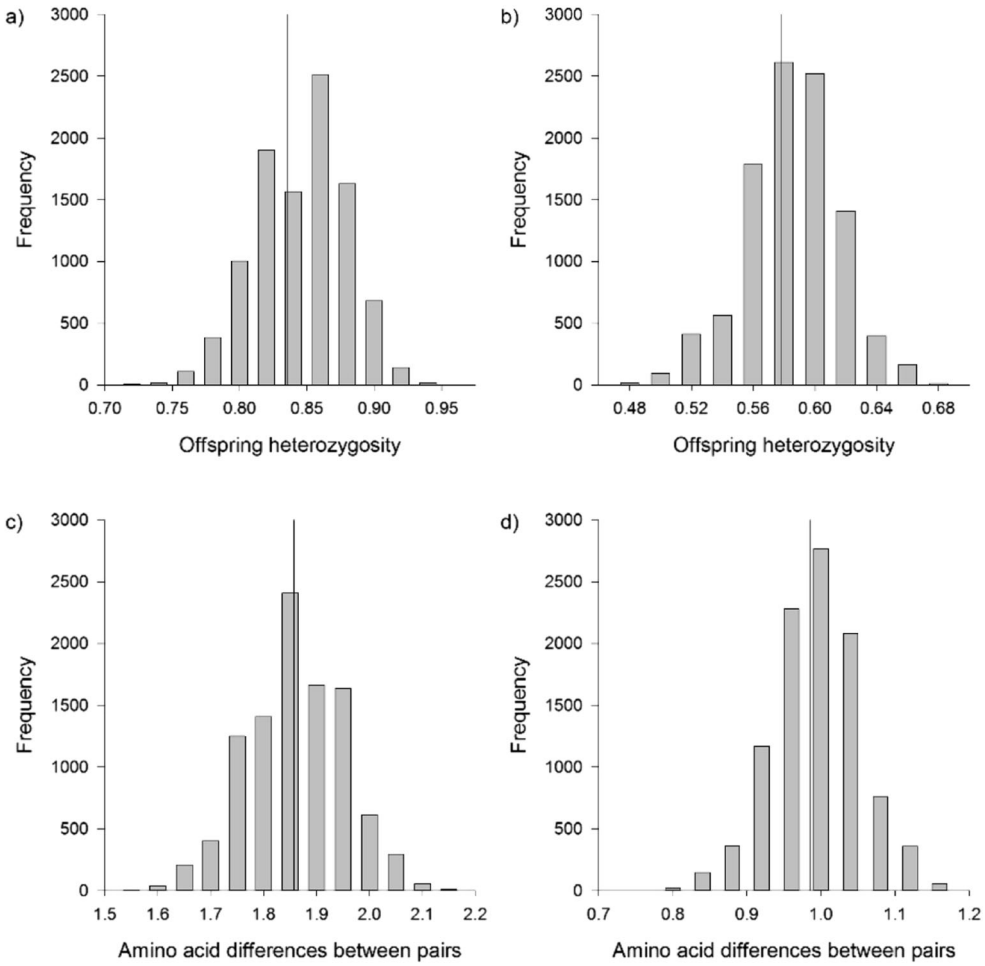

Fig. 1 Offspring heterozygosity and amino acid differences between pairs at the MHC class II peptide binding region under random mating in bluegill (*Lepomis macrochirus*). Panel **a** shows offspring heterozygosity based on the complete MHC amplicon, panel **b** shows offspring heterozygosity based only on key peptide binding residues, panel **c** shows amino acid differences between pairs based on the complete MHC amplicon and panel **d** shows amino acid differences between pairs based only on key peptide binding residues. Expected data are from Monte Carlo simulations and show the results of 10,000 replicates that randomly paired males and females. Observed values are shown with a vertical line

functional MHC locus. First, one of our sequences was identical to the consensus transcriptome sequence, and all of our sequences had high similarity to the consensus transcriptome sequence (> 97% identity), which indicates that the MHC locus we sequenced is expressed. Second, the sequences did not contain stop codons or frameshift mutations that would indicate that the gene encodes a non-functional protein (i.e. is a pseudogene). Third, our sequences had high identity with MHC class II putative peptide binding regions that have previously been characterized in other teleosts (88% identity with *M. saxatilis*, 87% identity with *Epinephelus coioides*). High identity between MHC sequences in bluegill and other fishes are consistent with this region being a conserved, functional gene. Our sequence data also suggest that the MHC II consists of a single locus in bluegill, with all individuals possessing either one or two unique alleles. Overall, our data suggest that the primers capture a single expressed MHC class II gene in bluegill.

