WYDZIAŁ BIOLOGII
i OCHRONY ŚRODOWISKA
Uniwersytet Łódzki

Stacjonarne Studia Doktoranckie
Ekologii i Ochrony Środowiska

Ewa Gabriela Pikus

Rola głównego kompleksu zgodności tkankowej w procesach synurbizacji ptaków

The role of the major histocompatibility complex in the processes of bird urbanization

Praca doktorska

wykonana w Katedrze Badania
Różnorodności Biologicznej,
Dydaktyki i Bioedukacji

Instytutu Ekologii i Ochrony Środowiska

pod kierunkiem
dr. hab. Piotra Miniasa, prof. UŁ

Łódź, 2021

Najserdeczniejsze podziękowania dla dr. hab. Piotra Miniasa, prof. UŁ za zaproszenie mnie do realizacji studiów doktoranckich pod jego opieką, merytoryczną pomoc i poświęcony czas oraz umożliwienie wzięcia udziału w pracach z tak fascynującym gatunkiem, jakim jest łyska.

Wszystkim Koleżankom i Kolegom z Katedry Badania Różnorodności Biologicznej, Dydaktyki i Bioedukacji bardzo dziękuję za nieustannie okazywaną serdeczność.

Dziękuję również moim rodzicom i przyjaciołom. W szczególności Mamie, za nieustającą wiarę we mnie oraz ś.p. Tacie, który przekazał mi zamiłowanie do nauki.

Spis treści

1 STRESZCZENIE 1

2 WSTĘP TEORETYCZNY 3

3 OBIEKT BADAŃ I METODY 8

3.1 OPIS GATUNKU 8

3.2 TEREN BADAŃ 9

3.3 METODY TERENOWE 10

3.4 METODY LABORATORYJNE 11

3.4.1 GENY GŁÓWNEGO UKŁADU ZGODNOŚCI TKANKOWEJ (MHC) 11

3.4.2 NEUTRALNA ZMIENNOŚĆ GENETYCZNA 12

3.5 ANALIZY STATYSTYCZNE I BIOINFORMATYCZNE 13

3.5.1 REKOMBINACJA, POLIMORFIZM I DOBÓR 13

3.5.2 ZWIĄZEK POLIMORFIZMU MHC KLASY I i II Z DOSTOSOWANIEM OSOBNIKÓW 13

3.5.3 ZRÓŻNICOWANIE GENETYCZNE POPULACJI 14

4 WYNIKI I WNIOSKI 15

5 BIBLIOGRAFIA 18

6 ROZDZIAŁ PIERWSZY 26

Pikus, E., Minias, P. Using de novo genome assembly and high-throughput sequencing to characterize the Major Histocompatibility Complex in a non-model rallid bird, the Eurasian coot *Fulica atra*. Scientific Reports (praca zgłoszona do druku).

7 ROZDZIAŁ DRUGI 56

Pikus, E., Dunn, P., Minias, P. High MHC diversity does not confer fitness advantage in a wild bird. Journal of Animal Ecology (praca zgłoszona do druku).

8 ROZDZIAŁ TRZECI 111

Pikus, E., Włodarczyk, R., Jedlikowski, J. and Minias, P., 2021. Urbanization processes drive divergence at the major histocompatibility complex in a common waterbird. PeerJ, *9*, e12264.

9 DOROBEK NAUKOWY 139

10 OŚWIADCZENIA WSPÓŁAUTORÓW 140

1 STRESZCZENIE

Cząsteczki głównego kompleksu zgodności tkankowej (ang. *major histocompatibility complex,* MHC) biorą udział w obronie przeciw patogenom wewnątrz- i zewnątrzkomórkowym (odpowiednio MHC klasy I i II). Geny MHC wykazują wysoki polimorfizm, szczególnie w obrębie miejsca wiążącego antygen (rowka), co umożliwia skuteczną obronę przed szerokim spektrum patogenów, zarówno na poziomie osobniczym jak i populacyjnym.

W niniejszej pracy, jako obiekt badań wykorzystano niemodelowy gatunek ptaka wodnego, łyskę *Fulica atra* (Rallidae, Gruifomes). Próbki krwi pobierano od ptaków z czterech populacji, z których dwie związane były z krajobrazem silnie zurbanizowanym (Łódź, Warszawa), a dwie z terenami pozamiejskimi. W efekcie, badaniami objęto dwa różne typy siedlisk, w których może występować różna presja patogenów. Celem pracy było zarówno scharakteryzowanie regionu MHC łyski, jak i określenie roli tych genów w procesach synurbizacji ptaków.

Przy pomocy sekwencjonowania całego genomu oraz celowanego genotypowania kluczowych eksonów (sekwencjonowanie wysokoprzepustowe) ustalono, że region MHC wykazuje u łyski nadzwyczaj wysoki poziom polimorfizmu i wysoką liczbę zduplikowanych genów. Jednocześnie wykazano, że dobór różnicujący i sygnał rekombinacji jest słabszy w przypadku genów MHC klasy I niż klasy II, co może wskazywać na różne trajektorie ewolucyjne tych genów.

Z kolei badania kondycji, ekspresji ornamentu oraz komponentów dostosowania (sukces reprodukcyjny) wykazały negatywny związek między różnorodnością genów MHC, a cechami związanymi z dostosowaniem. Wskazuje to na wysokie koszty ekspresji dużej liczby wariantów allelicznych MHC przez osobniki z populacji miejskich, gdzie fauna patogenów może charakteryzować się względnie niską różnorodnością.

Przeprowadzone badania wykazały również istotne zróżnicowanie międzypopulacyjne genów MHC klasy II w zależności od stopnia urbanizacji. Stwierdzono istotne zróżnicowanie genów MHC między populacjami miejskimi i pozamiejskimi, a populacje pozamiejskie charakteryzowały się najwyższą różnorodnością alleliczną MHC. Jednocześnie, nie wykryto istotnego neutralnego zróżnicowania genetycznego między populacjami, co sugeruje że zróżnicowanie MHC może być efektem adaptacji immunogenetycznych ptaków do życia w siedliskach miejskich.

Podsumowując, wyniki niniejszej rozprawy dostarczają nowych informacji o zróżnicowaniu genów MHC u ptaków i mechanizmów, które to zróżnicowanie kształtują, ale również zwiększają nasze zrozumienie mikroewolucyjnych adaptacji zwierząt do życia w silnie przekształconym krajobrazie miejskim.

SUMMARY

Molecules of the major histocompatibility complex (MHC) are involved in the defense against intra- and extracellular pathogens (MHC class I and II, respectively). In general, MHC genes show extreme polymorphism, especially within the antigen binding region (groove), which enables effective recognition of diverse pathogens, both at the individual and population level.

In this study, a non-model species of waterbird, the Eurasian coot *Fulica atra* (Rallidae, Gruifomes) was used as the research object. Blood samples were taken from birds from four populations, two associated with a highly urbanized landscape (Łódź, Warsaw), and the other two from non-urban areas. As a result, the study covered two different types of habitats that may be subject to different pathogen pressure. The aim of the study was both to characterize the MHC region of the coot and to determine the role of these genes in the processes of bird urbanization.

Using whole genome sequencing and targeted genotyping of key exons (high-throughput sequencing), it was found that the MHC region in coots shows an exceptional level of polymorphism and a high number of duplicated genes. At the same time, it was shown that diversifying selection and recombination signal were weaker at MHC class I than class II, which may indicate different evolutionary trajectories of these genes.

On the other hand, the study of condition, ornament expression and components of fitness (reproductive success) showed a negative relationship between the diversity of MHC genes and the fitness-related traits. This indicates that expressing large number of MHC allelic variants may impose high costs on individuals from urban populations, where pathogen fauna may have relatively low diversity.

The research also showed a significant inter-population differentiation of MHC class II genes depending on the degree of urbanization. Significant differentiation of MHC genes was found between urban and non-urban populations, and non-urban populations were characterized by the highest allelic diversity of the MHC. At the same time, no significant neutral genetic variation between populations was detected, suggesting that MHC variation may be the result of immunogenetic adaptation of birds to urban life.

Taken together, the results of this dissertation not only provide a new insight into the mechanisms that shape diversity of MHC genes in birds, but they also enhance our understanding of the microevolutionary adaptations of animals to highly transformed urban landscape.

2 WSTĘP TEORETYCZNY

W ciągu ostatnich dekad wpływ człowieka na ekosystemy naturalne zasadniczo się zmienił. Do niedawna termin „ekosystemy zdominowane przez człowieka" przywoływał obrazy nie tylko krajobrazów miejskich, ale także mozaiki małoobszarowych pól uprawnych oraz rozległych pastwisk. Jednak wraz ze wzrostem populacji ludzkiej i postępem technologicznym zakres i charakter antropogenicznej modyfikacji krajobrazu uległ drastycznemu pogłębieniu (Vitousek et al. 1997). W ciągu ostatnich 100 lat nastąpiła ogromna zmiana w globalnej strukturze populacji ludzkiej, przy czym większość ludzi żyje obecnie w środowiskach miejskich, a nie wiejskich (Cox et al. 2018). Według danych ONZ w 2030 roku na świecie będzie żyło 8,55 mld ludzi (ONZ 2019). Szacuje się, że w 2016 roku tereny miejskie zasiedlało 54,4 procent ludności świata. Ze względu na wzrost liczby ludności i migrację do 2030 roku tereny miejskie mają pomieścić 60 procent ludzi na całym świecie, a co trzeci mieszkaniec będzie mieszkał w miastach o powierzchni co najmniej pół miliona mieszkańców (Camacho-Valdez et al. 2019).

Upraszczając, proces urbanizacji jest skrają formą intensyfikacji użytkowania ziemi, powodującą redukcję i fragmentację siedlisk przyrodniczych (Sol et al. 2020). Zmiany środowiskowe związane z procesem urbanizacji obejmują wylesianie, ekspansję rolnictwa, eutrofizację, zanieczyszczenia powietrza i wody, hałas, izolację i utratę naturalnych siedlisk (Croci et al. 2008; Vitousek et al. 1997). Na niektóre obszary zdominowane przez człowieka zostały wprowadzone egzotyczne gatunki zwierząt bądź roślin (Grimm et al. 2008; Fitzherbert et al. 2008), które mogą wypierać gatunki rodzime (Faeth et al. 2005; McConkey et al. 2012). Te często ekstremalne przekształcenia ekosystemowe powodują szybki spadek bogactwa i liczebności gatunków notowany na całym świecie, jako że krajobraz miejski może wykluczać wiele gatunków zwierząt, szczególnie tych mniej plastycznych i bardziej wyspecjalizowanych ekologicznie (Concepción et al. 2015; Luniak, 2004). Urbanizacja wpływa więc na negatywnie na stan naturalnych populacji wielu gatunków, ale jednocześnie ekosystemy miejskie tworzą nowe nisze ekologiczne (Aouissi et al. 2017; Green et al. 2016; Croci et al. 2008). W efekcie, z jednej strony urbanizacja powoduje spadek bioróżnorodności i lokalne wymieranie populacji, z drugiej niektóre gatunki dobrze prosperują na terenach miejskich (McKinney 2006; Lowry et al. 2013). Coraz więcej gatunków przystosowuje się do nowych antropogenicznych warunków środowiskowych i kolonizuje miasta, pokonując bariery ekologiczne, demograficzne, pokarmowe i behawioralne (Luniak 2004). Stąd, ważnym jest, aby zrozumieć w jaki sposób populacje wykorzystują zasoby zarówno w środowiskach naturalnych, jak i zmodyfikowanych przez człowieka. Okazuje się, że obszary miejskie mogą być alternatywnym, a czasem nawet podstawowym miejscem życia gatunków, które potrafią wykorzystać nowe warunki panujące

w miastach, w tym dostęp do antropogenicznego pokarmu, łagodniejszy mikroklimat zimą oraz niższą presję ze strony naturalnych drapieżników (Luniak 2004). Z drugiej strony wysoka presja drapieżnicza ze strony psów i kotów może powodować problemy z eksploatacją zasobów miejskich przez dzikie zwierzęta (Tryjanowski et al. 2016). Badania wykazały, że koty domowe żyjące na wolności stanowią zagrożenie dla wielu populacji ptaków, w tym gatunków priorytetowych dla ochrony (Dauphiné & Cooper 2009). Kolejnym wyzwaniem związanym z procesem kolonizacji miast mogą być między-siedliskowe różnice w presji patogenów wzdłuż gradientu urbanizacji. Na zmiany w presji patogenów może mieć wpływ szereg cech środowiska miejskiego, w tym zwiększona obecność gatunków egzotycznych (Boal et al. 1998), gorsza jakość siedliska (Patz et al. 2004), potencjalna transmisja horyzontalna ze strony gatunków udomowionych (Lehrer 2010) bądź transmisja poprzez kontakt z ich odchodami (Bradley & Altizer 2007; Frenkel et al. 1995) oraz zmiany klimatyczne związane z występowaniem wysp ciepła (Cumming & Van Vuuren 2006).

Urbanizacja może wpływać negatywnie na występowanie naturalnych patogenów, ale z drugiej strony tempo ich transmisji między gospodarzami może w niektórych przypadkach być istotnie wyższe w krajobrazie miejskim (Łoś et al. 2020; Bradley & Altizer 2007). Podstawową barierą dla infekcji jest układ odpornościowy gospodarza, który odgrywa kluczową rolę w kształtowaniu się interakcji gospodarz – patogen (Combes 2001). Zatem obecność innego spektrum patogenów na terenach miejskich (w stosunku do siedlisk naturalnych) może wymagać niezbędnych adaptacji układu odpornościowego gospodarzy (Alcaide et al. 2008; Watson et al. 2017). Dodatkowo populacje miejskie mogą cierpieć z powodu dryfu genetycznego i chowu wsobnego (Lourenço et al. 2017), co z kolei może skutkować niskim poziomem różnorodności genetycznej, w tym różnorodności receptorów odpornościowych odpowiedzialnych za rozpoznawanie antygenów (Alcaide 2010; Strandet al. 2012; Sutton et al. 2011). Dryf genetyczny może utrwalać również mutacje upośledzające funkcje organizmu, w tym odpowiedź immunologiczną (Radwan et al. 2010). Ma to znaczenie szczególnie w populacjach o małej różnorodności genetycznej, które przeszły przez epizod tzw. wąskiego gardła genetycznego (Stervander et al. 2020). Z drugiej strony, przestrzennie niejednorodna presja selekcyjna ze strony patogenów może skutkować lokalną adaptacją (Loiseau et al. 2009). Ponieważ wielkoskalowe czynniki antropogeniczne mogą wpływać na funkcje odpornościowe zwierząt, coraz ważniejsze staje się zrozumienie źródeł zmienności immunologicznej w kontekście zmian środowiskowych (Acevedo-Whitehouse & Duffus 2009; Martin et al. 2010).

Komponenty układu odpornościowego odpowiedzialne za identyfikację obcych antygenów są kodowane przez dwie podstawowe rodziny genów: receptory Toll-podobne (ang. *toll-like receptors*, TLR) i główny układ zgodności tkankowej (MHC) (Schenten & Medzhitov 2011). Ogólnie rzecz biorąc,

TLR są związane z wrodzonym układem odpornościowym i są odpowiedzialne za identyfikację konserwatywnych antygenów (wzorce molekularne związane z patogenem, ang. *pathogen-associated molecular patterns*, PAMP), które są charakterystyczne dla szerokiego spektrum patogenów i pasożytów (Barton & Miedzhitow 2020). Z kolei cząsteczki MHC aktywują nabytą odpowiedź immunologiczną (Ekblom et al. 2007), prezentując komórkom układu odpornościowego specyficzne antygeny pochodzące z procesowania patogenów wewnątrzkomórkowych (MHC klasy I) lub zewnątrzkomórkowych (MHC klasy II) (Piertney & Oliver 2006). Antygeny związane z MHC klasy I pochodzą z białek wewnątrzkomórkowych patogenów przetworzonych na drodze proteasomalnej proteolizy i prezentowane są limfocytom T CD8+, co może skutkować bezpośrednią destrukcją lub uruchomieniem procesów apoptozy zakażonej komórki. Cząsteczki MHC klasy II prezentują limfocytom T CD4+ antygeny egzogenne pochodzące z białek zewnątrzkomórkowych patogenów przetworzonych na drodze endocytozy lub autofagii, co umożliwia indukcję odpowiedniego szlaku odpowiedzi immunologicznej (Steimle et al. 1993; Münz 2012; Duraes et al. 2015). Antygeny związane z cząsteczkami MHC klasy I i II to niewielkie peptydy długości 8–10 lub 15–20 aminokwasów. Wiązane są one przez najbardziej polimorficzny region cząsteczek MHC, tzw. rowek wiążący peptyd (ang. *peptide binding region*, PBR) (Duraes et al. 2015). Rowek wiążący peptyd o określonej sekwencji aminokwasów może wiązać się z ograniczoną liczbą obcych peptydów, a zatem liczba różnych cząsteczek MHC w organizmie określa spektrum patogenów, które mogą zostać rozpoznane poprzez mechanizmy nabytej obrony immunologicznej (O'Connor et al. 2016).

Z perspektywy ewolucyjnej i populacyjnej, dominującą cechą MHC jest ekstremalna różnorodność nukleotydowa w obszarze rowka wiążącego peptyd. Nawet niewielkie zmiany aminokwasowe w PBR mogą prowadzić do dużych różnic w repertuarze identyfikowanych peptydów i w konsekwencji diametralnie zmienić spektrum patogenów, przeciwko którym organizm skutecznie aktywuje nabytą odpowiedź immunologiczną (Dionne et al. 2007). W wyniku wyścigu zbrojeń gospodarz-patogen geny MHC wykazują najwyższy opisany u kręgowców poziom funkcjonalnego polimorfizmu (Gaudieri et al. 2000 ; Bernatchez & Landry 2003; Garrigan & Hedrick 2003) utrzymywanemu w populacjach na drodze doboru równoważącego (ang. *balancing selection;* Spurgin & Richardson 2010), a ich skład alleliczny na poziomie populacji powinien szybko reagować na zmiany w składzie fauny patogenów poprzez lokalną adaptację (Ekblom, et al. 2007; Miller et al. 2010) oraz warunkować dostosowanie osobników.

Zależności między polimorfizmem MHC a dostosowaniem mogą powstawać na drodze kilku alternatywnych mechanizmów. Po pierwsze, mechanizm przewagi rzadkich alleli zakłada (dobór negatywnie zależny od częstości), że osobniki z allelami MHC o niskich częstościach w populacji będą

posiadały wyższe dostosowanie, jako że ich układ odpornościowy może lepiej wykrywać i zwalczać patogeny, które ewoluowały aby uniknąć typowych (najbardziej rozpowszechnionych w populacji) alleli MHC gospodarza (Brouwer et al. 2010; Slade et al. 1992). W takiej sytuacji powinniśmy obserwować wpływ konkretnych alleli MHC na dostosowanie osobników (tzw. allele odpornościowe). Alternatywnie, najkorzystniejsza pod względem dostosowania może być maksymalna różnorodność alleli na poziomie osobniczym – zgodnie z mechanizmem przewagi heterozygotycznej większa liczba alleli ulegająca ekspresji w organizmie zwiększa spektrum rozpoznawanych antygenów, a tym samym zapewnia selektywną przewagę w zwalczaniu infekcji i poprawia dostosowanie osobnika (Doherty et al. 1975; McClelland et al. 2003). W takim wypadku można spodziewać się pozytywnej liniowej zależności między heterozygotycznością MHC i dostosowaniem. Zgodnie z tą hipotezą, geny MHC podlegały na drodze ewolucji procesowi intensywnej duplikacji, który umożliwił zwiększenie indywidualnej różnorodności allelicznej MHC. Z drugiej strony, wydaje się, że liczba loci MHC u poszczególnych gatunków (a tym samym liczba alleli u poszczególnych osobników) nie mogą zwiększać się w sposób nieograniczony. Zbyt duża różnorodność MHC może być niekorzystna ze względu na zwiększone ryzyko chorób autoimmunologicznych oraz zmniejszenie repertuaru antygenów rozpoznawanych przez receptory limfocytów T (ang. *T-cell receptor*, TCR) (Roved et al. 2018). Niedawne badania na nornicach rudych *Myodes glareolus* potwierdziły, że obniżenie repertuaru TCR wiąże się z wysoką osobniczą różnorodnością w obrębie genów MHC klasy I, ale nie wykazano podobnej zależności dla genów MHC klasy II (Migalska et al. 2019). Mechanizm ten był podstawą sformułowania hipotezy optymalności (Wegner et al. 2003), mówiącą że osobniki o optymalnej (pośredniej), a nie maksymalnej różnorodności alleli MHC będą osiągały najwyższe dostosowanie w populacjach (nieliniowa zależność między heterozygotycznością MHC i dostosowanie). Interpretacja mechanizmów warunkujących działanie doboru naturalnego na geny MHC jest szczególnym wyzwaniem, ponieważ różne jego rodzaje mogą działać w synergii i nie wykluczają się wzajemnie (Spurgin & Richardson 2010).

Kolejną charakterystyczną cechą genów MHC, jest wysoki poziom duplikacji oraz wysoki poziom podobieństwa między zduplikowanymi genami zarówno między gatunkami, jak również w ich obrębie (Miller & Lambert 2004). Przykładowo, liczba kopii genów MHC u ptaków kształtuje się w zakresie 1-33 dla klasy I i 1-23 dla klasy II (Minias et al. 2019), choć najnowsze badania wskazują, że w niektórych ptasich liniach ewolucyjnych może być ona znacznie wyższa (He et al. 2021). Podobieństwo między zduplikowanymi genami spowodowane jest głównie rozpowszechnionym u ptaków mechanizmem ewolucji zespołowej (ang. *concerted evolution*), w wyniku którego różne loci w obrębie gatunku ewoluują w podobny sposób, co powoduje ich homogenizację (Goebel et al. 2017; Reusch et al. 2004; Hess & Edwards 2002). Generuje to metodologiczne trudności w genotypowaniu

MHC u organizmów niemodelowych, szczególnie w projektowaniu starterów specyficznych dla pojedynczego locus. W efekcie, nie ma zwykle możliwości genotypowania MHC tradycyjnymi metodami Sangera i dopiero rozwój technik sekwencjonowania następnej generacji (ang. *next generation sequencing*, NGS) zdecydowanie rozwinął badania nad MHC, umożliwiając genotypowanie wielu zduplikowanych loci jednocześnie (Babik et al. 2010). Kolejnym krokiem milowym w badaniach MHC był rozwój technik związanych z sekwencjonowaniem całych genomów, które nie skupiają się genotypowaniu pojedynczych eksonów (jak ma to zazwyczaj miejsce w przypadku populacyjnych badań z wykorzystaniem metod NGS), ale dają obraz całych genów, a często nawet architektury całego regionu MHC (He et al. 2021).

Celem niniejszej pracy było:

1. Scharakteryzowanie regionu MHC u niemodelowego gatunku ptaka, łyski *Fulica atra* (Rallidae, Gruifomes) przy pomocy sekwencjonowania całego genomu oraz celowanego genotypowania kluczowych eksonów z wykorzystaniem sekwencjonowania wysokoprzepustowego.
2. Przetestowanie zależności między różnorodnością genów MHC, a cechami związanymi z dostosowaniem (kondycja, cechy reprodukcyjne, ekspresja ornamentów) u łyski.
3. Analiza różnorodności oraz zróżnicowania między-populacyjnego genów MHC łyski w gradiencie urbanizacji.

3 OBIEKT BADAŃ I METODY

3.1 OPIS GATUNKU

Łyska *Fulica atra* to gatunek pospolitego ptaka wodnego z rodziny chruścieli Rallidae (rząd żurawiowych Gruiformes), który dzieli się na cztery podgatunki. W Polsce można spotkać podgatunek nominatywny *F. a. atra* (Rycina 1), który występuj niemal w całe Europie, Azji Środkowej, Japonii oraz w północnej Afryce. Są to ptaki szeroko rozpowszechnione i licznie występujące w różnorodnych siedliskach wodnych takich jak jeziora, zbiorniki retencyjne czy wolno płynące rzeki, preferują wody słodkie, rzadko występują na wodach zasolonych, na miejsca gniazdowe najczęściej wybierają wody płytkie, z mulistym dnem i obfitą roślinnością szuwarową. Chętnie nurkują, zwykle żerują zarówno na lądzie jak i w wodzie, dieta jest różnorodna, częściowo roślinna, ale jej ważnym komponentem są również bezkręgowce takie jak ślimaki, małże i owady. Łyska jest gatunkiem średniej wielkości, długość ciała wynosi 36-38 cm, rozpiętości skrzydeł 70-80 cm, a waga mieści się w zakresie 600 – 1265 g. Dorosłe osobniki charakteryzują się czarnym upierzeniem oraz białym dziobem i białą tarczą czołową, brak jest wyraźnego dymorfizmu płciowego, choć samce z reguły są większe od samic. W okresie lęgowym, który trwa od marca do lipca, łyski łączą się w monogamiczne pary, są agresywne i terytorialne. Samica składa zwykle 6-10 jaj, a niektóre pary mogą wyprowadzić dwa lęgi w ciągu jednego sezonu. W Europie Środkowej część ptaków prowadzi osiadły tryb życia, jednak większość migruje na zimę w stronę zachodniej i południowej części kontynentu, w okresie nielęgowym chętnie łączy się w stada mogące liczyć setki bądź tysiące osobników. Okres migracyjny trwa od marca do kwietnia (przylot na tereny lęgowe) i od października do listopada (odlot na tereny zimowiskowe). Łyska należy do gatunków, które w ostatnich dziesięcioleciach skutecznie skolonizowały stale rozrastające się obszary miejskie na terenie całej Europy, co wiązało się nie tylko z szeregiem adaptacji behawioralnych, ale powinno także sprowokować dostosowanie się układu odpornościowego do fauny patogenów występujących w miastach.

Rycina 1. Łyska *Fulica atra*.

3.2 TEREN BADAŃ

Badania prowadzone były na czterech naturalnych populacjach łyski położonych w regionie Polski Centralnej. Dwie wybrane do badań populacje związane były z terenami zurbanizowanymi (Łódź i Warszawa), a dwie z terenów pozamiejskimi (Sarnów i Żeromin). Warszawa (52,259° N, 21.020° E) jest największym miastem w Polsce pod względem liczby ludności (1,79 mln mieszkańców w 2020 roku, https://stat.gov.pl) i powierzchni (517,24 km²); charakteryzuje się również dużą gęstością zaludnienia (3462 os./km^2). Miasto zostało skolonizowane przez łyski około połowy XX wieku lub nawet wcześniej. Stabilną populację lęgową łysek w obrębie silnie zurbanizowanych części Warszawy odnotowano w latach 1960-tych (Luniak et al. 1964), a w latach 1980-tych wielkość populacji została oszacowana na 190-210 par lęgowych (Luniak et al. 2001). Stosunkowo wczesnemu skolonizowaniu Warszawy przez łyski sprzyjała prawdopodobnie obecność dużej rzeki (Wisła) w centrum miasta. Z kolei Łódź (51.757° N, 19.493° E) jest trzecim co do wielkości miastem w Polsce pod względem liczby ludności (677 tysięcy mieszkańców w 2020 roku, https://stat.gov.pl), jej obszar administracyjny obejmuje 293,25 km², a gęstość zaludnienia wynosiła 2309 osób/km^2 w 2020 roku. W przeciwieństwie

do Warszawy, obszar aglomeracji łódzkiej został skolonizowany przez łyski stosunkowo niedawno. W latach 1994-2002 odnotowano około 20 miejsc gniazdowania łysek wyłącznie na obrzeżach Łodzi (Janiszewski et al. 2009). W czasie ostatnich dwóch dekad łyski skolonizowały również centrum miasta, a w okresie niniejszych badań liczebność populacji oszacowano na ok. 70 par lęgowych, w tym ok 40 par lęgowych w centrum miasta. Ze względu na różnice w czasie kolonizacji, populacje Warszawy i Łodzi zostały zaklasyfikowane odpowiednio jako stara i nowa populacja miejska. Kolejne dwie populacje łysek zlokalizowane były na obszarze stawów rybnych w Sarnowie (51,851° N, 19,109° E) i Żerominie (51,617° N, 19,607° E), które zapewniały rozległe siedliska lęgowe z obfitą roślinnością szuwarową, podobne w swojej strukturze do typowych siedlisk naturalnych tego gatunku. W obu lokalizacjach wstęp nieautoryzowanego personelu był ograniczony, co skutkowało względnie niskim stopniem antropopresji.

3.3 METODY TERENOWE

Prace terenowe wykonywane były w latach 2010-2020. Ptaki łapano za pomocą pułapek nagniazdowych lub ręcznie, podczas inkubacji na gnieździe lub podczas żerowania na brzegu. Każdy ptak został oznaczony metalową obrączką na lewym skoku oraz białą plastikową obrożą założoną na szyję. Każda obroża posiadała indywidualny kod alfanumeryczny, co pozwalało na łatwą identyfikację ptaków w terenie i uniknięcie wielokrotnego chwytania tych samych osobników. Podczas chwytania pobierano ok. 50 µl krwi do analiz genetycznych. Próbki krwi pobrano w okresie rozrodczym od 20-27 dorosłych osobników z populacji warszawskiej oraz obu populacji pozamiejskich (w sumie n = 73 osobniki), natomiast populacja łódzka poddana została długoterminowemu monitoringowi i pozyskano z niej próby krwi od ponad 200 osobników.

W ramach monitoringu populacji łódzkiej badano dodatkowo kondycję, ekspresję ornamentu oraz komponenty dostosowania (sukces reprodukcyjny) łyski. Podczas chwytania wykonywano podstawowe pomiary biometryczne, w tym długość skoku i całkowitą długość głowy (± 0,1 mm), mierzono także masę ciała (± 1 g). Mierzono również rozmiar (szerokość i długość, ± 0,1 mm) przypuszczalnego ornamentu (białej tarczki czołowej). W końcu, pobierano ok. 5 µl krwi do pomiaru całkowitego stężenia hemoglobiny. Masa ciała została użyta jako ogólna miara kondycji, ponieważ po odpowiedniej korekcie względem strukturalnej wielkości ciała może być wiarygodnym wskaźnikiem rezerw energetycznych (Peig & Green 2009). Całkowite stężenie hemoglobiny we krwi było wskaźnikiem stanu fizjologicznego, gdyż odzwierciedla zdolność organizmu do zaspokojenia zapotrzebowania na tlen, a jednocześnie koreluje z ogólnym stanem kondycyjnym i zdrowotnym osobników (Minias 2015). Z kolei rozmiar blaszki czołowej został użyty do wskaźnik ekspresji tego

ornamentu zarówno dla samców jak i samic. Mimo braku bezpośrednich informacji dotyczących roli blaszki czołowej w doborze płciowym u łyski, wykazano że ekspresja tego rodzaju struktur morfologicznych u innych gatunków chruścieli jest warunkowana poziomem testosteronu i wiarygodnie sygnalizuje status społeczny, zdolność dominacji, agresywność i kondycję osobników (Crowley & Magrath 2004; Alvarez et al. 2005; Dey et al. 2014). Wszystkie powyższe pomiary wykonano dla 114 schwytanych osobników.

W łódzkiej populacji badano również parametry reprodukcyjne łyski, w tym określano datę zniesienia pierwszego jaja, wielkość zniesienia, sukces klucia (0 – brak sukcesu, 1 – co najmniej jedno pisklę wyklute) i sukces lęgowy (liczba odchowanych piskląt). Dane te zebrano dla ponad 230 lęgów wyprowadzonych przez 106 oznakowanych osobników, od których pobrane zostały próbki krwi do analiz genetycznych.

3.4 METODY LABORATORYJNE

W pierwszej kolejności z próbek krwi wyizolowano genomowe DNA z wykorzystaniem zestawu GeneJET Genomic DNA Purification Kit (Fermentas, Thermo Fisher Scientific, Waltham, MA, USA). Płeć wszystkich schwytanych ptaków oznaczono molekularnie na drodze amplifikacji genów chromodomeny helikazy wiążącej DNA (CHD) zlokalizowanych na chromosomach płciowych (Griffiths et al. 1998).

3.4.1 GENY GŁÓWNEGO UKŁADU ZGODNOŚCI TKANKOWEJ (MHC)

Następnie zgenotypowano fragmenty genów MHC klasy I i II, które kodują rowek wiążący peptyd (PBR), odpowiedzialny za wiązanie antygenu. W celu zgenotypowania MHC klasy I użyto starterów zaprojektowanych pierwotnie dla ptaków szponiastych Accipitriformes (Alcaide et al. 2009) wiążących się z regionem flankującym intronu 2 i konserwatywnym regionem eksonu 4, co pozwoliło na amplifikację całego eksonu 3 kodującego domenę α_2 PBR. Całkowita długość amplikonu wynosiła 411 pz, w tym sekwencja eksonu 3 miała długość 276 pz. Z kolei w celu zgenotypowania MHC klasy II użyto starterów zaprojektowanych w trakcie wcześniejszych badań nad łyską (Alcaide et al. 2014), które wiążą się z regionem flankującym intronu 1 i konserwatywnym regionem eksonu 3, co pozwoliło na amplifikację całego eksonu 2 MHC klasy II, kodującego domenę β PBR (270 pz). W celu wykrycia pozytywnych amplifikacji wszystkie produkty PCR oceniano za pomocą elektroforezy na 2% żelu agarozowym. Ponieważ spodziewaliśmy się, że nasze startery mogą amplifikować więcej niż jeden locus MHC każdej klasy, w celu genotypowania osobników wykorzystaliśmy metody

sekwencjonowania następnej generacji (NGS). Wszystkie produkty PCR oczyszczono przy użyciu kulek magnetycznych AMPure XP (Beckman Coulter, Brea, CA, USA), a ich stężenia oceniono za pomocą zestawu do znakowania dsDNA Quant-iT PicoGreen (Thermo FisherScientific, Waltham, MA, USA). Biblioteki przygotowano oddzielnie dla MHC klasy I i klasy II z równomolowych ilości produktów PCR przy użyciu zestawu NEB-Next DNA Library Prep Master Mix Set (New England Biolabs, Ipswich, MA, USA) i zsekwencjonowano na 2 x 250 pz Illumina MiSeq Platforma. Procesowanie surowych danych uzyskanych z sekwencjonowania Illumina (filtrowanie, grupowanie i walidacja wariantów allelicznych) zostało wykonane z wykorzystaniem serwera AmpliSAS (Sebastian et al. 2016) i protokołów opracowanych przez Biedrzycką et al. (2017).

W celu dokładniejszej charakteryzacji architektury regionu MHC (określenie liczby kopii zduplikowanych genów) oraz polimorfizmu eksonów nie kodujących PBR, dla pojedynczego osobnika z naszej populacji zsekwencjonowaliśmy oraz złożyliśmy *de novo* cały genom. Biblioteki przygotowano z wykorzystaniem zestawu Chromium Genome Library & Gel Bead Kit v2 (10x Genomics, Pleasanton, CA, USA), natomiast sekwencjonowanie wykonano wykorzystując technikę Illumina NovaSeq 6000 (Illumina Inc., San Diego, CA, USA). W trakcie sekwencjonowania uzyskano niemal 900 milionów sparowanych odczytów, które zostały przeprocesowane i złożone w genom z wykorzystaniem programu Supernova v2.1.1 (10x Genomics). Pełen rozmiar uzyskanego genomu wynosił 1168 Mb, a wartość N50 (dla contigów) wynosiła 0.25 Mb. Anotacja genomu została wykonana z wykorzystaniem programu MAKER v3.01.1 (Holt & Yandell 2011) w oparciu o dostępne genomy innych gatunków chruścieli i mewowców Lari. Sekwencjonowanie, złożenie i anotacja genomu zostały wykonane w Carver Biotechnology Center na Uniwersytecie Illinois w Urbana-Champaign, USA. Pełne sekwencje genów MHC klasy I i II zostały wyekstrahowane z genomu z wykorzystaniem algorytmów BLAST, bazując na dostępnych w bazie GenBank sekwencjach pochodzących od innych ptaków niewróblowych.

3.4.2 NEUTRALNA ZMIENNOŚĆ GENETYCZNA

Aby uzyskać punkt odniesienia dla zmienności genów MHC, oszacowaliśmy również poziom neutralnej zmienności genetycznej w badanych przez nas populacjach. Po pierwsze, na poziomie populacyjnym zmienność neutralną określaliśmy z wykorzystaniem dziesięciu wysoce polimorficznych markerów mikrosatelitarnych (Rozdział 3, Tabela 1). Po drugie, aby oszacować heterozygotyczność osobników, dla których mierzone były cechy fenotypowe (kondycja, ekspresja ornamentu) oraz rozrodcze (Rozdział 2), wykorzystaliśmy sekwencjonowanie metodą ddRAD (ang. *double digest restriction-site association DNA sequencing*). Heterozygotyczność każdego osobnika została wyliczona

w oparciu o panel 14525 polimorfizmów pojedynczego nukleotydu (ang. *single nucleotide polymorphism*, SNP), szczegóły procesowania danych ddRAD podane są w Rozdziale 2.

3.5 ANALIZY STATYSTYCZNE I BIOINFORMATYCZNE

3.5.1 REKOMBINACJA, POLIMORFIZM I DOBÓR

W celu scharakteryzowania polimorfizmu MHC klasy I i II (dane NGS i genomowe) wykorzystano oprogramowanie DnaSP v.6.10.3 (Rozas et al. 2017). Polimorfizm sekwencji oceniono jako liczbę miejsc segregujących (polimorficznych), całkowitą liczbę mutacji, średnią różnorodność nukleotydów i średnią liczbę różnic nukleotydów. Aby określić mechanizmy odpowiedzialne za utrzymywanie polimorfizmu MHC w populacji przeanalizowany został sygnał doboru oraz rekombinacji (konwersji genów) w obrębie eksonu 3 MHC klasy I oraz eksonu 2 MHC klasy II (dane NGS). Aby określić siłę natężenia doboru obliczono wskaźnik dN/dS, odzwierciedlający proporcję niesynonimicznych (dN) do synonimicznych (dS) podstawień nukleotydowych. Sygnaturę dobru różnicującego (pozytywnego, $dN/dS > 1$) i oczyszczającego (negatywnego, $dN/dS < 1$) oceniono za pomocą dwóch podejść, wnioskowania Bayesowskiego (ang. *Fast Unconstrained Bayesian AppRoximation*, FUBAR) oraz metod maksymalnego prawdopodobieństwa (ang. *Fixed Effects Likelihood*, FEL), dostępnych na serwerze Datamonkey (Weaver et al. 2018). W celu identyfikacji miejsc podlegających epizodycznemu doborowi różnicującemu (sygnał widoczny jedynie dla części wariantów allelicznych) wykorzystany został model MEME (ang. *Mixed Effect Model of Evolution*). Sygnał rekombinacji określono przy użyciu oprogramowania RDP v.4.97, które łączy kilka podstawowych algorytmów opracowanych do wykrywania sekwencji rekombinowanych (Martin et al. 2015).

3.5.2 ZWIĄZEK POLIMORFIZMU MHC KLASY I i II Z DOSTOSOWANIEM OSOBNIKÓW

Zależność różnorodności genów MHC klasy I i II z cechami związanymi z dostosowaniem (fenotypowym i reprodukcyjnym) analizowano za pomocą uogólnionych mieszanych modeli liniowych (ang. *generalized linear mixed models,* GLMM) dostępnych w pakiecie *glmmADMB* (Skaug et al. 2012) działającym w środowisku statystycznym R (R Foundation for Statistical Computing, Wiedeń, Austria). Szczegółowa struktura modeli opisana została w Rozdziale 2.

3.5.3 ZRÓŻNICOWANIE GENETYCZNE POPULACJI

Całkowitą liczbę alleli oraz liczbę alleli specyficznych dla poszczególnych populacji (MHC klasy II i markery mikrosatelitarne) obliczono za pomocą oprogramowania GeneAlEx v.6.5. Dla markerów mikrosatelitarnych obliczono również średnią obserwowaną heterozygotyczność dla każdej populacji, podczas gdy dla MHC obliczono miary polimorfizmu (liczba miejsc segregujących, liczba różnic nukleotydów, różnorodność nukleotydów) dla wariantów allelicznych stwierdzanych w obrębie każdej populacji przy użyciu DnaSP v6.10.03. Wszystkie obliczenia powtórzono również dla wystandaryzowanej wielkości próby (20 osobników na populację).

Aby ocenić zróżnicowanie genetyczne między populacjami, zastosowaliśmy dwa podejścia metodyczne. Najpierw obliczyliśmy miary zróżnicowania D Josta (Jost, 2008) zarówno dla MHC (program SpadeR; Chao et al. 2016), jak i dla markerów mikrosatelitarnych (pakiet R diveRsity; Keenan et al. 2013). Dodatkowo dla markerów mikrosatelitarnych obliczone zostały wystandaryzowane wartości G'_{ST} (surowe wartości G_{ST} podzielone przez maksymalną zanotowaną wartość; Hedrick 2005). Po drugie, zastosowano algorytm grupowania Bayesowskiego dostępny w programie STRUCTURE (Pritchard et al. 2000) do określenia liczby klastrów genetycznych (K) i przypisania poszczególnych genotypów (osobników) do tych klastrów (analizy przeprowadzone oddzielnie dla MHC i markerów mikrosatelitarnych).

4 WYNIKI I WNIOSKI

Struktura regionu MHC jest wysoce złożona, co przejawia się przede wszystkim wysoką liczbą zduplikowanych genów. Aby lepiej zrozumieć skomplikowaną architekturę MHC u łyski połączyliśmy dwie metody badawcze, celowane genotypowanie kluczowych eksonów PBR na poziomie populacyjnym oraz sekwencjonowanie *de novo* całego genomu. Analiza uzyskanego genomu wykazała wysoki wskaźnik duplikacji genów w obrębie MHC klasy I (21 wyekstrahowanych kontigów zawierających sekwencję eksonów 1-5), co sugeruje obecność ok. 10 loci MHC-I u łyski. Z kolei analiza danych uzyskanych z sekwencjonowania eksonów na poziomie populacyjnym wskazała, że wykorzystywane przez nas startery amplifikowały co najmniej różnych 5 loci (max. 10 wariantów allelicznych stwierdzonych u pojedynczego osobnika). Oba zastosowane podejścia metodyczne wskazują, że region MHC łyski charakteryzuje się wyjątkowo złożoną strukturą w porównaniu do innych ptaków niewróblowych, gdzie liczba genów jest zwykle niewielka (poniżej trzech u większości zbadanych gatunków). Nasze badania wykazały również wysoki polimorfizm genów MHC klasy I i II, zarówno na poziomie różnorodności nukleotydowej, jak i bogactwa allelicznego (ok. 140-170 alleli wykrytych w obrębie każdej klasy w badanej populacji). Jest to najwyższy wykazany dotychczas poziom polimorfizmu MHC wśród ptaków niewróblowych. Mimo względnie podobnego poziomu zmienności, nasze analizy wykazały, że zarówno dobór różnicujący jak i sygnał rekombinacji jest słabszy w przypadku genów MHC klasy I, niż klasy II. W obrębie MHC klasy II wykryliśmy większą liczbę zdarzeń rekombinacyjnych i większy odsetek sekwencji rekombinowanych, geny te wykazywały się również wyższą proporcją podstawień niesynonimicznych do synonimicznych (dN/dS) w obrębie regionu PBR, a także wyższą liczbą kodonów z sygnałem doboru pozytywnego. Wyniki te sugerują obecność kontrastujących trajektorii ewolucyjnych oraz zróżnicowanych mechanizmów utrzymujących polimorfizm w obu klasach MHC u łyski.