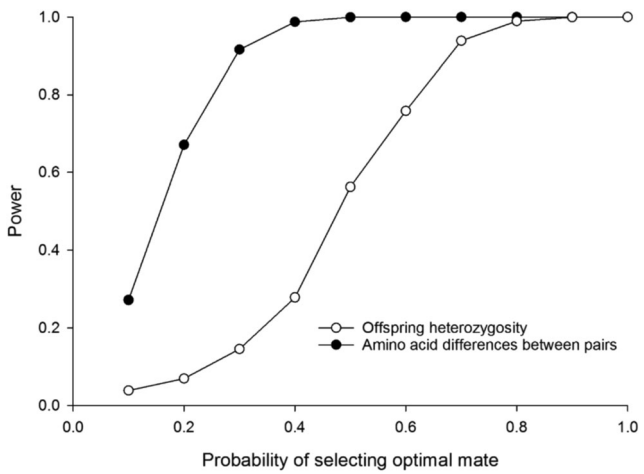

Fig. 2 Power to detect non-random mating at the MHC class II peptide binding region in bluegill (*Lepomis macrochirus*) as a function of the probability that individuals select an optimal mate based on either offspring heterozygosity or amino acid differences between pairs

Given the role of MHC sequence variation in the pathogen-binding characteristics of MHC proteins (Hedrick 2002), MHC genes have frequently been linked to positive selection and high levels of functional variation (Cohen 2002). In bluegill, we found strong evidence that the MHC is subject to positive selection. Using PAML, models of positive selection were found to explain the observed sequence data

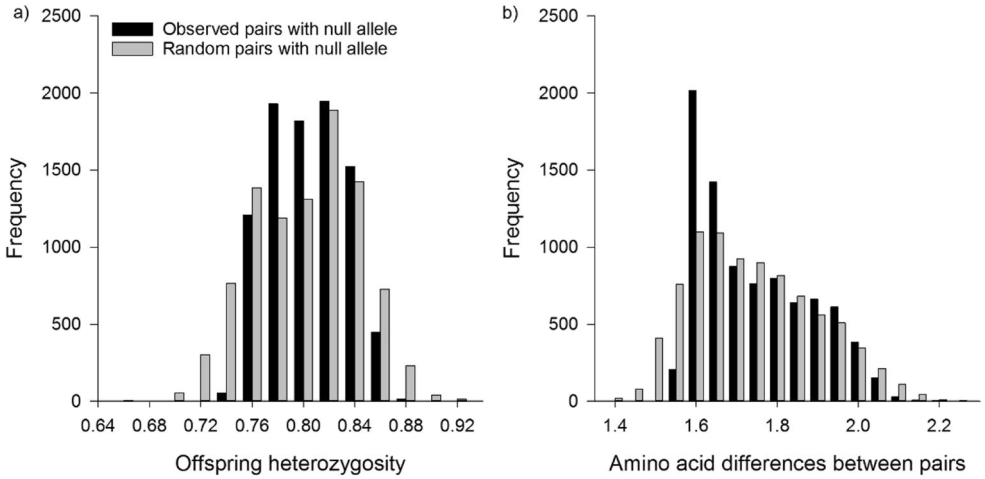

Fig. 3 Effects of a null allele at the MHC class II peptide binding region on MHC dissimilarity for observed and random mating pairs in bluegill (*Lepomis macrochirus*). Panel **a** shows offspring heterozygosity, and panel **b** shows amino acid differences between pairs. Data are from Monte Carlo simulations and show the results of 10,000 replicates that probabilistically incorporated a null allele for individuals that were homozygous at the MHC

significantly better than neutral models. We also observed a high ratio of non-synonymous to synonymous mutations (5:2), which is consistent with positive selection acting on the gene. These results are similar to a study of guppies, in which PAML models provided support for positive selection on the MHC class II gene (Fraser et al. 2010a). High ratios of non-synonymous to synonymous mutations have likewise been found in southern platyfish (*Xiphophorus maculatus*) (9:2, McConnell et al. 1998). Overall, the evidence of positive selection affecting the MHC class II gene in bluegill appears to be in line with what has been found in other fish species and indicates that natural selection has a major role in promoting functional variation at the MHC in bluegill and other fishes.