W łódzkiej populacji łysek przetestowano zależność między różnorodnością genów MHC klasy I i II, a komponentami dostosowania (reprodukcja) oraz cechami fenotypowymi związanymi z dostosowaniem (kondycja i ekspresja ornamentu). Badania wykazały, że wysoka różnorodność genów MHC klasy I i klasy II była związana z gorszą kondycją (niższe stężenie hemoglobiny we krwi), chociaż zależność ta była specyficzna dla płci i widoczna u samców w przypadku genów MHC klasy I oraz u samic w przypadku genów MHC klasy II. Łyski z wysoką różnorodnością MHC klasy II miały również słabszą ekspresją ornamentu (mniejsza blaszka czołowa) oraz później rozpoczynały lęgi. Termin składania jaj korelował także ze zróżnicowaniem MHC klasy I, gdyż osobniki ze średnią i wysoką liczbą alleli MHC rozpoczęły lęgi istotnie później niż osobniki z mniejszą liczbą alleli. W przypadku

wielkości zniesienia odnotowaliśmy odmienne wzorce zmienności dla MHC klasy I i II, gdzie ptaki o wysokiej różnorodności MHC klasy I oraz niskiej różnorodność MHC klasy II posiadały najmniejsze zniesienia. Nie wykazano jednak bezpośredniego związku liczby alleli MHC z sukcesem klucia oraz z podstawowym komponentem dostosowania jakim jest sukces rozrodczy. Nie znaleziono również istotnego związku między różnorodnością MHC a masą ciała (po korekcie uwzględniającej wielkość strukturalną osobników) oraz ogólnym poziomem heterozygotyczności badanych ptaków (oszacowanym na podstawie sekwencjonowania ddRAD) W zdecydowanej większości, uzyskane wyniki były niezgodne z hipotezą optymalności (zgodnie z którą najwyższe dostosowanie osiągają osobniki o pośrednim poziomie różnorodności MHC), gdyż nie zaobserwowano istotnego obniżenia cech związanych z dostosowaniem u osobników z niską liczbą alleli MHC. Przeprowadzone badania nie potwierdziły również mechanizmu przewagi heterozygot u łyski, zgodnie z którym dostosowanie osobników powinno wzrastać wraz z różnorodnością MHC. Badania zdają się jednak potwierdzać hipotezę związaną z mechanizmem uszczuplania repertuaru receptorów TCR, która zakłada, że duża różnorodność MHC może być niekorzystna i nie musi prowadzić do zwiększenia dostosowania osobników na drodze poszerzenia spektrum rozpoznawanych patogenów. Niniejsze wyniki sugerują, że koszty ekspresji dużej liczby alleli MHC mogą w pewnych sytuacjach przeważać nad korzyściami związanymi ze zróżnicowanym repertuarem MHC. Niewykluczone, że jest to zjawisko specyficznie związane z krajobrazem miejskim, gdzie różnorodność fauny patogenów jest często zubożona w porównaniu do naturalnych siedlisk, a co za tym idzie wysoka różnorodność MHC może nie być konieczna do zapewnienia skutecznej ochrony immunologicznej. Bez względu jednak na mechanizmy leżące u podstaw wykazanych zależności, prezentowane wyniki po raz pierwszy dostarczają spójnych dowodów na negatywne korelacje różnorodności MHC z szeregiem wskaźników związanych z dostosowaniem (np. kondycja, wielkość ornamentu i wielkość lęgów) w naturalnej populacji ptaków.

Celem prowadzonych badań było również ustalenie, czy procesy synurbizacyjne u łyski są związane ze zmianami w składzie allelicznym i różnorodności genów MHC klasy II. W tym celu przebadano ptaki z czterech populacji, z których dwie miały charakter miejski (Warszawa – stara populacja miejska, Łódź – nowa populacja miejska), a dwie związane były z terenami pozamiejskimi (stawy hodowlane w Sarnowie i Żerominie). We wszystkich badanych populacjach wykazano obecność 114 alleli MHC klasy II, co odpowiadało 113 unikalnym sekwencjom aminokwasowym. Przeprowadzone analizy wykazały istotne różnice w różnorodności MHC na poziomie osobniczym (liczba alleli MHC na osobnika) między populacjami. W szczególności, różnorodność MHC była niższa w starej populacji miejskiej (Warszawa), niż w nowej populacji miejskiej (Łódź) i jednej z populacji pozamiejskich. Populacje z terenów pozamiejskich charakteryzowały się także większą całkowitą liczbą alleli MHC (różnice na poziomie 34-48% po standaryzacji wielkości próby). Podobny wzorzec

zauważono w przypadku alleli MHC specyficznych dla populacji. Wyniki te wskazują, że kolonizacja miast przez łyskę związana była z utratą różnorodności MHC, być może na skutek efektu założyciela. Analiza wskaźników D Josta oraz analiza klastrów genetycznych wykazały najsilniejsze zróżnicowanie genów MHC między starą populacją miejską (Warszawa) a pozostałymi trzema populacjami. Jednocześnie analiza loci mikrosatelitarnych nie potwierdziła neutralnego zróżnicowania genetycznego między populacjami. Podsumowując uzyskane wyniki wskazują, że skład alleli MHC zależał od poziomu urbanizacji siedliska i czasu, jaki upłynął od kolonizacji danego ośrodka miejskiego, a obserwowane zróżnicowanie między populacjami miejskimi i pozamiejskimi mogło wynikać z procesów sortowania genotypów i lokalnej adaptacji. Sugeruje to, że procesy kolonizacji miast przez dzikie zwierzęta mogą wymagać zmian adaptacyjnych w puli genowej MHC w odpowiedzi na zmiany w składzie fauny patogenów i pasożytów charakterystycznych dla obszarów zurbanizowanych.

5 BIBLIOGRAFIA

Acevedo-Whitehouse, K., & Duffus, A. L. (2009). Effects of environmental change on wildlife health. *Philosophical Transactions of the Royal Society B*, *364*(1534), 3429-3438.

Alberti, M., Marzluff, J., & Hunt, V. M. (2017). Urban driven phenotypic changes: empirical observations and theoretical implications for eco-evolutionary feedback. *Philosophical Transactions of the Royal Society B*, *372*(1712), 20160029.

Alcaide, M. (2010). On the relative roles of selection and genetic drift in shaping MHC variation. *Molecular Ecology*, *19*, 3842-3844.

Alcaide, M., Edwards, S. V., Negro, J. J., Serrano, D., & Tella, J. L. (2008). Extensive polymorphism and geographical variation at a positively selected MHC class II B gene of the lesser kestrel (*Falco naumanni*). *Molecular Ecology*, *17*(11), 2652-2665.

Alcaide, M., Edwards, S. V., Cadahía, L., & Negro, J. J. (2009). MHC class I genes of birds of prey: isolation, polymorphism and diversifying selection. *Conservation Genetics*, *10*(5), 1349.

Alcaide, M., Munoz, J., Martínez-de la Puente, J., Soriguer, R., & Figuerola, J. (2014). Extraordinary MHC class II B diversity in a non-passerine, wild bird: the Eurasian Coot *Fulica atra* (Aves: Rallidae). *Ecology and Evolution*, *4*(6), 688-698.

Alvarez, F., S´nchez, C., & Angulo, S. (2005). The frontal shield of the moorhen: sex differences and relationship with body condition. *Ethology Ecology & Evolution*, *17*(2), 135-148.

Aouissi, H. A., Gasparini, J., Belabed, A. I., & Bouslama, Z. (2017). Impact of greenspaces in city on avian species richness and abundance in Northern Africa. *Comptes Rendus Biologies*, *340*(8), 394-400.

Babik, W. (2010). Methods for MHC genotyping in non-model vertebrates. *Molecular Ecology Resources*, *10*(2), 237-251.

Barton, G. M., & Medzhitov, R. (2002). Control of adaptive immune responses by Toll-like receptors. *Current Opinion in Immunology*, *14*(3), 380-383.

Bernatchez, L., & Landry, C. (2003). MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, *16*(3), 363-377.

Biedrzycka, A., Sebastian, A., Migalska, M., Westerdahl, H., & Radwan, J. (2017). Testing genotyping strategies for ultra-deep sequencing of a co-amplifying gene family: MHC class I in a passerine bird. *Molecular Ecology Resources*, *17*(4), 642-655.

Boal, C. W., Mannan, R. W., & Hudelson, K. S. (1998). Trichomoniasis in Cooper's hawks from Arizona. *Journal of Wildlife Diseases*, *34*(3), 590-593.

Bradley, C. A., & Altizer, S. (2007). Urbanization and the ecology of wildlife diseases. *Trends in Ecology & Evolution*, *22*(2), 95-102.

Brouwer, L., Barr, I., Van de Pol, M., Burke, T., Komdeur, J., & Richardson, D. S. (2010). MHC-dependent survival in a wild population: evidence for hidden genetic benefits gained through extra-pair fertilizations. *Molecular Ecology*, *19*(16), 3444-3455.

Camacho-Valdez, V., Tello-Alcaide, E. M., Wootton, A., & Valencia-Barrera, E. (2019). Land Use Change and Urban Ecosystem Services: A Case Study of Urban Wetlands in a Rapidly Sprawling City in the Highlands of Chiapas, Mexico. *Journal of Management & Sustainability*, *9*, 67.

Chao, A., Ma, K. H., Hsieh, T. C., & Chiu, C. H. (2016). User's guide for online program SpadeR (Species-richness prediction and diversity estimation in R). *National Tsing Hua University, Taiwan*, *88*.

Combes, C. (2001). Parasitism: the ecology and evolution of intimate interactions. *University of Chicago Press.*

Concepción, E. D., Moretti, M., Altermatt, F., Nobis, M. P., & Obrist, M. K. (2015). Impacts of urbanisation on biodiversity: the role of species mobility, degree of specialisation and spatial scale. *Oikos*, *124*(12), 1571-1582.

Cox, D. T., Shanahan, D. F., Hudson, H. L., Fuller, R. A., & Gaston, K. J. (2018). The impact of urbanization on nature dose and the implications for human health. *Landscape and Urban Planning*, *179*, 72-80.

Croci, S., Butet, A., & Clergeau, P. (2008). Does urbanization filter birds on the basis of their biological traits. *Condor*, *110*(2), 223-240.

Crowley, C. E., & Magrath, R. D. (2004). Shields of offence: signalling competitive ability in the dusky moorhen, *Gallinula tenebrosa*. *Australian Journal of Zoology*, *52*(5), 463-474.

Cumming, G. S., & Van Vuuren, D. P. (2006). Will climate change affect ectoparasite species ranges? *Global Ecology and Biogeography*, *15*(5), 486-497.

Dauphiné, N. I. C. O., & Cooper, R. J. (2009). Impacts of free-ranging domestic cats (*Felis catus*) on birds in the United States: a review of recent research with conservation and management recommendations. In *Proceedings of the fourth international partners in flight conference: tundra to tropics* (Vol. 205).

Doherty, P. C., & Zinkernagel, R. M. (1975). Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*, 256(5512), 50-52.

Duraes, F. V., Niven, J., Dubrot, J., Hugues, S., & Gannagé, M. (2015). Macroautophagy in endogenous processing of self-and pathogen-derived antigens for MHC class II presentation. *Frontiers in Immunology*, *6*, 459.

Ekblom, R., Saether, S. A., Jacobsson, P., Fiske, P., Sahlman, T., Grahn, M., ... & Höglund, J. (2007). Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*). *Molecular Ecology*, *16*(7), 1439-1451.

Faeth, S. H., Warren, P. S., Shochat, E., & Marussich, W. A. (2005). Trophic dynamics in urban communities. *BioScience*, *55*(5), 399-407.

Fitzherbert, E. B., Struebig, M. J., Morel, A., Danielsen, F., Brühl, C. A., Donald, P. F., & Phalan, B. (2008). How will oil palm expansion affect biodiversity? *Trends in Ecology & Evolution*, *23*(10), 538-545.

Frenkel, J. K., Hassanein, K. M., Hassanein, R. S., Brown, E., Thulliez, P., & Quintero-Nunez, R. (1995). Transmission of *Toxoplasma gondii* in Panama City, Panama: a five-year prospective cohort study of children, cats, rodents, birds, and soil. *American Journal of Tropical Medicine and Hygiene*, *53*(5), 458-468.

Garrigan, D., & Hedrick, P. W. (2003). Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution*, *57*(8), 1707-1722.

Gaudieri, S., Dawkins, R. L., Habara, K., Kulski, J. K., & Gojobori, T. (2000). SNP profile within the human major histocompatibility complex reveals an extreme and interrupted level of nucleotide diversity. *Genome Research*, *10*(10), 1579-1586.

Goebel, J., Promerová, M., Bonadonna, F., McCoy, K. D., Serbielle, C., Strandh, M., ... & Fumagalli, L. (2017). 100 million years of multigene family evolution: origin and evolution of the avian MHC class IIB. *BMC Genomics*, *18*(1), 1-9.

Green, T. L., Kronenberg, J., Andersson, E., Elmqvist, T., & Gomez-Baggethun, E. (2016). Insurance value of green infrastructure in and around cities. *Ecosystems*, *19*(6), 1051-1063.

Griffiths, R., Double, M. C., Orr, K., & Dawson, R. J. (1998). A DNA test to sex most birds. *Molecular Ecology*, *7*(8), 1071-1075.

Grimm, N. B., Faeth, S. H., Golubiewski, N. E., Redman, C. L., Wu, J., Bai, X., & Briggs, J. M. (2008). Global change and the ecology of cities. *Science*, *319*(5864), 756-760.

He, K., Minias, P., & Dunn, P. O. (2021). Long-read genome assemblies reveal extraordinary variation in the number and structure of MHC loci in birds. *Genome Biology and Evolution*, *13*(2), evaa270.

Hess, C. M., & Edwards, S. V. (2002). The Evolution of the Major Histocompatibility Complex in Birds: Scaling up and taking a genomic approach to the major histocompatibilty complex (MHC) of birds reveals surprising departures from generalities found in mammals in both large-scale structure and the mechanisms shaping the evolution of the MHC. *Bioscience*, *52*(5), 423-431.

Holt, C., & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, *12*(1), 491.

Janiszewski, T., Wojciechowski, Z., & Markowski, J. (2009). *Atlas ptaków lęgowych Łodzi*. Wydawnictwo Uniwersytetu Łódzkiego.

Jost, L. O. U. (2008). GST and its relatives do not measure differentiation. *Molecular Ecology*, *17*(18), 4015-4026.

Lehrer, E. W., Fredebaugh, S. L., Schooley, R. L., & Mateus-Pinilla, N. E. (2010). Prevalence of antibodies to *Toxoplasma gondii* in woodchucks across an urban–rural gradient. *Journal of Wildlife Diseases*, *46*(3), 977-980.

Loiseau, C., Richard, M., Garnier, S., Chastel, O., Julliard, R., Zoorob, R., & Sorci, G. (2009). Diversifying selection on MHC class I in the house sparrow (*Passer domesticus*). *Molecular Ecology*, *18*(7), 1331-1340.

Lourenço, A., Álvarez, D., Wang, I. J., & Velo-Antón, G. (2017). Trapped within the city: Integrating demography, time since isolation and population-specific traits to assess the genetic effects of urbanization. *Molecular Ecology*, *26*(6), 1498-1514.

Lowry, H., Lill, A., & Wong, B. B. (2013). Behavioural responses of wildlife to urban environments. *Biological Reviews*, *88*(3), 537-549.

Luniak, M. (2004). Synurbization–adaptation of animal wildlife to urban development. In *Proceedings 4th international urban wildlife symposium* (pp. 50-55). University of Arizona.

Luniak, M., Kalbarczyk, W., & Pawłowski, W. (1964). Birds of Warsaw. *Acta Ornithologica*, *8*, 175-285.

Łoś, A., Skórka, P., Strachecka, A., Winiarczyk, S., Adaszek, Ł., Winiarczyk, M., & Wolski, D. (2020). The associations among the breeding performance of *Osmia bicornis* L. (*Hymenoptera: Megachilidae*), burden of pathogens and nest parasites along urbanisation gradient. *Science of The Total Environment*, *710*, 135520.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, *1*(1), vev003.

Martin, L. B., Hopkins, W. A., Mydlarz, L. D., & Rohr, J. R. (2010). The effects of anthropogenic global changes on immune functions and disease resistance. *Annals of the New York Academy of Sciences*, *1195*(1), 129-148.

McClelland, E. E., Penn, D. J., & Potts, W. K. (2003). Major histocompatibility complex heterozygote superiority during coinfection. *Infection and Immunity*, 71(4), 2079-2086.

McConkey, K. R., Prasad, S., Corlett, R. T., Campos-Arceiz, A., Brodie, J. F., Rogers, H., & Santamaria, L. (2012). Seed dispersal in changing landscapes. *Biological Conservation*, *146*(1), 1-13.

McKinney, M. L. (2006). Urbanization as a major cause of biotic homogenization. *Biological Conservation*, *127*(3), 247-260.

Migalska, M., Sebastian, A., & Radwan, J. (2019). Major histocompatibility complex class I diversity limits the repertoire of T cell receptors. *Proceedings of the National Academy of Sciences*, 116(11), 5021-5026.

Miller, H. C., & Lambert, D. M. (2004). Gene duplication and gene conversion in class II MHC genes of New Zealand robins (*Petroicidae*). *Immunogenetics*, *56*(3), 178-191.

Minias, P., Pikus, E., Whittingham, L. A., & Dunn, P. O. (2019). Evolution of copy number at the MHC varies across the avian tree of life. *Genome Biology and Evolution*, *11*(1), 17-28.

Minias, P. (2015). The use of haemoglobin concentrations to assess physiological condition in birds: a review. *Conservation Physiology*, *3*(1), cov007.

Møller, A. P., & Saino, N. (2004). Immune response and survival. *Oikos*, *104*(2), 299-304.

Münz, C. (2012). Antigen processing for MHC class II presentation via autophagy. *Frontiers in Immunology*, *3*, 9.

O'Connor, E. A., Strandh, M., Hasselquist, D., Nilsson, J. Å., & Westerdahl, H. (2016). The evolution of highly variable immunity genes across a passerine bird radiation. *Molecular Ecology*, *25*(4), 977-989.

Patz, J. A., Daszak, P., Tabor, G. M., Aguirre, A. A., Pearl, M., Epstein, J., ... & Working Group on Land Use Change Disease Emergence. (2004). Unhealthy landscapes: policy recommendations on land use change and infectious disease emergence. *Environmental Health Perspectives*, *112*(10), 1092-1098.

Pedersen, A. B., & Greives, T. J. (2008). The interaction of parasites and resources cause crashes in a wild mouse population. *Journal of Animal Ecology*, *77*(2), 370-377.

Peig, J., & Green, A. J. (2009). New perspectives for estimating body condition from mass/length data: the scaled mass index as an alternative method. *Oikos*, *118*(12), 1883-1891.

Piertney, S. B., & Oliver, M. K. (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity*, *96*(1), 7-21.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.

Radwan, J., Zagalska-Neubauer, M., Cichoń, M., Sendecka, J., Kulma, K., Gustafsson, L., & Babik, W. (2012). MHC diversity, malaria and lifetime reproductive success in collared flycatchers. *Molecular Ecology*, 21(10), 2469-2479.

Radwan, J., Biedrzycka, A., & Babik, W. (2010). Does reduced MHC diversity decrease viability of vertebrate populations? *Biological Conservation*, *143*(3), 537-544.

Reusch, T. B., Schaschl, H., & Wegner, K. M. (2004). Recent duplication and inter-locus gene conversion in major histocompatibility class II genes in a teleost, the three-spined stickleback. *Immunogenetics*, *56*(6), 427-437.

Roved, J., Hansson, B., Tarka, M., Hasselquist, D., & Westerdahl, H. (2018). Evidence for sexual conflict over major histocompatibility complex diversity in a wild songbird. *Proceedings of the Royal Society B*, 285(1884), 20180841.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology & Evolution*, *34*(12), 3299-3302.

Schenten, D., & Medzhitov, R. (2011). The control of adaptive immune responses by the innate immune system. *Advances in Immunology*, *109*, 87-124.

Sebastian, A., Herdegen, M., Migalska, M., & Radwan, J. (2016). Amplisas: a web server for multilocus genotyping using next-generation amplicon sequencing data. *Molecular Ecology Resources*, *16*(2), 498-510.

Shochat, E., Warren, P. S., Faeth, S. H., McIntyre, N. E., & Hope, D. (2006). From patterns to emerging processes in mechanistic urban ecology. *Trends in Ecology & Evolution*, *21*(4), 186-191.

Slade, R. W., & McCallum, H. I. (1992). Overdominant vs. frequency-dependent selection at MHC loci. *Genetics*, *132*(3), 861-862.

Sol, D., Trisos, C., Múrria, C., Jeliazkov, A., González-Lagos, C., Pigot, A. L., ... & Pavoine, S. (2020). The worldwide impact of urbanisation on avian functional diversity. *Ecology Letters*, *23*(6), 962-972.

Spurgin, L. G., & Richardson, D. S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B*, *277*(1684), 979-988.

Steimle, V., Otten, L. A., Zufferey, M., & Mach, B. (1993). Complementation cloning of an MHC class II transactivator mutated in hereditary MHC class II deficiency (or bare lymphocyte syndrome). *Cell*, *75*(1), 135-146.

Stervander, M., Dierickx, E. G., Thorley, J., Brooke, M. D. L., & Westerdahl, H. (2020). High MHC gene copy number maintains diversity despite homozygosity in a Critically Endangered single-island endemic bird, but no evidence of MHC-based mate choice. *Molecular Ecology*, *29*(19), 3578-3592.

Strand, T. M., Segelbacher, G., Quintela, M., Xiao, L., Axelsson, T., & Höglund, J. (2012). Can balancing selection on MHC loci counteract genetic drift in small fragmented populations of black grouse? *Ecology & Evolution*, *2*(2), 341-353.

Sutton, J. T., Nakagawa, S., Robertson, B. C., & Jamieson, I. G. (2011). Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immunity genes. *Molecular Ecology*, *20*(21), 4408-4420.

Tryjanowski, P., Møller, A. P., Morelli, F., Biaduń, W., Brauze, T., Ciach, M., ... & Zduniak, P. (2016). Urbanization affects neophilia and risk-taking at bird-feeders. *Scientific Reports*, *6*(1), 28575.

ONZ. (2019). *Population Division Department of Economic and Social Affairs,* https://population.un.org/wpp/Download/Standard/Population/, 15.10.2021.

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., & Kosakovsky Pond, S. L. (2018). Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Molecular Biology & Evolution*, *35*(3), 773-777.

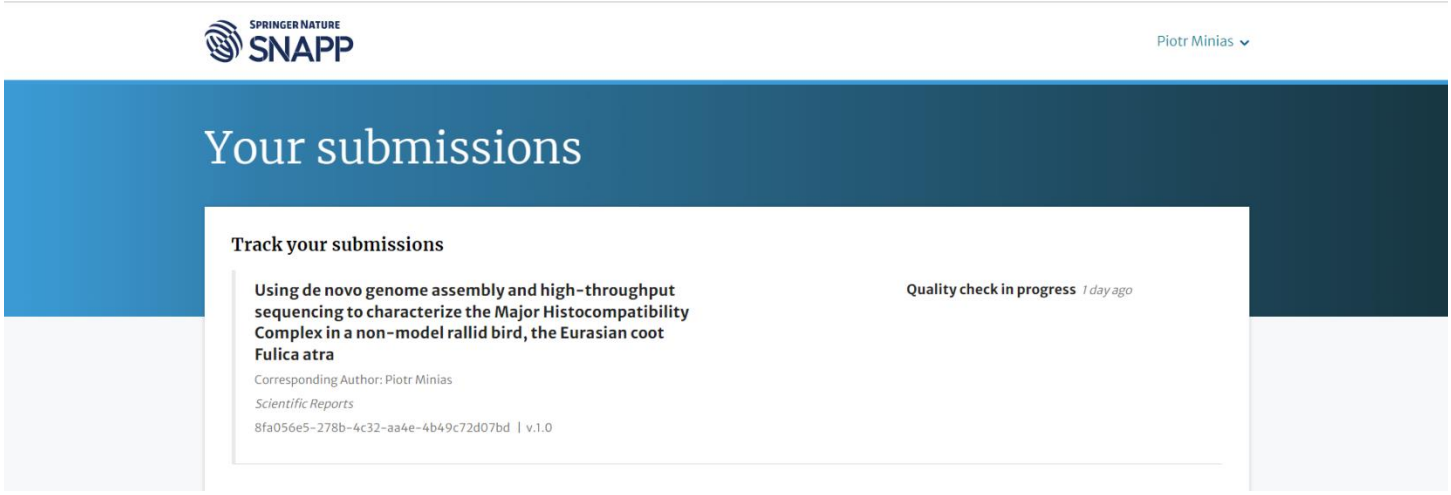
Vitousek, P. M., Mooney, H. A., Lubchenco, J., & Melillo, J. M. (1997). Human domination of Earth's ecosystems. *Science*, *277*(5325), 494-499.

Watson, H., Videvall, E., Andersson, M. N., & Isaksson, C. (2017). Transcriptome analysis of a wild bird reveals physiological responses to the urban environment. *Scientific Reports*, *7*(1), 44180.

Wegner, K. M., Kalbe, M., Kurtz, J., Reusch, T. B., & Milinski, M. (2003). Parasite selection for immunogenetic optimality. *Science*, 301(5638), 1343-1343.

6 ROZDZIAŁ PIERWSZY

PIKUS, E., MINIAS, P. USING DE NOVO GENOME ASSEMBLY AND HIGH-THROUGHPUT SEQUENCING TO CHARACTERIZE THE MAJOR HISTOCOMPATIBILITY COMPLEX IN A NON-MODEL RALLID BIRD, THE EURASIAN COOT *FULICA ATRA*. SCIENTIFIC REPORTS (PRACA ZGŁOSZONA DO DRUKU).



Ref: Submission ID 8fa056e5-278b-4c32-aa4e-4b49c72d07bd

Dear Dr Minias,

Thank you for submitting your manuscript to Scientific Reports.

Your manuscript is now at our initial Quality Check stage, where we look for adherence to the journal's submission guidelines, including any relevant editorial and publishing policies. If there are any points that need to be addressed prior to progressing we will send you a detailed email. Otherwise, your manuscript will proceed into peer review.

You can check on the status of your submission at any time by using the link below and logging in with the account you created for this submission:

https://researcher.nature.com/your-submissions?utm_source=submissions&utm_medium=email&utm_campaign=confirmation-email&journal_id=41598

Kind regards,

Peer Review Advisors
Scientific Reports

Springer Nature offers an open access support service to make it easier for our authors to discover and apply for APC funding. For further information please visit http://www.springernature.com/gp/open-research/funding

Using de novo genome assembly and high-throughput sequencing to characterize the Major Histocompatibility Complex in a non-model rallid bird, the Eurasian coot *Fulica atra*

Ewa Pikus, Piotr Minias

Department of Biodiversity Studies and Bioeducation, Faculty of Biology and Environmental Protection, University of Łódź, Banacha 1/3, 90-237 Łódź, Poland

Correspondence and requests for materials should be addressed to P.M. (email: pminias@op.pl) or E.P. (ewa.pikus@biol.uni.lodz.pl).

Abstract

Genes of the Major Histocompatibility Complex (MHC) form a key component of vertebrate adaptive immunity, as they code for molecules which bind antigens of intra- and extra-cellular pathogens (MHC class I and II, respectively) and present them to T cell receptors. In general, MHC genes are hyper-polymorphic and high MHC diversity is often maintained within natural populations (via balancing selection) and within individuals (via gene duplications). Because of its complex architecture with tandems of duplicated genes, characterization of MHC region in non-model vertebrate species still poses a major challenge. Here, we combined *de novo* genome assembly and high-throughput sequencing to characterize MHC polymorphism in a rallid bird species, the Eurasian coot *Fulica atra*. An analysis of genome assembly indicated high duplication rate at MHC-I (21 contigs with exons 1-5), which was also supported by targeted sequencing of peptide-binding exons (at least five MHC-I loci genotyped). We found high allelic richness at both MHC-I and MHC-II, although signature of diversifying selection and recombination (gene conversion) was much stronger at MHC-II. Our results indicate that Eurasian coot retains extraordinary polymorphism at both MHC classes (when compared to other non-passerine bird species), although they may be subject to different evolutionary mechanism.

The Major Histocompatibility Complex (MHC) is a complex multi-gene family playing an essential role in vertebrate adaptive immune response. The MHC genes code for key transmembrane molecules recognizing and binding foreign peptides in order to present them to immune cells, which initiates an immune response[1,2]. Two major classes of MHC genes can be distinguished, including class I genes (MHC-I) which are expressed on the surface of almost all nucleated somatic cells, and class II genes (MHC-II) which are expressed on the specialized cells of the immune system, such as B lymphocytes and macrophages[3]. Antigens recognized by MHC-I molecules originate from intracellular pathogen proteins processed via proteasomal proteolysis. They are presented to T CD8+ lymphocytes, resulting in a direct destruction or the initiation of apoptotic processes within infected cells. In turn, exogenous antigens derived from extracellular pathogen peptides processed by endocytosis or autophagy are presented by MHC-II molecules to T CD4+ lymphocytes. This allows for the trigger of a proper immune response pathway [4,5]. Antigens recognized by MHC-I and MHC-II comprise small peptides composed of 8-10 and 15-20 amino acid residues, respectively and they are bound by the most polymorphic region of MHC molecules, the peptide-binding groove[5]. The peptide-binding region (PBR) of any particular MHC allelic variant can bind a limited number of foreign peptides and, hence, the number of different MHC molecules expressed in the body determines an array of pathogens recognizable by the mechanisms of an adaptive immune defence[6].

An extreme polymorphism within the area of the MHC peptide-binding groove may be considered one of key evolutionary features of these genes. Due to host-pathogen arms race, MHC-I and MHC-II genes demonstrate the highest level of functional polymorphism known in vertebrates[3,7,8]. For example, thousands of different alleles were identified in the global human population at some MHC loci[9]. As for birds, the sedge warbler *Acrocephalus schoenobaenus* with over 3500 MHC-I alleles[10] and common yellowthroat *Geothlypis trichas* with almost 1000 MHC-II alleles recorded within a single population[11] may provide examples of species with exceptionally high MHC polymorphism. It is possible to distinguish three basic mechanisms of the natural (balancing) selection responsible for the maintenance of MHC polymorphism in vertebrate populations. First, the mechanism of heterozygote advantage assumes that heterozygous individuals are able to recognize a broader scope of antigens, which should provide them with fitness benefits over homozygotes, especially when exposed to diverse pathogen faunas[12,13]. Second, parasites or pathogens may evolve quickly to avoid host immune defences. In this scenario, a negative frequency-dependent selection can maintain MHC polymorphism via favouring rare alleles (rare-allele advantage hypothesis). Alleles that increase in frequency gradually lose their selective advantage, as pathogens tend to evolve towards the avoidance of the most common immune barriers of their hosts (including the most frequent MHC alleles)[13,14]. Third, high polymorphism of MHC genes

in populations may be generated and maintained in response to pathogen-driven selection that varies in space and time (fluctuating selection)[15]. All these mechanisms are expected to produce an apparent excess of non-synonymous (amino acid altering) mutations (dN) over synonymous (silent) mutations (dS) within the region of the peptide-binding groove. Thus, high values of dN/dS ratio (>1) provide a molecular signature of balancing selection acting on the MHC[16,17].

Extensive gene duplication constitutes the next characteristic feature of MHC genes in birds[18]. For instance, the number of gene copies has been reported to vary between 1-33 for MHC-I and 1-23 for MHC-II[19]. However, the latest genomic studies indicate that duplication rate in some avian evolutionary lines may be even greater[20]. Duplicated genes may retain similar molecular features or even share identical alleles, which is mainly due to the processes of concerted evolution, where different loci evolve non-independently, leading to their homogenization within species[21,22,23]. This triggers methodological difficulties in the MHC genotyping, pertaining particularly to difficulties in designing locus-specific primers in non-model taxa. A common usage of conserved (multi-locus) primers preclude application of traditional Sanger sequencing methods for MHC genotyping and only after the development of new generation sequencing (NGS) methods a decisive breakthrough in the MHC studies took place[24]. Another milestone in the MHC research was reached via development of techniques used for the sequencing and assembling high-quality genomes. Genomic approaches are not restricted to sequencing single exons (as it frequently is with high-throughput MHC genotyping at the population scale), but provide resolution of complete genes, or may even provide information on the architecture of the entire MHC region[20].

The aim of our paper was to characterize the MHC region in a non-model species of a wild bird, the Eurasian coot *Fulica atra*. This is a medium-size nonpasserine waterbird from the rail family (Rallidae), which has a broad geographical range (spanning from Europe through Asia to Australia) and a large global population of ca. 5.3-6.5 million individuals[25]. We used both standard high-throughput sequencing (NGS) to characterize polymorphism and selection at the key MHC exons coding for PBR (within Central European population), as well as genomic techniques to characterize the architecture of the MHC region. To date, polymorphism of MHC-II has already been examined in the Iberian population of our study species[26], therefore our primary aim was to comprehensively characterize the MHC-I genes and to compare the levels of polymorphism and signature of selection between both MHC classes, as they may show distinct evolutionary trajectories in birds[27].

Material and Methods

Sample collection

Fieldwork took place in central Poland, mostly in the city of Łódź (51° 45' N, 19° 28' E) and non-urban areas located nearby. Blood samples were collected from adult birds (n = 283) captured mostly during the reproductive season (March-July) between 2012 to 2019. We caught birds at nests or while feeding on the shore using noose traps made from monofilament nylon. All birds were ringed with metal rings (tarsus) and plastic collars (neck) to enhance identification of individuals in the field and avoid recaptures. From each captured bird we took 100 µl of blood from a tarsal vein into 96% ethanol and stored the samples in 5°C until DNA isolation. We extracted genomic DNA using GeneJET Genomic DNA Purification Kit (Fermentas, Thermo Fisher Scientific, Waltham, 136 MA, USA) according to the manufacturer's protocol. Bird capturing and blood sampling was performed by the permissions of the Local Bioethical Commission for Experiments on Animals in Łodź (nos 40/ŁB 620/2012 and 15/ŁB/2016) and complied with current laws of Poland Act on Nature Conservation from 16 April 2004, Journal of Laws from 2004, No. 92, item 880).

Genome sequencing and assembly

To get an insight into the architecture of the MHC region (gene copy numbers) and polymorphism of non-PBR exons in the Eurasian coot we generated de novo genome assembly for a single individual from our study population. For this purpose, one Chromium linked-read library was constructed with the Chromium Genome Library Kit & Gel Bead Kit v2 (10x Genomics, Pleasanton, CA, USA) and sequenced on an SP lane on an Illumina NovaSeq 6000 instrument (Illumina Inc., San Diego, CA, USA) at the Carver Biotechnology Center at the University of Illinois at Urbana-Champaign. A total of 889,998,744 paired-end reads (2x150nt) were generated, demultiplexed, and assembled with Supernova v2.1.1 (10x Genomics), setting maximum reads used to n = 600,000,000 and all other parameters as default. Supernova output was converted to *pseudohap* FASTA format for downstream processing. The genome assembly was filtered for duplicate contigs and scaffolds using the *dedupe* script from BBMap v38.36 (Bushnell 2014 [28]). A custom vector-screening script was employed to remove residual sequencing adapters and vector sequence. No contaminant non-bird sequences were detected in the genome assembly as assessed by BlobTools v0.9.19.6 (Laetsch and Blaxter 2017 [29]). Contigs and scaffolds less than 1kb were filtered from the assembly prior to acceptance at NCBI under BioProject PRJNA633903 and GenBank accession GCA_013372525.1. Genome

completeness was estimated at 93.3% (1.3% duplicated), as assessed by BUSCO v3.0.1 (Seppey et al. 2019) [30] using the Aves odb9 lineage of 4,915 orthologs. Total sequence length of the genome assembly was 1,168 Mb, scaffold N50 was 6.4 Mb, while contig N50 was 0.25 Mb.

The genome assembly was annotated by employing three rounds of MAKER v3.01.1 (Holt and Yandell 2011) [31]. In the first round of MAKER annotation, gene models were predicted using homology searches from the following lines of evidence: transcriptome assembly sequences from NCBI for Okinawa rail *Gallirallus okinawae* (ICPP01000000) and laughing gull *Leucophaeus atricilla* (GFNV00000000), as well as proteins from available NCBI RefSeq sequences (Gruiformes) and the SwissProt database. Two MAKER *ab initio* gene predictors, SNAP and Augustus, were trained using gene models predicted from comparative evidence output from the first round. A second round of MAKER utilized the trained models for Augustus and SNAP *ab initio* gene prediction. A final third round of gene prediction used re-trained models for SNAP and Augustus from the second round of output.

Retrieving MHC sequences from genomic data

To retrieve MHC class I and class II from genome assembly we performed BLAST searches using available non-passerine MHC sequences. First, to retrieve contigs with MHC we BLASTed concatenated exonic sequences containing exons 1-5 for MHC-I and exons 1-4 for MHC-II. The remaining exons were not included because of short length (<35 bp). For both MHC classes we used available GenBank sequences from the Chinese egret *Egretta eulophotes* (KY511591 and KC282841). We also checked blasting results using sequences from other non-passerine species (KC282841 from red knot *Calidris canutus* and KC205115 from golden pheasant *Chrysolophus pictus* for MHC-I; HM070250 from Mallard *Anas platyrhynchos* and AB872444 from crested ibis *Nipponia nippon* for MHC-II), but they produced very consistent results. Second, we aimed to retrieve all available PBR sequences, which are coded by exons 2 and 3 of the same gene at MHC-I and by exons 2α and 2β coded by separate genes at MHC-II (Bjorkman et al., 1987; Brown et al., 1988 [32,33]). For this purpose, we first retrieved PBR exon sequences from longer MHC contigs and used them to repeat blastn searches (a single exon per search) within the genome assembly.

MHC genotyping (high-throughput sequencing)

To get a better resolution of MHC polymorphism and the mechanisms that may contribute to its maintenance (recombination and selection) in the Eurasian coot, we genotyped key MHC regions

(selected PBR-coding exons) in all captured individuals (283 individuals genotyped at MHC-I, 230 genotyped at MHC-II). Population genotyping focused on a single exon per MHC class (exon 3 at MHC-I and exon 2β at MHC-II), as these exons are traditionally targeted in avian MHC research (allowing direct comparisons across species) and their polymorphism is expected to be well representative for the entire PBR region (e.g. Minias et al. 2021 [34]). To genotype MHC-I exon 3, we used primers MHCI-int2F (5'- CATTTCCCTYGTGTTTCAGG-3') and MHCI-ex4R (3'-GGGTAGAAGCCGTGAGCRC-5'), which were originally designed for accipitrid birds (Alcaide et al., 2009 [35]). Primer MHCI-int2F binds to the conserved flanking region of intron 2 and primer MHCI-ex4R binds to the conserved region of exon 4. The length of the entire amplicon was 411 bp, including almost entire exon 3 (273 bp out of 276 bp). Species-specific primers Fuat-Ex2Fw (5'-CTGACCRGCCTCCCTGCA-3') and Fuat-Ex2Rv (5'-TTGTGCCAYACACCCACC-3') were used to amplify MHC-II. These two primers were originally designed for the Eurasian coot (Alcaide et al., 2014 [26]) and they successfully amplify the entire MHC-II exon 2 (270 bp), binding to the flanking regions of intron 1 and 2. In each PCR reaction we used fusion primers with Illumina Nextera Transposase adapter sequences (Illumina Corp., San Diego, CA, USA) and 7-bp barcodes to identify the samples. PCR amplifications were carried out in a final volume of 20 µl containing 20-80 ng genomic DNA (1 µl of DNA isolate), 10 µl of 2X HotStarTaq Plus MasterMix Kit (Qiagen, Venlo, The Netherlands), 8 µl of deionized water and 0.5 µl of each primer. PCR protocols followed Alcaide et al. (2009) [35] for MHC-I and Alcaide et al. (2014) [26] for MHC-II, although in both cases the number of PCR cycles was reduced to 25 to suppress the formation of artificial chimeras, which could confound the correct interpretation of Illumina sequencing results. The effects of PCR reactions were confirmed for each sample by visual examination of band intensities on 2% agarose gel electrophoresis. To purify PCR products we used AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA) and concentration estimates were quantified using Quant-iT PicoGreen dsDNA marking kit (Thermo FisherScientific, Waltham, MA, USA). Separate libraries for MHC class I and II were prepared using equimolar concentrations of purified PCR products and NEB-Next DNA Library Prep Master Mix Set for Illumina (New England Biolabs, Ipswich, MA, USA). Both libraries were sequenced on the 2 x 250bp Illumina MiSeq platform.

In the processing of raw Illumina data we used an online webserver, the Amplicon Sequencing Analysis Tools (AmlpiSAT) (Sebastian et al., 2016 [36]), and followed recommendations by Biedrzycka et al. (2017b) [37]. In the first step we used the Amplicon Sequencing MERGing (AmpliMERGE) tool, which merges paired-end reads, optimizing their overlapping lengths according to amplicon data (Magoc and Salzberg, 2011 [38]). Next, we used the Amplicon Sequencing Assignment (AmpliSAS) tool, which performs read demultplexing, variant clustering and putative allele filtering based on user-specified criteria. For the clustering step (identification of reads resulting from genotyping errors

and clustering them with reads identified as true alleles) we used default AmpliSAS settings for Illumina data, including a substitution error rate of 1%, an indel error rate of 0.001% and the minimum dominant frequency of 25%. Finally, we used AmpliSAS to filter for clusters that are likely to be artefacts, including chimeras and other low-frequency artefacts (>3%) that were retained through the clustering step. Samples with amplicon depth of less than 300 reads were excluded from the analyses and the maximum amplicon depth was, by default, set to 5000 reads because of AmpliSAS performance reasons. The average amplicon depth prior to the processing was 4453 ± 66 [SE] reads for MHC-I and 2616 ± 97 [SE] reads for MHC-II. We obtained validated MHC-I and MHC-II genotypes for 270 and 220 individuals, respectively. Technical reproducibility of validated sequences was 93.7%, as estimated using 36 technical replicates (i.e. samples for which two amplicons were obtained in independent PCR reactions and sequenced). To align all unique MHC class I and II sequences we used Geneious v10.0.5 (Biomatters Ltd., Auckland, New Zealand). We removed intron regions from the alignments and we inferred alleles based on the exon fragments only.

Recombination

Recombination signal at the MHC-I exon 3 and MHC-II exon 2 (NGS data) was searched for using RDP v.4.97 software, which implements different algorithms developed specifically to detect recombinant sequences (Martin et al. 2015) [39]. We used seven basic algorithms (Maxchi, BootScan, Genconv, SiScan, RDP, Chimaera, and 3Seq) and ran all the analyses using default settings with statistical significance threshold of P = 0.05 and Bonferroni correction for multiple comparisons. To quantify recombination signal we calculated the number of different recombination events, number of recombinant sequences, and number of breakpoints within 100 nucleotide window. A recombination event was recognized when supported by two or more algorithms, while events recognized by a single algorithm were discarded. Presence of recombination hot and cold spots was tested with the local hot/cold-spot test (1000 permutations), as implemented in RDP software.

Sequence polymorphism and selection

We used DnaSP v.6.10.3 software (Rozas et al. 2017)[40] to characterize MHC class I and II polymorphism (NGS and genome assembly data). We assessed sequence polymorphism as the number of polymorphic sites, total number of mutations, average nucleotide diversity, and average number of nucleotide differences. To quantify the signature of selection at the MHC-I exon 3 and MHC-II exon 2 (NGS data) we calculated the dN/dS ratios, which reflect the relative rate of nonsynonymous (amino

acid altering) to synonymous (silent) nucleotide substitutions (per non-synonymous and synonymous site, respectively). Positive (diversifying) selection is detected when new allelic variants are promoted, which means that nonsynonymous substitutions accumulate faster than synonymous substitutions (dN/dS > 1) and similar pattern is expected under pathogen-driven balancing selection, when multiple alleles are maintained within populations. In contrast, negative (purifying) selection removes most nonsynonymous substitutions, which thus accumulate slower than synonymous ones (dN/dS < 1). Finally, similar rates of nonsynonymous and synonymous substitutions (dN/dS ≈ 1) indicate neutral evolution and no detectable signature of selection. We measured codon-specific signature of positive (diversifying) and negative (purifying) selection using two approaches, Bayesian inference (Fast Unconstrained Bayesian AppRoximation, FUBAR) and maximum likelihood (Fixed Effects Likelihood, FEL), implemented in HyPhy software available at the Datamonkey webserver (Weaver et al. 2018) [41]. We used 0.95 posterior probability (FUBAR) and 0.05 P value (FEL) thresholds to identify sites that may have experienced pervasive (apparent across all alleles) diversifying or purifying selection. We also used Mixed Effect Model of Evolution (MEME) to identify sites subject to episodic (apparent across a subset of alleles) diversifying selection. We used default settings and input trees inferred from the alignments in all the analyses. Selection analyses were performed on alignments lacking recombinant sequences, because recombination can mask true phylogenetic relationships between allelic variants (allele tree topology) and, thus, lead to erroneous estimates of the nucleotide substitution rates (Anisimova et al. 2003)[42]. Positions of positively selected sites were compared with putative PBR sites in non-passerine birds (as identified based on the global analysis of selection at the avian MHC; Minias et al. 2018)[27] and humans (based on the crystallographic structure of MHC molecules by Saper et al. 1991 and Brown et al. 1993)[43,44]. Positions of positively selected sites at MHC-II exon 2 were also compared with previous data from Iberian coot population (Alcaide et al. 2014)[26]. To quantitatively assess an agreement between these positions we calculated intra-class correlation (ICC) coefficients in the *irr* R package (Gamer et al. 2012)[45].

Results

Genome assembly

Our BLAST searches of MHC-I exons 1-5 retrieved 32 contigs containing all five exons and another 12 searches containing four exons (either 1-4 or 2-5). Visual inspection of these sequences retained 27 functional haplotypes, out of which 21 contained exons 1-5 (Fig. S1 in the Electronic Supplementary Material). All five exons showed similar level of polymorphism, although exon 4 had noticeably lower nucleotide diversity from the remaining exons (Table 1). Also, the number of retrieved haplotypes (n = 27) suggested the presence of over ten duplicated MHC-I loci in the Eurasian coot. However, when we BLASTed only exon 2 or exon 3 against genome assembly, we retrieved 76 and 82 unique functional alleles, respectively. Although it is likely that some of these sequences may represent genotyping artefacts, overall they provide support for high duplication rate at MHC-I in our study species. BLAST searches for MHC-II exons 1-4 retrieved only one contig containing a single sequence of α and β chain (Fig. S1). Using PBR exons for blast searches yielded similar results (one sequence of each exon retrieved).

Polymorphism of PBR exons

High-throughput sequencing of PBR exons revealed high level of allelic richness in our study coot population, as in total we detected 165 allelic variants of MHC-I exon 3 (n = 270 individuals genotyped) and 147 allelic variants of MHC-II exon 2 (n = 220 individuals) (Table 1). All allelic variants were functional (no stop codons or frameshift mutations), providing no evidence of pseudogenization. We recorded a relatively minor frequency (14.3%) of variants with an indel mutation at MHC-II exon 2 (one-codon deletion at position 84 as marked at Fig. 1), but they all retained functionality. The maximum number of allelic variants recorded per individual was ten at MHC-I and six at MHC-II, indicating that we genotyped at least five MHC-I and three MHC-II loci. Most frequently we recorded two MHC-I and three MHC-II variants per individual (21.5% and 44.4%, respectively) (Fig. 2). The total number of segregating sites and total number of mutations were higher at MHC-I exon 3, but MHC-II exon 2 had slightly higher nucleotide diversity, suggesting similar levels of polymorphism at MHC class I and II.

Recombination

We found evidence for much stronger recombination signal at the MHC class II than class I. At MHC-I we identified three recombination events and the total number of recombinant sequences was 31 (18.8 % of all sequences). In contrast, there were 17 recombination events recognized at MHC-II and the number of recombinant was over twice higher (n = 68, 45.9 % of all sequences). The mean number of breakpoints per 100 nucleotide window ranged from 1 to 4 for the MHC-I, and from 4 to 21 for the MHC-II (Fig. 3). Consistently, three recombination hotspots were detected at MHC-II, while none was detected at MHC-I (Fig. 3).

Selection

Our analyses provided evidence for a much stronger positive (diversifying) selection at MHC-II than MHC-I in the Eurasian coot. Bayesian methods (FUBAR) identified 15 sites under pervasive diversifying selection within MHC-II exon 2, whereas only 7 sites were recognized as under pervasive diversifying selection within MHC-I exon 3 (Table 2, Fig. 1). Maximum-likelihood approach (FEL) provided similar results (17 and 8 sites pervasive diversifying selection at MHC class II and class I, respectively). The number of sites under episodic diversifying selection (MEME) was also higher at MHC-II (n = 22) than MHC-I (n = 13) (Table 2, Fig. 1). At the same time, the number of sites under negative (purifying) selection was much lower at MHC-II when compared with MHC-I (5 vs. 16 sites, Table 2). Stronger diversifying selection at MHC-II than MHC-I was also inferred from the analysis of nucleotide substitution rates (dN/dS), as measured across putative PBR sites recognized in non-passerine birds (2.83 vs. 1.73) or humans (3.37 vs. 1.64) (Table 2). Positions of positively selected sites at MHC-II showed a moderately high agreement with positions of putative PBR sites in non-passerines (ICC = 0.55) and humans (ICC = 0.47), and with positions of sites previously identified as under positive selection in the Iberian coot population (ICC = 0.59) (all $P < 0.001$). Agreement rate at MHC-I was much lower (ICC = 0.28, $P = 0.003$ for putative non-passerine PBR) or non-significant ($P = 0.90$ for human PBR).

Discussion

In order to characterize the MHC in a non-model bird species, the Eurasian coot, we have combined two molecular approaches, i.e. genome assembly and high-throughput sequencing at the population level. The analysis of genome sequences revealed the presence of an unexpectedly large number of MHC-I loci, as contrasted to the other nonpasserine bird species. This finding confirmed previous genomic analyses of MHC architecture in birds, indicating that the genome assembly approach may produce a better resolution of the MHC region compared to the population-wide genotyping of single exons[20]. On the other hand, our screening of MHC polymorphism at the population level provided support for contrasting evolutionary trajectories at both MHC classes in the Eurasian coot. Although we found a similar allelic and nucleotide diversity at MHC-I and MHC II genes, the mechanisms responsible for maintenance of this variation clearly differed, as MHC-II showed stronger signal of positive (diversifying) selection and recombination (gene conversion).

In general, MHC studies in birds demonstrated that most species of nonpasserines seem to have a smaller number of both MHC-I and MHC-II loci compared to passerines, in which the rate of MHC gene duplication is considerably greater[19]. Our study indicates that coot genome may contain more MHC-I loci (ca. 10 loci suggested by the analysis genome assembly, at least 5 loci indicated by targeted exon sequencing) than genomes of most nonpasserine bird species studied thus far. For example, birds from the Galliformes order (landfowl), such as the domestic chicken *Gallus gallus*, common quail *Coturnix coturnix*, wild turkey *Meleagris gallopavo* and golden pheasant *Chrysolophus pictus*, have long been recognized to have a compact MHC containing between one and three MHC-I genes[46,47,48]. In fact, the domestic chicken is known to have the most compact MHC region ever reported (the minimal essential MHC), containing only one dominantly-expressed locus at each class[49,50]. A recent broad-scale comparative analysis confirmed that the vast majority (90%) of nonpasserine species have only three or fewer MHC loci of a given class[19]. Higher numbers of MHC loci were recorded in nonpasserines only exceptionally, e.g. eight gene copies of MHC-I or MHC-II were found in blue petrel *Halobaena caerulea*[51] and blakiston's fish owl *Ketupa blakistoni*[52], and the maximum level of MHC expansion among nonpasserines was recorded in tufted duck *Aythya fuligula* and carmine bee-eater *Merops nubicus* (up to 11-12 functional MHC-I loci)[20]. Contrasted with general patterns of MHC architecture in non-passerine birds, the number of MHC-I loci (5-10) in the Eurasian coot seems to be exceptional and further research is needed to examine whether this is a conserved feature of the MHC across rail family.

In contrast to nonpasserine birds, passerines often demonstrate considerably larger numbers of MHC genes, with the average numbers of 7.5 MHC-I and 5 MHC-II loci per species[19]. To date, population-wide genotyping of PBR exons in birds has revealed the presence of up to 33 MHC-I loci in the sedge warbler[10] and 22 MHC-II loci in the common yellowthroat[53]. However, this approach is expected to underestimate the total number of MHC loci, as it assumes heterozygosity at each locus, and reliable information on MHC copy number variation has been obtained for only a handful of model species with relatively simple architecture of the MHC region[47,54,55]. This issues have been largely alleviated by advancements in third-generation sequencing (TGS) techniques, which allow much more reliable reconstruction of complex MHC regions, even in passerines. For example, recent analyses of TGS-based genomes provided support for rapid MHC expansion in manakins (Pipridae) with up to 180 MHC-II loci recorded in the golden-collared manakin *Manacus vitellinus*[20].

A relatively large number of MHC loci in the Eurasian coot may suggest that the species has been exposed to a greater pathogen-driven selection over its evolutionary history, when compared to other nonpasserines. However, the processes of adaptive MHC gene duplications in response to pathogen and parasite pressure may have a complex nature. On one hand, a comparative study conducted across 54 divergent avian species showed that blood parasite diversity negatively covaried with the number of MHC-I loci, suggesting their effective eradication from hosts with broader spectrum of MHC allelic variants[56]. On the other hand, a positive correlation between the MHC-II gene copy number and helminth richness has been reported in nonpasserines, which may reflect an evolutionary (historical) pressure of parasitic faunas on the MHC expansion[57]. It is possible that an apparent discrepancy between these studies is due to different evolutionary trajectories of MHC-I and MHC-II or due to distinct evolutionary processes shaping MHC architecture in major avian lineages (the first study was based mainly on passerines[56]). Irrespective of these differences, pathogens and parasites seem to constitute a leading force that govern the evolution of MHC architecture and duplication processes within this region. At the same time, there are scarcely any studies testing for evolutionary associations between pathogen diversity and MHC copy numbers in vertebrate taxa other than birds. Also, most of this research focused on associations of pathogens or parasites with MHC polymorphism and allelic diversity, rather than with gene copy numbers. For instance, high parasite diversity was associated with greater diversity of MHC-II alleles in some mammalian clades, including rodents, bats and ungulates[58,59]. Despite empirical evidence supporting associations of duplication processes at the MHC with pathogen-driven selection, we acknowledge that pathogen pressure per se seems insufficient to fully explain a huge MHC gene copy number variation, which is observed among divergent vertebrate lineages. Simulation studies indicate that the evolution

of MHC numbers may also be driven by the inherent costs of expressing multiple allelic variants[60], such as the risk of autoimmune diseases or the depletion of T lymphocyte receptor (TLR) reservoir[61,62].

Despite extensive duplications at the MHC-I in the Eurasian coot, our analyses failed to find any convincing evidence for pseudogenization processes. All MHC-I and MHC-II PBR exon sequences retrieved in our study were functional and we found no allelic variants with stop codons or frameshift mutations. Although some haplotypes retrieved from our genome assembly showed the signs of non-functionality, they could be most likely attributed to errors in genotyping or assembly procedures. In accordance with our findings, the previous study on the MHC-II in another (Iberian) population of the Eurasian did not reveal the presence of pseudogenes[26]. In general, the "birth and death" evolution model assumes that over evolutionary times some copies of MHC genes should preserve their primary functions, some others my get duplicated and gain novel functions, while some others turn into pseudogenes via non-functional mutations (e.g. indels)[63,64]. Here, we found a relatively minor frequency of variants with a one-codon deletion at MHC-II exon 2, but they all retained functionality. In general, passerines have a much more complex MHC architecture, with many extremely polymorphic and duplicated genes, but also with long introns and pseudogenes. In the great reed warbler *Acrocephalus arundinaceus* 25% MHC-I allelic variants presumably originated from pseudogenes, as they contained a 5-bp deletion in exon 3, which leads to the shift of the reading frame[65]. Similarly, 20% MHC-II sequences were identified as pseudogenic in the red-winged blackbird *Agelaius phoenicus*[66], while in the house finch *Carpodacus mexicanus* frameshift mutations were recorded in both MHC-II exon 2 and 3[67]. In nonpasserines, occurrence of pseudogenes have rarely been reported and non-functional allelic variants, if present, usually showed minor frequencies[68,50,69].

Our study showed that MHC expansion in the Eurasian coot was accompanied with high levels of allelic diversity and, in total, we retrieved 165 MHC-I and 147 MHC-II alleles within the Central European population. Even a greater level of the MHC-II allelic polymorphism was previously revealed in the Iberian population of this species (265 alleles), yet these analyses were based on the sample several times greater than ours (902 vs. 283 individuals)[26]. Our findings in combination with previous research[26] indicate that European coots show the highest degree of MHC-I and MHC-II allelic polymorphism ever reported in nonpasserines. Within non-Passeriformes, a relatively high level of MHC polymorphism (though markedly lower than in coots) was found primarily at class II genes. For example, 109 MHC-II alleles were detected in the Mediterranean population of the great flamingo *Phoenicopterus roseus*[70], and 103 allelic variants were described in a population of lesser kestrels *Falco naumanni*[71]. Yet, most of nonpasserine species demonstrate a considerably smaller allelic polymorphism at the MHC (particularly MHC-I) genes. For instance, the MHC-I allelic richness

was estimated at 47 alleles in black-tailed godwit *Limosa limosa*[72], 38 alleles in red-billed gull *Chroicocephalus novaehollandiae scopulinus*[69] and 36 alleles in red knot *Calidris canutus*[73]. Compared to the Eurasian coot, the higher levels of MHC polymorphism have been demonstrated only in passerine birds, reaching hundreds or thousands alleles in some populations[10,11,74]. A comparison of MHC allelic richness between different coot populations showed that most MHC-II allelic variants (86%) found in our (Central-European) population were previously described in the Iberian population[26], despite the fact that both populations are spatially separated (the vast majority of our birds do not reach the Iberian Peninsula even during winter[75]). This indicates a strong homogenization of the MHC pool at a relatively large geographical scale and a minor significance of local adaptation processes in shaping MHC polymorphism in coots. This, in turn, may suggest a relatively strong spatial homogeneity in the pressure of extracellular pathogens on this species.

Population-wide genotyping of key PBR exons in the Eurasian coot indicated that an overall level of polymorphism was similar between MHC-I and MHC-II genes. Although the number of alleles and segregation sites, as well as the total number of mutations were slightly greater for MHC-I, we observed slightly higher nucleotide diversity at the MHC-II. Despite this similarity, our analyses revealed a considerably stronger signature of positive selection and recombination at the MHC-II than MHC-I. At MHC-I we detected only 8 sites under pervasive positive selection, which was half the number of positively selected sites at the MHC-II (n=17). The rate of non-synonymous to synonymous nucleotide substitutions within the putative PBR sites was also greater within MHC-II, indicating a stronger pathogen-driven diversifying selection. Finally, we found a greater number of recombination events and a greater percentage of recombinant sequences at the MHC-II and, in general, recombination mechanisms are known to effectively generate MHC variation under strong pathogen pressure[76]. All these results seem to suggest that, in an evolutionary context, extracellular parasites might have exerted a stronger selective pressure on the MHC-II genes in coots, when compared with the intracellular pathogens, whose antigens are recognized by the MHC-I. A similar pattern (i.e. stronger diversifying selection at MHC-II than MHC-I) was previously described in other nonpasserine lineages, e.g. Procellariformes[50] and Phoenicopteriformes, as well as in a large-scale analysis of selection at the nonpasserine MHC[27]. An opposite pattern was recorded in passerines, where diversifying selection is usually stronger at the MHC-I[27]. In general, the evolutionary trajectories of MHC-I and MHC-II genes in passerine and nonpasserine birds may differ, suggesting a contrasting pressure by extra- and intracellular pathogens[27]. As the diversity of ecological niches for extracellular parasites should increase along with the structural size of their hosts[77], these differences could possibly be linked to greater body sizes of nonpasserines, which are thus likely to interact with more diverse faunas of extracellular parasites.

To sum up, the combination of complementary molecular approaches, i.e. targeted exon genotyping via high-throughput sequencing and de novo genome assembly, allowed us to obtain a high-quality resolution of MHC polymorphism in a non-model bird species from rail family. Our study provided novel insights into the evolution of key pathogen-recognition genes of the adaptive immunity in a poorly researched lineage of birds and revealed some unique features of the MHC in our study species (extraordinary duplication rate and allelic richness). Finally, our comparisons of selection and recombination processes clearly indicated that the polymorphism of MHC-I and MHC-II genes in birds may be governed by distinct mechanisms, thereby providing evidence for variation in the evolutionary trajectories of both MHC classes.

Acknowledgements We thank Radosław Włodarczyk for participation in fieldwork and Amelia Chyb for assistance in laboratory analyses.

Funding The study was financially supported by the research grant of the National Science Centre in Poland (2020/38/E/NZ8/00143).

Conflicts of interest/Competing interests The authors declare that they have no conflict or competing interests.

Availability of data and material Raw data have been deposited in GenBank (whole genome shotgun sequences: JABXFB010000001-JABXFB010031348; whole genome shotgun scaffolds: MU065399-MU069429; whole genome assembly: GCA_013372525.1; high-throughput MHC-I sequences: TBA; novel high-throughput MHC-II sequences: MT74863–MT748786).

References

1. Klein, J. Natural history of the major histocompatibility complex. *Wiley* (1986).
2. Trowsdale, J. "Both man & bird & beast": comparative organization of MHC genes. *Immunogenetics* **41**, 1-17 (1995).
3. Bernatchez, L., & Landry, C. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *J. Evol. Biol.* **16**, 363-377 (2003).
4. Steimle, V., Otten, L. A., Zufferey, M., & Mach, B. Complementation cloning of an MHC class II transactivator mutated in hereditary MHC class II deficiency (or bare lymphocyte syndrome). *Cell* **75**, 135-146 (1993).
5. Duraes, F. V., Niven, J., Dubrot, J., Hugues, S., & Gannagé, M. Macroautophagy in endogenous processing of self-and pathogen-derived antigens for MHC class II presentation. *Front. Immunol.* **6**, (2015).
6. O'Connor, E. A., Strandh, M., Hasselquist, D., Nilsson, J. Å., & Westerdahl, H. The evolution of highly variable immunity genes across a passerine bird radiation. *Mol. Ecol.* **25**, 977-989 (2016).
7. Gaudieri, S., Dawkins, R. L., Habara, K., Kulski, J. K., & Gojobori, T. SNP profile within the human major histocompatibility complex reveals an extreme and interrupted level of nucleotide diversity. *Genome Res.* **10**, 1579-1586 (2000).
8. Garrigan, D., & Hedrick, P. W. Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution* **57**, 1707-1722 (2003).
9. Robinson, J., Soormally, A. R., Hayhurst, J. D., & Marsh, S. G. The IPD-IMGT/HLA Database – new developments in reporting HLA variation. *Hum. Immunol.* **77**, 233-237 (2016).
10. Biedrzycka, A., O'Connor, E., Sebastian, A., Migalska, M., Radwan, J., Zając, T. et al. Extreme MHC class I diversity in the sedge warbler (*Acrocephalus schoenobaenus*); selection patterns and allelic divergence suggest that different genes have different functions. *BMC Evol. Biol.* **17**, 159 (2017).
11. Bollmer, J. L., Dunn, P. O., Freeman-Gallant, C. R., & Whittingham, L. A. Social and extra-pair mating in relation to major histocompatibility complex variation in common yellowthroats. *Proc. R. Soc. B* **279**, 4778-4785 (2012).
12. Penn, D. J., & Potts, W. K. The evolution of mating preferences and major histocompatibility complex genes. *Am. Nat.* **153**, 145-164 (1999).
13. Hedrick, P. W. Pathogen resistance and genetic variation at MHC loci. *Evolution* **56**, 1902-1908 (2002).
14. Apanius, V., Penn, D., Slev, P. R., Ruff, L. R., & Potts, W. K. The nature of selection on the major histocompatibility complex. *Crit. Rev. Immunol*. **17** (1997).
15. Sommer, S. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front. Zool.* **2**, 1-18 (2005).
16. Hughes, A. L. & Yeager, M. Natural selection at major histocompatibility complex loci of vertebrates. *Ann. Rev. Genet.* **32**, 415-435 (1998).
17. Furlong, R. F., & Yang, Z. Diversifying and purifying selection in the peptide binding region of DRB in mammals. *J. Mol. Evol.* **66**, 384-394 (2008).
18. Miller, H. C., & Lambert, D. M. Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae). *Immunogenetics* **56**, 178-191 (2004).
19. Minias, P., Pikus, E., Whittingham, L. A., & Dunn, P. O. Evolution of copy number at the MHC varies across the avian tree of life. *Genome Biol. Evol.* **11**, 17-28 (2019).

20. He, K., Minias, P., & Dunn, P. O. Long-read genome assemblies reveal extraordinary variation in the number and structure of MHC loci in birds. *Genome Biol. Evol.* **1**, evaa270 (2021).

21. Goebel, J., Promerová, M., Bonadonna, F., McCoy, K. D., Serbielle, C., Strandh, M. et al. 100 million years of multigene family evolution: origin and evolution of the avian MHC class IIB. *BMC Genomics* **18**, 1-9 (2017).

22. Reusch, T. B., Schaschl, H., & Wegner, K. M. Recent duplication and inter-locus gene conversion in major histocompatibility class II genes in a teleost, the three-spined stickleback. *Immunogenetics* **56**, 427-437 (2004).

23. Hess, C. M. & Edwards, S. V. The evolution of the Major Histocompatibility Complex in birds. *BioScience* **52**, 423-431 (2002).

24. Babik, W. Methods for MHC genotyping in non-model vertebrates. *Mol. Ecol. Res.* **10,** 237-251 (2010).

25. BirdLife International (2021) Species factsheet: *Fulica atra*. Downloaded from http://www.birdlife.org on 26/05/2021.

26. Alcaide, M., Munoz, J., Martínez-de la Puente, J., Soriguer, R. & Figuerola, J. Extraordinary MHC class II B diversity in a non-passerine, wild bird: the Eurasian Coot *Fulica atra* (Aves: Rallidae). *Ecol. Evol.* **4**, 688-698 (2014).

27. Minias, P., Pikus, E., Whittingham, L. A., & Dunn, P. O. A global analysis of selection at the avian MHC. *Evolution* **72**, 1278-1293 (2018).

28. Bushnell, B. BBMap: a fast, accurate, splice-aware aligner. *Ernest Orlando Lawrence Berkeley National Laboratory* (2014).

29. Laetsch, D. R. & Blaxter, M. L. BlobTools: Interrogation of genome assemblies. *F1000Research* **6**, 1287 (2017).

30. Seppey, M., Manni, M., & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness. *Gene Prediction* (pp. 227-245). Humana, New York, NY (2019).

31. Holt, C., & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).

32. Bjorkman, P., J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. T., & Wiley, D. C. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329,** 506-512 (1987).

33. Brown, J. H., Jardetzky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., & Strominger, J. L. et al. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364**, 33-39 (1993).

34. Minias, P., He, K., & Dunn, P. O. The strength of selection is consistent across both domains of the MHC class I peptide-binding groove in birds. *BMC Ecol. Evol.* **21**, 80 (2021).

35. Alcaide, M., Edwards, S. V., Cadahía, L., & Negro, J. J. MHC class I genes of birds of prey: isolation, polymorphism and diversifying selection. *Conserv. Genet.* **10**, 1349 (2009).

36. Sebastian, A., Herdegen, M., Migalska, M., & Radwan, J. amplisas: A web server for multilocus genotyping using next-generation amplicon sequencing data. *Mol. Ecol. Res.* **16**, 498-510 (2016).

37. Biedrzycka, A., Sebastian, A., Migalska, M., Westerdahl, H., & Radwan, J. Testing genotyping strategies for ultra-deep sequencing of a co-amplifying gene family: MHC class I in a passerine bird. *Mol. Ecol. Res.* **17**, 642-655 (2017).

38. Magoč, T., & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957-2963 (2011).

39. Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, vev003 (2015).

40. Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P. & Ramos-Onsins, S.E. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299-3302 (2017).

41. Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., & Kosakovsky Pond, S. L. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol. Biol. Evol.* **35**, 773-777 (2018).

42. Anisimova, M., Nielsen, R., & Yang, Z. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* **164**, 1229-1236 (2003).

43. Saper, M. A., Bjorkman, P., & Wiley, D. C. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J. Mol. Biol.* **219**, 277-319 (1991).

44. Brown, J. H., Jardetzky, T., Saper, M. A., Samraoui, B., Bjorkman, P. J., & Wiley, D. C. A hypothetical model of the foreign antigen binding site of class II histocompatibility molecules. *Nature* **332**, 845-850 (1988).

45. Gamer, M., Lemon, J., Gamer, M. M., Robinson, A., & Kendall's, W. (2012). *irr*: various coefficients of interrater reliability and agreement. *R package ver. 0.84.* Available at https://CRAN.R-project.org/package=irr.

46. Zeng, Q. Q., Zhong, G. H., He, K., Sun, D. D., & Wan, Q. H. Molecular characterization of classical and nonclassical MHC class I genes from the golden pheasant (*Chrysolophus pictus*). *Int. J. Immunogenet.* **43**, 8-17 (2016).

47. Chaves, L. D., Krueth, S. B., & Reed, K. M. Characterization of the turkey MHC chromosome through genetic and physical mapping. *Cytogenet. Genome. Res.*, **117**, 213-220 (2007).

48. Shiina, T., Shimizu, S., Hosomichi, K., Kohara, S., Watanabe, S., Hanzawa, K., et al. Comparative genomic analysis of two avian (quail and chicken) MHC regions. *J. Immunol.* **172**, 6751-6763 (2004).

49. Kaufman, J., Milne, S., Göbel, T. W., Walker, B. A., Jacob, J. P., Auffray, C. et al. The chicken B locus is a minimal essential major histocompatibility complex. *Nature* **401**, 923-925 (1999).

50. Kaufman, J., Völk, H. E. I. N. E. R., & Wallny, H. J. A" minimal essential Mhc" and an" unrecognized Mhc": two extremes in selection for polymorphism. *Immunol. Rev.* **143**, 63-88 (1995).

51. Strandh, M., Lannefors, M., Bonadonna, F., & Westerdahl, H. Characterization of MHC class I and II genes in a subantarctic seabird, the blue petrel, *Halobaena caerulea* (Procellariiformes). *Immunogenetics* **63**, 653-666 (2011).

52. Kohyama, T. I., Omote, K., Nishida, C., Takenaka, T., Saito, K., Fujimoto, S. et al. Spatial and temporal variation at major histocompatibility complex class IIB genes in the endangered Blakiston's fish owl. *Zool. Lett.* **1**, 1-12 (2015).

53. Dunn, P. O., Bollmer, J. L., Freeman-Gallant, C. R., & Whittingham, L. A. MHC variation is related to a sexually selected ornament, survival, and parasite resistance in common yellowthroats. *Evolution* **67**, 679-687 (2013).

54. Chen, L. C., Lan, H., Sun, L., Deng, Y. L., Tang, K. Y., & Wan, Q. H. Genomic organization of the crested ibis MHC provides new insight into ancestral avian MHC structure. *Sci. Rep.* **5**, 7963 (2015).

55. Balakrishnan, C. N., Ekblom, R., Völker, M., Westerdahl, H., Godinez, R., Kotkiewicz, H. et al. Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biol.* **8**, 29 (2010).

56. Vincze, O., Loiseau, C., & Giraudeau, M. Avian blood parasite richness decreases with major histocompatibility complex class I loci number. *Biol. Lett.* **17**, 20210253 (2021).

57. Minias, P., Gutiérrez, J. S., & Dunn, P. O. Avian major histocompatibility complex copy number variation is associated with helminth richness. *Biol. Lett.* **16**, 20200194 (2020).

58. Winternitz, J. C., Minchey, S. G., Garamszegi, L. Z., Huang, S., Stephens, P. R., & Altizer, S. Sexual selection explains more functional variation in the mammalian major histocompatibility complex than parasitism. *Proc. R. Soc. B* **280**, 20131605 (2013).

59. Goüy de Bellocq, J., Charbonnel, N., & Morand, S. Coevolutionary relationship between helminth diversity and MHC class II polymorphism in rodents. *J. Evol. Biol.* **21**, 1144-1150 (2008).

60. Bentkowski, P., & Radwan, J. Evolution of major histocompatibility complex gene copy number. *PLoS Comput. Biol.* **15**, e1007015 (2019).

61. Migalska, M., Sebastian, A., & Radwan, J. Major histocompatibility complex class I diversity limits the repertoire of T cell receptors. *Proc. Natl. Acad. Sci. USA* **116**, 5021-5026 (2019).

62. Nowak, M. A., Tarczy-Hornoch, K., & Austyn, J. M. The optimal number of major histocompatibility complex molecules in an individual. *Proc. Natl. Acad. Sci. USA* **89**, 10896-10899 (1992).

63. van Oosterhout, C. A new theory of MHC evolution: beyond selection on the immune genes. *Proc. R. Soc. B* **276**, 657-665 (2009).

64. Nei, M., Gu, X., & Sitnikova, T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci*. *USA* **94**, 7799-7806 (1997).

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

66. Edwards, S. V., Gasper, J., & March, M. Genomics and polymorphism of Agph-DAB1, an Mhc class II B gene in red-winged blackbirds (*Agelaius phoeniceus*). *Mol. Biol. Evol.* **15**, 236-250 (1998).

67. Hess, C. M., Gasper, J., Hoekstra, H. E., Hill, C. E., & Edwards, S. V. MHC class II pseudogene and genomic signature of a 32-kb cosmid in the house finch (*Carpodacus mexicanus*). *Genome Res*. **10**, 613-623 (2000).

68. Gangoso, L., Alcaide, M., Grande, J. M., Muñoz, J., Talbot, S. L., Sonsthagen, S. A. et al. Colonizing the world in spite of reduced MHC variation. *J. Evol. Biol.* **25**, 1438-1447 (2012).

69. Cloutier, A., Mills, J. A., & Baker, A. J. Characterization and locus-specific typing of MHC class I genes in the red-billed gull (*Larus scopulinus*) provides evidence for major, minor, and nonclassical loci. *Immunogenetics* **63**, 377-394 (2011).

70. Gillingham, M. A., Béchet, A., Courtiol, A., Rendón-Martos, M., Amat, J. A. & Samraoui, B. et al. Very high MHC Class IIB diversity without spatial differentiation in the Mediterranean population of greater Flamingos. *BMC Evol. Biol.* **17**, 56 (2017).

71. Alcaide, M., Edwards, S. V., Negro, J. J., Serrano, D., & Tella, J. L. Extensive polymorphism and geographical variation at a positively selected MHC class II B gene of the lesser kestrel (*Falco naumanni*). *Mol. Ecol.* **17**, 2652-2665 (2008).

72. Pardal, S., Drews, A., Alves, J. A., Ramos, J. A., & Westerdahl, H. Characterization of MHC class I in a long distance migratory wader, the Icelandic black-tailed godwit. *Immunogenetics* **69**, 463-478 (2017).

73. Buehler, D. M., Verkuil, Y. I., Tavares, E. S., & Baker, A. J. Characterization of MHC class I in a long-distance migrant shorebird suggests multiple transcribed genes and intergenic recombination. *Immunogenetics* **65**, 211-225 (2013).

74. Sepil, I., Moghadam, H. K., Huchard, E., & Sheldon, B. C. Characterization and 454 pyrosequencing of Major Histocompatibility Complex class I genes in the great tit reveal complexity in a passerine system. *BMC Ecol. Evol.* **12**, 68 (2012).

75. Chyb, A., Jedlikowski, J., Włodarczyk, R., & Minias, P. Consistent choice of landscape urbanization level across the annual cycle in a migratory waterbird species. *Sci. Rep.* **11**, 836 (2021).

76. Spurgin, L. G., Van Oosterhout, C., Illera, J. C., Bridgett, S., Gharbi, K., Emerson, B. C., et al. Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Mol. Ecol.* **20**, 5213-5225 (2011).

77. Poulin, R. Phylogeny, ecology, and the richness of parasite communities in vertebrates. *Ecol. Monogr.* **65**, 283-302 (1995).

Table 1. Polymorphism of MHC class I and MHC class II in the Eurasian coot, as inferred from genome assembly and high-throughput (Illumina sequencing).

MHC	Method	Exon	Length (bp)	No. sequences	No. segregating sites	Total no. mutations	Nucleotide diversity
Class I	Genome assembly	1	64	25	38	56	0.158
		2	264	27	130	160	0.107
		3	276	27	131	166	0.122
		4	273	27	90	103	0.051
		5	105	23	43	54	0.116
	Illumina sequencing	3	273	165	181	251	0.103
Class II	Illumina sequencing	2	270	147	103	143	0.118

Table 2. Selection signature at the MHC class I exon 3 and MHC class II exon 2.

MHC	Exon	No. of sites			Nucleotide substitution rates (*dN*/*dS*)		
		Pervasive negative selection (FUBAR/FEL)	Pervasive positive selection (FUBAR/FEL)	Episodic positive selection (MEME)	All sites	Non-passerines PBR	Human PBR
Class I	3	13/16	7/8	13	1.702	1.73	1.64
Class II	2	4/5	15/17	22	0.932	2.83	3.37

Figure 1. Alignments of amino acid sequences of MHC class I exon 3 (A) and MHC class II exon 2 (B) in the Eurasian coot. Five most frequent variants of each exon are shown. Dots indicate amino acids identical with the reference consensus sequence. Sites under pervasive and episodic positive selection are marked with dark and light red, sites under pervasive negative selection are marked with blue. Pervasive and episodic selection was assessed with FEL/FUBAR and MEME, respectively, using non-recombinant sequences only. Variation in selection parameter (*dN* – *dS*) is shown above the alignments. Putative peptide-binding residues (PBR) of non-passerine birds (based on the global analysis of selection at the avian MHC by Minias et al. 2018) and humans (based on the crystallographic structure of MHC molecules by Saper et al. 1991 and Brown et al. 1993) are indicated with crosses (+) and large dots (•) at the top of each panel. Positively selected sites previously recognized at MHC class II exon 2 by Alcaide et al. (2014) are indicated with asterisks (*).

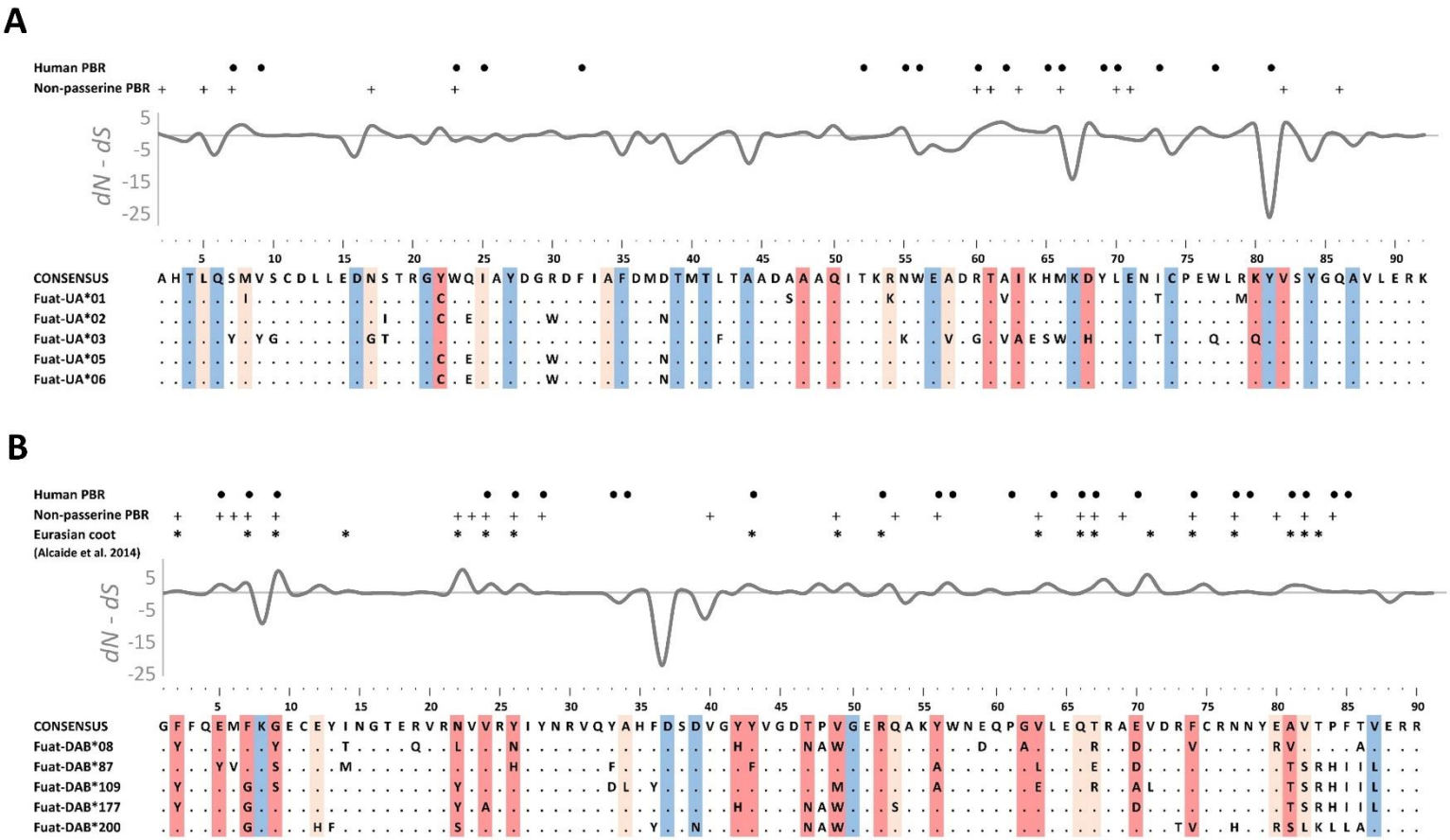


Figure 2. The number of MHC class I (A) and MHC class II (B) allelic variants recorded per individual in the Eurasian coot.

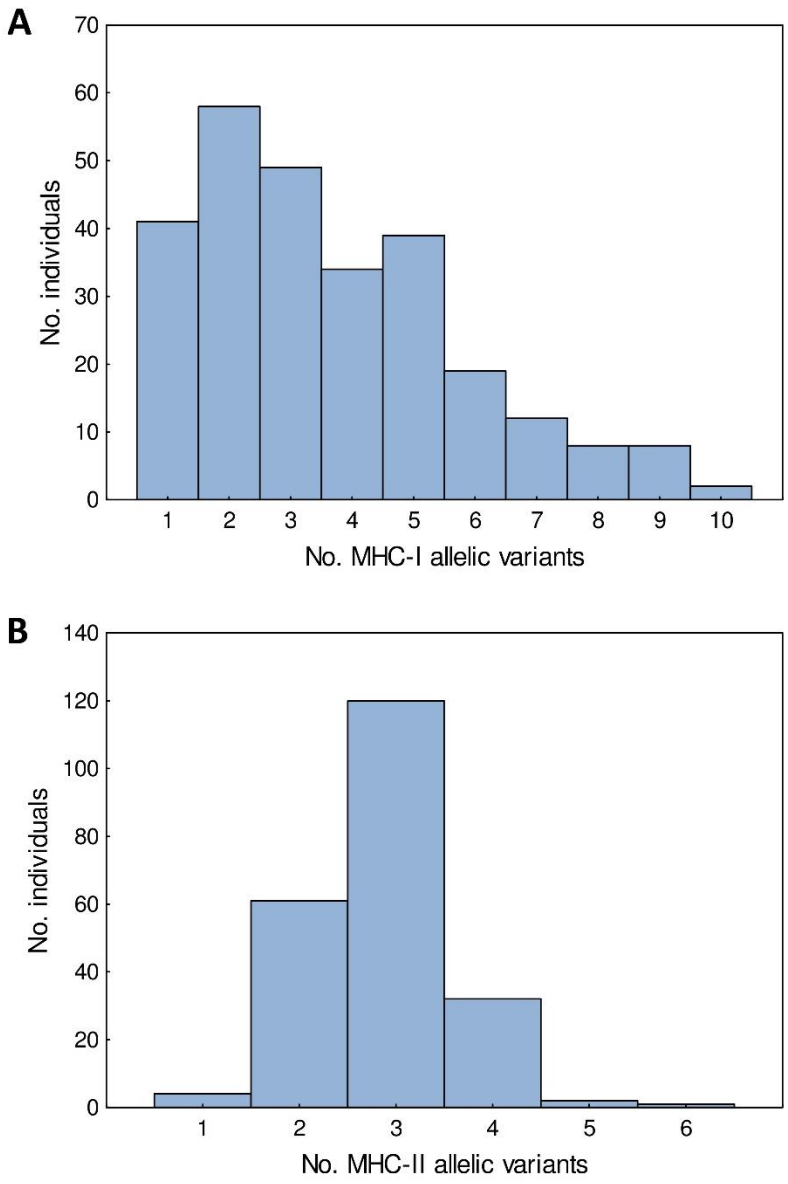


Figure 3. Recombination signal at the MHC class I exon 3 (A) and class II exon 2 (B). Black line indicates the number of breakpoints detectable within a 100 nucleotide (nt) window. Dark and light grey areas indicate the 95% and 99% confidence intervals for the expected degrees of breakpoint clustering in the absence of recombination hot- and cold-spots (as assessed with the local hot/cold spot test). Positions in alignment where black line emerges above the grey areas indicate a recombination hot-spot, while positions where it drops below the grey areas indicate a recombination cold-spot.