In bluegill, tests for Hardy-Weinberg equilibrium at the MHC showed a significant excess of homozygotes, which might be explained by several processes. First, this homozygote excess may have resulted from population subdivision or inbreeding. If the homozygote excess at the MHC is caused by population subdivision or inbreeding, then we would also expect a similar homozygote excess at other genes. However, previous studies using microsatellite genetic markers in Lake Opinicon bluegill have shown no evidence of a genome-wide homozygote excess (Neff 2001; Garner and Neff 2013), suggesting that population subdivision or inbreeding cannot explain the MHC genotype data. Second, the homozygote excess at the MHC could result from non-random mating for high MHC homozygosity in offspring. Our simulations indicated that the observed mating patterns were likely random with respect to MHC heterozygosity ($p = 0.84$) and that the observed mating patterns should lead to much higher MHC heterozygosity than was observed in the parents (the observed heterozygosity was 53%, whereas 84% offspring heterozygosity was expected for our observed mating pairs). Third, lower than expected MHC heterozygosity could result from a post-copulatory fertilization advantage for MHC-similar sperm (e.g. Yeates et al. 2009; Gasparini et al. 2015). However, based on the observed mating pairs, we would still expect high heterozygosity even if the eggs were always fertilized by her partner's more MHC-similar sperm (observed heterozygosity was 53%, 73% heterozygosity would be expected if the eggs were always fertilized by the more MHC-similar sperm). Non-random fertilization is thus unlikely to explain the low MHC heterozygosity. Instead, the homozygote excess at the MHC II in bluegill is most consistent with the presence of a null allele, which would not be amplified during PCR due to, for example, a mismatch in the primer-binding sequence. Null alleles have previously been indicated in a number of studies of MHC in species that include bighorn sheep (*Ovis canadensis*), Atlantic salmon, and black-throated blue warbler (*Dendroica caerulescens*) (Gutierrez-Espeleta et al. 2001; Smith et al. 2005; Dionne et al. 2007). In bluegill, a null allele present at an estimated frequency of 24% is the most likely explanation for the observed excess of homozygosity at the MHC.

Mating patterns

MHC-based mating patterns have been reported in many taxa, including multiple fishes (recently reviewed by Kamiya et al. 2014). Here, we provide the first test of the effect of the MHC on mating patterns in bluegill. We found that the observed mating pairs between parental males and females did not differ from the expectations based on random mating with regard to either offspring heterozygosity or amino acid differences between pairs. No evidence for non-random mating was found regardless of whether MHC differences were based on the entire amplicon sequence or only on the key residues of the peptide binding region identified by Brown et al. (1993). In contrast, significant effects of MHC on mating patterns have previously been identified in other fishes, including for example a preference for high amino acid differences between pairs in Atlantic salmon (Landry et al. 2001), and chinook salmon (Neff et al. 2008). These differences in mating patterns could arise from species and environment-specific factors such as the nature of pathogen-

mediated selection on the MHC, the opportunity for mate choice in a mating system and the relative importance of mate choice for other factors (e.g. parental care). We also found no evidence that bluegill disproportionately mated with MHC-similar partners, as for example as a mechanism to avoid hybridization with pumpkinseed. Initial tests of our MHC primers showed that they were effective in pumpkinseed, and that many of the MHC alleles observed in pumpkinseed were identical to alleles observed in bluegill. Consequently, the MHC does not appear to provide a discrete cue that could be used to avoid hybridization between these species. It is also possible that post-copulatory mechanisms to achieve MHC complementarity operate in bluegill. For example, increased fertilization success for MHC-compatible sperm has been observed in three-spined stickleback (Lenz et al. 2018), although a similar study of whitefish (*Coregonus* sp.) found no evidence that MHC affected gamete fusion (Wedekind et al. 2004). Regardless, the absence of pre-copulatory MHC-based mating patterns in a fish appears to be uncommon, though there are a number of examples of mammals where a similar lack of evidence for MHC-based mating patterns has been found (e.g. Kuduk et al. 2014; Liu et al. 2017).

Given the frequency with which MHC-mediated mating patterns have been observed in fishes, it is important to consider whether its absence in bluegill could be explained by limitations in our study design. For example, it is possible that non-random mating with respect to the MHC class II was not observed in bluegill due to low statistical power to detect an effect. Indeed, Hoover and Nevitt (2016) highlighted the importance evaluating power in MHC studies in order to prevent misinterpreting results due to the high variability found at MHC genes. They outlined the importance of a sample size (and therefore statistical power) large enough to detect both the entire variation of MHC genes in a population, as well as the true population mean. Given our sample size, our analysis showed that we had 80% power to detect mate choice if females were able to choose MHC-optimal males at least 25% of the time. A mating preference of this strength would be associated with an increase in the average amino acid differences between mates of about 0.3 amino acids. In a population of Atlantic salmon, Landry et al. (2001) found that mate choice increased the number of amino acid differences between mating pairs at key residues of the MHC by about 0.2 amino acids. In chinook salmon, non-random mating was associated with an increase in amino acid differences between pairs of about 0.3 amino acids (Neff et al. 2008). Our power analysis thus indicates that we had high power to detect an MHC-mediated mate preference of a similar magnitude to those that have been observed previously. It is also possible that the MHC class I instead influences mate choice in bluegill, as for example is the case in some other species (reviewed by Kamiya et al. 2014), although characterizing and testing the MHC class I was beyond the scope of our study.

Another potential challenge that may have reduced our power to detect non-random mating with respect to the MHC is the presence of a null allele, which we estimated to be present at a frequency of 24% in our population. To address this concern, we used a novel simulation approach in which we incorporated the null allele into our Monte Carlo simulations to assess if its presence would alter our conclusions. We found that a null allele was unlikely to have significantly altered our results, and that after accounting for the presence of the null allele, the offspring heterozygosity and amino acid differences between pairs still did not differ significantly from expectations under random mating. Although null alleles have previously been identified in studies of the MHC (Dionne et al. 2007), our study represents one of the first to incorporate a null allele at the MHC into tests of assortative mating. That the incorporation of a null allele into our mating simulations had no effect on our conclusions suggests that the presence of a null allele does not compromise the ability to quantify mating patterns with respect to the MHC.

An under-studied question surrounding MHC-mediated mating patterns is the role of extra-pair mating. In bluegill, extra-pair mating is driven by cuckolder males that use alternative reproductive tactics to gain paternity in the nests of parental males (Gross 1982). These alternative reproductive tactics may enhance the opportunity for MHC-mediated mate choice by enabling females to mate with a cuckolder male for genetic benefits (MHC compatibility) and a parental male for direct benefits (parental care). Anecdotally, we observed similar offspring heterozygosity and amino acid dissimilarity in parental-female pairs and cuckolder-female pairs, albeit we captured only a small number of cuckolder males during mating so we cannot rule out MHC-mediated mate choice between cuckolders and females. Alternatively, the presence of male alternative reproductive tactics could reduce the benefit for MHC-mediated mate choice when cuckolder males mate indiscriminately with respect to MHC compatibility. By undermining female preference for a mate (parental males in the bluegill system), these cuckolder males may actually impede the evolution of MHC-mediated mate choice in a system. However, this impediment may clearly be overcome in some cases, as for example in Atlantic salmon in which cuckolder-type males appear to mate indiscriminately with respect to the MHC, but female mating preferences still lead to an overall increase in MHC dissimilarity in their offspring (Consuegra and Garcia de Leaniz 2008). Regardless, the role of extra-pair mating and particularly the presence of specialized cuckolder reproductive tactics in the evolution of MHC-mediated mating patterns remains a rich area for further enquiry.

Acknowledgments We thank S. Colborne, M. Stefan, R. Knapp and A. Cunha for assistance with field collections. Two anonymous reviewers and the editor provided helpful comments that improved the manuscript. The Queen's University Biological Station provided facilities and equipment to support field collections.

Funding information This study was funded by the Natural Sciences and Engineering Council of Canada by grant RGPIN-2017-06045.

Data availability The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All applicable international, national and/or institutional guidelines for the care and use of animals were followed. All procedures were in accordance with the ethical standards of the Western University Animal Care Committee (Permit 2010-214).