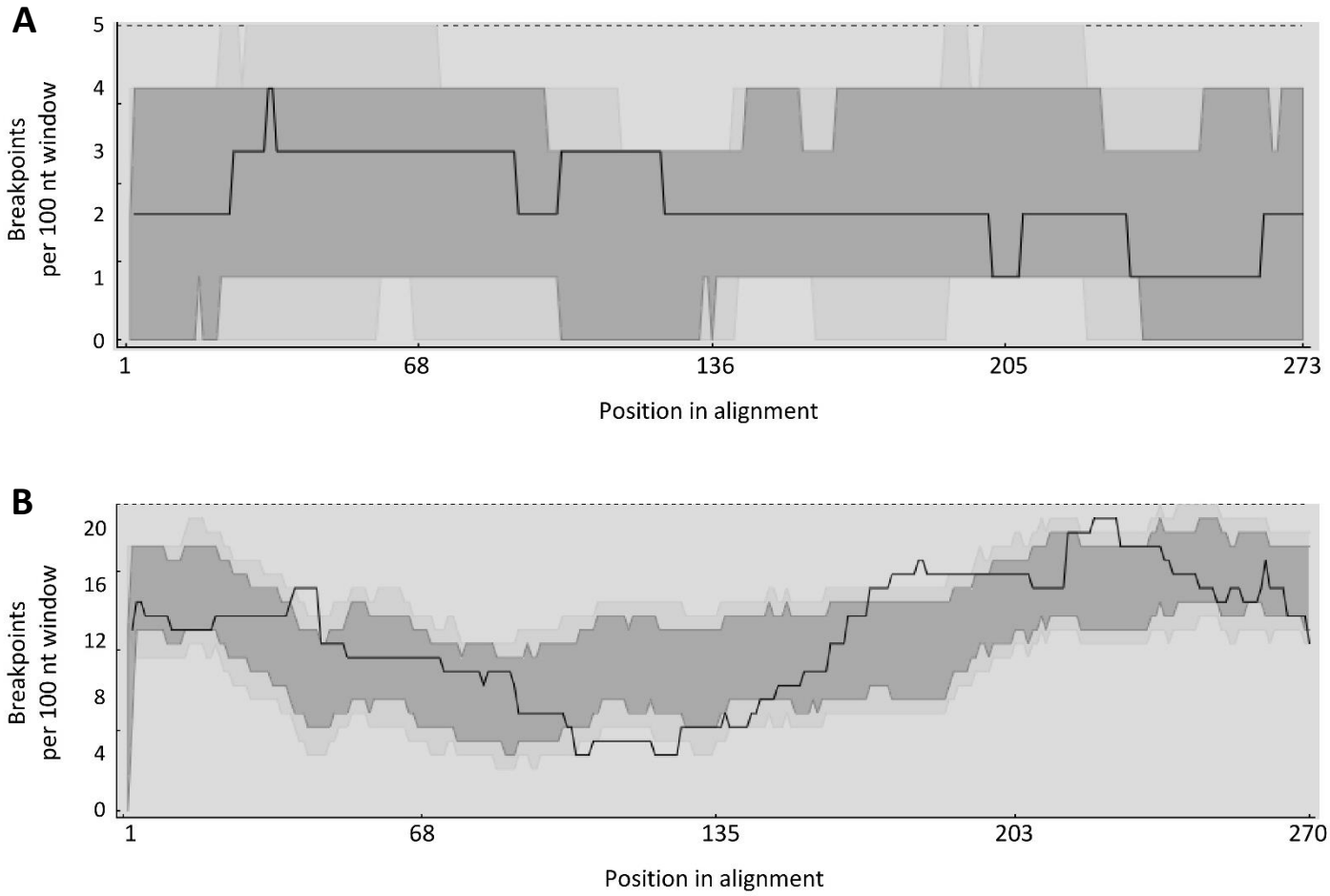
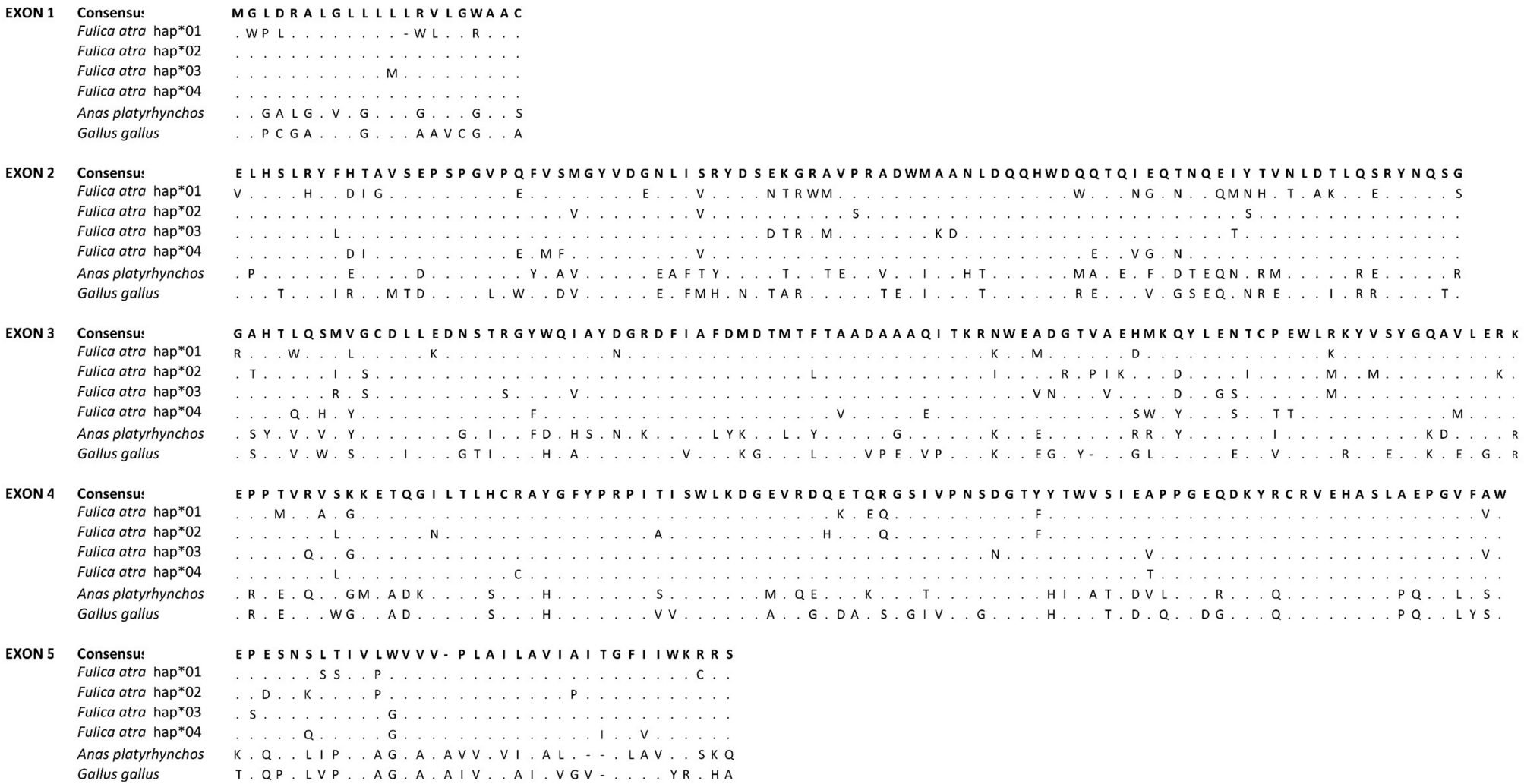
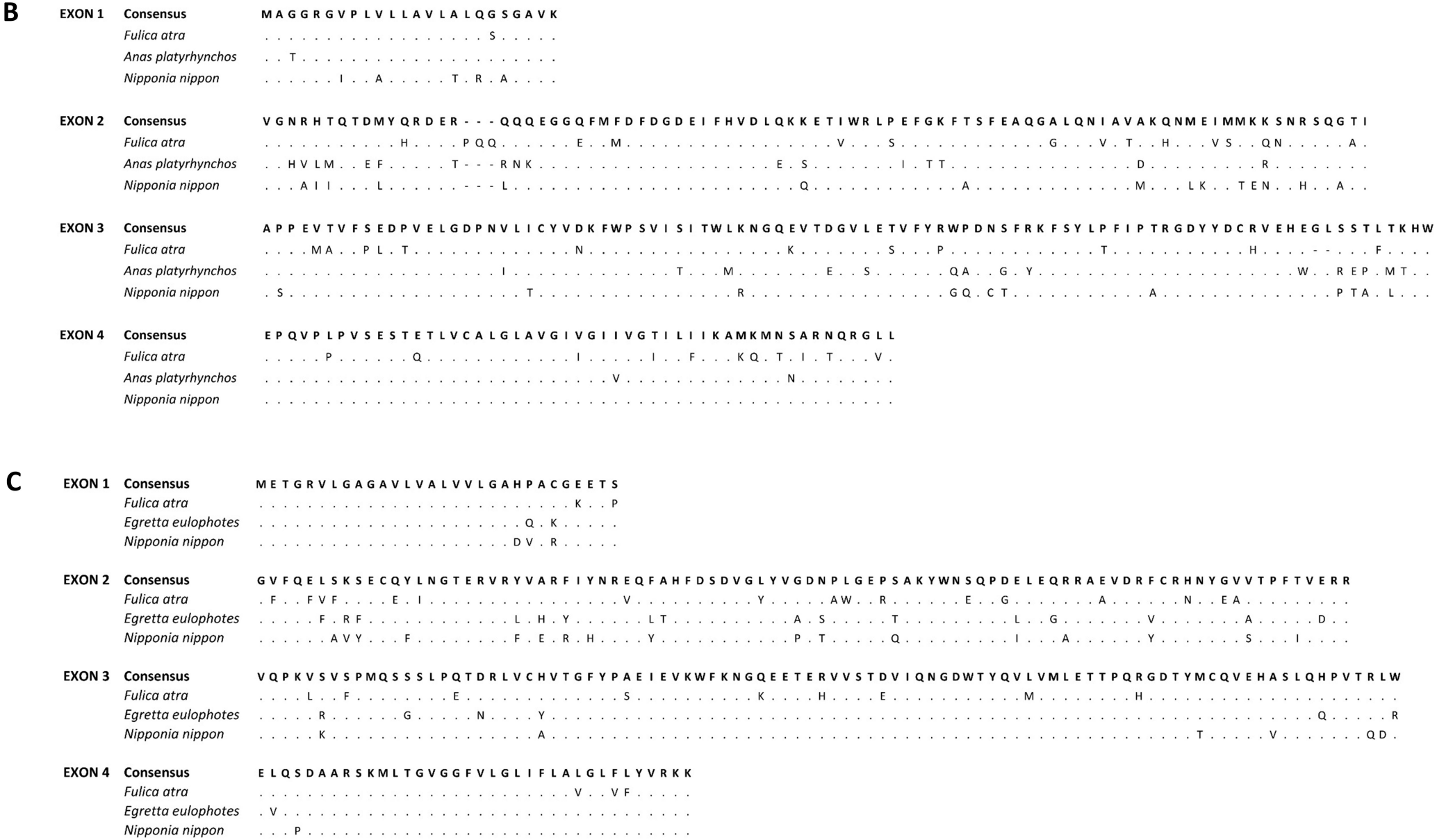


Figure S1. Alignments of amino acid MHC-I (A), MHC-IIA (B), and MHC-IIB (C) sequences of the Eurasian coot *Fulica atra*, as retrieved from *de novo* genome assembly. Four randomly selected MHC-I haplotypes were shown. Sequences from the mallard *Anas platyrhynchos* (MH218846 for MHC-I, HM070250 for MHC-IIA), chicken *Gallus gallus* (AF459830 for MHC-I), crested ibis *Nipponia nippon* (AB872444 for MHC-IIA and MHC-IIB), and Chinese egret *Egretta eulophotes* (KC282841 for MHC-IIB) were added for reference.

A




B
EXON 1 Consensus M A G G R G V P L V L L A V L A L Q G S G A V K
Fulica atra . . . . . S . . . . .
Anas platyrhynchos . . T . . . . . .
Nipponia nippon . . . . . I . . A . . . . . T . R . A . . . .
EXON 2 Consensus V G N R H T Q T D M Y Q R D E R - - - Q Q Q E G G Q F M F D F D G D E I F H V D L Q K K E T I W R L P E F G K F T S F E A Q G A L Q N I A V A K Q N M E I M M K K S N R S Q G T I
Fulica atra . . . . . H . . . . P Q Q . . . . . E . . M . . . . . V . . . S . . . . . G . . . V . T . . H . . . V S . . Q N . . . . . A .
Anas platyrhynchos . . H V L M . . E F . . . . . T - - - R N K . . . . . E . S . . . . . I . T T . . . . . D . . . . . R . . . . . .
Nipponia nippon . . . A I I . . . L . . . . . - - - L . . . . . Q . . . . . A . . . . . M . . . L K . . T E N . . H . . A . .
EXON 3 Consensus A P P E V T V F S E D P V E L G D P N V L I C Y V D K F W P S V I S I T W L K N G Q E V T D G V L E T V F Y R W P D N S F R K F S Y L P F I P T R G D Y Y D C R V E H E G L S S T L T K H W
Fulica atra . . . . M A . . P L . T . . . . . N . . . . . K . . . . . S . . . P . . . . . T . . . . . H . . . . - - . . . . F . . . .
Anas platyrhynchos . . . . . I . . . . . T . . . M . . . . . E . . S . . . . . Q A . . G . Y . . . . . W . . R E P . M T . .
Nipponia nippon . S . . . . . T . . . . . R . . . . . G Q . C T . . . . . A . . . . . P T A . L . . .
EXON 4 Consensus E P Q V P L P V S E S T E T L V C A L G L A V G I V G I I V G T I L I I K A M K M N S A R N Q R G L L
Fulica atra . . . . . P . . . . . Q . . . . . I . . . . . I . . F . . . K Q . T . I . T . . . V .
Anas platyrhynchos . . . . . V . . . . . N . . . . . .
Nipponia nippon . . . . . .
C
EXON 1 Consensus M E T G R V L G A G A V L V A L V V L G A H P A C G E E T S
Fulica atra . . . . . K . . P
Egretta eulophotes . . . . . Q . K . . . . .
Nipponia nippon . . . . . D V . R . . . . .
EXON 2 Consensus G V F Q E L S K S E C Q Y L N G T E R V R Y V A R F I Y N R E Q F A H F D S D V G L Y V G D N P L G E P S A K Y W N S Q P D E L E Q R R A E V D R F C R H N Y G V V T P F T V E R R
Fulica atra . F . . F V F . . . E . I . . . . . V . . . . . Y . . . . . A W . . R . . . . . E . . G . . . . . A . . . . . N . . E A . . . . . .
Egretta eulophotes . . . . . F . R F . . . . . L . H . Y . . . . . L T . . . . . A . S . . . . . T . . . . . L . . G . . . . . V . . . . . A . . . . . D . .
Nipponia nippon . . . . . A V Y . . . F . . . . . F . E . R . H . . . . Y . . . . . P . T . . . . . Q . . . . . I . . . A . . . . . Y . . . . . S . . . I . . . .
EXON 3 Consensus V Q P K V S V S P M Q S S S L P Q T D R L V C H V T G F Y P A E I E V K W F K N G Q E E T E R V V S T D V I Q N G D W T Y Q V L V M L E T T P Q R G D T Y M C Q V E H A S L Q H P V T R L W
Fulica atra . . . . L . . F . . . . . E . . . . . S . . . . . K . . . . H . . . . E . . . . . M . . . . . H . . . . . .
Egretta eulophotes . . . . . R . . . . . G . . . . . N . . . . Y . . . . . Q . . . . . R
Nipponia nippon . . . . . K . . . . . A . . . . . T . . . . . V . . . . . Q D .
EXON 4 Consensus E L Q S D A A R S K M L T G V G G F V L G L I F L A L G L F L Y V R K K
Fulica atra . . . . . V . . V F . . . . .
Egretta eulophotes . V . . . . . .
Nipponia nippon . . . P . . . . . .

7 ROZDZIAŁ DRUGI

PIKUS, E., DUNN, P., MINIAS, P. HIGH MHC DIVERSITY DOES NOT CONFER FITNESS ADVANTAGE IN A WILD BIRD. JOURNAL OF ANIMAL ECOLOGY (PRACA ZGŁOSZONA DO DRUKU).

Journal of Animal Ecology

High MHC diversity does not confer fitness advantage in a wild bird

Journal:	*Journal of Animal Ecology*
Manuscript ID	Draft
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Pikus, Ewa; Uniwersytet Łódzki, Department of Biodiversity Studies and Bioeducation Dunn, Peter; University of Wisconsin-Milwaukee, Biological Sciences Minias, Piotr; Uniwersytet Łódzki, Department of Biodiversity Studies and Bioeducation
Key-words:	birds, fitness, heterozygote advantage, major histocompatibility complex, optimality hypothesis

High MHC diversity does not confer fitness advantage in a wild bird

Ewa Pikus[1], Peter O. Dunn[2], Piotr Minias[1]

[1] Department of Biodiversity Studies and Bioeducation, Faculty of Biology and Environmental Protection, University of Łódź, Banacha 1/3, 90-237 Łódź, Poland

[2] Behavioral and Molecular Ecology Group, Department of Biological Sciences, University of Wisconsin-Milwaukee

Author for correspondence

Piotr Minias

Email: pminias@op.pl

Abstract

1. Genes of the major histocompatibility complex (MHC) encode the antigen binding molecules and are an integral part of the acquired immune response of vertebrates. In general, high individual MHC diversity is expected to increase fitness by broadening the spectrum of pathogens recognized by the immune system, in accordance with the heterozygote advantage mechanism. On the other hand, the optimality hypothesis assumes that individuals with optimal (intermediate), rather than maximum diversity of the MHC will achieve the highest fitness because of inherent costs associated with expressing diverse MHC alleles.

2. Here, we tested for associations between individual diversity of the MHC class I and class II genes (binding antigens of intra- and extra-cellular pathogens, respectively) and a range of fitness-related traits (condition, ornament expression and reproduction) in an urban population of the Eurasian coot *Fulica atra*.

3. Contrary to our expectation, we found that high allelic diversity of MHC genes (both class I and II) was associated with poorer condition (lower blood haemoglobin concentrations), weaker expression of the putative ornament (smaller frontal shield), later onset of breeding and smaller clutches in the same individuals. An analysis of functional MHC allele clusters (supertypes) provided further support for negative associations of MHC diversity with fitness-related traits and we identified several specific MHC class I and II supertypes that were associated with lower phenotypic quality or reproduction. Little empirical support was found for optimality hypothesis and clutch size was the only trait showing non-liner association with MHC class II diversity.

4. Our results suggest that the costs of expressing high number of MHC alleles outweighed any benefits associated with diverse MHC repertoire, which could be driven by depauperate pathogen diversity in an urban landscape. To the best of our knowledge, this is one of the first studies providing consistent evidence for negative associations of MHC diversity with a range of fitness-related traits in a natural avian population.

Keywords: birds, fitness, heterozygote advantage, major histocompatibility complex, optimality hypothesis.

1. Introduction

The family of major histocompatibility complex (MHC) genes forms the most polymorphic region within vertebrate genome and it plays a critical role in the response of the acquired immune system (Geraghty et al., 2002). MHC molecules bind fragments of peptides derived from the processing of intracellular (MHC class I) or extracellular (MHC class II) pathogens and display them for recognition by the appropriate T lymphocytes, which initiates the production of antibodies or the destruction of pathogen-infected cells (Janeway et al., 2001). As a result of the co-evolutionary arms race between pathogens and their hosts, the antigen-binding domains of MHC molecules are subjected to strong balancing (diversifying) selection, which leads to the maintenance of high diversity of these genes in natural populations (Spurgin & Richardson, 2010).

There are several non-exclusive mechanisms of balancing selection that promote maintenance of high MHC diversity within natural populations, and these processes can be further reinforced by sexual selection (Spurgin & Richardson, 2010, Ejsmond et al., 2014). In general, females may prefer to mate with males that have “good genes”, i.e. alleles that can directly improve the genetic quality of their offspring (Brouwer et al., 2010; Dunn et al., 2013). This kind of mating mechanism may be linked to the rare allele hypothesis (negative frequency-dependent selection), where individuals with low frequency MHC alleles gain a fitness advantage, as their immune system may better detect and counteract novel pathogens with mutations that avoid the most common MHC alleles of the host (Brouwer et al., 2010; Gillingham et al., 2017). Another hypothesis proposes that the maximum diversity of alleles at the individual level may be the most favourable in terms of fitness. According to the heterozygote advantage mechanism (overdominant selection), a greater number of alleles expressed within an individual increases the spectrum of antigens recognized and, thus, provides a selective advantage in pathogen recognition, thereby improving fitness (Doherty & Zinkernagel, 1975; McClelland et al., 2003). Consistently, MHC genes of various vertebrate lineages have been subject to extensive duplication processes throughout their evolution, which enhanced an increased individual allelic MHC repertoire (Axtner & Sommer 2007). Although it is usually difficult to separate the mechanisms responsible for maintaining high MHC diversity at the population level (Spurgin & Richardson 2010), the mechanism of heterozygote advantage predicts a positive relationship between MHC diversity and fitness-related traits across individuals (henceforth referred to as MHC diversity hypothesis).

The mechanisms of heterozygote advantage do not necessarily consider any upper limits on the individual MHC allele richness. However, high MHC diversity at individual level may be

unfavourable due to an increased risk of autoimmune diseases and a reduction in the repertoire of antigens recognized by T lymphocyte receptors (TCR) (Roved et al., 2018). Recent studies on bank voles *Myodes glareolus* confirmed that the reduction of TCR repertoire is associated with high individual variability within MHC class I genes, but no similar relationship has been demonstrated for MHC class II genes (Migalska et al., 2019). These mechanisms are directly associated with the hypothesis of optimality (Wegner et al., 2003), assuming that intermediate rather than maximum diversity of MHC alleles is optimal and associated with the highest fitness within populations.

So far, research on the association between individual MHC diversity and fitness components in birds have yielded mixed results. On the one hand, positive linear relationships between MHC diversity and reproduction or survival have been demonstrated in species such as the Egyptian vulture *Neophron percnopterus* (Agudo et al., 2012), Magellanic penguin *Spheniscus magellanicus* (Knafler et al., 2012), and common yellowthroat *Geothlypis trichas* (Dunn et al., 2013). Individual MHC diversity was also reported to be positively correlated with the expression of sexual ornaments (Hale et al., 2009; Whittingham et al., 2015), and negatively associated with the prevalence or intensity of infection by different types of parasites (Radwan et al., 2012; Slade et al. 2017). On the other hand, many studies demonstrated lack of significant relationships (e.g., Ekblom et al., 2004; Karlsson et al., 2015; Bateson et al., 2016), suggesting that the link between the MHC and fitness is not a common phenomenon in natural bird populations. However, most research has tested for linear relationships between MHC diversity and fitness-related traits, while the attempts to test the assumptions of the optimality hypothesis in birds have been fewer and mostly inconclusive or negative (e.g., Radwan et al., 2012; Sepil et al., 2013b; Biedrzycka et al., 2018). Limited support for MHC optimality in birds clearly contrasted with the results of similar studies in fish, especially three-spined stickleback *Gasterosteus aculeatus*, which provided one of the most convincing empirical evidence for optimal MHC diversity so far (Wegner et al., 2003; Kurtz et al., 2004; Kalbe et al., 2009).

The aim of this study was to test the associations between the diversity of MHC genes and fitness-related traits (condition, reproduction, and ornament expression) in the Eurasian coot *Fulica atra*, a common waterbird. For this purpose, we genotyped MHC class I and class II in over 100 individuals from a recently established (possibly bottlenecked) urban population in central Poland and, at the same time, collected information about almost 200 reproductive episodes of genotyped individuals. We tested for both positive linear and non-linear associations between MHC diversity and fitness-related traits, as predicted by the MHC diversity and optimality hypotheses, respectively. We also expected that the patterns of associations between the MHC and fitness may differ between

classes I and class II genes, as the costs of expressing a large number of MHC alleles may vary between the both classes (Migalska et al., 2019).

2. Materials and Methods

(a) Study site and general methodology

The fieldwork took place in an urban area of Łódź (51° 45’ N, 19° 27’ E), one of the largest cities in Poland (680 000 inhabitants; 293.25 km²). In 2010-2020 we monitored the entire coot population within the administrative borders of the city (ca. 30-60 breeding pairs). During this period we captured and sampled blood from 114 adult individuals. All birds were captured with noose traps made from monofilament nylon or by hand, while incubating on the nest or while feeding on the shore (at the pre-laying stage or in the chick rearing period). Each bird was marked with a metal ring (left tarsus) and a plastic neck collar with individual alphanumeric code that allowed easy identification of individuals in the field. At capture, we took basic measurements, including tarsus length and total head length, both measured with callipers (± 0.1 mm). Body mass was measured with an electronic balance (± 1 g). We also measured the size of the putative non-plumage ornament, the frontal shield (see below for details). Finally, we collected ca. 5 µl of blood for the measurement of the total blood haemoglobin concentration (as an indicator of physiological condition, see below), while 50 µl of blood was collected into 96% ethanol for genetic analyses: molecular sexing, MHC genotyping, and double digest restriction-site association DNA (ddRAD) sequencing. Genomic DNA was extracted from all blood samples using GeneJET Genomic DNA Purification Kit (Fermentas, Thermo Fisher Scientific, Waltham, MA, USA) and Bio-Trace DNA Purification Kit (EURx, Gdańsk, Poland), following manufacturers’ protocols. Molecular sexing followed methodology developed by Griffiths et al. (1998), according to the protocol described in Minias et al. (2015a). Sex ratio of our sample was roughly equal (58 females and 56 males).

(b) Phenotypic traits

In the analyses we used three traits that were expected to be associated with phenotypic quality of birds:

i) Body mass was used as a general measure of condition, as it may be a reliable indicator of energy reserves, when appropriately corrected for structural body size (Peig & Green, 2009). To estimate structural body size, we calculated the first principal component (PC1) from two size measurements, tarsus and total head length, and included it as a covariate in the modelling.

ii) Total blood haemoglobin concentration was used as a measure of physiological condition. Haemoglobin concentration is a key indicator of blood oxygen-carrying capacity and it reflects the potential of an organism to satisfy its oxygen demands. In birds, it has been reported to correlate with other measures of condition, diet quality, parasite prevalence and survival across a wide range of taxa (reviewed in Minias 2015b). To measure total blood haemoglobin concentration we used a portable HemoCue Hb 201+ photometer (HemoCue, Ängeholm, Sweden), which was previously shown to reliably measure this trait in avian blood (Velguth et al., 2010). Absorbance of blood (directly proportional to haemoglobin concentration) was measured using disposable HemoCue microcuvettes.

iii) Frontal shield size was used as a measure of a putative ornament expression. Eurasian coots are sexually monochromatic in plumage, but both sexes have a conspicuous and sexually dimorphic (larger in males; Minias, 2015a) white fleshy frontal shield that extends from the bill onto the head crown. This kind of morphological structure is typical for many rallid species and it increases in size prior to the breeding season in a testosterone-dependent manner (Eens et al., 2000). Frontal shield size has been reported to signal dominance, social status, fighting ability and condition in other rallids (Crowley & Magrath 2004; Alvarez et al., 2005; Dey et al., 2014), although similar information is lacking for the Eurasian coot (but see Minias et al., 2020 for the correlation between female frontal shield size and the quality of eggshell pigmentation). We used callipers to measure length and width of the frontal shield (± 0.1 mm) and calculated the first principal component (PC1) from these two measurements, which was used as an indicator of ornament size.

(c) Reproductive traits

During the study period we recorded 232 reproductive events (clutches) by 106 marked individuals. We recorded up to seven reproductive events per individual, but 67% of individuals had only ≤2 reproductive events recorded. Most of the events represented first clutches (82.8%), but there were also renest clutches after brood failure (12.9%) and second clutches (4.3%) included in the dataset. For each reproductive event, laying of the first egg (laying date) was assigned to the successive five-day periods beginning from 20 March (earliest laying date recorded within our study population). For most of the events, we also recorded clutch size (n = 171), hatching success (coded as hatched versus non-hatched clutches, n = 225) and breeding success (number of fully grown offspring recorded ca. 1.5 month from hatching; n = 204).

(d) MHC genotyping

We genotyped the MHC class I and class II gene fragments coding for the peptide-binding region of the molecule, which is directly involved in antigen binding. To genotype MHC class I we used two degenerate primers *MHCI-int2F* (5'- CATTTCCCTYGTGTTTCAGG-3') and *MHCI-ex4R* (3'-GGGTAGAAGCCGTGAGCRC-5') originally designed for accipitrid birds (Alcaide et al., 2009). These two primers bind to the flanking region of intron 2 and the conserved region of exon 4, and successfully amplify the entire MHC class I exon 3. For MHC class II genotyping we used two primers *Fuat-Ex2Fw* (5'-CTGACCRGCCTCCCTGCA-3') and *Fuat-Ex2Rv* (5'-TTGTGCCAYACACCCACC-3') originally designed for the Eurasian coot (Alcaide et al., 2014). The primers bind to the flanking region of intron 1 and the conserved region of exon 3 and successfully amplify the entire MHC class II exon 2 (270 bp). All PCR amplifications followed original protocols (Alcaide et al. 2009; 2014). Each PCR sample contained 1 µl of template DNA corresponding to approximately 20-80 ng of DNA, 10 µl of 2X HotStarTaq Plus MasterMix Kit (Qiagen, Venlo, The Netherlands), 8 µl of deionized water and 0.5 µl of each primer (total volume of 20 µl). In all PCRs we used fusion primers, with Illumina Nextera Transposase adapter sequences (Illumina Corp., San Diego, CA, USA), and 7-bp barcodes for sample identification. We used a relatively low number of amplification cycles (n = 25) to reduce the risk of chimera formation. To detect positive amplifications all PCR products were evaluated by electrophoresis on 2% agarose gel. All PCR products were purified using AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA) and their concentrations were assessed with Quant-iT PicoGreen dsDNA marking kit (Thermo FisherScientific, Waltham, MA, USA). The libraries were prepared separately for MHC class I and class II from equimolar quantities of PCR products using NEB-Next DNA Library Prep Master Mix

Set for Illumina (New England Biolabs, Ipswich, MA, USA) and sequenced on the 2 x 250bp Illumina MiSeq platform.

(e) Processing of Illumina data and MHC allele validation

Raw data from Illumina MiSeq sequencing were analysed and processed using the Amplicon Sequencing Analysis Tools (AmlpiSAT) webserver (Sebastian et al., 2016), following recommendations for Illumina data processing developed by Biedrzycka et al. (2017). First, paired-end sequences were merged using the Amplicon Sequencing MERGing tool (AmpliMERGE), which operates based on the FLASH algorithm with optimum overlapping parameters (Magoc & Salzberg, 2011). The stages of de-multiplexing, clustering, and filtering of merged reads were conducted in the Amplicon Sequencing Assignment (AmpliSAS) tool. At the clustering stage we used default parameters for Illumina data, including 1% substitution errors, 0.001% indel errors, and 25% minimum dominant frequency. During the filtering stage, chimeras and low frequency sequences (>3%) were discarded, as recommended for the general processing of Illumina data (Kircher et al., 2011). The processing was conducted for amplicons with the minimum depth of 300 reads and the maximum amplicon depth was set to 5000 reads because of AmpliSAS performance reasons (by default excluding additional reads from analyses). The average amplicon depth was 1706 ± 73 [SE] (prior to processing) and the average number of reads for validated sequences was 1398 ± 64 [SE] reads per sample. To determine the reproducibility of our genotyping and processing approaches, a total of 36 technical replicates (samples from the same individuals amplified in independent PCR reactions) were genotyped. The technical reproducibility of validated sequences was high (93.7 %). All unique MHC class I and class II sequences were aligned using Geneious v10.0.0.5 (Biomatters Ltd., Auckland, New Zealand). Intron regions were removed from the alignments and alleles were inferred based on the exon fragments only. We retrieved 1-10 MHC class I and 1-6 MHC class II alleles per individual, indicating that our primers targeted at least five and three MHC class I and class II loci, respectively. Although the total number of MHC genes is not known in our study species (as in most non-model avian taxa), it has been reported to be generally low in non-passerine birds (compared to passerines) and only exceptionally exceeds five genes per class (Minias et al., 2019). Individual MHC allelic diversity was calculated as the number of MHC class I or class II alleles recorded per individual and categorized into three groups of low (1-2 class I or class II alleles), intermediate (3-5 class I alleles or 3 class II alleles) and high (6-10 class I alleles or 4-6 class II alleles) diversity. This categorization allowed us to deal with non-normal distribution of allele numbers and to effectively test for non-linear associations of MHC diversity with fitness-related traits (according to the key assumptions of the optimality

hypothesis). We found the traditional approach to test the optimality hypothesis (adding a quadratic term of MHC diversity to the models) ineffective, as significant quadratic terms could indicate one-directional (positive or negative) non-linear associations and may not necessarily provide support for local trait maxima or minima under intermediate MHC diversity.

(f) MHC supertype clustering

As different MHC alleles may have similar peptide-binding properties and focusing on functional MHC allele clusters (supertypes) may be more biologically relevant (Biedrzycka et al., 2018), we have performed allele clustering for MHC class I and II sequences detected in our study population. For this purpose we used k-means clustering and discriminant function analysis of principal components (DAPC) implemented in *adegenet* package (Jombart, 2008) developed for the R statistical environment (R Foundation for Statistical Computing, Vienna, Austria). In the analyses we followed a protocol described by Biedrzycka et al. (2018) and details of allele clustering are presented in the Electronic Supplementary Material (ESM). The MHC supertype diversity was calculated as the number of supertypes detected within an individual. Since analyses of MHC allelic diversity provided little support for non-linear associations with fitness-related traits (see results for details), we avoided categorization of these data and tested for linear associations only.

(g) ddRAD sequencing

Allelic diversity at the MHC may not only be governed by evolutionary or population processes that specifically target these genes, but may also be affected by the mechanisms operating at the level of the entire genome (e.g. inbreeding). To test whether any possible associations of MHC diversity with fitness were actually driven by genome-wide diversity, we used ddRAD sequencing to estimate genome-wide heterozygosity of all captured individuals. The ddRAD libraries were prepared at the Texas A&M AgriLife Genomics Facility (USA), which followed a protocol recommended by Peterson et al. (2012). Briefly, we used PstI (5'-CTGCA/G-3') and MboI (5'-/GATC-3') restriction enzymes for DNA digestion and selected 300-600 bp fragments for Illumina HiSeq 2500v4 sequencing. Approximately 2.3 million reads per sample were generated and processed with Stacks 2.0 software (Catchen et al., 2013). The processing included the following steps: *i*) trimming and demultiplexing raw reads; *ii*) removing low quality reads (based onPhred33 quality score) and reads with uncalled sites; *iii*) calling

single nucleotide polymorphism (SNP) with de novo approach (a minimum coverage depth of 3 reads per stack and a maximum of 3 mismatches between any two alleles in the population and within individuals). The processing stage produced 868 297 loci with an effective sample coverage of 9x (SD = 1.8x, range: 5.0 – 14.7x), which were then exported into a variant call format (VCF) file. Genome-wide heterozygosity was calculated across 14 525 SNPs that were called in at least 80% of individuals (on average 12 831 ± 98 [SE] SNPs per sample) using VCFtools (Danecek et al., 2011). Observed heterozygosity values ranged from 0.168 to 0.223 (on average 0.203 ± 0.001 [SE]).

(h) Statistical analyses

Associations of MHC class I and class II diversity with fitness-related (phenotypic and reproductive) traits were analysed using the generalized linear mixed models (GLMMs), as implemented in the *glmmADMB* R package (Skaug et al., 2012). First, we conducted analyses at the level of MHC allelic diversity (number of alleles). All phenotypic (blood haemoglobin concentration, body mass, and frontal shield size) and reproductive (laying date, clutch size, hatching success, breeding success) traits were entered as response variables in separate models. Measures of MHC class I and class II allelic diversity were entered as three-level (low, intermediate, high) fixed factors in each model, while sex was entered as an additional fixed factor and genome-wide heterozygosity as a covariate. In the analyses of phenotypic traits we also entered capture date and body size (PC1) as covariates. Hour of measurement was entered as an additional covariate only in the analysis of blood haemoglobin concentration to control for diurnal variation in this trait. In the analyses of clutch size and hatching/breeding success we entered laying date (covariate) and brood status (fixed factor) to control for variation in reproductive performance between first, renest, and second clutches. The analysis of laying dates was conducted for first clutches only. We also tested for the effects of interactions between MHC diversity (class I and class II) and sex on each phenotypic and reproductive trait, to assess whether these associations are not sex-specific. All interactions were non-significant and removed from the models, except for the analysis of blood haemoglobin concentration (Tables S1 and S2 in the ESM). Thus, the final models for blood haemoglobin concentration were run separately for each sex. Second, we ran the analyses at the level of MHC supertype diversity (number of supertypes). The models had similar structure, except for the fact that MHC class I and class II supertype diversity were included as a covariates. We also tested for the effects of specific MHC supertypes on each fitness-related trait (not possible for alleles because of their lower frequencies), and these models were run separately for MHC class I and II. Because of numerous predictors (12 MHC class I and 14 MHC class II supertypes, see results for details), we reduced these models by

removing highly non-significant ($p > 0.15$) supertype effects and corrected p values for multiple comparisons (false discovery rate; Benjamini & Hochberg, 1995). Finally, we tested whether associations of MHC supertype diversity with fitness-related traits were not primarily driven by the occurrence of specific supertypes and for this purpose we combined both types of predictors (supertype diversity and presence of specific supertypes) within the same models. The effects of individual identity and year were entered as random factors in each GLMM to avoid pseudoreplication resulting from repeated measurements of the same individuals (e.g. across years) and to control for inter-annual variation in all the traits. Hatching success (binary trait) was analysed with a binomial distribution, while breeding success was analysed with a zero-inflated Poisson distribution. All other analyses were run with Gaussian distributions for the response variables. All values are reported as means ± SE.

3. Results

We found that MHC diversity was associated with physiological condition and ornament expression in the Eurasian coot, but in the opposite direction of our predictions. Birds with a high number of MHC class I and class II alleles were in poorer condition (had significantly lower blood haemoglobin concentrations) when compared with birds that had a low number of MHC alleles, although these associations were sex-specific (see Table S1 for significant MHC-sex interaction) and apparent either in males (MHC class I) or females (MHC class II) (Table 2; Fig. 1A,2A). Also, coots with high MHC class II diversity had smaller ornaments, the frontal shield (Table 3; Fig. 2B). No significant association was found between MHC diversity and body mass (after correction for structural size of individuals) (Table S3). Analysis of reproductive traits revealed significant relationships of MHC diversity with laying date and clutch size. Laying date was associated with MHC class I diversity, as individuals with intermediate and high number of MHC alleles started breeding significantly later (on average 5.47 ± 2.59 and 7.34 ± 3.18 days later, respectively) than individuals with low number of alleles (Table 4; Fig. 1B). Also, coots with high MHC class I diversity had smaller clutches than birds with low MHC diversity, even after controlling for variation in laying date (Table 4, Fig. 1C). In contrast, individuals with intermediate MHC class II diversity had clutch size significantly higher than individuals with low diversity (Table 4; Fig. 2C), whereas no significant differences in clutch size were recorded between birds that had low and high MHC class II diversity (Table 4). Associations of MHC diversity with both laying and clutch size were apparent in both sexes, as indicated by non-significant MHC-sex interactions (Table S2). In contrast to laying date and clutch size, we found no evidence for significant

associations between MHC diversity and hatching or breeding success (Table S4). Also, we found no statistical support for associations of genome-wide heterozygosity with either phenotypic traits (condition, ornament expression) or reproductive traits (laying date, clutch size, hatching and breeding success) of coots (Tables 2-4, S3, S4).

Allele clustering indicted the presence of 12 MHC class I and 14 MHC class II supertypes (18.2 ± 4.8 and 10.6 ± 1.1 alleles per supertype, respectively) in our study population. After controlling for the false discovery rate we found that nearly half of all MHC supertypes (46.2%) showed significant associations with different phenotypic or reproductive traits. In particular, we identified five MHC class I supertypes that were associated with reduced body mass and haemoglobin concentration, delayed laying and smaller clutches, while seven MHC class II supertypes were associated with reduced condition (body mass and blood haemoglobin) and lower breeding success (Tables S5-S18). None of the supertypes was associated with stronger expression of fitness-related traits (Tables S5-S18). An analysis of linear associations between the number of MHC supertypes and fitness-related traits indicated that high supertype diversity was linked with weaker expression of several phenotypic and reproductive traits. Specifically, the number of MHC class I supertypes was negatively associated with male haemoglobin concentration and positively associated with laying date (indicating delayed breeding), while the number of MHC class II supertypes showed negative associations with male body mass and female haemoglobin concentration (Tables S19, S20). Most of these associations remained significant after controlling for the effects of particular supertypes (Tables S21-S24).

4. Discussion

This study examined the relationship between the diversity of MHC genes and fitness-related traits (condition, ornament expression, reproduction) in a recently established urban population of the Eurasian coot. Contrary to expectation, we showed that high allelic diversity of MHC class I or class II genes was associated with poorer condition (lower blood haemoglobin concentrations), weaker expression of the putative ornament (smaller frontal shield), later onset of breeding, and smaller clutches. An analysis of functional allele clusters (supertypes) provided further support for negative associations between MHC diversity and fitness-related traits and we showed that most of these relationships could not be explained by the presence of specific maladaptive supertypes. Finally, we found little support for the optimality hypothesis, as only clutch size showed non-linear associations with MHC class II diversity (smaller clutches recorded in individuals with low than intermediate level

of MHC diversity). Overall, our results indicate that high MHC allelic diversity is not favourable for fitness in our study coot population.