References

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 364:33–39

Cohen S (2002) Strong positive selection and habitat-specific amino acid substitution patterns in MHC from an estuarine fish under intense pollution stress. Mol Biol Evol 19:1870–1880

Consuegra S, Garcia de Leaniz C (2008) MHC-mediated mate choice increases parasite resistance in salmon. Proc R Soc Lond B 275: 1397–1403

Dionne M, Miller KM, Dodson JJ, Caron F, Bernatchez L (2007) Clinal variation in MHC diversity with temperature: evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. Evolution 61:2154–2164

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

Forsberg LA, Dannewitz J, Petersson E, Grahn M (2007) Influence of genetic dissimilarity in the reproductive success and mate choice of brown trout—females fishing for optimal MHC dissimilarity. J Evol Biol 20:1859–1869

Fraser BA, Ramnarine IW, Neff BD (2010a) Selection at the MHC class IIB locus across guppy (*Poecilia reticulata*) populations. Heredity 104:155–167

Fraser BA, Ramnarine IW, Neff BD (2010b) Temporal variation at the MHC class IIb in wild populations of the guppy (*Poecilia reticulata*). Evolution 64:2086–2096

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 11:296–296

Garner SR, Neff BD (2013) Alternative male reproductive tactics drive asymmetrical hybridization between sunfishes (*Lepomis* spp.). Biol Lett 9:20130658

Garner SR, Bortoluzzi RN, Heath DD, Neff BD (2010) Sexual conflict inhibits female mate choice for major histocompatibility complex dissimilarity in Chinook salmon. Proc R Soc Lond B 277:885–894

Gasparini C, Congiu L, Pilastro A (2015) Major histocompatibility complex similarity and sexual selection: different does not always mean attractive. Mol Ecol 24:4286–4295

Gross MR (1979) Cuckoldry in sunfishes (*Lepomis*: Centrarchidae). Can J Zool 57:1507–1509

Gross MR (1982) Sneakers, satellites and parentals: polymorphic mating strategies in North American sunfishes. Z Tierpsychol 60:1–26

Gutierrez-Espeleta GA, Hedrick PW, Kalinowski ST, Garrigan D, Boyce WM (2001) Is the decline of desert bighorn sheep from infectious disease the result of low MHC variation? Heredity 86:439–450

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95–98

Hedrick PW (2002) Pathogen resistance and genetic variation at MHC loci. Evolution 56:1902–1908

Hoover B, Nevitt G (2016) Modeling the importance of sample size in relation to error in MHC-based mate-choice studies on natural populations. Integr Comp Biol 56:925–933

Immler S, Hamilton MB, Poslusny NJ, Birkhead TR, Epifanio JM (2011) Post-mating reproductive barriers in two unidirectionally hybridizing sunfish (Centrarchidae: *Lepomis*). J Evol Biol 24:111–120

Kalinowski ST, Taper ML (2006) Maximum likelihood estimation of the frequency of null alleles at microsatellite loci. Conserv Genet 7:991–995

Kamiya T, O'Dwyer K, Westerdahl H, Senior A, Nakagawa S (2014) A quantitative review of MHC-based mating preference: the role of diversity and dissimilarity. Mol Ecol 23:5151–5163

Kirkpatrick M (1996) Genes and adaptation: a pocket guide to the theory. In: Rose MR, Lauder GV (eds) Adaptation. Academic Press, San Diego, pp 125–146

Konkle BR, Philipp DP (1992) Asymmetric hybridization between two species of sunfishes (*Lepomis*: Centrarchidae). Mol Ecol 1:215–222

Kotiaho JS, LeBas NR, Puurtinen M, Tomkins JL (2008) On the resolution of the lek paradox. Trends Ecol Evol 23:1–3

Kuduk K, Babik W, Bellemain E, Valentini A, Zedrosser A, Taberlet P, Kindberg J, Swenson JE, Radwan J (2014) No evidence for the effect of MHC on male mating success in the brown bear. PLoS One 9:e113414

Landry C, Garant D, Duchesne P, Bernatchez L (2001) 'Good genes as heterozygosity': the major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*). Proc R Soc Lond B 268: 1279–1285

Langefors A, Lohm J, Grahn M, Andersen O, von Schantz T (2001) Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. Proc R Soc Lond B 268:479–485

Lenz TL, Eizaguirre C, Scharsack JP, Kalbe M, Milinski M (2009) Disentangling the role of MHC-dependent 'good genes' and 'compatible genes' in mate-choice decisions of three-spined sticklebacks *Gasterosteus aculeatus* under semi-natural conditions. J Fish Biol 75:2122–2142