So far, a positive relationship between individual MHC diversity and fitness has been empirically supported in several bird species, as high allelic diversity of the MHC was associated with lower pathogen or parasite prevalence (Radwan et al., 2012), increased survival (Westerdahl et al., 2005; Worley et al., 2010) or high reproductive success (Brouwer et al., 2010). On the other hand, many bird species showed no significant correlation between MHC diversity and infection rate (Loiseau et al., 2008, Sutton et al. 2016) or fitness (Ekblom et al., 2004; Radwan et al., 2012; Karlsson et al., 2015; Bateson et al., 2016). The relative scarcity of linear associations may be due to the inherent costs of MHC alleles expression, including an increased risk of autoimmune diseases or depleted repertoire of TCRs that recognize MHC-peptide complexes (Roved et al., 2018). In the three-spined stickleback *Gasterosteus aculeatus* individuals with a small and large diversity of MHC class II alleles had lower resistance to parasites such as *Schistocephalus solidus* tapeworms and microsporidia *Glugea anomala,* compared to individuals with an intermediate number of MHC alleles (consistent with the optimality hypothesis; Kurtz et al., 2004). This non-linear pattern may arise from the contradictory mechanisms of heterozygote advantage (a low ability to bind diverse antigens by individuals with low MHC diversity) and TCR repertoire depletion (a low ability to recognize MHC-peptide complexes by TCRs of individuals with high MHC diversity). Our results were not consistent with the optimality hypothesis, as we did not observe a considerable reduction in fitness-related traits of individuals with the lowest MHC diversity (except for the relationship between MHC class II diversity and clutch size). At the same time, we found virtually no support for the diversity hypothesis at the MHC genes in our study coot population, as individuals with high number of MHC alleles or supertypes were characterized by lower phenotypic quality or reproduction. Our analyses showed that this pattern was not primarily driven by the presence of specific resistance/susceptibility supertypes, which could be non-randomly associated with haplotypes carrying different numbers of MHC variants. Alternatively, it is possible that the negative associations of MHC diversity with fitness-related traits could reflect depauperate pathogen diversity in an urban landscape, as urbanization is known to reduce the abundance of many wildlife parasites (Bradley & Altizer, 2007). Under such scenario, large number of MHC variants expressed within individuals may not bring the usual benefits in terms of pathogen recognition, while incurring standard costs, e.g. in terms of autoimmunity or TCR depletion. Although we lacked any empirical data on pathogen and parasite communities of coots from our urban study site and from adjacent wildland, our results clearly show that high diversity of the MHC can under certain conditions be unfavourable and may not necessarily lead to increased fitness in natural animal populations.

The TCR depletion hypothesis was recently tested in the bank vole *Myodes glareolus* (Migalska et al., 2019) and it was shown that the higher diversity of MHC class I genes was associated with a reduced TCR repertoire, although similar relationship was not observed at the MHC class II genes (Migalska et al., 2019). As far as we are aware, the mechanism of TCR depletion has not been unequivocally demonstrated in birds so far, but the assumptions of the optimality hypothesis have been tested in several avian species, leading to mixed conclusions. For example, the mechanism of optimal MHC diversity has not been supported in female collared flycatchers *Ficedula albicollis*, as the prevalence of malaria infections showed a negative linear relationship with functional diversity of the MHC class II and there was no evidence for associations between MHC diversity and female survival or reproductive success (Radwan et al., 2012). Research conducted on the black-legged kittiwake *Rissa tridactyla* chicks showed a positive correlation between the diversity of MHC class II genes and chick growth rate and survival, as well as tick clearance in females but not in males (Pineaux et al., 2020). Despite the differences between the sexes that may have arisen from sex-specific intensity of parasite infections, the results were not consistent with the assumptions of the optimality hypothesis (Pineaux et al., 2020). Similarly, no support for non-linear relationships between MHC diversity and individual fitness or intensity of malaria infections was found in a population of the great tit *Parus major* (Sepil et al., 2013a). In contrast, empirical evidence for the optimality hypothesis in birds is scarce. A study on the Swedish population of the great reed warbler *Acrocephalus arundinaceus* showed a non-linear (consistent with the optimality hypothesis) relationship between the number of MHC alleles and the occurrence of infection with certain strains of malaria (Westerhald et al., 2005). Other studies on the great reed warbler also seem to partially support the optimality hypothesis, as there was a non-linear relationship between female reproductive success and the diversity of MHC genes, although no such relationship was found in males (Roved et al., 2018). Taking all this into account, any robust conclusions on the generality of linear and non-linear relationships between individual MHC diversity and fitness are currently difficult to draw without further studies in a broader phylogenetic context.

In our study, individual MHC diversity correlated with various characteristics that are expected be related to fitness (condition, ornament expression, reproductive phenology and clutch size), but we did not find any direct relationship between the number of MHC alleles and the key fitness component, i.e. the reproductive success. Phenotypic traits such as condition or expression of ornaments can be directly associated with individual immunocompetence (Koch et al., 2017), and therefore they may show relatively strong associations with the MHC. So far, some of the strongest evidence for associations between ornament expression and MHC diversity comes from studies on the common yellowthroat, in which MHC class II diversity correlated positively with the expression of

both carotenoid (yellow bib brightness) and melanin-based (black facial mask size) ornaments (Dunn et al. 2013; Whittingham et al., 2015). Initial phases of the reproductive cycle (e.g., the date of egg laying or the number of eggs laid) should also be more directly dependent on individual health status and condition (and thus indirectly on the abilities of the immune system) than reproductive success, which is more environmentally dependent (e.g., on the intensity of predation pressure or weather conditions) (e.g. Hasselquist et al., 2001; Descamps et al., 2011). Hence, the influence of MHC diversity on reproductive success may be much more effectively masked by the impact of external conditions not directly linked to the phenotypic characteristics of individuals or their genetic constitution. In the case of our urban coot population, this impact may be even greater than for populations nesting in natural habitats. Availability of traditional breeding habitat (patches of reed vegetation) is highly limited in urban landscape, thus the nests are often poorly sheltered and more exposed to urban predators (mainly corvids, but also domestic animals such as dogs) and unfavourable weather conditions. As a result, effect sizes of the association between MHC diversity and reproductive output are expected to be low and possibly difficult to detect under limited sample sizes. So far, the relationship between MHC and bird reproduction has mainly been reported for reproductive traits associated with the early stages of breeding cycle, e.g., egg mass (Hale et al., 2009), clutch size (Bonneaud et al., 2004), or hatching success (Knafler et al., 2012, Hoover et al., 2018), while correlations with fledging success were reported only sporadically (e.g., Agudo et al., 2012).

In conclusion, high individual diversity of the MHC genes is traditionally expected to increase fitness by broadening the spectrum of pathogens recognized by the immune system (e.g. due to the heterozygote advantage mechanism). That said, our results suggest that in some populations any advantages of high MHC diversity may be effectively masked by the costs of MHC allele expression (e.g. the depletion of TCR repertoire). The relative balance between these two contrasting mechanisms should produce a scenario where optimal (intermediate) MHC diversity is associated with the highest fitness. However, our results suggest that the costs of expressing high number of MHC alleles may, under certain conditions, outweigh the benefits which broad repertoire of MHC alleles should confer in terms of pathogen recognition, resulting in a negative relationship between MHC diversity and fitness-related traits. To the best of our knowledge, this is one of the first studies providing consistent evidence for a negative association between MHC diversity and a range of phenotypic and reproductive traits in a natural population.

Acknowledgements. We thank Radosław Włodarczyk for his help with the fieldwork and Jacek Radwan for his constructive comments on the manuscript.

Authors' contributions. EP and PM designed the study; EP and PM collected the data; EP conducted laboratory analyses; all authors analysed the data; EP wrote the manuscript; all authors revised the manuscript.

Data Availability Statement. Data are archived in the Dryad Digital Repository (TBA). All MHC sequences have been deposited in GenBank (accession nos: TBA).

Funding. The study was financially supported by the research grant of the National Science Centre in Poland (2020/38/E/NZ8/00143).

Ethics. All applicable institutional and/or national guidelines for the care and use of animals were followed during the study and all experiments were conducted by permission of the Local Bioethical Commission for Experiments on Animals.

Competing interests. We declare no competing interests.

References

Agudo, R., Carrete, M., Alcaide, M., Rico, C., Hiraldo, F., & Donazar, J. A. (2012). Genetic diversity at neutral and adaptive loci determines individual fitness in a long-lived territorial bird. *Proceedings of the Royal Society B*, *279*(1741), 3241-3249.

Alcaide, M., Munoz, J., Martínez-de la Puente, J., Soriguer, R., & Figuerola, J. (2014). Extraordinary MHC class II B diversity in a non-passerine, wild bird: the Eurasian Coot *Fulica atra* (Aves: Rallidae). *Ecology and Evolution*, *4*(6), 688-698.

Alcaide, M., Edwards, S. V., Cadahía, L., & Negro, J. J. (2009). MHC class I genes of birds of prey: isolation, polymorphism and diversifying selection. *Conservation Genetics*, *10*(5), 1349-1355.

Alvarez, F., Sánchez, C., & S. Angulo (2005). The frontal shield of the moorhen: sex differences and relationship with body condition. *Ethology, Ecology and Evolution*, *17*(2), 135-148.

Axtner, J., & Sommer, S. (2007). Gene duplication, allelic diversity, selection processes and adaptive value of MHC class II DRB genes of the bank vole, *Clethrionomys glareolus*. *Immunogenetics*, *59*(5), 417-426.

Bateson, Z. W., Hammerly, S. C., Johnson, J. A., Morrow, M. E., Whittingham, L. A., & Dunn, P. O. (2016). Specific alleles at immune genes, rather than genome-wide heterozygosity, are related to

immunity and survival in the critically endangered Attwater's prairie-chicken. *Molecular Ecology*, *25*(19), 4730-4744.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, 57 (1), 289–300.

Biedrzycka, A., Bielański, W., Ćmiel, A., Solarz, W., Zając, T., Migalska, M., Sebastian, A., Westerdahl, H., & Radwan, J. (2018). Blood parasites shape extreme major histocompatibility complex diversity in a migratory passerine. *Molecular Ecology*, *27*(11), 2594-2603.

Bonneaud, C., Mazuc, J., Chastel, O., Westerdahl, H., & Sorci, G. (2004). Terminal investment induced by immune challenge and fitness traits associated with major histocompatibility complex in the house sparrow. *Evolution*, *58*(12), 2823-2830.

Bradley, C. A., & Altizer, S. (2007). Urbanization and the ecology of wildlife diseases. *Trends in Ecology and Evolution*, *22*(2), 95-102.

Brouwer, L., Barr, I., Van De Pol, M., Burke, T., Komdeur, J. A. N., & Richardson, D. S. (2010). MHC-dependent survival in a wild population: evidence for hidden genetic benefits gained through extra-pair fertilizations. *Molecular Ecology*, *19*(16), 3444-3455.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140.

Crowley, C. E., & R. D. Magrath (2004). Shields of offence: signalling competitive ability in the dusky moorhen, *Gallinula tenebrosa*. *Australian Journal of Zoology, 52*(5), 463-474.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.

Descamps, S., Bêty, J., Love, O. P., & Gilchrist, H. G. (2011). Individual optimization of reproduction in a long-lived migratory bird: a test of the condition-dependent model of laying date and clutch size. *Functional Ecology*, *25*(3), 671-681.

Dey, C. J. , Dale, J. & Quinn, J. S. (2014). Manipulating the appearance of a badge of status causes changes in true badge expression. *Proceedings of the Royal Society B*, *281*(1775), 20132680.

Doherty, P. C., & Zinkernagel, R. M. (1975). Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*, *256*(5512), 50-52.

Dunn P.O., Bollmer J.L., Freeman-Gallant C.R. & Whittingham L.A. 2013. MHC variation is related to a sexually selected ornament, survival and parasite resistance in the common yellowthroat. *Evolution*, *67*(3), 679-687.

Eens, M., Van Duyse,E., Berghman, L., & Pinxten, R. (2000). Shield characteristics are testosterone-dependent in both male and female moorhens. *Hormones and Behaviour*, *37*(2), 126-134.

Ejsmond, J.M., Radwan, J., & Wilson, A. B. (2014). Sexual selection and the evolutionary dynamics of the major histocompatibility complex. *Proceedings of the Royal Society B*, *281*(1796), 20141662.

Ekblom, R., Saether, S. A., Grahn, M., Fiske, P., Kålås, J. A., & Höglund, J. (2004). Major histocompatibility complex variation and mate choice in a lekking bird, the great snipe (*Gallinago media*). *Molecular Ecology*, *13*(12), 3821-3828.

Geraghty, D. E., Daza, R., Williams, L. M., Vu, Q., & Ishitani, A. (2002). Genetics of the immune response: identifying immune variation within the MHC and throughout the genome. *Immunological Reviews*, *190*(1), 69-85.

Gillingham, M. A., Béchet, A., Courtiol, A., Rendón-Martos, M., Amat, J. A., Samraoui, B., Onmus O., Sommer, S., & Cézilly, F. (2017). Very high MHC Class IIB diversity without spatial differentiation in the mediterranean population of greater Flamingos. *BMC Evolutionary Biology*, 17(1), 1-10.

Grieves L. A., Gloor G. B., Bernards M. A., & MacDougall-Shackleton E. A. 2019. Songbirds show odour-based discrimination of similarity and diversity at the major histocompatibility complex. *Animal Behaviour*, *158*, 131-138.

Griffiths, R., Double, M. C., Orr, K., & Dawson, R. J. (1998). A DNA test to sex most birds. *Molecular Ecology*, *7*(8), 1071-1075.

Hale, M. L., Verduijn, M. H., Møller, A. P., Wolff, K., & Petrie, M. (2009). Is the peacock's train an honest signal of genetic quality at the major histocompatibility complex? *Journal of Evolutionary Biology*, *22*(6), 1284-1294.

Hasselquist, D., Wasson, M. F., & Winkler, D. W. (2001). Humoral immunocompetence correlates with date of egg-laying and reflects work load in female tree swallows. *Behavioral Ecology*, *12*(1), 93-97.

Hoover, B., Alcaide, M., Jennings, S., Sin, S. Y. W., Edwards, S. V., & Nevitt, G. A. (2018). Ecology can inform genetics: Disassortative mating contributes to MHC polymorphism in Leach's storm-petrels (*Oceanodroma leucorhoa*). *Molecular Ecology*, *27*(16), 3371-3385.

Janeway Jr, C. A., Travers, P., Walport, M., & Shlomchik, M. J. (2001). The major histocompatibility complex and its functions. In *Immunobiology: The Immune System in Health and Disease.* 5th edition. Garland Science.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics, 24(11), 1403-1405.

Kalbe, M., Eizaguirre, C., Dankert, I., Reusch, T. B., Sommerfeld, R. D., Wegner, K. M., & Milinski, M. (2009). Lifetime reproductive success is maximized with optimal major histocompatibility complex diversity. *Proceedings of the Royal Society B*, *276*(1658), 925-934.

Karlsson, M., Schroeder, J., Nakagawa, S., Smith, H. G., Burke, T., & Westerdahl, H. (2015). House sparrow *Passer domesticus* survival is not associated with MHC-I diversity, but possibly with specific MHC-I alleles. *Journal of Avian Biology*, *46*(2), 167-174.

Kircher, M., Heyn, P., & Kelso, J. (2011). Addressing challenges in the production and analysis of Illumina sequencing data. *BMC Genomics*, *12*(1), 382.

Knafler, G. J., Clark, J. A., Boersma, P. D., & Bouzat, J. L. (2012). MHC diversity and mate choice in the magellanic penguin, *Spheniscus magellanicus*. *Journal of Heredity*, *103*(6), 759-768.

Koch, R. E., Josefson, C. C., & Hill, G. E. (2017). Mitochondrial function, ornamentation, and immunocompetence. *Biological Reviews*, *92*(3), 1459-1474.

Kurtz, J., Kalbe, M., Aeschlimann, P. B., Häberli, M. A., Wegner, K. M., Reusch, T. B., & Milinski, M. (2004). Major histocompatibility complex diversity influences parasite resistance and innate immunity in sticklebacks. *Proceedings of the Royal Society of London. Series B*, *271*(1535), 197-204.

Loiseau, C., Zoorob, R., Garnier, S., Birard, J., Federici, P., Julliard, R., & Sorci, G. (2008). Antagonistic effects of a Mhc class I allele on malaria-infected house sparrows. *Ecology Letters*, *11*(3), 258-265.

McClelland, E. E., Penn, D. J., & Potts, W. K. (2003). Major histocompatibility complex heterozygote superiority during coinfection. *Infection and immunity*, *71*(4), 2079-2086.

Migalska, M., Sebastian, A., & Radwan, J. (2019). Major histocompatibility complex class I diversity limits the repertoire of T cell receptors. *Proceedings of the National Academy of Sciences USA*, *116*(11), 5021-5026.

Minias, P., Gómez, J., & Włodarczyk, R. (2020). Egg spottiness reflects female condition, physiological stress, and ornament expression in a common rallid species. *Auk*, *137*(4), ukaa054.

Minias, P., Pikus, E., Whittingham, L. A., & Dunn, P. O. (2019). Evolution of copy number at the MHC varies across the avian tree of life. *Genome Biology and Evolution*, *11*(1), 17-28.

Minias, P. (2015a). Sex determination of adult Eurasian coots (*Fulica atra*) by morphometric measurements. *Waterbirds*, *38*(2), 191-194.

Minias, P. (2015b). The use of haemoglobin concentrations to assess physiological condition in birds: a review. *Conservation Physiology*, *3*(1), cov007.

Peig, J., & Green, A. J. (2009). New perspectives for estimating body condition from mass/length data: the scaled mass index as an alternative method. *Oikos*, *118*(12), 1883-1891.

Pineaux, M., Merkling, T., Danchin, E., Hatch, S., Duneau, D., Blanchard, P., & Leclaire, S. (2020). Sex and hatching order modulate the association between MHC-II diversity and fitness in early-life stages of a wild seabird. *Molecular Ecology*, *29*, 3316-3329.

Radwan, J., Zagalska-Neubauer, M., Cichoń, M., Sendecka, J., Kulma, K., Gustafsson, L., & Babik, W. (2012). MHC diversity, malaria and lifetime reproductive success in collared flycatchers. *Molecular Ecology*, *21*(10), 2469-2479.

Radwan, J., Biedrzycka, A., & Babik, W. (2010). Does reduced MHC diversity decrease viability of vertebrate populations? *Biological Conservation*, *143*(3), 537-544.

Roved, J., Hansson, B., Tarka, M., Hasselquist, D., & Westerdahl, H. (2018). Evidence for sexual conflict over major histocompatibility complex diversity in a wild songbird. *Proceedings of the Royal Society B*, *285*(1884), 20180841.

Sepil, I., Lachish, S., Hinks, A. E., & Sheldon, B. C. (2013a). Mhc supertypes confer both qualitative and quantitative resistance to avian malaria infections in a wild bird population. *Proceedings of the Royal Society B*, *280*(1759), 20130134.

Sepil, I., Lachish, S., & Sheldon, B. C. (2013b). MHC-linked survival and lifetime reproductive success in a wild population of great tits. *Molecular Ecology*, *22*(2), 384-396.

Skaug, H., Fournier, D., Nielsen, A., Magnusson, A., & Bolker, B. (2012). Generalized linear mixed models using AD Model Builder. R package version c0.7.2.12. https://github.com/bbolker/glmmadmb

Slade, J. W., Sarquis-Adamson, Y., Gloor, G. B., Lachance, M. A., & MacDougall-Shackleton, E. A. (2017). Population differences at MHC do not explain enhanced resistance of song sparrows to local parasites. *Journal of Heredity*, *108*(2), 127-134.

Spurgin, L. G., & Richardson, D. S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B*, *277*(1684), 979-988.

Stervander, M., Dierickx, E. G., Thorley, J., Brooke, M. D. L., & Westerdahl, H. (2020). High MHC gene copy number maintains diversity despite homozygosity in a critically endangered single-island endemic bird, but no evidence of MHC-based mate choice. *Molecular Ecology*, *29*(19), 3578-3592.

Sutton, J. T., Castro, I., Robertson, B. C., Tompkins, D. M., Stanton, J. A. L., & Jamieson, I. G. (2016). MHC genetic diversity and avian malaria prevalence in Mokoia Island saddlebacks. *New Zealand Journal of Ecology*, *40*(3), 351-360.

Velguth, K. E., Payton, M. E., & Hoover, J. P. (2010). Relationship of hemoglobin concentration to packed cell volume in avian blood samples. *Journal of Avian Medicine and Surgery*, *24*(2), 115-121.

Wegner, K. M., Kalbe, M., Kurtz, J., Reusch, T. B., & Milinski, M. (2003). Parasite selection for immunogenetic optimality. *Science*, *301*(5638), 1343-1343.

Westerdahl, H., Waldenström, J., Hansson, B., Hasselquist, D., von Schantz, T., & Bensch, S. (2005). Associations between malaria and MHC genes in a migratory songbird. *Proceedings of the Royal Society B*, *272*(1571), 1511-1518.

Whittingham, L. A., Freeman-Gallant, C. R., Taff, C. C., & Dunn, P. O. (2015). Different ornaments signal male health and MHC variation in two populations of a warbler. *Molecular Ecology*, *24*(7), 1584-1595.

Worley, K., Collet, J., Spurgin, L. G., Cornwallis, C., Pizzari, T., & Richardson, D. S. (2010). MHC heterozygosity and survival in red junglefowl. *Molecular Ecology*, *19*(15), 3064-3075.

Table 1. Characteristics of fitness-related (phenotypic and reproductive) traits in the study population of the Eurasian coot.

Trait category	Trait	Sample size (no. measurements / no. individuals)	Mean	SE	Range (min – max)
Phenotypic	Blood haemoglobin concentration (g/l)	129 / 108	150.0	1.6	104 - 201
	Body mass (g)	129 / 108	767.7	9.4	559 - 1017
	Frontal shield size (mm)	117 / 98	20.0	0.3	15.2 – 31.4
Reproductive	Laying date (day of year)	191 / 105	110.8	1.2	80 - 170
	Clutch size (n eggs)	171 / 99	7.95	0.13	4 - 14
	Hatching success	225 / 103	0.77	0.03	0 - 1
	Breeding success	204 / 96	3.50	0.18	0 - 9

Table 2. Associations between haemoglobin concentration and the number of MHC class I and class II alleles in male and female Eurasian coots. Both models included haemoglobin concentration as the response and the number of both class I and II alleles (included separately, each coded with three levels), genome-wide heterozygosity, body size, capture date, and time of sampling (hour) as predictors. Bird identity and year were included as random factors in each model. Significant predictors are marked in bold.

Trait	Predictors	Estimate ± SE	z value	P
Haemoglobin concentration (males)	**Intercept**	**128.66 ± 42.28**	**3.04**	**0.002**
	MHC class I alleles (intermediate vs. low)	-7.37 ± 4.50	1.64	0.10
	MHC class I alleles (high vs. low)	**-18.57 ± 4.85**	**3.83**	**<0.001**
	MHC class II alleles (intermediate vs. low)	2.40 ± 4.91	0.49	0.62
	MHC class II alleles (high vs. low)	-11.77 ± 7.73	1.52	0.13
	Genome-wide heterozygosity	274.74 ± 180.61	1.52	0.13
	Body size	1.12 ± 2.31	0.48	0.63
	Capture date	**-0.20 ± 0.05**	**3.92**	**<0.001**
	Hour	0.38 ± 0.86	0.45	0.65
Haemoglobin concentration (females)	**Intercept**	**152.01 ± 44.34**	**3.43**	**0.001**
	MHC class I alleles (intermediate vs. low)	7.23 ± 5.27	1.37	0.17
	MHC class I alleles (high vs. low)	3.70 ± 6.45	0.57	0.57
	MHC class II alleles (intermediate vs. low)	**-10.67 ± 5.01**	**2.13**	**0.033**
	MHC class II alleles (high vs. low)	**-17.41 ± 7.30**	**2.39**	**0.017**
	Genome-wide heterozygosity	177.36 ± 208.72	0.85	0.40
	Body size	-5.04 ± 3.26	1.55	0.12
	Capture date	**-0.31 ± 0.08**	**3.80**	**<0.001**
	Hour	-0.39 ± 1.01	0.39	0.70

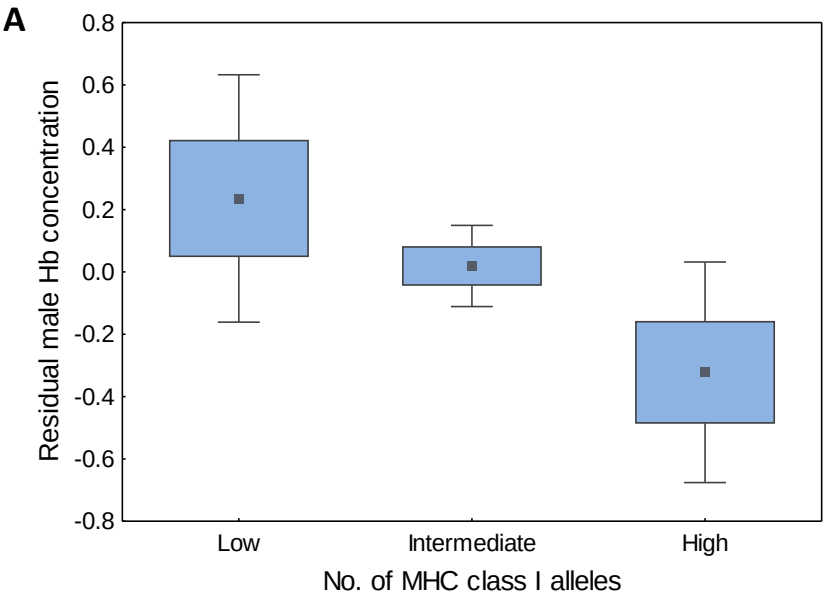
Table 3. Associations between frontal shield size and the number of MHC class I and class II alleles in the Eurasian coot. The model frontal shield size as the response and the number of both class I and II alleles (included separately, each coded with three levels), genome-wide heterozygosity, sex, body size, and capture date. Bird identity and year were included as random factors in each model. Significant predictors are marked in bold.

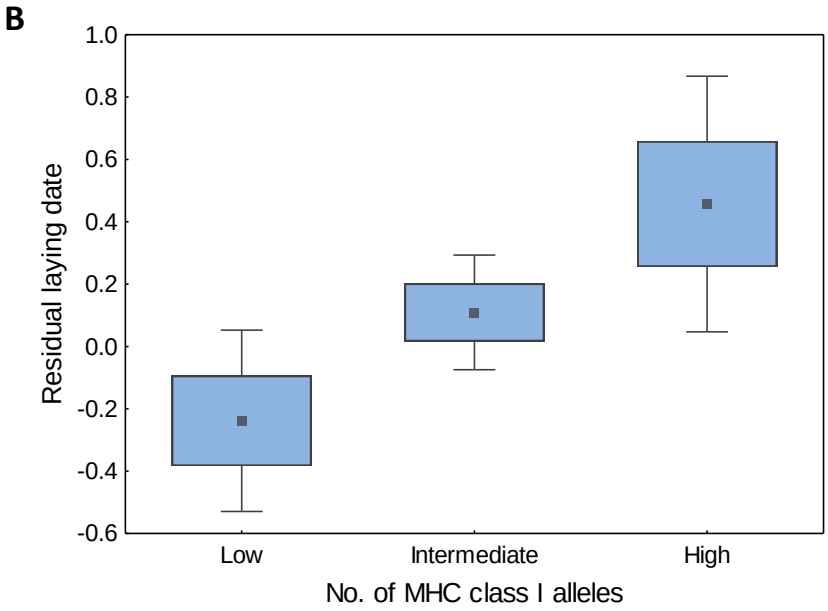
Trait	Predictors	Estimate ± SE	z value	P
Frontal shield size	Intercept	3.15 ± 2.00	1.58	0.11
	MHC class I alleles (intermediate vs. low)	-0.16 ± 0.23	0.69	0.49
	MHC class I alleles (high vs. low)	-0.01 ± 0.28	0.02	0.99
	MHC class II alleles (intermediate vs. low)	-0.24 ± 0.23	1.06	0.29
	MHC class II alleles (high vs. low)	**-0.86 ± 0.38**	**2.25**	**0.024**
	Genome-wide heterozygosity	-7.53 ± 9.61	0.78	0.43
	Sex (male vs. female)	**1.17 ± 0.35**	**3.34**	**0.001**
	Body size	0.20 ± 0.13	1.51	0.13
	Capture date	**-0.014 ± 0.003**	**4.36**	**<0.001**

Table 4. Associations between two reproductive traits (laying date and clutch size) and the number of MHC class I and class II alleles in the Eurasian coot. The first model included laying date as the response and the number of both class I and II alleles (included separately, each coded with three levels), sex and genome-wide heterozygosity as predictors. The second model included clutch size as the response and the same predictors with brood status (three levels) and laying date added. Bird identity and year were included as random factors in each model. Significant predictors are marked in bold.

Trait	Predictors	Estimate ± SE	z value	P
Laying date	**Intercept**	**118.29 ± 22.71**	**5.21**	**<0.001**
	MHC class I alleles (intermediate vs. low)	**5.51 ± 2.59**	**2.13**	**0.033**
	MHC class I alleles (high vs. low)	**7.39 ± 3.18**	**2.33**	**0.020**
	MHC class II alleles (intermediate vs. low)	1.31 ± 2.63	0.50	0.62
	MHC class II alleles (high vs. low)	-3.09 ± 4.57	0.67	0.50
	Genome-wide heterozygosity	-61.75 ± 114.60	0.54	0.59
	Sex (male vs. female)	0.54 ± 2.39	0.23	0.82
Clutch size	**Intercept**	**13.15 ± 2.51**	**5.23**	**<0.001**
	MHC class I alleles (intermediate vs. low)	0.16 ± 0.25	0.65	0.52
	MHC class I alleles (high vs. low)	**-0.88 ± 0.31**	**2.81**	**0.005**
	MHC class II alleles (intermediate vs. low)	**0.60 ± 0.26**	**2.34**	**0.019**
	MHC class II alleles (high vs. low)	0.33 ± 0.43	0.76	0.45
	Genome-wide heterozygosity	-2.46 ± 10.62	0.23	0.82
	Sex (male vs. female)	-0.03 ± 0.23	0.13	0.90
	Brood status (first vs. second)	0.02 ± 1.01	0.02	0.98
	Brood status (renest vs. second)	0.58 ± 1.02	0.57	0.57
	Laying date	**-0.04 ± 0.01**	**6.22**	**<0.001**

Figure 1. Associations of MHC class I allele numbers with male blood haemoglobin (Hb) concentration (A), laying date (B), and clutch size (C) in the Eurasian coot. Data presented as residuals from general linear mixed models (Tables 1 and 2). Central point – mean, box – SE, whiskers – 95% confidence intervals.





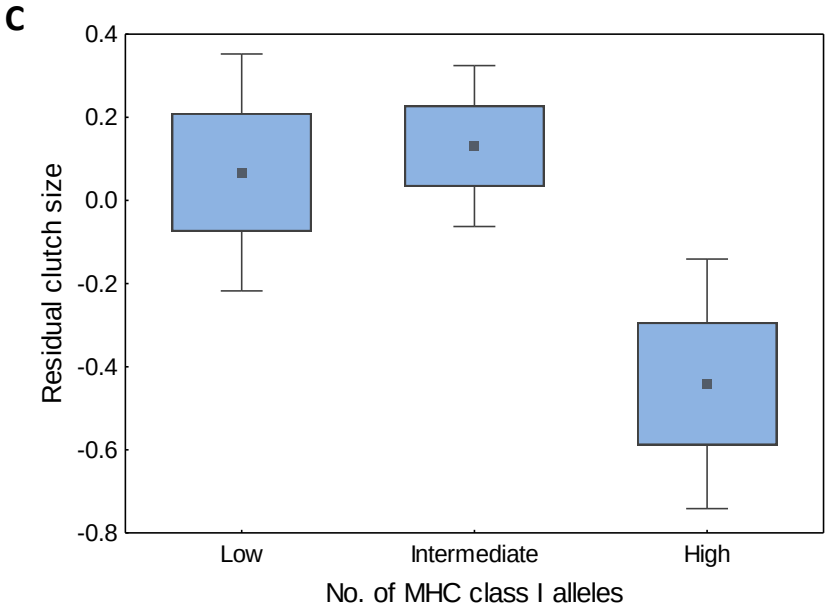
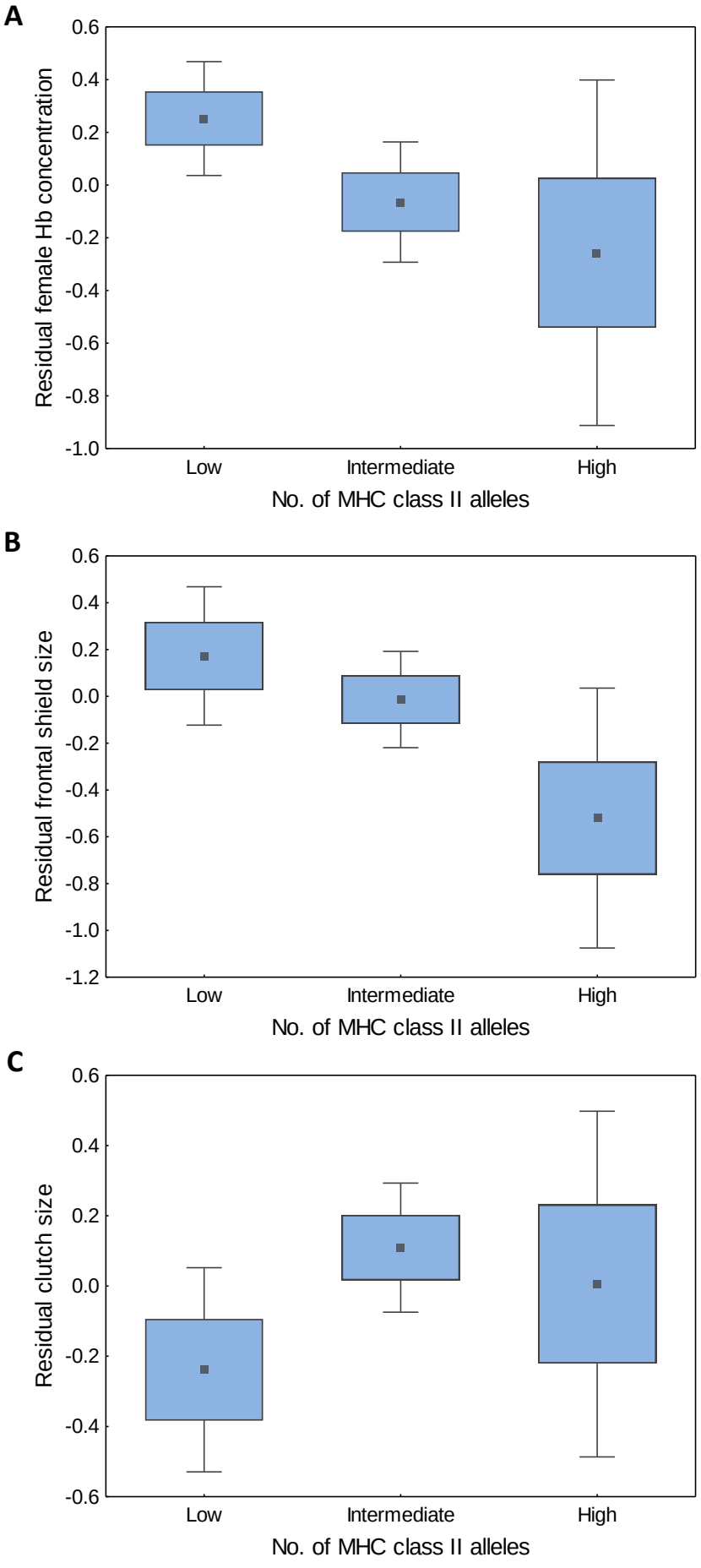


Figure 2. Associations of MHC class II allele numbers with female blood haemoglobin (Hb) concentration (A), frontal shield size (B), and clutch size (C) in the Eurasian coot. Data presented as residuals from general linear mixed models (Tables 1 and 2). Central point – mean, box – SE, whiskers – 95% confidence intervals.



Electronic Supplementary Material

High MHC diversity does not confer fitness advantage in a wild bird

Ewa Pikus[1], Peter O. Dunn[2], Piotr Minias[1]

[1] Department of Biodiversity Studies and Bioeducation, Faculty of Biology and Environmental Protection, University of Łódź, Banacha 1/3, 90-237 Łódź, Poland

[2] Behavioral and Molecular Ecology Group, Department of Biological Sciences, University of Wisconsin-Milwaukee

Author for correspondence

Piotr Minias

Email: pminias@op.pl

Supplementary methods

Supertype clustering

Supertype inference was performed separately for MHC class I and class II alleles using k-means clustering and discriminant function analysis of principal components (DAPC) implemented in *adegenet* R package (Jombart, 2008), following methodology by Biedrzycka et al. (2018). The analysis was limited to the positively selected sites (as identified for the Eurasian coot by Pikus and Minias 2021), which are likely corresponding to the peptide-binding region of the MHC molecules. Since the number of sites under pervasive positive selection was much lower at class I (n = 8) than class II (n = 18), we used sites under both episodic and pervasive selection for MHC class I (n = 15) (Pikus and Minias, 2021). To infer supertypes, we first identified the most probable number of clusters based on Bayesian information criterion (BIC), choosing the minimal number of clusters that preceded any increase in BIC values. This approach indicated the presence of 12 clusters for MHC class I and 14 clusters for MHC class II (Fig. S1). The number of principal components (PCs) in DAPC analysis was selected based on the maximum α-score (identified with the *optim.a.score* function), resulting in seven PCs in each analysis (Fig. S2). We used all available discriminant functions (n = 7) and re-run the entire procedure ten times to assess repeatability of supertype clustering. Repeatability was high for both MHC class I ($R = 0.813$, $P < 0.001$) and MHC class II ($R = 0.827$, $P < 0.001$), as assessed with kappa value (Fleiss, 1971) implemented in the *irr* R package (Gamer et al., 2012). The mean assignment probabilities (across all runs) were high for MHC class I (0.992 ± 0.001 [SE]) and MHC class II (0.983 ± 0.002 [SE]).

References

Biedrzycka, A., Bielański, W., Ćmiel, A., Solarz, W., Zając, T., Migalska, M., Sebastian, A., Westerdahl, H., Radwan, J. (2018). Blood parasites shape extreme major histocompatibility complex diversity in a migratory passerine. *Molecular Ecology*, 27(11), 2594-2603.

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2012). irr: Various coefficients for interrater reliability and agreement. R package ver. 0.84. Available at: https://CRAN.R-project.org/package=irr

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403-1405.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378-382.

Pikus, E., Minias, P. (2021) Using de novo genome assembly and high-throughput sequencing to characterize the Major Histocompatibility Complex in a non-model rallid bird, the Eurasian coot *Fulica atra*. *PLoS ONE* (submitted).

Table S1. Models assessing the effects of interactions between the number of MHC alleles (class I and class II) and sex on three phenotypic traits (body mass, haemoglobin concentration and frontal shield size) in the Eurasian coot. Bird identity and year were included as random factors in each model. Wald χ^2 statistics were used to assess statistical significance and significant predictors are marked in bold.

Trait	Predictors	W	P
Body mass	**Intercept**	**295.12**	**<0.001**
	MHC class I alleles	2.80	0.25
	MHC class II alleles	0.79	0.67
	Genome-wide heterozygosity	3.33	0.068
	Sex	**9.13**	**0.003**
	Body size	**18.90**	**<0.001**
	Capture date	2.93	0.087
	MHC class I alleles * Sex	3.78	0.15
	MHC class II alleles * Sex	0.96	0.62
Haemoglobin concentration	Intercept	**24.37**	**<0.001**
	MHC class I alleles	2.94	0.23
	MHC class II alleles	**9.29**	**0.010**
	Genome-wide heterozygosity	1.73	0.19
	Sex	**5.61**	**0.018**
	Body size	0.84	0.36
	Capture date	**32.73**	**<0.001**
	Hour	0.10	0.76
	MHC class I alleles * Sex	**12.92**	**0.002**
	MHC class II alleles * Sex	2.74	0.25
Frontal shield size	Intercept	2.71	0.10
	MHC class I alleles	1.52	0.47
	MHC class II alleles	4.96	0.084
	Genome-wide heterozygosity	0.54	0.46
	Sex	3.73	0.053
	Body size	1.55	0.21
	Capture date	**22.80**	**<0.001**
	MHC class I alleles * Sex	2.35	0.31
	MHC class II alleles * Sex	2.25	0.33

Table S2. Models assessing the effects of interactions between the number of MHC alleles (class I and class II) and sex on four reproductive traits (laying date, clutch size, hatching and breeding success) in the Eurasian coot. Bird identity and year were included as random factors in each model. Wald χ^2 statistics were used to assess statistical significance and significant predictors are marked in bold.

Trait	Predictors	W	P
Laying date	**Intercept**	**23.27**	**<0.001**
	MHC class I alleles	**1.94**	**0.38**
	MHC class II alleles	**3.46**	**0.18**
	Genome-wide heterozygosity	0.21	0.64
	Sex	0.37	0.54
	MHC class I alleles * Sex	0.93	0.63
	MHC class II alleles * Sex	3.11	0.21
Clutch size	**Intercept**	**25.79**	**<0.001**
	MHC class I alleles	3.96	0.14
	MHC class II alleles	1.29	0.52
	Genome-wide heterozygosity	0.07	0.79
	Sex	0.01	0.91
	Brood status	2.31	0.31
	Laying date	**35.67**	**<0.001**
	MHC class I alleles * Sex	0.36	0.83
	MHC class II alleles * Sex	0.45	0.80
Hatching success	Intercept	1.02	0.31
	MHC class I alleles	0.09	0.96
	MHC class II alleles	0.75	0.69
	Genome-wide heterozygosity	0.63	0.43
	Sex	0.50	0.48
	Brood status	1.69	0.43
	Laying date	0.36	0.55
	MHC class I alleles * Sex	0.64	0.72
	MHC class II alleles * Sex	0.09	0.96
Breeding success	Intercept	0.35	0.55
	MHC class I alleles	2.06	0.36
	MHC class II alleles	0.03	0.99
	Genome-wide heterozygosity	2.49	0.11
	Sex	1.02	0.31
	Brood status	0.98	0.61
	Laying date	**7.72**	**0.005**
	MHC class I alleles * Sex	0.90	0.64
	MHC class II alleles * Sex	0.56	0.76

Table S3. Associations between body mass and the number of MHC class I and class II alleles in the Eurasian coot. The model included body mass as the response and the number of both class I and II alleles (included separately, each coded with three levels), genome-wide heterozygosity, sex, body size, and capture date as predictors. Bird identity and year were included as random factors. Significant predictors are marked in bold.

Predictors	Estimate ± SE	z value	P
Intercept	**30.47 ± 1.83**	**16.64**	**<0.001**
MHC class I alleles (intermediate vs. low)	-0.04 ± 0.21	0.18	0.86
MHC class I alleles (high vs. low)	-0.07 ± 0.26	0.28	0.78
MHC class II alleles (intermediate vs. low)	-0.14 ± 0.22	0.63	0.53
MHC class II alleles (high vs. low)	-0.10 ± 0.33	0.32	0.75
Genome-wide heterozygosity	-14.54 ± 8.91	1.63	0.10
Sex (male vs. female)	**2.03 ± 0.33**	**6.07**	**<0.001**
Body size	**0.55 ± 0.13**	**4.22**	**<0.001**
Capture date	-0.006 ± 0.003	1.76	0.078

Table S4. Associations between two reproductive traits (hatching and breeding success) and the number of MHC class I and class II alleles in the Eurasian coot. The first model included hatching success as the response and the number of both class I and II alleles (included separately, each coded with three levels), genome-wide heterozygosity, sex, brood status (three levels), and laying date as predictors. The second model included breeding success as the response and the same predictors. Bird identity and year were included as random factors in each model. Significant predictors are marked in bold.

Trait	Predictors	Estimate ± SE	z value	P
Hatching success	Intercept	4.61 ± 4.17	1.11	0.27
	MHC class I alleles (intermediate vs. low)	-0.19 ± 0.39	0.47	0.64
	MHC class I alleles (high vs. low)	-0.14 ± 0.46	0.29	0.77
	MHC class II alleles (intermediate vs. low)	-0.52 ± 0.41	1.26	0.21
	MHC class II alleles (high vs. low)	0.62 ± 0.84	0.74	0.46
	Genome-wide heterozygosity	-15.69 ± 18.36	0.85	0.39
	Sex (male vs. female)	0.16 ± 0.36	0.44	0.66
	Brood status (first vs. second)	-0.67 ± 1.19	0.57	0.57
	Brood status (renest vs. second)	0.08 ± 1.22	0.06	0.95
	Laying date	0.006 ± 0.010	0.61	0.55
Breeding success	Intercept	0.93 ± 1.20	0.77	0.44
	MHC class I alleles (intermediate vs. low)	0.03 ± 0.12	0.28	0.78
	MHC class I alleles (high vs. low)	0.22 ± 0.14	1.55	0.12
	MHC class II alleles (intermediate vs. low)	-0.10 ± 0.13	0.82	0.41
	MHC class II alleles (high vs. low)	-0.06 ± 0.18	0.35	0.73
	Genome-wide heterozygosity	7.37 ± 4.98	1.48	0.14
	Sex (male vs. female)	0.05 ± 0.11	0.46	0.65
	Brood status (first vs. second)	0.11 ± 0.34	0.32	0.75
	Brood status (renest vs. second)	0.28 ± 0.32	0.88	0.38
	Laying date	**-0.011 ± 0.004**	**2.73**	**0.006**

Table S5. Associations between body mass and presence of specific MHC class I supertypes in the Eurasian coot. Bird identity and year were included as random factors in each model. The results of full and reduced model are shown. Significant predictors are marked in bold (at $P < 0.05$) or with asterisks (after correction for the false discovery rate).

Predictors	Estimate ± SE	z value	P
Full model			
Intercept	**31.47 ± 1.93**	**16.27**	**<0.001***
Supertype MHC1_1	0.15 ± 0.22	0.67	0.51
Supertype MHC1_2	0.12 ± 0.28	0.44	0.66
Supertype MHC1_3	**-0.79 ± 0.26**	**-3.03**	**<0.001***
Supertype MHC1_4	0.27 ± 0.24	1.13	0.26
Supertype MHC1_5	0.28 ± 0.32	0.86	0.39
Supertype MHC1_6	0.06 ± 0.33	0.18	0.86
Supertype MHC1_7	-0.39 ± 0.27	-1.47	0.14
Supertype MHC1_8	-0.26 ± 0.25	-1.04	0.30
Supertype MHC1_9	0.18 ± 0.19	0.95	0.34
Supertype MHC1_10	0.22 ± 0.29	0.76	0.45
Supertype MHC1_11	-0.02 ± 0.34	-0.05	0.96
Supertype MHC1_12	0.10 ± 0.28	0.37	0.71
Genome-wide heterozygosity	**-19.35 ± 9.29**	**-2.08**	**0.037**
Sex (male vs. female)	**2.07 ± 0.36**	**5.82**	**<0.001***
Body size	**0.50 ± 0.14**	**3.66**	**<0.001***
Capture date	**-0.008 ± 0.003**	**-2.39**	**0.020**
Reduced model			
Intercept	**31.51 ± 1.83**	**17.25**	**<0.001***
Supertype MHC1_3	**-0.66 ± 0.25**	**-2.67**	**0.008***
Supertype MHC1_7	-0.20 ± 0.24	-0.82	0.41
Genome-wide heterozygosity	**-18.95 ± 8.76**	**-2.16**	**0.031***
Sex (male vs. female)	**1.99 ± 0.33**	**6.04**	**<0.001***
Body size	**0.53 ± 0.13**	**4.16**	**<0.001***
Capture date	**-0.007 ± 0.003**	**-2.06**	**0.039***

Table S6. Associations between body mass and presence of specific MHC class II supertypes in the Eurasian coot. Bird identity and year were included as random factors in each model. The results of full and reduced model are shown. Significant predictors are marked in bold (at P < 0.05) or with asterisks (after correction for the false discovery rate).

Predictors	Estimate ± SE	z value	P
Full model			
Intercept	**28.82 ± 1.79**	**16.13**	**<0.001***
Supertype MHC2_1	**-0.6 ± 0.27**	**-2.25**	**0.024**
Supertype MHC2_2	-0.33 ± 0.26	-1.27	0.20
Supertype MHC2_3	**-0.94 ± 0.39**	**-2.43**	**0.015**
Supertype MHC2_4	-0.24 ± 0.31	-0.77	0.44
Supertype MHC2_5	0.15 ± 0.52	0.29	0.77
Supertype MHC2_6	0.14 ± 0.26	0.54	0.59
Supertype MHC2_7	-0.39 ± 0.23	-1.70	0.088
Supertype MHC2_8	-0.59 ± 0.33	-1.78	0.075
Supertype MHC2_9	0.25 ± 0.31	0.83	0.41
Supertype MHC2_10	0.40 ± 0.26	1.52	0.13
Supertype MHC2_11	0.02 ± 0.22	0.07	0.94
Supertype MHC2_12	-0.25 ± 0.30	-0.82	0.41
Supertype MHC2_13	0.19 ± 0.30	0.63	0.53
Supertype MHC2_14	0.09 ± 0.21	0.42	0.67
Genome-wide heterozygosity	-4.85 ± 8.79	-0.55	0.58
Sex (male vs. female)	**1.80 ± 0.31**	**5.81**	**<0.001***
Body size	**0.52 ± 0.12**	**4.24**	**<0.001***
Capture date	-0.005 ± 0.003	-1.71	0.086
Reduced model			
Intercept	**28.96 ± 1.80**	**16.09**	**<0.001***
Supertype MHC2_1	**-0.66 ± 0.19**	**-3.44**	**0.001***
Supertype MHC2_3	**-0.93 ± 0.37**	**-2.51**	**0.012***
Supertype MHC2_7	**-0.40 ± 0.19**	**-2.12**	**0.034**
Supertype MHC2_8	**-0.56 ± 0.28**	**-2.02**	**0.043**
Supertype MHC2_10	0.33 ± 0.24	1.33	0.18
Genome-wide heterozygosity	-5.97 ± 8.74	-0.68	0.49
Sex (male vs. female)	**1.89 ± 0.31**	**6.03**	**<0.001***
Body size	**0.48 ± 0.12**	**3.81**	**<0.001***
Capture date	-0.005 ± 0.003	-1.48	0.14

Table S7. Associations between blood haemoglobin concentration and presence of specific MHC class I supertypes in the Eurasian coot. Bird identity and year were included as random factors in each model. The results of full and reduced model are shown. Significant predictors are marked in bold (at P < 0.05) or with asterisks (after correction for the false discovery rate).

Predictors	Estimate ± SE	z value	P
Full model			
Intercept	**143.23 ± 30.81**	**4.65**	**<0.001***
Supertype MHC1_1	2.26 ± 3.51	0.65	0.52
Supertype MHC1_2	1.45 ± 4.43	0.33	0.74
Supertype MHC1_3	-1.78 ± 4.29	-0.41	0.68
Supertype MHC1_4	-0.57 ± 3.97	-0.14	0.89
Supertype MHC1_5	-5.65 ± 5.12	-1.10	0.27
Supertype MHC1_6	-4.87 ± 5.09	-0.96	0.34
Supertype MHC1_7	-8.16 ± 4.35	-1.88	0.060
Supertype MHC1_8	-0.92 ± 4.03	-0.23	0.82
Supertype MHC1_9	-4.66 ± 3.04	-1.53	0.13
Supertype MHC1_10	-7.74 ± 4.70	-1.65	0.10
Supertype MHC1_11	-8.63 ± 5.42	-1.59	0.11
Supertype MHC1_12	6.15 ± 4.38	1.40	0.16
Genome-wide heterozygosity	175.91 ± 148.85	1.18	0.24
Sex (male vs. female)	**16.82 ± 5.55**	**3.03**	**0.002***
Body size	-1.49 ± 2.10	-0.71	0.48
Capture date	**-0.25 ± 0.05**	**-5.49**	**<0.001***
Reduced model			
Intercept	**139.65 ± 30.04**	**4.65**	**<0.001***
Supertype MHC1_7	**-9.40 ± 4.17**	**-2.26**	**0.024***
Supertype MHC1_9	-3.33 ± 2.94	-1.13	0.26
Supertype MHC1_10	-7.37 ± 4.65	-1.58	0.11
Supertype MHC1_11	-8.39 ± 5.45	-1.54	0.12
Genome-wide heterozygosity	197.96 ± 143.43	1.38	0.17
Sex (male vs. female)	**16.62 ± 5.29**	**3.14**	**<0.001***
Body size	**-1.45 ± 0.48**	**-0.71**	**0.020***
Capture date	**-0.25 ± 0.04**	**-5.66**	**<0.001***

Table S8. Associations between blood haemoglobin concentration and presence of specific MHC class II supertypes in the Eurasian coot. Bird identity and year were included as random factors in each model. The results of full and reduced model are shown. Significant predictors are marked in bold (at P < 0.05) or with asterisks (after correction for the false discovery rate).

Predictors	Estimate ± SE	z value	P
Full model			
Intercept	**125.55 ± 29.53**	**4.25**	**<0.001***
Supertype MHC1_1	-0.40 ± 4.29	-0.09	0.93
Supertype MHC1_2	-2.99 ± 4.15	-0.72	0.47
Supertype MHC1_3	**-13.52 ± 6.33**	**-2.14**	**0.033**
Supertype MHC1_4	-8.12 ± 5.07	-1.60	0.11
Supertype MHC1_5	2.93 ± 8.26	0.35	0.72
Supertype MHC1_6	-5.94 ± 4.09	-1.45	0.15
Supertype MHC1_7	**-8.01 ± 3.75**	**-2.13**	**0.033**
Supertype MHC1_8	**-16.27 ± 5.14**	**-3.16**	**0.002***
Supertype MHC1_9	**-9.92 ± 4.98**	**-1.99**	**0.046**
Supertype MHC1_10	-5.48 ± 4.37	-1.26	0.21
Supertype MHC1_11	-5.95 ± 3.61	-1.65	0.10
Supertype MHC1_12	**-13.62 ± 4.75**	**-2.87**	**0.004***
Genome-wide heterozygosity	-1.16 ± 4.95	-0.23	0.82
Sex (male vs. female)	-1.49 ± 3.54	-0.42	0.67
Body size	**336.46 ± 142.62**	**2.36**	**0.018**
Capture date	**15.35 ± 4.93**	**3.12**	**0.002***
Reduced model			
Intercept	**119.71 ± 28.68**	**4.17**	**<0.001***
Supertype MHC2_3	**-11.70 ± 5.83**	**-2.00**	**0.045**
Supertype MHC2_4	-6.75 ± 4.70	-1.44	0.15
Supertype MHC2_6	-4.80 ± 3.76	-1.28	0.20
Supertype MHC2_7	**-7.48 ± 3.10**	**-2.41**	**0.016***
Supertype MHC2_8	**-14.62 ± 4.45**	**-3.28**	**0.001***
Supertype MHC2_9	**-9.25 ± 3.97**	**-2.33**	**0.020***
Supertype MHC2_11	-4.89 ± 3.14	-1.56	0.12
Supertype MHC2_12	**-13.86 ± 3.86**	**-3.59**	**<0.001***
Genome-wide heterozygosity	**345.81 ± 140.84**	**2.46**	**0.01***
Sex (male vs. female)	**15.38 ± 4.96**	**3.10**	**0.002***
Body size	-1.67 ± 1.91	-0.87	0.38
Capture date	**-0.26 ± 0.04**	**-5.98**	**<0.001***

Table S9. Associations between frontal shield size and presence of specific MHC class I supertypes in the Eurasian coot. Bird identity and year were included as random factors in each model. The results of full and reduced model are shown. Significant predictors are marked in bold (at P < 0.05) or with asterisks (after correction for the false discovery rate).

Predictors	Estimate ± SE	z value	P
Full model			
Intercept	3.90 ± 2.17	1.79	0.073
Supertype MHC1_1	-0.13 ± 0.24	-0.53	0.59
Supertype MHC1_2	-0.38 ± 0.34	-1.14	0.26
Supertype MHC1_3	-0.37 ± 0.31	-1.18	0.24
Supertype MHC1_4	-0.27 ± 0.26	-1.02	0.31
Supertype MHC1_5	0.11 ± 0.36	0.31	0.76
Supertype MHC1_6	0.20 ± 0.34	0.60	0.55
Supertype MHC1_7	-0.09 ± 0.29	-0.29	0.77
Supertype MHC1_8	-0.14 ± 0.26	-0.54	0.59
Supertype MHC1_9	0.16 ± 0.21	0.77	0.44
Supertype MHC1_10	0.31 ± 0.33	0.96	0.34
Supertype MHC1_11	-0.36 ± 0.39	-0.92	0.36
Supertype MHC1_12	0.22 ± 0.31	0.71	0.48
Genome-wide heterozygosity	-11.62 ± 10.32	-1.13	0.26
Sex (male vs. female)	**1.53 ± 0.39**	**3.96**	**<0.001***
Body size	0.10 ± 0.14	0.66	0.51
Capture date	**-0.014 ± 0.004**	**-4.09**	**<0.001***

Table S10. Associations between frontal shield size and presence of specific MHC class II supertypes in the Eurasian coot. Bird identity and year were included as random factors in each model. The results of full and reduced model are shown. Significant predictors are marked in bold (at P < 0.05) or with asterisks (after correction for the false discovery rate).

Predictors	Estimate ± SE	z value	P
Full model			
Intercept	3.02 ± 2.03	1.49	0.14
Supertype MHC2_1	-0.09 ± 0.30	-0.29	0.77
Supertype MHC2_2	-0.23 ± 0.29	-0.79	0.43
Supertype MHC2_3	-0.32 ± 0.40	-0.81	0.42
Supertype MHC2_4	0.06 ± 0.35	0.17	0.86
Supertype MHC2_5	-0.58 ± 0.63	-0.92	0.36
Supertype MHC2_6	-0.13 ± 0.31	-0.44	0.66
Supertype MHC2_7	-0.22 ± 0.26	-0.85	0.40
Supertype MHC2_8	-0.24 ± 0.38	-0.63	0.53
Supertype MHC2_9	-0.24 ± 0.35	-0.68	0.50
Supertype MHC2_10	-0.15 ± 0.28	-0.52	0.60
Supertype MHC2_11	-0.40 ± 0.26	-1.57	0.12
Supertype MHC2_12	**-0.68 ± 0.34**	**-1.99**	**0.046**
Supertype MHC2_13	-0.35 ± 0.36	-0.98	0.32
Supertype MHC2_14	0.16 ± 0.26	0.64	0.52
Genome-wide heterozygosity	-5.99 ± 9.95	-0.6	0.55
Sex (male vs. female)	**1.20 ± 0.35**	**3.43**	**0.001***
Body size	0.20 ± 0.14	1.45	0.15
Capture date	**-0.014 ± 0.003**	**-4.18**	**<0.001***
Reduced model			
Intercept	2.81 ± 1.96	1.44	0.15
Supertype MHC2_11	-0.34 ± 0.21	-1.63	0.10
Supertype MHC2_12	**-0.57 ± 0.27**	**-2.11**	**0.034**
Genome-wide heterozygosity	-6.16 ± 9.43	-0.65	0.51
Sex (male vs. female)	**1.16 ± 0.34**	**3.42**	**0.001***
Body size	0.20 ± 0.13	1.52	0.13
Capture date	**-0.015 ± 0.003**	**-4.51**	**<0.001***

Table S11. Associations between laying date and presence of specific MHC class I supertypes in the Eurasian coot. Bird identity and year were included as random factors in each model. The results of full and reduced model are shown. Significant predictors are marked in bold (at P < 0.05) or with asterisks (after correction for the false discovery rate).

Predictors	Estimate ± SE	z value	P
Full model			
Intercept	**109.3 ± 24.76**	**4.41**	**<0.001***
Supertype MHC1_1	-0.54 ± 2.74	-0.20	0.84
Supertype MHC1_2	4.58 ± 3.64	1.26	0.21
Supertype MHC1_3	-0.79 ± 3.54	-0.22	0.82
Supertype MHC1_4	0.8 ± 3.28	0.24	0.81
Supertype MHC1_5	-0.43 ± 3.91	-0.11	0.91
Supertype MHC1_6	**12.62 ± 4.89**	**2.58**	**0.010**
Supertype MHC1_7	4.17 ± 3.50	1.19	0.23
Supertype MHC1_8	-1.67 ± 3.11	-0.54	0.59
Supertype MHC1_9	-2.53 ± 2.44	-1.04	0.30
Supertype MHC1_10	**8.33 ± 3.97**	**2.10**	**0.036**
Supertype MHC1_11	-3.46 ± 4.11	-0.84	0.40
Supertype MHC1_12	3.05 ± 3.77	0.81	0.42
Genome-wide heterozygosity	-15.22 ± 125.61	-0.12	0.90
Sex (male vs. female)	-0.64 ± 2.44	-0.26	0.79
Reduced model			
Intercept	**110.06 ± 22.88**	**4.81**	**<0.001***
Supertype MHC1_6	**12.43 ± 4.88**	**2.55**	**0.011***
Supertype MHC1_10	**9.85 ± 3.84**	**2.56**	**0.010***
Genome-wide heterozygosity	-5.25 ± 113.59	-0.05	0.96
Sex (male vs. female)	0.19 ± 2.37	0.08	0.94

Table S12. Associations between laying date and presence of specific MHC class II supertypes in the Eurasian coot. Bird identity and year were included as random factors in each model. The results of full and reduced model are shown. Significant predictors are marked in bold (at $P < 0.05$) or with asterisks (after correction for the false discovery rate).

Predictors	Estimate ± SE	z value	P
Full model			
Intercept	**116.75 ± 24.54**	**4.76**	**<0.001***
Supertype MHC2_1	-2.4 ± 3.58	-0.67	0.50
Supertype MHC2_2	-0.32 ± 3.38	-0.1	0.92
Supertype MHC2_3	2.69 ± 4.65	0.58	0.56
Supertype MHC2_4	-4.01 ± 4.17	-0.96	0.34
Supertype MHC2_5	-0.27 ± 8.19	-0.03	0.97
Supertype MHC2_6	-0.15 ± 3.42	-0.05	0.96
Supertype MHC2_7	2.59 ± 3.16	0.82	0.41
Supertype MHC2_8	2.36 ± 4.16	0.57	0.57
Supertype MHC2_9	3.64 ± 4.32	0.84	0.40
Supertype MHC2_10	-3.92 ± 3.69	-1.06	0.29
Supertype MHC2_11	0.05 ± 3.09	0.02	0.99
Supertype MHC2_12	-5.5 ± 4.09	-1.34	0.18
Supertype MHC2_13	0.69 ± 4.01	0.17	0.86
Supertype MHC2_14	-5.17 ± 3.01	-1.72	0.086
Genome-wide heterozygosity	-21.9 ± 121.37	-0.18	0.86
Sex (male vs. female)	0.57 ± 2.61	0.22	0.83
Reduced model			
Intercept	**110.84 ± 23.38**	**4.74**	**<0.001***
Supertype MHC2_14	-3.28 ± 2.68	-1.22	0.22
Genome-wide heterozygosity	2.57 ± 116.53	0.02	0.98
Sex (male vs. female)	0.49 ± 2.42	0.20	0.84

Table S13. Associations between clutch size and presence of specific MHC class I supertypes in the Eurasian coot. Bird identity and year were included as random factors in each model. The results of full and reduced model are shown. Significant predictors are marked in bold (at $P < 0.05$) or with asterisks (after correction for the false discovery rate).

Predictors	Estimate ± SE	z value	P
Full model			
Intercept	**13.06 ± 2.66**	**4.91**	**<0.001***
Supertype MHC1_1	0.23 ± 0.25	0.89	0.37
Supertype MHC1_2	0.3 ± 0.36	0.84	0.4
Supertype MHC1_3	-0.42 ± 0.33	-1.27	0.2
Supertype MHC1_4	0.33 ± 0.31	1.07	0.29
Supertype MHC1_5	0.11 ± 0.37	0.29	0.77
Supertype MHC1_6	0.12 ± 0.48	0.25	0.81
Supertype MHC1_7	0.48 ± 0.35	1.38	0.17
Supertype MHC1_8	**-1.18 ± 0.3**	**-3.96**	**<0.001***
Supertype MHC1_9	-0.06 ± 0.24	-0.24	0.81
Supertype MHC1_10	-0.16 ± 0.41	-0.37	0.71
Supertype MHC1_11	-0.7 ± 0.40	-1.75	0.08
Supertype MHC1_12	0.12 ± 0.37	0.32	0.75
Genome-wide heterozygosity	2.27 ± 11.54	0.20	0.84
Sex (male vs. female)	-0.01 ± 0.23	-0.05	0.96
Brood status (first vs. second)	-0.49 ± 0.98	-0.5	0.62
Brood status (renest vs. second)	0.3 ± 0.99	0.30	0.76
Laying date	**-0.05 ± 0.01**	**-6.47**	**<0.001***
Reduced model			
Intercept	**11.64 ± 2.49**	**4.67**	**<0.001***
Supertype MHC1_8	**-1.11 ± 0.30**	**-3.75**	**<0.001***
Supertype MHC1_11	-0.64 ± 0.38	-1.66	0.10
Genome-wide heterozygosity	8.86 ± 10.53	0.84	0.40
Sex (male vs. female)	0.11 ± 0.22	0.50	0.62
Brood status (first vs. second)	-0.25 ± 0.98	-0.25	0.80
Brood status (renest vs. second)	0.46 ± 1.00	0.46	0.65
Laying date	**-0.04 ± 0.01**	**-6.48**	**<0.001***

Table S14. Associations between clutch size and presence of specific MHC class II supertypes in the Eurasian coot. Bird identity and year were included as random factors in each model. The results of full and reduced model are shown. Significant predictors are marked in bold (at P < 0.05) or with asterisks (after correction for the false discovery rate).

Predictors	Estimate ± SE	z value	P
Full model			
Intercept	**14.33 ± 2.65**	**5.40**	**<0.001***
Supertype MHC2_1	0.07 ± 0.37	0.20	0.84
Supertype MHC2_2	0.52 ± 0.33	1.57	0.12
Supertype MHC2_3	-0.48 ± 0.46	-1.03	0.30
Supertype MHC2_4	-0.53 ± 0.45	-1.18	0.24
Supertype MHC2_5	**-1.94 ± 0.85**	**-2.27**	**0.023**
Supertype MHC2_6	0.58 ± 0.32	1.82	0.070
Supertype MHC2_7	-0.31 ± 0.30	-1.03	0.30
Supertype MHC2_8	0.50 ± 0.40	1.26	0.21
Supertype MHC2_9	-0.14 ± 0.42	-0.33	0.74
Supertype MHC2_10	-0.08 ± 0.36	-0.23	0.82
Supertype MHC2_11	-0.05 ± 0.31	-0.15	0.88
Supertype MHC2_12	0.47 ± 0.41	1.15	0.25
Supertype MHC2_13	0.19 ± 0.37	0.53	0.60
Supertype MHC2_14	0.24 ± 0.29	0.84	0.40
Genome-wide heterozygosity	-6.04 ± 11.28	-0.54	0.59
Sex (male vs. female)	-0.03 ± 0.25	-0.13	0.90
Brood status (first vs. second)	-0.40 ± 1.01	-0.40	0.69
Brood status (renest vs. second)	0.37 ± 1.05	0.36	0.72
Laying date	**-0.04 ± 0.01**	**-6.27**	**<0.001***
Reduced model			
Intercept	**13.05 ± 2.52**	**5.17**	**<0.001***
Supertype MHC2_2	0.44 ± 0.27	1.64	0.10
Supertype MHC2_5	-1.57 ± 0.81	-1.94	0.052
Supertype MHC2_6	0.49 ± 0.28	1.76	0.078
Genome-wide heterozygosity	-1.11 ± 10.59	-0.10	0.92
Sex (male vs. female)	-0.13 ± 0.24	-0.52	0.60
Brood status (first vs. second)	0.01 ± 1.00	0.01	0.99
Brood status (renest vs. second)	0.80 ± 1.03	0.77	0.44
Laying date	**-0.04 ± 0.01**	**-6.30**	**<0.001***

Table S15. Associations between hatching success and presence of specific MHC class I supertypes in the Eurasian coot. Bird identity and year were included as random factors in each model. The results of full and reduced model are shown. Significant predictors are marked in bold (at $P < 0.05$) or with asterisks (after correction for the false discovery rate).

Predictors	Estimate ± SE	z value	P
Full model			
Intercept	0.67 ± 4.42	0.15	0.90
Supertype MHC1_1	0.17 ± 0.42	0.41	0.68
Supertype MHC1_2	-0.24 ± 0.60	-0.41	0.68
Supertype MHC1_3	1.43 ± 0.81	1.76	0.078
Supertype MHC1_4	0.18 ± 0.55	0.33	0.75
Supertype MHC1_5	0.43 ± 0.63	0.69	0.49
Supertype MHC1_6	0.05 ± 0.91	0.06	0.96
Supertype MHC1_7	-0.07 ± 0.57	-0.13	0.90
Supertype MHC1_8	**-1.01 ± 0.46**	**-2.21**	**0.027**
Supertype MHC1_9	-0.39 ± 0.38	-1.02	0.31
Supertype MHC1_10	0.03 ± 0.62	0.05	0.96
Supertype MHC1_11	0.44 ± 0.68	0.65	0.52
Supertype MHC1_12	0.31 ± 0.65	0.48	0.63
Genome-wide heterozygosity	5.00 ± 19.93	0.25	0.80
Sex (male vs. female)	0.12 ± 0.38	0.32	0.75
Brood status (first vs. second)	-0.80 ± 1.25	-0.64	0.52
Brood status (renest vs. second)	0.04 ± 1.26	0.03	0.98
Laying date	0.004 ± 0.011	0.33	0.74
Reduced model			
Intercept	0.51 ± 4.31	0.12	0.91
Supertype MHC1_3	1.45 ± 0.78	1.87	0.062
Supertype MHC1_8	**-1.02 ± 0.44**	**-2.32**	**0.021**
Genome-wide heterozygosity	4.14 ± 18.61	0.22	0.82
Sex (male vs. female)	0.14 ± 0.36	0.40	0.69
Brood status (first vs. second)	-0.90 ± 1.22	-0.74	0.46
Brood status (renest vs. second)	-0.17 ± 1.24	-0.13	0.89
Laying date	0.01 ± 0.01	0.49	0.62

Table S16. Associations between hatching success and presence of specific MHC class II supertypes in the Eurasian coot. Bird identity and year were included as random factors in each model. The results of full and reduced model are shown. Significant predictors are marked in bold (at $P < 0.05$) or with asterisks (after correction for the false discovery rate).

Predictors	Estimate ± SE	z value	P
Full model			
Intercept	7.91 ± 4.68	1.69	0.091
Supertype MHC2_1	-0.31 ± 0.55	-0.56	0.57
Supertype MHC2_2	0.43 ± 0.53	0.82	0.41
Supertype MHC2_3	0.42 ± 0.74	0.57	0.57
Supertype MHC2_4	-0.24 ± 0.64	-0.37	0.71
Supertype MHC2_5	-0.23 ± 1.44	-0.16	0.87
Supertype MHC2_6	-0.43 ± 0.50	-0.85	0.40
Supertype MHC2_7	0.08 ± 0.49	0.17	0.87
Supertype MHC2_8	1.66 ± 0.86	1.92	0.055
Supertype MHC2_9	-0.07 ± 0.66	-0.11	0.91
Supertype MHC2_10	0.59 ± 0.63	0.94	0.35
Supertype MHC2_11	-0.02 ± 0.51	-0.04	0.97
Supertype MHC2_12	-0.23 ± 0.60	-0.38	0.70
Supertype MHC2_13	0.06 ± 0.64	0.09	0.93
Supertype MHC2_14	0.67 ± 0.49	1.36	0.17
Genome-wide heterozygosity	-31.35 ± 20.19	-1.55	0.12
Sex (male vs. female)	0.09 ± 0.41	0.22	0.82
Brood status (first vs. second)	-0.97 ± 1.24	-0.78	0.44
Brood status (renest vs. second)	-0.02 ± 1.245	-0.02	0.99
Laying date	0.002 ± 0.012	0.20	0.84
Reduced model			
Intercept	7.49 ± 4.42	1.70	0.090
Supertype MHC2_8	**1.60 ± 0.77**	**2.07**	**0.038**
Genome-wide heterozygosity	-28.84 ± 19.13	-1.51	0.13
Sex (male vs. female)	0.14 ± 0.36	0.40	0.69
Brood status (first vs. second)	-1.01 ± 1.19	-0.84	0.40
Brood status (renest vs. second)	-0.16 ± 1.22	-0.13	0.90
Laying date	0.003 ± 0.010	0.25	0.80

Table S17. Associations between breeding success and presence of specific MHC class I supertypes in the Eurasian coot. Bird identity and year were included as random factors in each model. The results of full and reduced model are shown. Significant predictors are marked in bold (at P < 0.05) or with asterisks (after correction for the false discovery rate).

Predictors	Estimate ± SE	z value	P
Full model			
Intercept	2.01 ± 1.36	1.49	0.14
Supertype MHC1_1	0.13 ± 0.12	1.06	0.29
Supertype MHC1_2	0.29 ± 0.19	1.49	0.14
Supertype MHC1_3	-0.18 ± 0.16	-1.11	0.27
Supertype MHC1_4	0.20 ± 0.15	1.35	0.18
Supertype MHC1_5	0.08 ± 0.16	0.50	0.61
Supertype MHC1_6	0.02 ± 0.26	0.08	0.94
Supertype MHC1_7	**-0.41 ± 0.19**	**-2.17**	**0.030**
Supertype MHC1_8	0.10 ± 0.17	0.56	0.57
Supertype MHC1_9	0.00 ± 0.11	0.01	0.99
Supertype MHC1_10	**0.46 ± 0.19**	**2.46**	**0.014**
Supertype MHC1_11	0.03 ± 0.17	0.18	0.86
Supertype MHC1_12	-0.09 ± 0.19	-0.51	0.61
Genome-wide heterozygosity	-0.29 ± 5.99	-0.05	0.96
Sex (male vs. female)	-0.06 ± 0.11	-0.59	0.55
Brood status (first vs. second)	-0.21 ± 0.34	0.62	0.53
Brood status (renest vs. second)	0.19 ± 0.32	0.60	0.55
Laying date	**-0.010 ± 0.004**	**-2.43**	**0.015**
Reduced model			
Intercept	0.61 ± 1.45	0.54	0.59
Supertype MHC1_7	-0.24 ± 0.16	-1.55	0.12
Supertype MHC1_10	**0.41 ± 0.18**	**2.32**	**0.020**
Genome-wide heterozygosity	6.79 ± 4.88	1.39	0.17
Sex (male vs. female)	0.00 ± 0.10	-0.05	0.96
Brood status (first vs. second)	0.24 ± 0.32	0.75	0.45
Brood status (renest vs. second)	0.22 ± 0.31	0.71	0.48
Laying date	**-0.008 ± 0.003**	**-2.24**	**0.025**

Table S18. Associations between breeding success and presence of specific MHC class II supertypes in the Eurasian coot. Bird identity and year were included as random factors in each model. The results of full and reduced model are shown. Significant predictors are marked in bold (at P < 0.05) or with asterisks (after correction for the false discovery rate).

Predictors	Estimate ± SE	z value	P
Full model			
Intercept	0.09 ± 1.28	0.07	0.95
Supertype MHC2_1	0.12 ± 0.16	0.73	0.47
Supertype MHC2_2	-0.11 ± 0.14	-0.78	0.43
Supertype MHC2_3	-0.31 ± 0.23	-1.31	0.19
Supertype MHC2_4	-0.09 ± 0.18	-0.53	0.60
Supertype MHC2_5	-0.63 ± 0.83	-0.76	0.45
Supertype MHC2_6	**-0.41 ± 0.18**	**-2.26**	**0.024**
Supertype MHC2_7	0.02 ± 0.14	0.17	0.87
Supertype MHC2_8	0.14 ± 0.18	0.77	0.44
Supertype MHC2_9	0.00 ± 0.18	0.01	0.99
Supertype MHC2_10	0.02 ± 0.16	0.14	0.89
Supertype MHC2_11	0.12 ± 0.14	0.87	0.38
Supertype MHC2_12	0.11 ± 0.17	0.63	0.53
Supertype MHC2_13	0.02 ± 0.18	0.13	0.90
Supertype MHC2_14	0.05 ± 0.14	0.34	0.74
Genome-wide heterozygosity	5.86 ± 5.22	1.12	0.26
Sex (male vs. female)	0.14 ± 0.12	1.20	0.23
Brood status (first vs. second)	0.35 ± 0.35	1.02	0.31
Brood status (renest vs. second)	0.22 ± 0.33	0.67	0.50
Laying date	-0.003 ± 0.004	-0.75	0.45
Reduced model			
Intercept	0.19 ± 1.14	0.16	0.87
Supertype MHC2_6	**-0.43 ± 0.15**	**-2.88**	**0.004***
Genome-wide heterozygosity	6.89 ± 4.79	1.44	0.15
Sex (male vs. female)	0.09 ± 0.10	0.85	0.39
Brood status (first vs. second)	0.28 ± 0.32	0.88	0.38
Brood status (renest vs. second)	0.20 ± 0.31	0.67	0.51
Laying date	-0.004 ± 0.004	-1.20	0.23

Table S19. Associations between three phenotypic traits (body mass, haemoglobin concentration and frontal shield size) and the number of MHC class I and class II supertypes in the Eurasian coot. Bird identity and year were included as random factors. Significant predictors are marked in bold

Trait	Predictors	Estimate ± SE	z value	P
Body mass (males)	**Intercept**	**32.79 ± 2.99**	**10.97**	**<0.001**
	MHC class I supertypes	0.09 ± 0.09	0.95	0.34
	MHC class II supertypes	**-0.47 ± 0.21**	**-2.25**	**0.024**
	Genome-wide heterozygosity	-11.67 ± 13.34	-0.87	0.38
	Body size	**0.66 ± 0.20**	**3.35**	**0.001**
	Capture date	-0.007 ± 0.005	-1.49	0.14
Body mass (females)	**Intercept**	**31.00 ± 2.54**	**12.21**	**<0.001**
	MHC class I supertypes	**-0.16 ± 0.09**	**-1.78**	**0.076**
	MHC class II supertypes	0.26 ± 0.17	1.47	0.14
	Genome-wide heterozygosity	-22.9 ± 12.5	-1.83	0.067
	Body size	**0.45 ± 0.19**	**2.39**	**0.017**
	Capture date	0.000 ± 0.005	-0.08	0.93
Haemoglobin concentration	**Intercept**	**136.37 ± 38.61**	**3.53**	**<0.001**
(males)	**MHC class I supertypes**	**-4.57 ± 1.17**	**-3.89**	**<0.001**
	MHC class II supertypes	-0.95 ± 2.84	-0.34	0.74
	Genome-wide heterozygosity	313.89 ± 177.23	1.77	0.077
	Body size	0.04 ± 2.31	0.02	0.99
	Capture date	**-0.21 ± 0.05**	**-4.31**	**<0.001**
Haemoglobin concentration	**Intercept**	**156.27 ± 44.02**	**3.55**	**<0.001**
(females)	MHC class I supertypes	0.38 ± 1.55	0.25	0.81
	MHC class II supertypes	**-8.60 ± 2.90**	**-2.97**	**0.003**
	Genome-wide heterozygosity	265.33 ± 213.69	1.24	0.21
	Body size	-4.22 ± 3.17	-1.33	0.18
	Capture date	**-0.36 ± 0.08**	**-4.59**	**<0.001**
Frontal shield size	Intercept	3.28 ± 2.03	1.62	0.11
	MHC class I supertypes	-0.03 ± 0.07	-0.40	0.69
	MHC class II supertypes	-0.20 ± 0.15	-1.37	0.17
	Genome-wide heterozygosity	-6.97 ± 9.94	-0.70	0.48
	Sex (male vs. female)	**1.29 ± 0.35**	**3.69**	**<0.001**
	Body size	0.17 ± 0.13	1.23	0.22
	Capture date	**-0.014 ± 0.003**	**-4.27**	**<0.001**

Table S20. Associations between four reproductive traits (laying date, clutch size, hatching and breeding success) and the number of MHC class I and class II supertypes in the Eurasian coot. Bird identity and year were included as random factors. Significant predictors are marked in bold

Trait	Predictors	Estimate ± SE	z value	P
Laying date	**Intercept**	**117.19 ± 23.63**	**4.96**	**<0.001**
	MHC class I supertypes	**1.95 ± 0.83**	**2.35**	**0.019**
	MHC class II supertypes	-3.2 ± 1.79	-1.79	0.074
	Genome-wide heterozygosity	-15.03 ± 118.27	-0.13	0.90
	Sex (male vs. female)	0.21 ± 2.46	0.08	0.93
Clutch size	**Intercept**	**12.81 ± 2.65**	**4.83**	**<0.001**
	MHC class I supertypes	-0.06 ± 0.08	-0.72	0.47
	MHC class II supertypes	0.20 ± 0.17	1.17	0.24
	Genome-wide heterozygosity	-1.27 ± 11.23	-0.11	0.91
	Sex (male vs. female)	0.01 ± 0.24	0.06	0.95
	Brood status (first vs. second)	-0.16 ± 1.04	-0.16	0.87
	Brood status (renest vs. second)	0.53 ± 1.06	0.50	0.62
	Laying date	**-0.04 ± 0.01**	**-5.80**	**<0.001**
Hatching success	Intercept	4.96 ± 4.26	1.16	0.24
	MHC class I supertypes	0.02 ± 0.12	0.19	0.85
	MHC class II supertypes	0.09 ± 0.26	0.35	0.73
	Genome-wide heterozygosity	-19.13 ± 18.27	-1.05	0.29
	Sex (male vs. female)	0.03 ± 0.35	0.08	0.94
	Brood status (first vs. second)	-0.75 ± 1.21	-0.62	0.54
	Brood status (renest vs. second)	-0.03 ± 1.22	-0.02	0.98
	Laying date	0.004 ± 0.011	0.37	0.71
Breeding success	Intercept	0.77 ± 1.22	0.63	0.53
	MHC class I supertypes	0.03 ± 0.04	0.82	0.41
	MHC class II supertypes	-0.02 ± 0.07	-0.27	0.79
	Genome-wide heterozygosity	7.46 ± 5.07	1.47	0.14
	Sex (male vs. female)	0.02 ± 0.11	0.21	0.83
	Brood status (first vs. second)	0.12 ± 0.34	0.34	0.73
	Brood status (renest vs. second)	0.25 ± 0.32	0.76	0.45
	Laying date	**-0.010 ± 0.004**	**-2.50**	**0.012**

Table S21. Association between male haemoglobin concentration and the number of MHC class I supertypes in the Eurasian coot, while accounting for the presence of specific supertypes. Bird identity and year were included as random factors. Significant predictors are marked in bold

Predictors	Estimate ± SE	z value	P
Intercept	**114.29 ± 39.51**	**2.89**	**0.004**
MHC class I supertypes	**-4.49 ± 1.81**	**-2.48**	**0.013**
Supertype MHC1_7	-1.12 ± 5.56	-0.20	0.84
Supertype MHC1_9	5.09 ± 4.55	1.12	0.26
Supertype MHC1_10	-4.90 ± 6.13	-0.80	0.42
Supertype MHC1_11	-10.07 ± 6.97	-1.44	0.15
Genome-wide heterozygosity	**409.90 ± 184.26**	**2.22**	**0.026**
Body size	-1.57 ± 2.43	-0.64	0.52
Capture date	**-0.21 ± 0.05**	**-4.29**	**<0.001**

Table S22. Association between laying date and the number of MHC class I supertypes in the Eurasian coot, while accounting for the presence of specific supertypes. Bird identity and year were included as random factors. Significant predictors are marked in bold

Predictors	Estimate ± SE	z value	P
Intercept	**113.28 ± 22.98**	**4.93**	**<0.001**
MHC class I supertypes	0.52 ± 0.91	0.57	0.57
Supertype MHC1_6	-2.81 ± 1.72	-1.63	0.10
Supertype MHC1_10	**11.06 ± 5.05**	**2.19**	**0.029**
Genome-wide heterozygosity	**9.19 ± 4.17**	**2.21**	**0.027**
Sex (male vs. female)	10.53 ± 114.75	0.09	0.93

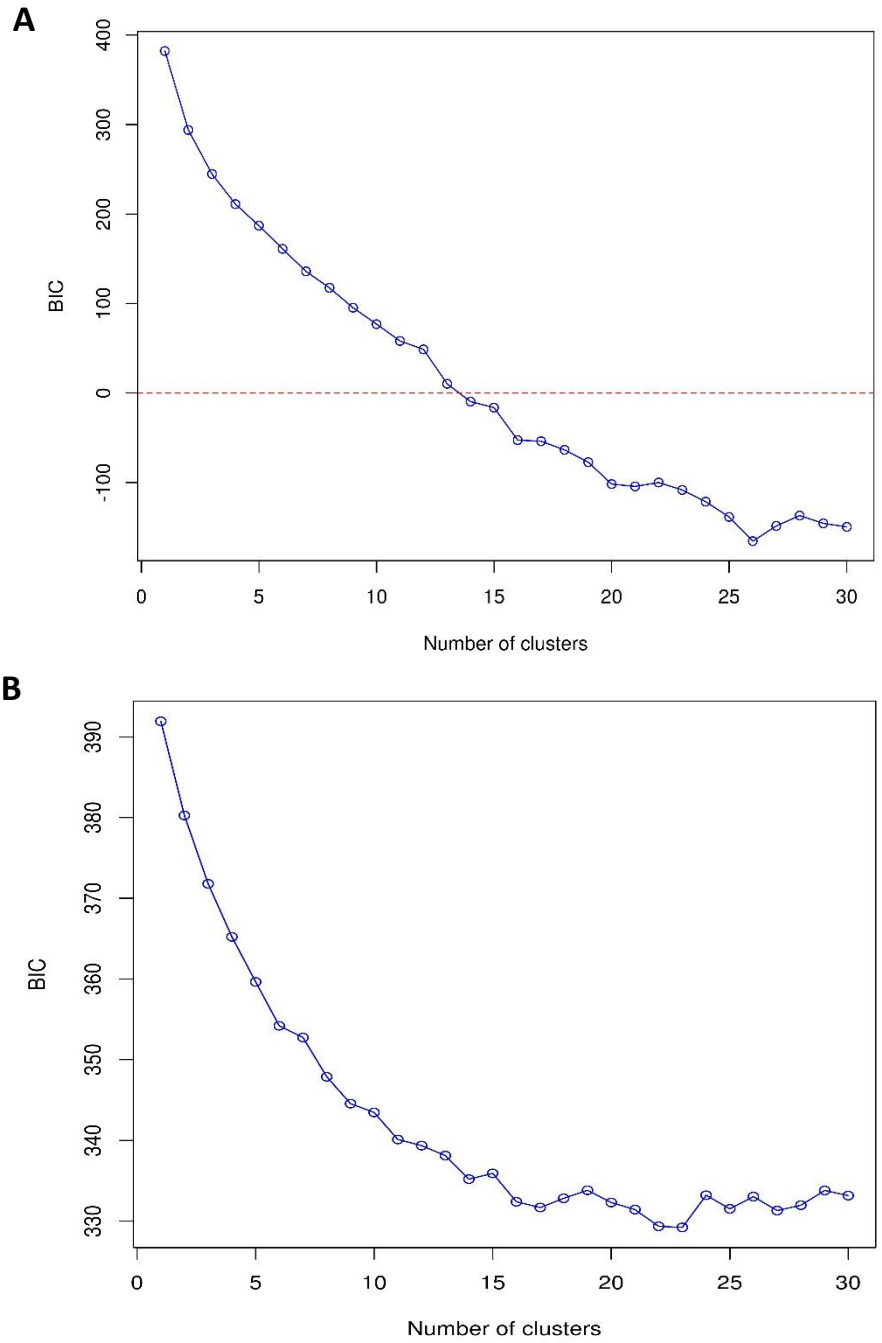
Table S23. Association between male body mass and the number of MHC class II supertypes in the Eurasian coot, while accounting for the presence of specific supertypes. Bird identity and year were included as random factors. Significant predictors are marked in bold

Predictors	Estimate ± SE	z value	P
Intercept	**30.51 ± 2.79**	**10.95**	**<0.001**
MHC class II supertypes	**-0.44 ± 0.20**	**-2.19**	**0.029**
Supertype MHC2_1	**-0.87 ± 0.29**	**-2.98**	**0.003**
Supertype MHC2_3	-0.32 ± 0.98	-0.32	0.75
Supertype MHC2_7	-0.14 ± 0.27	-0.51	0.61
Supertype MHC2_8	-0.45 ± 0.42	-1.07	0.28
Supertype MHC2_10	**0.69 ± 0.31**	**2.21**	**0.027**
Genome-wide heterozygosity	2.18 ± 12.80	0.17	0.86
Body size	**0.51 ± 0.18**	**2.92**	**0.004**
Capture date	-0.007 ± 0.004	-1.60	0.11

Table S24. Association between female haemoglobin concentration and the number of MHC class II supertypes in the Eurasian coot, while accounting for the presence of specific supertypes. Bird identity and year were included as random factors. Significant predictors are marked in bold

Predictors	Estimate ± SE	z value	P
Intercept	**127.23 ± 42.88**	**2.97**	**0.003**
MHC class II supertypes	**-6.77 ± 2.82**	**-2.40**	**0.016**
Supertype MHC2_3	-11.28 ± 6.38	-1.77	0.077
Supertype MHC2_7	-7.54 ± 4.51	-1.67	0.095
Supertype MHC2_8	-11.53 ± 5.99	-1.93	0.054
Supertype MHC2_9	-1.65 ± 6.60	-0.25	0.80
Supertype MHC2_12	-9.12 ± 5.36	-1.70	0.089
Genome-wide heterozygosity	**424.27 ± 208.90**	**2.03**	**0.042**
Body size	-3.47 ± 3.14	-1.11	0.27
Capture date	**-0.35 ± 0.08**	**-4.59**	**<0.001**

Fig. S1. Bayesian Information Criterion (BIC) values for different numbers of clusters of MHC class I (A) and MHC class II (B) alleles in the Eurasian coot.