Lenz TL, Hafer N, Samonte IE, Yeates SE, Milinski M (2018) Cryptic haplotype-specific gamete selection yields offspring with optimal MHC immune genes. Evolution 72:2478–2490

Liu HY, Xue F, Gong J, Wan QH, Fang SG (2017) Limited polymorphism of the functional MHC class II B gene in the black-spotted frog (*Pelophylax nigromaculatus*) identified by locus-specific genotyping. Ecol Evol 7:9860–9868

McConnell TJ, Godwin UB, Norton SF, Nairn RS, Kazianis S, Morizot DC (1998) Identification and mapping of two divergent, unlinked major histocompatibility complex class II B genes in *Xiphophorus* fishes. Genetics 149:1921–1934

Murphy K, Travers P, Walport M, Janeway C (2012) Janeway's immunobiology. Garland Science, New York

Neff BD (2001) Genetic paternity analysis and breeding success in bluegill sunfish (*Lepomis macrochirus*). J Hered 92:111–119

Neff BD (2004) Stabilizing selection on genomic divergence in a wild fish population. Proc Natl Acad Sci USA 101:2381–2385

Neff BD, Gross MR (2001) Dynamic adjustment of parental care in response to perceived paternity. Proc R Soc Lond B 268:1559–1565

Neff BD, Pitcher TE (2005) Genetic quality and sexual selection: an integrated framework for good genes and compatible genes. Mol Ecol 14:19–38

Neff BD, Fu P, Gross MR (2000) Microsatellite multiplexing in fish. Trans Am Fish Soc 129:584–593

Neff BD, Garner SR, Heath JW, Heath DD (2008) The MHC and non-random mating in a captive population of Chinook salmon. Heredity 101:175–185

Partridge CG, MacManes MD, Knapp R, Neff BD (2016) Brain transcriptional profiles of male alternative reproductive tactics and females in bluegill sunfish. PLoS One 11:e0167509

Penn DJ, Damjanovich K, Potts WK (2002) MHC heterozygosity confers a selective advantage against multiple-strain infections. Proc Natl Acad Sci USA 99:11260–11264

Potts WK, Manning CJ, Wakeland EK (1991) Mating patterns in semi-natural populations of mice influenced by MHC genotype. Nature 352:619–621

Promerová M, Králová T, Bryjová A, Albrecht T, Bryja J (2013) MHC class IIB exon 2 polymorphism in the grey partridge (*Perdix perdix*) is shaped by selection, recombination and gene conversion. PLoS One 8:e69135

Reusch TBH, Häberli MA, Aeschlimann PB, Milinski M (2001) Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. Nature 414:300–302

Sebastian A, Herdegen M, Migalska M, Radwan J (2016) AMPLISAS: a web server for multilocus genotyping using next-generation amplicon sequencing data. Mol Ecol Resour 16:498–510

Smith SB, Webster MS, Holmes RT (2005) The heterozygosity theory of extra-pair mate choice in birds: a test and a cautionary note. J Avian Biol 36:146–154

Stoltz JA, Neff BD (2006) Male size and mating tactic influence proximity to females during sperm competition in bluegill sunfish. Behav Ecol Sociobiol 59:811–818

Tregenza T, Wedell N (2000) Genetic compatibility, mate choice and patterns of parentage: invited review. Mol Ecol 9:1013–1027

Wedekind C, Walker M, Portmann J, Cenni B, Müller R, Binz T (2004) MHC-linked susceptibility to a bacterial infection, but no MHC-linked cryptic female choice in whitefish. J Evol Biol 17:11–18

Worley K, Collet J, Spurgin LG, Cornwallis C, Pizzari T, Richardson DS (2010) MHC heterozygosity and survival in red junglefowl. Mol Ecol 19:3064–3075

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–1118

Yeates SE, Einum S, Fleming IA, Megens HJ, Stet RJM, Hindar K, Holt WV, Van Look KJW, Gage MJG (2009) Atlantic salmon eggs favour sperm in competition that have similar major histocompatibility alleles. Proc R Soc Lond B 276:559–566

Zhao M, Wang Y, Shen H, Li C, Chen C, Luo Z, Wu H (2013) Evolution by selection, recombination, and gene duplication in MHC class I genes of two *Rhacophoridae* species. BMC Evol Biol 13:113