8 ROZDZIAŁ TRZECI

PIKUS, E., WŁODARCZYK, R., JEDLIKOWSKI, J. AND MINIAS, P., 2021. URBANIZATION PROCESSES DRIVE DIVERGENCE AT THE MAJOR HISTOCOMPATIBILITY COMPLEX IN A COMMON WATERBIRD. PEERJ, *9*, E12264.

Urbanization processes drive divergence at the major histocompatibility complex in a common waterbird

Ewa Pikus[1], Radosław Włodarczyk[1], Jan Jedlikowski[2] and Piotr Minias[1]

[1] Department of Biodiversity Studies and Bioeducation, University of Łódź, Łódź, Poland
[2] Biological and Chemical Research Centre, University of Warsaw, Warsaw, Poland

ABSTRACT

Urban sprawl is one of the most common landscape alterations occurring worldwide, and there is a growing list of species that are recognised to have adapted to urban life. To be successful, processes of urban colonization by wildlife require a broad spectrum of phenotypic (*e.g.*, behavioural or physiological) adjustments, but evidence for genetic adaptations is much scarcer. One hypothesis proposes that different pathogen-driven selective pressures between urban and non-urban landscapes leads to adaptations in host immune genes. Here, we examined urbanization-related differentiation at the key pathogen-recognition genes of vertebrate adaptive immunity-the major histocompatibility complex (MHC)-in a common waterbird, the Eurasian coot (*Fulica atra*). Samples were collected from an old urban population (established before the 1950s), a new urban population (established in the 2000s), and two rural populations from central Poland. We found strong significant divergence (as measured with Jost's D) at the MHC class II between the old urban population and the remaining (new urban and rural) populations. Also, there was a moderate, but significant divergence at the MHC between the new urban population and two rural populations, while no divergence was found between the two rural populations. The total number of MHC alleles and the number of private (population-specific) MHC alleles was lower in old urban populations, as compared to the rural ones. These patterns of differentiation at the MHC were not consistent with patterns found for neutral genetic markers (microsatellites), which showed few differences between the populations. Our results indicate that MHC allele composition depended on the level of anthropogenic disturbance and the time which passed since urban colonization, possibly due to the processes of genotype sorting and local adaptation. As such, our study contributes to the understanding of genetic mechanisms associated with urbanization processes in wildlife.

Subjects Ecology, Genetics, Molecular Biology, Zoology, Population Biology
Keywords Local adaptation, Eurasian coot, Major histocompatibility complex, Population differentiation, Urbanization, MHC, Urban evolution

INTRODUCTION

For the first time in our history more than a half of the world's human population lives in cities, and by 2030 every third resident is expected to live in a city of at least a half-million people (*Camacho-Valdez et al., 2009*; *McDonald, Kareiva & Forman, 2008*). Such intense

Submitted 28 April 2021
Accepted 16 September 2021
Published 5 October 2021

Corresponding author
Piotr Minias,
piotr.minias@biol.uni.lodz.pl

Academic editor
James Baxter-Gilbert

Additional Information and Declarations can be found on page 17

DOI **10.7717/peerj.12264**

urbanization entails huge investments in urban infrastructure, which significantly changes the natural landscape and, at the same time, creates novel urban ecosystems and novel ecological niches (*Aouissi et al., 2017*; *Croci, Butet & Clergeau, 2008*; *Green et al., 2016*). Many animal species may not be able to adapt and reproduce in highly altered landscapes outside of their natural niches, particularly those less plastic and more ecologically specialized, and thus they are often excluded from urban environments (*Concepción et al., 2015*; *Luniak, 2004*). Despite the challenges posed by urban environments, more and more animal species are adapting to new anthropogenic environmental conditions and colonizing cities-overcoming the ecological, demographic, and behavioural barriers (*Luniak, 2004*). Urban areas may be an attractive habitat alternative for species that can take advantage of these novel ecosystems, including benefiting from access to anthropogenic food, milder climate during the winter, and lower pressure from natural predators (*Davis, Malas & Minor, 2014*; *Luniak, 2004*; *Minias & Janiszewski, 2016*). Conversely, the use of urban resources by wild animals often comes at some costs (*e.g.*, exposure to novel pathogens).

Anthropogenic environmental change may promote the emergence of novel pathogens through the transportation and introduction of infectious agents or hosts to new environments, through manipulation of local ecosystems which can favour the proliferation or prolonged survival of infectious agents or by facilitating new host–pathogen interactions (*Miller et al., 2002*). Relatively poor biodiversity of urban areas leads to the simplified regulation processes of animal communities (*Adams, Van Druff & Luniak, 2005*), and thus, habitat-related differences in pathogen pressure may pose a challenge during the process of urban colonization. Urbanization generally reduces the abundance and diversity of parasites, but on the other hand parasite or pathogen transmission can, in some cases, increase among urban-adapted hosts (*Bradley & Altizer, 2007*; *Brearley et al., 2013*). For example, a recently established urban population of dark-eyed juncos (*Junco hyemalis*) was found to host significantly higher numbers of ectoparasites and had more pox lesion scars compared to non-urban populations from montane habitats (*Whittaker et al., 2012*). Additionally, urban wildlife can be exposed to novel parasites found in the faeces of domestic animals, such as the infectious oocytes of the protozoan parasite *Toxoplasma gondii*, the causative agent of toxoplasmosis (*Bradley & Altizer, 2007*; *Frenkel et al., 1995*). The differences in the composition of pathogen faunas between urban and non-urban areas may require necessary adaptations of the immune system (*Alcaide et al., 2008*; *Watson et al., 2017*). Urban bullfinches (*Loxigilla barbadensis*) are known to have enhanced immunocompetence compared to non-urban conspecifics (*Audet, Ducatez & Lefebvre, 2016*) and transcriptomic analysis of great tits (*Parus major*) has provided strong evidence for the upregulation of adaptive immune responses in urban individuals (*Watson et al., 2017*). However, genetic variation in urban colonizing wildlife may not only be driven by selective processes, such as local adaptation (*Loiseau et al., 2009*), but also by stochastic processes, like genetic drift (*Lourenço et al., 2017*). Genetic bottlenecks may occur during the establishment of novel urban populations (*i.e.*, the founder effect), which can result in a low level of genetic

diversity, including the diversity of the immune receptors responsible for antigen recognition (*Alcaide, 2010*; *Sutton et al., 2011*; *Strand et al., 2012*).

Two families of pathogen-recognition genes play a key role in innate and adaptive immunity: toll-like receptors (TLR) and the major histocompatibility complex (MHC) (*Schenten & Medzhitov, 2011*). In general, TLRs are responsible for identifying conservative antigens (pathogen-associated molecular patterns, PAMP) which are characteristic of a wide spectrum of pathogens and parasites (*Barton & Medzhitov, 2020*). In turn, MHC molecules activate adaptive immune response (*Ekblom et al., 2007*) by presenting specific intercellular (MHC class I) or extracellular (MHC class II) antigens to T cells (*Piertney & Oliver, 2006*). MHC genes are the most polymorphic genes in vertebrates and the majority of polymorphism is concentrated in the peptide-binding region (PBR) which binds peptides originating from pathogen processing (*i.e.*, from the proteasomal proteolysis or lysosomal hydrolysis of pathogens) (*Hess & Edwards, 2002*; *Hughes & Yeager, 1998*; *Münz, 2012*). Even minor amino acid alterations in PBR may lead to large differences in the repertoire of the peptides recognized and, as a consequence, change the spectrum of the pathogens against which an organism effectively activates the immune response (*Dionne et al., 2007*). In general, pathogen-driven balancing selection-acting through the mechanisms of heterozygote advantage, negative frequency-dependent selection, and fluctuating selection-maintains extraordinary diversity of MHC in natural populations; unlike the relatively low variable and structurally conservative TLRs (*Barton & Medzhitov, 2020*; *Borghans, Beltman & De Boer, 2004*). Thus, MHC allelic composition within populations should quickly respond *via* local adaptation to changes in the composition of pathogens and parasite fauna (*Ekblom et al., 2007*; *Miller, Allendorf & Daugherty, 2010*).

So far, the strongest evidence for associations between immune genes and urbanization comes from mammals. For example, significant genetic variation at neutral and immune gene (MHC and TLR) linked microsatellite loci was found between populations of bobcats (*Lynx rufus*) whose habitats differed in urbanization level (*Serieys et al., 2015*). Red foxes (*Vulpes vulpes*) that colonized Zurich, Switzerland, showed depleted neutral and functional (immune-linked) genetic diversity following a founder event (*DeCandia et al., 2019a*). Also, significant differentiation between fox populations separated by natural and anthropogenic barriers was found at both types of markers, with evidence of selection acting on MHC-linked markers (*DeCandia et al., 2019a*). In contrast to the older colonization of Zurich by red foxes, there was little support for balancing selection maintaining diversity at the MHC-linked loci of coyotes (*Canis latrans*) after the recent colonization of the New York metropolitan area, USA (*DeCandia et al., 2019b*). Differentiation of MHC genes has been detected in birds at the landscape level, but little research has tested for the effects of urbanization. For example, significant between-population differentiation at the MHC class II genes contrasted with the relatively homogeneous distribution of microsatellite alleles in the lesser kestrel (*Falco naumanni*; *Alcaide et al., 2008*) and great snipe (*Gallinago media*; *Ekblom et al., 2007*). These contrasting patterns were primarily attributed to variation in habitat structure and

processes of local adaptation to site-specific pathogen faunas (*Alcaide et al., 2008*; *Ekblom et al., 2007*).

Our study aims to examine the differentiation and diversity of MHC genes in a common waterbird, the Eurasian coot (*Fulica atra*, Rallidae, Gruifomes), from areas that varied in urbanization level. For this purpose, we genotyped MHC class II genes in four populations of the Eurasian coot from central Poland: (1) an old urban population (Warszawa; urban area colonized by coots before 1950); (2) a new urban population (Łódź; urban area colonized by coots at the beginning of the 2000s); and (3) two non-urban populations (Sarnów and Żeromin). Since the process of urban colonization by the Eurasian coot is still in progress in central Europe, the species may be considered a good model to study genetic adaptations across urban and non-urban populations with different histories. Here, we test whether urbanization processes are related to changes in the MHC gene pool and if they lead to reduced immunogenetic diversity at the population level. We hypothesized that differences in urbanization level altered the allelic composition of MHC genes between urban and non-urban Eurasian coot populations, with more pronounced differences observed in the old *versus* new urban population. We additionally hypothesized that the process of urban colonization would be associated with reduced MHC diversity, either due to local adaptation (no parallel reduction in neutral diversity expected) or demographic history and genetic drift (parallel reduction in neutral diversity expected).

MATERIALS AND METHODS

Study area

Data were collected in two urban (Warszawa and Łódź) and two rural (Sarnów and Żeromin) populations of Eurasian coot from central Poland during 2012–2018 (Fig. 1). Warszawa (52.259° N, 21.020° E) is the largest city in Poland in terms of population size (1.78 million inhabitants in 2020; according to data by Statisctics Poland, Poland, https://stat.gov.pl/en/), and area (517.24 km^2); it is also characterized by high population density (3,462 people/km^2 in 2020). The city was colonized by coots in the mid-20th century, with a stable breeding population of coots being reported in highly urbanized parts of Warszawa as early as in the 1960s (*Luniak, Kalbarczyk & Pawłowski, 1964*). During the 1980s there were already 190–210 breeding pairs (*Luniak et al., 2001*) in the city. The relatively early establishment of the urban population in Warszawa was likely facilitated by the presence of a large river (Vistula) in the city centre. Łódź (51.757° N, 19.493° E) is the third largest city in Poland in terms of population size (672,185 inhabitants in 2020; Statisctics Poland, Poland), and its administrative area covers 293.25 km^2 with a population density of 2,309 people/km^2 in 2020. The urban areas of Łódź were colonized by coots relatively recently, with only ~20 coot nesting sites being recorded exclusively on the outskirts of Łódź during surveys in 1994–2002 (*Janiszewski, Wojciechowski & Markowski, 2009*). Since then, several locations in the city centre have been colonized. During the period of this study, the coot population size in Łódź was estimated at ~70 breeding pairs, including ~40 pairs breeding in the city centre. Because of evident differences in the timing of colonization, Warszawa and Łódź populations are

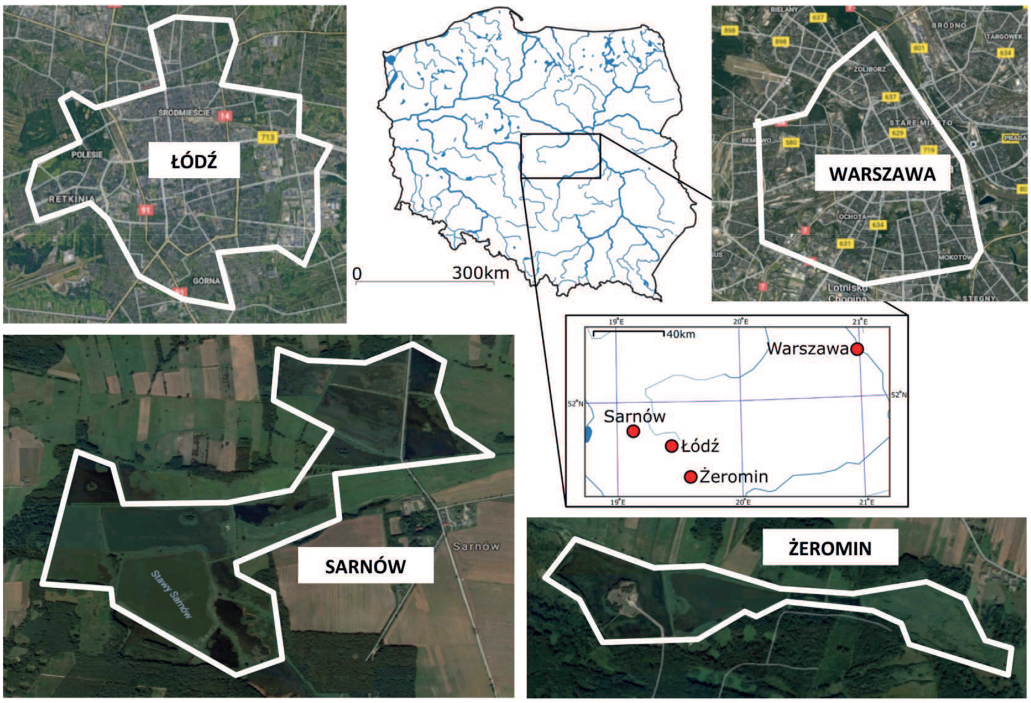


Figure 1 Location and land cover of sampling sites. Borders of sampling areas are marked with white lines. Satellite map data © 2021 Google, CNES/Airbus, MGGP Aero, Maxar technologies.
Full-size DOI: 10.7717/peerj.12264/fig-1

henceforth referred to as old and new urban populations, respectively. In the natural environment, coots inhabit a variety of standing waterbodies, such as lakes, field ponds, oxbow lakes, but also in the reservoirs of artificial origin, including fish ponds, clay pit ponds, peat lakes, storage reservoirs, or channels. Rural populations of coots were located at fish ponds in Sarnów (51.851° N, 19.109° E) and Żeromin (51.617° N, 19.607° E), which provided semi-natural nesting habitat with extensive reed areas (Fig. 1). Trespassing of unauthorized personnel was restricted at both sites, resulting in low human presence and a relatively low degree of anthropogenic pressure.

Sample collection and DNA isolation

Blood samples were taken during the reproductive season from 20–30 adult breeding individuals per population (n = 103 individuals in total). Birds were caught with noose traps put in the nests and made from monofilament nylon. Several birds were caught by hand (exclusively in urban populations). Approximately 75 µl of blood was taken from a tarsal vein from each bird and stored in 96% ethanol at 2 °C until further analysis. Each bird was ringed and marked with a plastic neck collar, which allowed us to avoid recapturing the same individuals. Genomic DNA was isolated from blood samples with GeneJET Genomic DNA Purification Kit (Fermentas, Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's protocol. All methods were carried out in accordance with the current laws of Poland (Act on Nature Conservation from 16 April 2004, Journal of Laws from 2004, No. 92, item 880) and were approved by the Local

Bioethical Commission for Experiments on Animals in Łodź (nos 40/ŁB 620/2012 and 15/ŁB/2016).

Microsatellite genotyping and variation

To assess the level of neutral genetic variation in our study populations, we used ten microsatellite loci (Table S1). Nine microsatellite markers were originally developed for other Rallidae species: Tm18, Tm20, Tm27, Tm38 (Tasmanian native-hen, *Tribonyx mortierii*) (*Buchan, 2000*), Kira8, Kira9, Kira10, Kira16 (king rail, *Rallus elegans*) (*Brackett et al., 2013*), and B106 (black rail, *Laterallus jamaicensis*) (*Spautz, Nur & Stralberg, 2005*). The last marker, TG03002, was highly conserved and amplified successfully in a wide range of passerine and non-passerine bird species (*Dawson et al., 2010*). PCR amplifications followed original protocols and were conducted in a total volume of 20 µl containing ten µl of 2X DreamTaq PCR Master Mix (DreamTaq DNA Polymerase, 2X DreamTaq buffer with four mM $MgCl_2$, and 0.4 mM of each dNTP) (ThermoFisher Scientific, Waltham, MA, USA), eight µL of deionized water, 0.5 µL of each primer (final concentration of 1.25 µM) and one µL of genomic DNA solution (20–80 ng of DNA). Fragment size analysis was conducted with ABI/Hitachi 3,500 (Applied Biosystems, Foster City, CA, USA) sequencer. The size of alleles was assessed with GeneScan TM 600 LIZ Standard (Applied Biosystems, Foster City, CA, USA) using Geneious 7.1.7 software (Biomatters, Auckland, New Zealand). All loci were successfully genotyped in all individuals. To exclude any major genotyping errors allele calling was checked twice and genotyping of technical replicates showed high (>95%) reproducibility of alleles.

We found no deviation from the Hardy–Weinberg equilibrium at any locus in any population (all $p > 0.05$ after Bonferroni correction), as assessed with the exact tests (*Guo & Thompson, 1992*) that were run with the Markov chain method (chain length: 1,000,000; dememorization: 1,00,000) in Arlequin 1.3.5.2 software (*Excoffier & Lischer, 2010*). Also, no evidence for linkage disequilibrium was found between any pair of loci (all $p > 0.05$ after Bonferroni correction), as tested with FSTAT 2.9.3 software (*Goudet, 1995*). The frequency of null alleles was relatively low (max. 0.052 at Kira-16), as checked with Cervus 3.0.3 (*Kalinowski, Taper & Marshall, 2007*). Micro-Checker 2.2.3 (*Van Oosterhout et al., 2004*) did not indicate genotyping errors due to null alleles, short allele dominance (large allele dropout), or stuttering. The observed and expected heterozygosity was calculated with GeneAlEx v.6.5 software (*Peakall & Smouse, 2006*; *Smouse & Peakall, 2012*).

MHC genotyping

MHC genotyping focused on class II genes, which were previously reported to show high level of polymorphism in natural coot populations (*Alcaide et al., 2014*).
For amplifications we used species-specific primers, Fuat-Ex2Fw (5′-CTGACCR GCCTCCCTGCA-3′) and Fuat-Ex2Rv (5′-TTGTGCCAYACACCCACC-3′), which amplify the entire MHC class II exon 2 (270 bp), binding with the neighbouring regions of intron 1 and 2 (*Alcaide et al., 2014*). Our targeted region (exon 2) codes for one of two domains that form the peptide-binding groove of the MHC molecule and, thus, have a

direct impact on antigen recognition. PCR amplifications were done in a total volume of 20 µl containing ten µl of 2X HotStarTaq Plus MasterMix Kit (HotStarTaq Plus DNA Polymerase, PCR Buffer with 0 mM $MgCl_2$, and 400 µM of each dNTP) (Qiagen, Venlo, Netherlands), eight µl of deionized water, 0.5 µl of each primer (final concentration of 1.25 µM), and one µl of genomic DNA solution (20–80 ng of DNA). To allow identification of samples, amplifications were completed using fusion primers containing Illumina Nextera Transposase adapter sequences (Illumina Corp., San Diego, CA, USA), seven-bp barcodes that indicated sample identity, and original MHC primers. Following *Alcaide et al. (2014)*, conditions for PCR reactions consisted of an initial denaturation (95 °C for 5 min) followed by 24 cycles of denaturation (95 °C for 60 s), annealing (60 °C for 40 s) and elongation (72 °C for 60 s), and a final extension (72 °C for 10 min). The number of PCR cycles was set to 24 to reduce the risk of producing chimeras. All PCR products were purified, and their concentration was assessed with a Quant-iT PicoGreen dsDNA marking kit (Thermo Fisher Scientific, Waltham, MA, USA). A library was prepared from equimolar quantities of PCR products using NEB-Next DNA Library Prep Master Mix Set for Illumina (New England Biolabs, Ipswich, MA, USA) and sequenced on the 2 × 250 bp Illumina MiSeq platform.

Processing of Illumina data and MHC allele validation

Raw Illumina MiSeq data were processed using Amplicon Sequencing Analysis Tools (AmpliSAT) web server (*Sebastian et al., 2016*) and processing algorithms recommended by *Biedrzycka et al. (2017)*. To merge pair-ended reads we used FLASH algorithm (*Magoc & Salzberg, 2011*) with optimum overlapping parameters (determined based on amplicon data), as implemented in AmpliMERGE tool. Then, AmpliSAS tool was used for de-multiplexing, clustering, and filtering of reads. Default parameters for Illumina data were used for clustering (1% substitution errors, 0.001% indel errors, and 25% minimum dominant frequency), while in the filtering step we discarded chimeras and sequences that had less than 3% frequency. Minimum amplicon depth was set to 300 reads, while maximum amplicon depth was set to 5,000 reads (due to processing limitations of AmpliSAS). Before data processing, the average sequencing (amplicon) depth was 1,706 ± 73 [SE] per sample, while the average number of reads for validated alleles was 1,398 ± 64 [SE] reads per sample. Reproducibility of alleles was assessed using 25 technical replicates (independent PCR amplifications from the same individuals). All reads had 273 or 276 bp length and the difference in sequence length was due to a single-codon deletion, previously reported for MHC class II exon 2 of the Eurasian coot (*Alcaide et al., 2014*). All validated alleles were aligned in Geneious v10.0.5 (Biomatters Ltd., Auckland, New Zealand). Intron regions (two codons) were removed from the alignments, retaining full length (267/270 bp) MHC class II exons 2 sequences, which are referred to as alleles in all further analyses.

Statistical analyses

Basic measures of allelic diversity within populations (*i.e.*, total number of alleles and the number of private alleles per population) were calculated for MHC and microsatellite loci

using GeneAlEx v.6.5 software. For microsatellite loci we also calculated mean observed heterozygosity for each population using the same software, while for the MHC we calculated population-specific sequence polymorphism measures (number of segregating sites, number of nucleotide differences, nucleotide diversity) using DnaSP v6.10.03 (*Rozas et al., 2017*). We used rarefaction to standardize all population-specific measures of allelic diversity for a population size of 20 individuals, which was the minimum sample size available for any of the study populations (Tables S2, S3 in File S2). Differences in the individual MHC diversity (*i.e.*, number of alleles per individual) and individual microsatellite heterozygosity between populations were tested with general linear mixed models (GLMM) in *glmmADMB* package (*Skaug et al., 2012*) developed for the R v3.6.3 statistical environment (R Foundation for Statistical Computing, Vienna, Austria). The MHC diversity and microsatellite heterozygosity were entered as response variables in separate GLMMs, while population was entered as a four-level fixed factor. In the analysis of the MHC we also added microsatellite heterozygosity as a covariate to control for neutral genetic variation. The effect of year was included in both models as a random factor to control for inter-annual variation.

In order to assess genetic differentiation between populations, we used two major approaches. First, we calculated pairwise Jost's D measures of differentiation for both MHC and microsatellite data (*Jost, 2008*). Calculations of Jost's D are based on the effective number of alleles (*Jost, 2008*), instead of expected heterozygosity, as in the case of F_{ST} and similar statistics (*Meirmans & Hedrick, 2011*), which was unknown for multilocus MHC data. Jost's D estimates for MHC data were calculated in the *SpadeR* R package (*Chao et al., 2016*) and for the purpose of analyses each allele was coded as present or absent (as in a dominant marker). For microsatellite data, Jost's D values were estimated using bootstrapping (1,000 permutation) in the *diveRsity* R package (*Keenan et al., 2013*). Additionally, we calculated standardized G'_{ST} values (*i.e.*, pairwise G_{ST} values divided by the maximum G_{ST} value; *Hedrick, 2005*), but this was done exclusively for microsatellite loci. A correlation between pairwise Jost's D values for MHC and microsatellites was tested using Pearson product-moment correlation coefficient.

Second, we used Bayesian clustering algorithm implemented in the program STRUCTURE (*Pritchard, Stephens & Donnelly, 2000*) to infer the number of genetic clusters (K) in the data and to assign individual genotypes to these clusters (separately for MHC and microsatellite data). Since it was not possible to assign MHC alleles to loci, we encoded each allele as a dominant biallelic locus, following *Herdegen, Babik & Radwan (2014)* and ran the analysis of MHC data as for dominant markers (*Falush, Stephens & Pritchard, 2007*). We set the number of tested K values between one and five. We used the admixture model of ancestry, correlated model of allele frequencies, and sampling location as prior information. The number of Markov chain Monte Carlo iterations was set to 5,00,000 and the length of burn-in period was set to 2,00,000. Each analysis was replicated ten times. The number of genetic clusters for each type of marker was inferred by a comparison of mean posterior probabilities for different K values $L(K)$ (also referred to as likelihood scores), as implemented in the Structure Harvester 0.6.94

(*Earl & von Holdt, 2012*). We preferred L(*K*) against the Δ*K* algorithm (*Evanno, Regnaut & Goudet, 2005*), since the latter does not test for $K = 1$ and, thus, it may overestimate the number of genetic clusters under the homogeneous genetic structure (*Janes et al., 2017*). Multiple runs of clustering analysis were averaged with Clumpp 1.1.2 (*Jakobsson & Rosenberg, 2007*) and the final output was visualized with Distruct 1.1 (*Rosenberg, 2004*). All values are reported as means ± SE.

RESULTS

MHC and microsatellite polymorphism

After clustering and filtering of Illumina sequences, we found that 115 MHC class II alleles were retained across all populations. These corresponded to 113 unique amino acid sequences. Reproducibility of alleles after processing was 100% (see File S2 for allele calling across 25 technical replicates). Between one and six alleles were recorded per individual, being consistent with the presence of three MHC class II loci in the Eurasian coot. Most individuals (53.4%, $n = 103$) had three MHC alleles and the mean number of MHC alleles was 2.89 ± 0.08 per individual.

Most microsatellite loci showed moderate polymorphism (3–17 alleles per locus), except for Kira-16, where 48 alleles were recorded (Table S1). The observed heterozygosity (Ho) of microsatellites was 0.59–0.66 per population (Table S3) and 0.09–0.88 per locus (Table S1).

Reduced MHC diversity in the old urban population

We found significant differences in individual MHC diversity (*i.e.*, the number of MHC alleles per individual) between the populations (Fig. 2), as MHC diversity was significantly lower in the old urban population (Warszawa), when compared with the new urban population (Łódź; $\beta = -0.60 \pm 0.21$, $P = 0.005$, Table 1) and one of the rural populations (Żeromin; $\beta = -0.52 \pm 0.23$, $P = 0.029$, Table 1). There were no significant differences between the new urban population and both rural populations (Table 1). Rural populations were characterized by the highest total number of MHC alleles, as well as the highest number of private MHC alleles (Table S2). A similar pattern was revealed after the standardization of the sample sizes between populations, showing that the total number of MHC alleles was 34–48% lower in urban than rural populations (Table S2, Fig. 3A). Despite higher allelic diversity in rural populations, all measures of MHC sequence polymorphism (*i.e.*, number of segregating sites, number of nucleotide differences, and nucleotide diversity) were similar across urban and rural populations (Table S2). In contrast to the MHC, we found no differences in the individual heterozygosity at the microsatellite loci between the populations (Table 1; Fig. 2), although the difference between the old urban population (Warszawa) and one of the rural populations (Żeromin) was only marginally non-significant (lower heterozygosity in Warszawa; $\beta = -0.070 \pm 0.039$, $P = 0.072$, Table 1). The total number of alleles and the number of private alleles at microsatellite loci showed little variation across all four populations (Table S3, Fig. 3B).

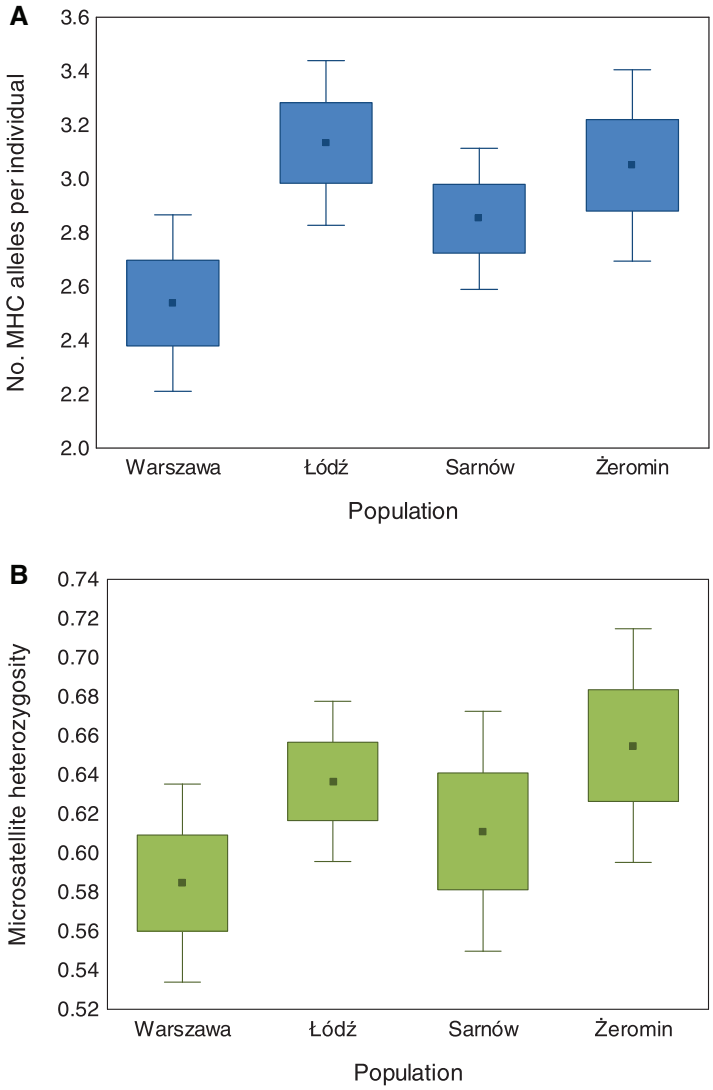


Figure 2 Differences in the number of MHC class II alleles per individual (A) and individual heterozygosity at microsatellite loci (B) between old urban (Warszawa), new urban (Łódź) and two rural (Sarnów and Żeromin) populations of the Eurasian coot. Statistical significance for pairwise comparisons is reported in Table 2. Full-size DOI: 10.7717/peerj.12264/fig-2

Significant MHC differentiation between urban and rural populations

We found strong population differentiation at the MHC and the average pairwise Jost's D was 0.441. The strongest differentiation at MHC genes was recorded between the old urban population (Warszawa) and all other populations, although the new urban population (Łódź) also showed significant differentiation from both rural populations (Table 2). The two rural populations showed no significant differentiation at the MHC (Jost's D = 0.142 ± 0.152; Table 2). No significant population differentiation was found at the microsatellite loci, as measured with Jost's D and G'_{ST} (Table 2). Pairwise Jost's D values for MHC and microsatellites did not show significant correlation (r = 0.47, n = 6, p = 0.35).

Table 1 Differences in the individual MHC diversity (the number of MHC class II alleles per individual) and individual microsatellite heterozygosity between populations.

Marker	Predictor	Estimate ± SE	t	*P*
MHC class II	**Intercept**	**2.58 ± 0.38**	**6.82**	**<0.001**
	Population (Warszawa *vs.* Łódź)	**−0.60 ± 0.21**	**−2.87**	**0.005**
	Population (Warszawa *vs.* Sarnów)	−0.32 ± 0.21	−1.49	0.14
	Population (Warszawa *vs.* Żeromin)	**−0.52 ± 0.23**	**−2.22**	**0.029**
	Population (Łódź *vs.* Sarnów)	0.28 ± 0.21	1.38	0.17
	Population (Łódź *vs.* Żeromin)	0.08 ± 0.22	0.37	0.71
	Population (Sarnów *vs.* Żeromin)	−0.20 ± 0.22	−0.88	0.38
	Microsatellite heterozygosity	−0.08 ± 0.59	−0.14	0.89
Microsatellites	**Intercept**	**0.584 ± 0.026**	**22.89**	**<0.001**
	Population (Warszawa *vs.* Łódź)	−0.052 ± 0.035	−1.49	0.14
	Population (Warszawa *vs.* Sarnów)	−0.026 ± 0.036	−0.74	0.46
	Population (Warszawa *vs.* Żeromin)	−0.070 ± 0.039	−1.82	0.072
	Population (Łódź *vs.* Sarnów)	0.026 ± 0.034	0.74	0.46
	Population (Łódź *vs.* Żeromin)	−0.018 ± 0.038	−0.49	0.63
	Population (Sarnów *vs.* Żeromin)	−0.044 ± 0.038	−1.14	0.26

Notes:
(Warszawa–old urban population, Łódź–new urban population, Sarnów and Żeromin–rural populations). Year was included as a random factor.
Significant predictors are marked in bold.

MHC data suggests two genetic clusters

Our Bayesian assignment of individuals to genetic clusters also provided support for greater MHC differentiation between populations (when compared with microsatellites). Two genetic clusters were inferred for the MHC genes, as based on the mean posterior probabilities L(*K*), although Δ*K* were similar for *K* = 2 and *K* = 3 (Fig. 4). All individuals from the new urban and both rural populations were assigned to a single genetic cluster with very high assignment probabilities (0.933 ± 0.002; yellow cluster, Fig. 4). The old urban population clustered exclusively to itself, indicating differentiation from the remaining populations (blue cluster, Fig. 4), although average assignment probabilities to this cluster were only moderately high (0.738 ± 0.020). Two genetic clusters were also inferred for microsatellite loci, as indicated by both L(*K*) and Δ*K* values (Fig. 4). The pattern was highly contrasting with the MHC, as a similar proportion of individuals from each population was assigned to each cluster, providing no support for neutral genetic differentiation between populations (Fig. 4).

DISCUSSION

Our study investigated MHC class II gene divergence between Eurasian coots sampled in urban and non-urban environments, where different pathogen pressure may occur. We found significant differentiation of MHC genes between both urban populations and the populations inhabiting non-urban areas. We also demonstrated that the strongest differentiation of MHC genes occurred between the old urban population (Warszawa) and the three others-the new urban (Łódź) and two non-urban populations (Sarnów and

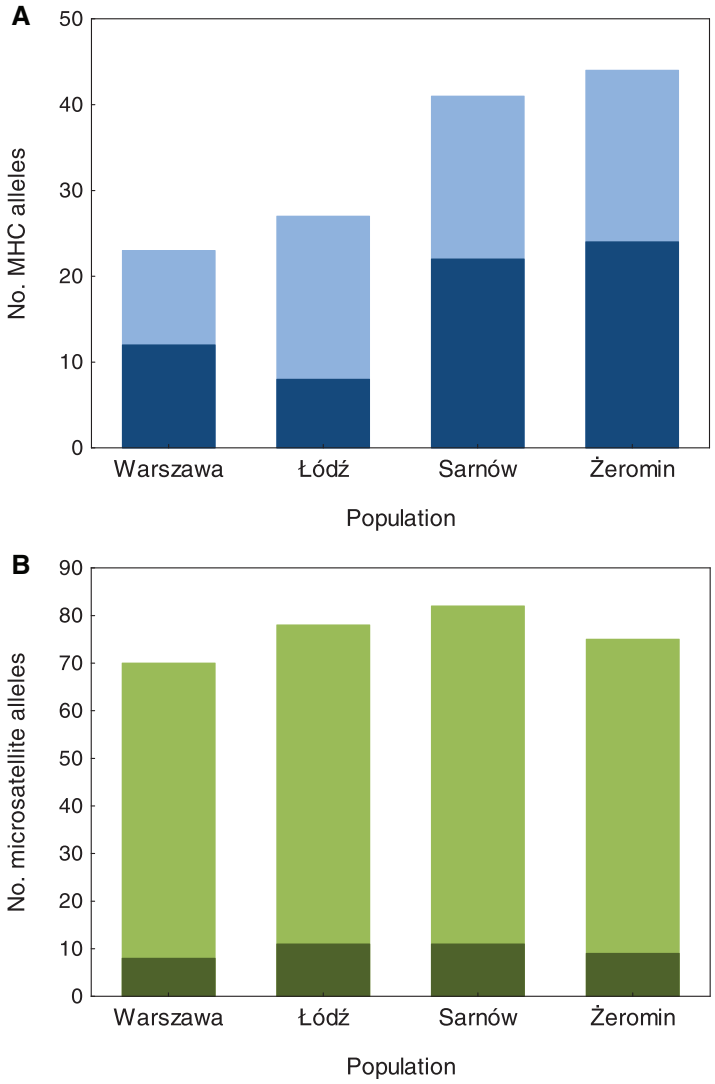


Figure 3 The number of MHC class II (A) and microsatellite (B) alleles recorded in old urban (Warszawa), new urban (Łódź) and two rural (Sarnów and Żeromin) populations of the Eurasian coot. All estimates are shown for a standardized population size of 20 individuals. Dark colours represent private (population-specific) alleles, while light colours represent alleles shared between populations. Full-size DOI: 10.7717/peerj.12264/fig-3

Żeromin). Finally, there was an apparent loss in MHC allele diversity in both urban populations when compared to rural ones. These findings were not consistent with the patterns of neutral genetic diversity and differentiation between the populations (across a panel of microsatellite loci), suggesting that processes shaping MHC variation were not primarily driven by demographic processes.

MHC genes are an important model for studying the adaptive genetic differentiation in wildlife populations, as co-evolution with pathogens is commonly referred to as the main mechanism responsible for the occurrence of MHC polymorphism (*Spurgin & Richardson, 2010*). Our results are suggestive of adaptive differentiation at the MHC in urban populations, supported by the lack of genetic differentiation at neutral loci. These

Table 2 Genetic differentiation between urban and rural Eurasian coot populations at the MHC and microsatellite loci, as measured with pairwise Jost's D (below diagonal) and G'_{ST} (above diagonal) values.

Marker	Population	Warszawa (old urban)	Łódź (new urban)	Sarnów (rural)	Żeromin (rural)
MHC class II	Warszawa	–	NA	NA	NA
	Łódź	**0.602**	–	NA	NA
	Sarnów	**0.588**	**0.277**	–	NA
	Żeromin	**0.623**	**0.413**	0.142	–
Microsatellites	Warszawa	–	0.019	0.009	0.025
	Łódź	0.012	–	0.020	0.020
	Sarnów	0.005	0.006	–	0.019
	Żeromin	0.017	0.018	0.005	–

Note:
Significant comparisons ($P < 0.05$) are indicated in bold.

results were inconsistent with our previous research, showing significant microsatellite differentiation of urban and non-urban coots in Poland (*Minias et al., 2017*). Yet, it is important to note that our earlier analyses were conducted on a larger number of populations using a directly paired (urban *vs.* adjacent non-urban) framework. This was possible due to the other type of material collected (*i.e.*, moulted feathers), which did not allow efficient MHC genotyping because of low quality DNA extracts. Here, we used blood sampled from captured individuals, which on one hand allowed us to obtain reliable MHC genotypes, but on the other hand it limited our sampling and reduced statistical power to detect inter-population differentiation. It must be acknowledged, however, that the results on microsatellite and MHC variation within this study were obtained under identical sampling scheme and, thus, can be reliably compared. Taking this into account, contrasting patterns for neutral genetic variation (a lack of differentiation) and adaptive immunogenetic variation (a significant differentiation at the MHC) suggest that the latter cannot be solely explained by stochastic mechanisms associated with genetic drift or demographic processes, which may take place during urban colonization.

The urbanization process may be associated with a random reduction in genetic diversity as a result of the bottleneck effect (*i.e.*, when a group of individuals colonizing novel areas is genetically not representative for the source population; *Evans et al., 2010*). Traces of such processes should be, however, detected both at the level of neutral and adaptive genetic markers, which was not confirmed in our study. Not only was there no significant differentiation in microsatellite loci found between our coot populations, but we also observed little differences in the diversity of these loci between the populations. The only pairwise difference in microsatellite heterozygosity that was close to the significance threshold indicated a tendency for a lower heterozygosity in the old urban populations when compared to a non-urban one. This may suggest that demographic processes, such as recent genetic bottleneck or founder effect, could have contributed to the patterns of genetic variation that we observed between our study coot population, but this contribution was rather low and unlikely to explain strong differentiation at the MHC genes. Alternatively, MHC alleles may show lower average frequencies within

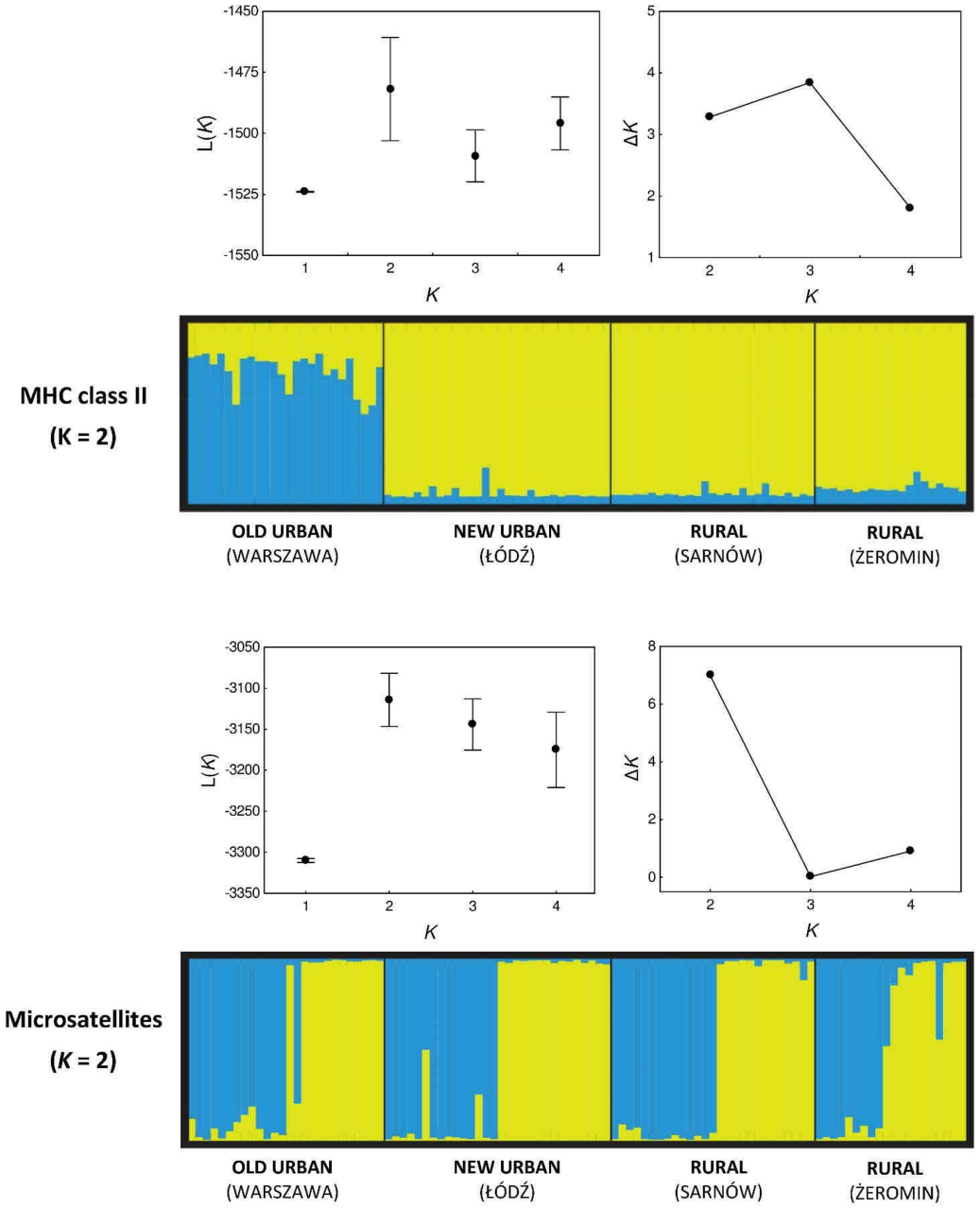


Figure 4 Bayesian assignment of Eurasian coots from two urban and two rural populations to genetic clusters. Each bar represents the estimated proportional posterior probability of each individual belonging to each cluster, as marked with different colours. Genetic clusters were inferred separately for MHC class II genes and ten microsatellite markers. Mean (±SD) posterior probability L(K) and Δ*K* values for different number of clusters (*K*) are shown above each assignment plot.

Full-size DOI: 10.7717/peerj.12264/fig-4

populations when compared to microsatellites and, for this reason, they may be more prone to loss during urban colonization processes, which can promote stronger differentiation at the MHC between urban and non-urban populations. However, it was

unfeasible to calculate true frequencies of MHC alleles without locus-specific data and, thus, we were unable to directly test this hypothesis.

Our results suggest that MHC differentiation between populations could stem from the processes acting specifically on these genes. The processes of urban colonization by wildlife may require adaptations of key immune receptors to the composition of local pathogen and parasite faunas. Congruent with this hypothesis, our analyses showed differences in the MHC allele composition between populations that differed in the level of urbanization. There are several non-exclusive mechanisms which may explain this pattern, although we could not effectively distinguish between these alternative scenarios with our data. First, differences in the allelic composition of MHC between the new urban population (Łódź) and the non-urban populations (putative source populations for the colonization of Łódź; *Minias et al., 2017*) may suggest that the influx on non-urban individuals into cities could have proceeded through the sorting of MHC genotypes. This mechanism assumes that only individuals with specific MHC genotypes, pre-adapted to the urban pathogen fauna, can effectively settle, survive and raise offspring in the urban environment. Similar mechanisms are observed at the phenotype level, where birds with particular behavioural features (*e.g.*, more aggressive, bolder, more explorative or more resistant to stress) are more likely to colonize novel habitats, including urban areas (*Luniak, 2004*; *Evans et al., 2010*; *Minias, Jedlikowski & Włodarczyk, 2018*). However, it must be stressed that our study did not directly test this hypothesis and further research into the potential for genotype sorting is required.

Alternatively, differentiation at the MHC between the old and new urban populations may suggest the occurrence of local adaptations, where MHC repertoire becomes gradually adapted to the local pathogen pressure (*Lan et al., 2019*). In accordance with our results, the strength of the local MHC adaptations in urban populations should be correlated with the time of the exposure of a given population to the urban pathogen fauna (under similar strength of pathogen-driven selection in both populations), and, therefore, indirectly with the time at which a given urban population was established. At the same time, the old urban population showed the lowest individual MHC variability (*i.e.*, the average number of alleles per individual). This may suggest an adaptation to a specific (perhaps relatively homogenous) assemblage of urban pathogens and it is consistent with empirical studies showing altered composition of parasite faunas in an urbanized landscape and a negative association of parasite richness with urbanization level (*Calegaro-Marques & Amato, 2014*; *Łoś et al., 2020*). We, however, explicitly acknowledge that no pathogen data were collected in our study, and we cannot draw any firm conclusions with this respect. Also, we cannot rule out that this effect can simply have arisen due to different strength of pathogen-driven selection in each environment and may not be directly associated with the timing of population establishment.

Our capability to infer the exact mechanisms responsible for the observed MHC differentiation was further impaired by some methodological limitations. First, we had no replicates for old and new urban populations and, thus, our analyses yielded no statistical power to directly test for the effect of population history (*i.e.*, timing of urban colonization) on MHC differentiation. Also, our populations were not evenly distributed

in space, as the old urban population was located relatively far from the remaining three (new urban and two rural) populations. This may have suggested that any genetic differentiation of the old urban population could have arisen due to isolation by distance, although this scenario has not been confirmed by the analysis of microsatellite data (*i.e.*, the lack of significant differentiation between populations).

In general, maintenance of local adaptations may be facilitated by limited gene flow between populations. For example, research on burrowing owls (*Athene cunicularia*) has shown that there is limited gene flow between adjacent urban and rural populations (*i.e.*, enhancing local adaptation), but not between different urban populations (*Mueller et al., 2018*). Basic factors that can reduce gene flow between urban and non-urban populations include marked differences in habitat structure and intensity of anthropogenic pressure, but also differences in local climate conditions or predatory pressure (*e.g.*, in the composition of predator faunas; *Edelaar, Siepielski & Clobert, 2008*). For instance, the shift from high genotypic variation in rural landscapes to strongly favoured genotypes in urban settings was observed in wood frogs (*Lithobates sylvaticus*), as a results of strong habitat differences which diminished gene flow and amplified genetic drift (*Homola et al., 2019*). This habitat variation is often reflected by high behavioural divergence between urban and non-urban individuals (*Shochat, 2004*), which may further reduce migration rate of individuals (genes) across landscapes. Indeed, behavioural studies conducted in our study coot populations revealed different spectra of behaviours between the urban and non-urban birds. It was demonstrated that coots from the old urban population (Warszawa) were more aggressive while defending the nest, less sensitive to stress and the presence of humans, and used food of anthropogenic origin better than individuals from the new urban population (Łódź) and the non-urban populations (*Minias, Jedlikowski & Włodarczyk, 2018*). It seems that this kind of behavioural differentiation may constitute one of the mechanisms promoting the maintenance of local genetic adaptations (also at the level of MHC) in the urban areas. On the other hand, an increasing number of studies show that local adaptations can be maintained despite high gene flow (reviewed in *Tigano & Friesen, 2016*). Preservation of local adaptations under gene flow is facilitated by several genomic processes (*e.g.*, rearrangements of genomic architecture, linkage with already diverged loci, or mechanisms suppressing recombination) that maintain clusters of adaptive loci (*Tigano & Friesen, 2016*). Since our analyses of neutral (microsatellite) genetic variation indicated no apparent barriers for gene flow between our coot populations, it seems possible that some of these mechanisms could contribute to the maintenance of local adaptations at the MHC genes. So far, evidence of local adaptation to urbanization in the face of gene flow was found in a highly mobile insect pollinator, the red-tailed bumblebee (*Bombus lapidarius*; *Theodorou et al., 2018*).

In conclusion, our study showed significant between-population differentiation of MHC repertoire depending on the level of urbanization, likely as a result of non-neutral processes (genotype sorting or local adaptation). Currently, there is a scarcity of studies testing for immunogenetic adaptations to urban habitats in birds and our results provide an important contribution to this topic, as well as advancing our more general understanding of genetic mechanisms associated with urbanization processes in wildlife.

Although our results revealed associations of MHC allele composition with habitat urbanization and the time which passed since urban colonization, we are aware that our conclusions based on a limited number of populations should be confirmed using longitudinal data collected at broader geographical scale and across a larger number of urban and non-urban population replicates. We also acknowledge that future research should complement our findings with pathogen data from different habitats and, ideally, should examine fitness consequences associated with different MHC composition in urban and non-urban environments. Nevertheless, we believe that our study provides important foundational information that future work can build upon, not only in coots, but also pertaining to other urban colonizing wildlife. As our knowledge of how species are evolutionarily responding to increasing rates of urbanization grows, knowing the role immunogenetic adaptations play will provide key insights into urbanized taxa and what it takes to survive in a human world.

ACKNOWLEDGEMENTS

We are grateful to the Botanical Garden in Łódź for permission to carry out fieldwork. We thank two anonymous reviewers for constructive comments on the earlier draft of the manuscript.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Ewa Pikus conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Radosław Włodarczyk performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Jan Jedlikowski performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Piotr Minias conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

Animal Ethics

The following information was supplied relating to ethical approvals (*i.e.*, approving body and any reference numbers):

All methods were carried out in accordance with the current laws of Poland (Act on Nature Conservation from 16 April 2004, Journal of Laws from 2004, No. 92, item 880)

and were approved by the Local Bioethical Commission for Experiments on Animals in Łodź (nos 40/ŁB 620/2012 and 15/ŁB/2016).

Field Study Permissions

The following information was supplied relating to field study approvals (*i.e.*, approving body and any reference numbers):

Field experiments were approved by the Local Bioethical Commission for Experiments on Animals in Łodź (nos 40/ŁB 620/2012 and 15/ŁB/2016). Access to non-urban sites was granted by private land owners (Agricultural and Fishing Farms at Sarnów and Żeromin).

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

All novel MHC sequences are available in GenBank: MT74863–MT748786. All other alleles were previously described by *Alcaide et al. (2014)*.

Data Availability

The following information was supplied regarding data availability:

The raw genetic data (MHC and microsatellites) are available in the Supplementary File.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.12264#supplemental-information.

REFERENCES

Adams LW, Van Druff LW, Luniak M. 2005. Managing urban habitats and wildlife. In: Braun CE, ed. *Techniques for Wildlife Investigations and Management*. Sixth Edition. Bethesda, Maryland, USA: The Wildlife Society, 714–739.

Alcaide M. 2010. On the relative roles of selection and genetic drift in shaping MHC variation. *Molecular Ecology* **18(19)**:3842–3844 DOI 10.1111/j.1365-294X.2010.04772.x.

Alcaide M, Edwards SV, Negro JJ, Serrano D, Tella JL. 2008. Extensive polymorphism and geographical variation at a positively selected MHC class II B gene of the lesser kestrel (*Falco naumanni*). *Molecular Ecology* **17(11)**:2652–2665 DOI 10.1111/j.1365-294X.2008.03791.x.

Alcaide M, Munoz J, Martínez-de la Puente J, Soriguer R, Figuerola J. 2014. Extraordinary MHC class II B diversity in a non-passerine, wild bird: the Eurasian Coot *Fulica atra* (Aves: Rallidae). *Ecology and Evolution* **4(6)**:688–698 DOI 10.1002/ece3.974.

Aouissi HA, Gasparini J, Belabed AI, Bouslama Z. 2017. Impact of greenspaces in city on avian species richness and abundance in Northern Africa. *Comptes Rendus Biologies* **340(8)**:394–400 DOI 10.1016/j.crvi.2017.07.002.

Audet JN, Ducatez S, Lefebvre L. 2016. The town bird and the country bird: problem solving and immunocompetence vary with urbanization. *Behavioral Ecology* **27(2)**:637–644 DOI 10.1093/beheco/arv201.

Barton GM, Medzhitov R. 2020. Control of adaptive immune responses by toll-like receptors. *Current Opinion in Immunology* **14(3)**:380–383 DOI 10.1016/S0952-7915(02)00343-6.

Biedrzycka A, O'Connor E, Sebastian A, Migalska M, Radwan J, Zając T, Bielański W, Solarz W, Westerdahl H. 2017. Extreme MHC class I diversity in the sedge warbler (*Acrocephalus schoenobaenus*); selection patterns and allelic divergence suggest that different

genes have different functions. *BMC Evolutionary Biology* **17(1)**:159 DOI 10.1186/s12862-017-0997-9.

Borghans JA, Beltman JB, De Boer RJ. 2004. MHC polymorphism under host-pathogen coevolution. *Immunogenetics* **55(11)**:732–739 DOI 10.1007/s00251-003-0630-5.

Brackett CL, Maley JM, Brumfield RT, McRae SB. 2013. Characterization of microsatellite loci for a threatened species, the king rail, *Rallus elegans*, using a next-generation sequencing protocol. *Conservation Genetics Resources* **5(4)**:1189–1191 DOI 10.1007%2Fs12686-013-9999-0.

Bradley CA, Altizer S. 2007. Urbanization and the ecology of wildlife diseases. *Trends in Ecology & Evolution* **22(2)**:95–102 DOI 10.1016/j.tree.2006.11.001.

Brearley G, Rhodes J, Bradley A, Baxter G, Seabrook L, Lunney D, Liu Y, McAlpine C. 2013. Wildlife disease prevalence in human-modified landscapes. *Biological Reviews* **88(2)**:427–442 DOI 10.1111/brv.12009.

Buchan JC. 2000. Behavioural and genetic aspects of mate-sharing in the Tasmanian native hen (*Gallinula mortierii*) and dusky moorhen (*Gallinula tenebrosa*). PhD thesis, University of Queensland, Brisbane, Australia.

Calegaro-Marques C, Amato SB. 2014. Urbanization breaks up host-parasite interactions: a case study on parasite community ecology of rufous-bellied thrushes (*Turdus rufiventris*) along a rural-urban gradient. *PLOS ONE* **9(7)**:e103144 DOI 10.1371/journal.pone.0103144.

Camacho-Valdez V, Tello-Alcaide EM, Wootton A, Valencia-Barrera E. 2009. Land use change and urban ecosystem services: a case study of urban wetlands in a rapidly sprawling city in the highlands of Chiapas. *Mexico Journal of Management and Sustainability* **9(1)**:67–81 DOI 10.5539/jms.v9n1p67.

Chao A, Ma KH, Hsieh TC, Chiu CH. 2016. SpadeR: species-richness prediction and diversity estimation in R. R package ver. 0.1.1. *Available at https://CRAN.R-project.org/package=SpadeR*.

Concepción ED, Moretti M, Altermatt F, Nobis MP, Obrist MK. 2015. Impacts of urbanisation on biodiversity: the role of species mobility, degree of specialisation and spatial scale. *Oikos* **124(12)**:1571–1582 DOI 10.1111/oik.02166.

Croci S, Butet A, Clergeau P. 2008. Does urbanization filter birds on the basis of their biological traits. *Condor* **110(2)**:223–240 DOI 10.1525/cond.2008.8409.

Davis AY, Malas N, Minor ES. 2014. Substitutable habitats? The biophysical and anthropogenic drivers of an exotic bird's distribution. *Biological Invasions* **16(2)**:415–427 DOI 10.1007/s10530-013-0530-z.

Dawson DA, Horsburgh GJ, Küpper C, Stewart IR, Ball AD, Durrant KL, Hansson B, Bacon I, Bird S, Klein A. 2010. New methods to identify conserved microsatellite loci and develop primer sets of high cross-species utility-as demonstrated for birds. *Molecular Ecology Resources* **10(3)**:475–494 DOI 10.1111/j.1755-0998.2009.02775.x.

DeCandia AL, Brzeski KE, Heppenheimer E, Caro CV, Camenisch G, Wandeler P, Driscoll C, vonHoldt BM. 2019a. Urban colonization through multiple genetic lenses: the city-fox phenomenon revisited. *Ecology and Evolution* **9(4)**:2046–2060 DOI 10.1002/ece3.4898.

DeCandia AL, Henger CS, Krause A, Gormezano LJ, Weckel M, Nagy C, Munshi-South J, Vonholdt BM. 2019b. Genetics of urban colonization: neutral and adaptive variation in coyotes (*Canis latrans*) inhabiting the New York metropolitan area. *Journal of Urban Ecology* **5(1)**:3490 DOI 10.1093/jue/juz002.

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(9)**:2154–2164 DOI 10.1111/j.1558-5646.2007.00178.x.

Earl DA, von Holdt BM. 2012. Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. *Conservation Genetics Resources* **4(2)**:359–361 DOI 10.1007/s12686-011-9548-7.

Edelaar P, Siepielski AM, Clobert J. 2008. Matching habitat choice causes directed gene flow: a neglected dimension in evolution and ecology. *Evolution* **62(10)**:2462–2472 DOI 10.1111/j.1558-5646.2008.00459.x.

Ekblom R, Saether SA, Jacobsson P, Fiske P, Sahlman T, Grahn M, Kålås JA, Höglund J. 2007. Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*). *Molecular Ecology* **16(7)**:1439–1451 DOI 10.1111/j.1365-294X.2007.03281.x.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software Structure: a simulation study. *Molecular Ecology* **14(8)**:2611–2620 DOI 10.1111/j.1365-294X.2005.02553.x.

Evans KL, Hatchwell BJ, Parnell M, Gaston KJA. 2010. A conceptual framework for the colonisation of urban areas: the blackbird *Turdus merula* as a case study. *Biological Reviews* **85(3)**:643–667 DOI 10.1111/j.1469-185X.2010.00121.x.

Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10(3)**:564–567 DOI 10.1111/j.1755-0998.2010.02847.x.

Falush D, Stephens M, Pritchard JK. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7(4)**:574–578 DOI 10.1111/j.1471-8286.2007.01758.x.

Frenkel JK, Hassanein KM, Hassanein RS, Brown E, Thulliez P, Quintero-Nunez R. 1995. Transmission of *Toxoplasma gondii* in Panama City, Panama: a 5-year prospective cohort study of children, cats, rodents, birds, and soil. *American Journal of Tropical Medicine and Hygiene* **53(5)**:458–468 DOI 10.4269/ajtmh.1995.53.458.

Goudet J. 1995. FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity* **86(6)**:485–486 DOI 10.1093/oxfordjournals.jhered.a111627.

Green TL, Kronenberg J, Andersson E, Elmqvist T, Gomez-Baggethun E. 2016. Insurance value of green infrastructure in and around cities. *Ecosystems* **19(6)**:1051–1063 DOI 10.1007/s10021-016-9986-x.

Guo SW, Thompson EA. 1992. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* **48(2)**:361–372 DOI 10.2307/2532296.

Hedrick PW. 2005. A standardized genetic differentiation measure. *Evolution* **59(8)**:1633–1638 DOI 10.1111/j.0014-3820.2005.tb01814.x.

Herdegen M, Babik W, Radwan J. 2014. Selective pressures on MHC class II genes in the guppy (*Poecilia reticulata*) as inferred by hierarchical analysis of population structure. *Journal of Evolutionary Biology* **27(11)**:2347–2359 DOI 10.1111/jeb.12476.

Hess CM, Edwards SV. 2002. The evolution of the major histocompatibility complex in birds. *BioScience* **52(5)**:423–431 DOI 10.1641/0006-3568(2002)052[0423: TEOTMH]2.0.CO;2.

Homola JJ, Loftin CS, Cammen KM, Helbing CC, Birol I, Schultz TF, Kinnison MT. 2019. Replicated landscape genomics identifies evidence of local adaptation to urbanization in wood frogs. *Journal of Heredity* **110(6)**:707–719 DOI 10.1093/jhered/esz041.

Hughes AL, Yeager M. 1998. Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics* **32(1)**:415–435 DOI 10.1146/annurev.genet.32.1.415.

Janes JK, Miller JM, Dupuis JR, Malenfant RM, Gorrell JC, Cullingham CI, Andrew RL. 2017. The $K = 2$ conundrum. *Molecular Ecology* **26(14)**:3594–3602 DOI 10.1111/mec.14187.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23(14)**:1801–1806 DOI 10.1093/bioinformatics/btm233.

Janiszewski T, Wojciechowski Z, Markowski J. 2009. *Atlas ptaków lęgowych Łodzi*. Łódź, Poland: Łódź Univ. Press. [In Polish with English summary].

Jost L. 2008. G_{ST} and its relatives do not measure differentiation. *Molecular Ecology* **17(18)**:4015–4026 DOI 10.1111/j.1365-294X.2008.03887.x.

Kalinowski ST, Taper ML, Marshall TC. 2007. Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* **16(5)**:1099–1106 DOI 10.1111/j.1365-294X.2007.03089.x.

Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA. 2013. diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution* **4(8)**:782–788 DOI 10.1111/2041-210X.12067.

Lan H, Zhou T, Wan QH, Fang SG. 2019. Genetic diversity and differentiation at structurally varying MHC haplotypes and microsatellites in bottlenecked populations of endangered Crested Ibis. *Cells* **8(4)**:377 DOI 10.3390/cells8040377.

Loiseau C, Richard M, Garnier S, Chastel O, Julliard R, Zoorob R, Sorci G. 2009. Diversifying selection on MHC class I in the house sparrow (*Passer domesticus*). *Molecular Ecology* **18(7)**:1331–1340 DOI 10.1111/j.1365-294X.2009.04105.x.

Lourenço A, Álvarez D, Wang IJ, Velo-Antón G. 2017. Trapped within the city: integrating demography, time since isolation and population-specific traits to assess the genetic effects of urbanization. *Molecular Ecology* **26(6)**:1498–1514 DOI 10.1111/mec.14019.

Luniak M. 2004. Synurbization–adaptation of animal wildlife to urban development. In: *Proceedings 4th International Urban Wildlife Symposium Tucson*. 50–55.

Luniak M, Kalbarczyk W, Pawłowski W. 1964. Birds of Warsaw. *Acta Ornithologica* **8**:198–199.

Luniak M, Kozłowski P, Nowicki W, Plit J. 2001. *Birds of Warsaw (1962–2000)*. Warszawa, Poland: Polish Academy of Sciences. In Polish with English summary.

Łoś A, Skórka P, Strachecka A, Winiarczyk S, Adaszek Ł, Winiarczyk M, Wolski D. 2020. The associations among the breeding performance of Osmia bicornis L. (Hymenoptera: Megachilidae), burden of pathogens and nest parasites along urbanisation gradient. *Science of the Total Environment* **710(3)**:135520 DOI 10.1016/j.scitotenv.2019.135520.

Magoc T, Salzberg SL. 2011. Flash: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27(21)**:2957–2963 DOI 10.1093/bioinformatics/btr507.

McDonald RI, Kareiva P, Forman RT. 2008. The implications of current and future urbanization for global protected areas and biodiversity conservation. *Biological Conservation* **141(6)**:1695–1703 DOI 10.1016/j.biocon.2008.04.025.

Meirmans PG, Hedrick PW. 2011. Assessing population structure: FST and related measures. *Molecular Ecology Resources* **11(1)**:5–18 DOI 10.1111/j.1755-0998.2010.02927.x.

Miller HC, Allendorf F, Daugherty CH. 2010. Genetic diversity and differentiation at MHC genes in island populations of tuatara (*Sphenodon* spp.). *Molecular Ecology* **19(18)**:3894–3908 DOI 10.1111/j.1365-294X.2010.04771.x.

Miller MA, Gardner IA, Kreuder C, Paradies DM, Worcester KR, Jessup DA, Dodd E, Harris MD, Ames JA, Packham AE, Conrad PA. 2002. Coastal freshwater runoff is a risk factor for *Toxoplasma gondii* infection of southern sea otters (*Enhydra lutris nereis*). *International Journal for Parasitology* **32(8)**:997–1006 DOI 10.1016/S0020-7519(02)00069-3.

Minias P, Janiszewski T. 2016. Territory selection in the city: can birds reliably judge territory quality in a novel urban environment? *Journal of Zoology* **300(2)**:120–126 DOI 10.1111/jzo.12362.

Minias P, Jedlikowski J, Włodarczyk R. 2018. Development of urban behaviour is associated with time since urbanization in a reed-nesting waterbird. *Urban Ecosystems* **21(6)**:1021–1028 DOI 10.1007/s11252-018-0781-7.

Minias P, Włodarczyk R, Minias A, Dziadek J. 2017. How birds colonize cities: genetic evidence from a common waterbird, the Eurasian coot. *Journal of Avian Biology* **48(8)**:1095–1103 DOI 10.1111/jav.01334.

Mueller JC, Kuhl H, Boerno S, Tella JL, Carrete M, Kempenaers B. 2018. Evolution of genomic variation in the burrowing owl in response to recent colonization of urban areas. *Proceedings of the Royal Society B: Biological Sciences* **285(1878)**:20180206 DOI 10.1098/rspb.2018.0206.

Münz C. 2012. Antigen processing for MHC class II presentation via autophagy. *Frontiers in Immunology* **3**:9 DOI 10.3389/fimmu.2012.00009.

Peakall ROD, Smouse PE. 2006. Genetic analysis in excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6(1)**:288–295 DOI 10.1111/j.1471-8286.2005.01155.x.

Piertney SB, Oliver MK. 2006. The evolutionary ecology of the major histocompatibility complex. *Heredity* **96(1)**:7–21 DOI 10.1038/sj.hdy.6800724.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155(2)**:945–959 DOI 10.1093/genetics/155.2.945.

Rosenberg NA. 2004. Distruct, a program for the graphical display of population structure. *Molecular Ecology Notes* **4(1)**:37–138 DOI 10.1046/j.1471-8286.2003.00566.x.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* **34(12)**:3299–3302 DOI 10.1093/molbev/msx248.

Schenten D, Medzhitov R. 2011. The control of adaptive immune responses by the innate immune system. *Advances in Immunology* **109(11)**:87–124 DOI 10.1016/B978-0-12-387664-5.00003-0.

Sebastian A, Herdegen M, Migalska M, Radwan J. 2016. Amplisas: a web server for multilocus genotyping using next-generation amplicon sequencing data. *Molecular Ecology Resources* **16(2)**:498–510 DOI 10.1111/1755-0998.12453.

Serieys LE, Lea A, Pollinger JP, Riley SP, Wayne RK. 2015. Disease and freeways drive genetic change in urban bobcat populations. *Evolutionary Applications* **8(1)**:75–92.

Shochat E. 2004. Credit or debit? Resource input changes population dynamics of city-slicker birds. *Oikos* **106(3)**:622–626 DOI 10.1111/j.0030-1299.2004.13159.x.

Skaug H, Fournier D, Nielsen A, Magnusson A, Bolker B. 2012. GlmmADMB: generalized linear mixed models using AD model builder. R package ver. 0.7.2.12. *Available at https://github.com/bbolker/glmmadmb*.

Smouse RPP, Peakall R. 2012. GenAlEx 6.5: genetic analysis in excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28(19)**:2537–2539 DOI 10.1093/bioinformatics/bts460.

Spautz H, Nur N, Stralberg D. 2005. California black rail (*Laterallus jamaicensis coturniculus*) distribution and abundance in relation to habitat and landscape features in the San Francisco Bay Estuary. USDA Forest Service General Technical Report PSW-GTR-191, Washington, D.C.: USDA Forest Service, 465–468.

Spurgin LG, Richardson DS. 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences* **277(1684)**:979–988 DOI 10.1098/rspb.2009.2084.

Strand TM, Segelbacher G, Quintela M, Xiao L, Axelsson T, Höglund J. 2012. Can balancing selection on MHC loci counteract genetic drift in small fragmented populations of black grouse? *Ecology and Evolution* **2(2)**:341–353 DOI 10.1002/ece3.86.

Sutton JT, Nakagawa S, Robertson BC, Jamieson IG. 2011. Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immunity genes. *Molecular Ecology* **20(21)**:4408–4420 DOI 10.1111/j.1365-294X.2011.05292.x.

Theodorou P, Radzevičiūtė R, Kahnt B, Soro A, Grosse I, Paxton RJ. 2018. Genome-wide single nucleotide polymorphism scan suggests adaptation to urbanization in an important pollinator, the red-tailed bumblebee (*Bombus lapidarius* L.). *Proceedings of the Royal Society B: Biological Sciences* **285(1877)**:20172806 DOI 10.1098/rspb.2017.2806.

Tigano A, Friesen VL. 2016. Genomics of local adaptation with gene flow. *Molecular Ecology* **25(10)**:2144–2164 DOI 10.1111/mec.13606.

Van Oosterhout C, Hutchinson WF, Wills DP, Shipley P. 2004. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4(3)**:535–538 DOI 10.1111/j.1471-8286.2004.00684.x.

Watson H, Videvall E, Andersson MN, Isaksson C. 2017. Transcriptome analysis of a wild bird reveals physiological responses to the urban environment. *Scientific Reports* **7(1)**:44180 DOI 10.1038/srep44180.

Whittaker DJ, Dapper AL, Peterson MP, Atwell JW, Ketterson ED. 2012. Maintenance of MHC class IIB diversity in a recently established songbird population. *Journal of Avian Biology* **43(2)**:109–118 DOI 10.1111/j.1600-048X.2012.05504.x.

Supplementary File 1

Urbanization processes drive divergence at the Major Histocompatibility Complex in a common waterbird

Ewa Pikus[1], Radosław Włodarczyk[1], Jan Jedlikowski[2], Piotr Minias[1]

[1] Department of Biodiversity Studies and Bioeducation, Faculty of Biology and Environmental Protection, University of Łódź, Banacha 1/3, 90-237 Łódź, Poland

[2] Faculty of Biology, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland

Corresponding author:

Piotr Minias, pminias@op.pl

Table S1 Microsatellite loci used to assess neutral variation in the Eurasian coot. Repeat unit size, size range, number of alleles, observed heterozygosity (H_o) and expected heterozygosity (H_e) are reported for each marker.

Locus	Forward primer sequences (5'–3')	Reverse primer sequences (5'–3')	Repeat unit size (bp)	Size range (bp)	N alleles	H_o	H_e
Tm18	CTGAACATTTCAAGAGTCCTCA	GCAGAGCTGACCACTCAGAAC	2	137-149	3	0.703	0.539
Tm20	TGCACAGATGAGAAGGAGGGC	GCCTGCTCACACTAGCAGAAC	2	127-133	3	0.090	0.085
Tm27	CTAGGTCACCTTTACAGTAGTC	CAATACCATGTATAGAAATGTTGG	2	131-173	17	0.836	0.846
Tm38	GATAGATGGAGCGTAAAATGCA	AAGAAAGCAAAATGTCTGTATTGG	2	103-113	6	0.504	0.511
Kira8	GCTTTGCATCTTGCCTTAAA	GTGACACTGATACAGTGTGCCT	2	307-331	16	0.879	0.901
Kira9	TGATCTGGGCAGGCTTCTAC	GTCGAATAATGGCAGCAATG	2	163-173	5	0.520	0.499
Kira10	CCAAGTACCATCTGCGAAGC	AACCCGAACGAGAGATGTGA	2	121-145	12	0.785	0.774
Kira16	CCAGGTGAAACTCTGCATT	ACAGTTGTGATGTGGCTGGA	2	299-419	48	0.858	0.930
B106	TAGTGCTCTCAGGAAAGACTTG	CTCTTCCAGAAGCTGTAGTTG	2	168-204	15	0.777	0.788
TG03002	TCTTGCCTTTTTGGTATGAGTAT	TACAAAGCACTGTGGAGCAG	2	118-126	5	0.265	0.257

Table S2 Measures of MHC class II diversity in old urban (Warszawa), new urban (Łódź) and two rural (Sarnów and Żeromin) populations of the Eurasian coot. Estimates given in parentheses were standardized for a population size of 20 individuals using rarefaction.

Population	No. individuals	No. alleles	No. private alleles	No. segregating sites	No. nucleotide differences	Nucleotide diversity
Warszawa	26 (20)	27 (23)	14 (12)	93 (90)	32.4 (32.5)	0.121 (0.122)
Łódź	30 (20)	37 (27)	14 (8)	93 (90)	31.2 (32.0)	0.117 (0.119)
Sarnów	27 (20)	52 (41)	30 (22)	96 (93)	31.3 (32.5)	0.117 (0.122)
Żeromin	20 (20)	44 (44)	24 (24)	92 (92)	31.4 (31.4)	0.118 (0.118)

Table S3 Measures of microsatellite diversity in old urban (Warszawa), new urban (Łódź) and two rural (Sarnów and Żeromin) populations of the Eurasian coot. Estimates given in parentheses were standardized for a population size of 20 individuals using rarefaction.

Population	No. individuals	No. alleles	No. private alleles	Observed heterozygosity (H_o)
Warszawa	26 (20)	76 (70)	8 (8)	0.585 (0.580)
Łódź	30 (20)	88 (78)	13 (11)	0.637 (0.655)
Sarnów	27 (20)	88 (82)	13 (11)	0.611 (0.635)
Żeromin	20 (20)	75 (75)	7 (9)	0.655 (0.655)

9 DOROBEK NAUKOWY

Pozostałe publikacje w czasopismach z listy *Journal Citations Reports*:

Minias, P., Janiszewska, A., Pikus, E., Zadworny, T., & Anderwald, D. (2021). MHC reflects fine-scale habitat structure in white-tailed eagles *Haliaeetus albicilla*. *Journal of Heredity*, *112*(4), 335-345.

Rybaczek, D., Musiałek, M. W., Vrána, J., Petrovská, B., Pikus, E. G., & Doležel, J. (2021). Kinetics of DNA repair in Vicia faba meristem regeneration following replication stress. *Cells*, *10*(1), 88.

Minias, P., Pikus, E., & Anderwald, D. (2019). Allelic diversity and selection at the MHC class I and class II in a bottlenecked bird of prey, the White-tailed Eagle. *BMC Evolutionary Biology*, *19*(1), 2.

Minias, P., Pikus, E., Whittingham, L. A., & Dunn, P. O. (2019). Evolution of copy number at the MHC varies across the avian tree of life. *Genome Biology & Evolution*, *11*(1), 17-28.

Minias, P., Pikus, E., Whittingham, L. A., & Dunn, P. O. (2018). A global analysis of selection at the avian MHC. *Evolution*, *72*(6), 1278-1293.

Wiśnik, E., Pikus, E., Duchnowicz, P., & Koter-Michalak, M. (2017). Tolerance of monocytes and macrophages in response to bacterial endotoxin. *Postępy Higieny i Medycyny Doświadczalnej*, *71*, 176-185.

10 OŚWIADCZENIA WSPÓŁAUTORÓW

Łódź, 30.11.2021

mgr inż. Ewa Pikus

Katedra Badania Różnorodności Biologicznej,
Dydaktyki i Bioedukacji
Uniwersytet Łódzki

OŚWIADCZENIE O WKŁADZIE W POWSTANIE PRACY

1. Pikus, E., Minias, P. Using de novo genome assembly and high-throughput sequencing to characterize the Major Histocompatibility Complex in a non-model rallid bird, the Eurasian coot *Fulica atra*. Scientific Reports (praca zgłoszona do druku).

Niniejszym oświadczam, że jestem wiodącym autorem powyższej pracy, a mój wkład w jej powstanie polegał na udziale w projektowaniu hipotez badawczych, udziale w badaniach terenowych, udziale w projektowaniu eksperymentów, wykonaniu prac laboratoryjnych, zebraniu danych literaturowych, udziale w analizowaniu danych oraz przygotowaniu rycin i tabel, pisaniu i redagowaniu treści manuskryptu oraz powiązanych z nim treści dodatkowych.

Swój udział ilościowy w powstanie pracy oceniam na 60%.

2. Pikus, E., Dunn, P., Minias, P. High MHC diversity does not confer fitness advantage in a wild bird. Journal of Animal Ecology (praca zgłoszona do druku).

Niniejszym oświadczam, że jestem wiodącym autorem powyższej pracy, a mój wkład w jej powstanie polegał na udziale w projektowaniu hipotez badawczych, udziale w badaniach terenowych, udziale w projektowaniu eksperymentów, wykonaniu prac laboratoryjnych - ustalenie metodyki badawczej oraz przeprowadzenie eksperymentów, udziale w analizowaniu danych oraz przygotowaniu rycin i tabel, zebraniu danych literaturowych, pisaniu i redagowaniu treści manuskryptu oraz powiązanych z nim materiałów dodatkowych.

Swój udział ilościowy w powstanie pracy oceniam na 50%.

3. Pikus, E., Włodarczyk, R., Jedlikowski, J. and Minias, P., 2021. Urbanization processes drive divergence at the major histocompatibility complex in a common waterbird. PeerJ, *9*, p.e12264.

Niniejszym oświadczam, że jestem wiodącym autorem powyższej pracy, a mój wkład w jej powstanie polegał na udziale w projektowaniu hipotez badawczych, udziale w badaniach terenowych, udziale w projektowaniu eksperymentów, wykonaniu prac laboratoryjnych - ustaleniu metodyki badawczej oraz przeprowadzeniu eksperymentów, udziale w analizowaniu danych oraz przygotowaniu rycin i tabel, zebraniu danych literaturowych, pisaniu oraz redagowaniu treści manuskryptu.

Swój udział ilościowy w powstanie pracy oceniam na 50%.

Ewa Pikus

Podpis

Łódź, 30.11.2021

dr hab. Piotr Minias, prof. UŁ

Katedra Badania Różnorodności Biologicznej,
Dydaktyki i Bioedukacji
Uniwersytet Łódzki

OŚWIADCZENIE O WKŁADZIE W POWSTANIE PRACY

1. Pikus, E., Minias, P. Using de novo genome assembly and high-throughput sequencing to characterize the Major Histocompatibility Complex in a non-model rallid bird, the Eurasian coot *Fulica atra*. Scientific Reports (praca zgłoszona do druku).

Niniejszym oświadczam, że jestem współautorem powyższej pracy, a mój wkład w jej powstanie polegał na udziale w projektowaniu hipotez badawczych, udziale w badaniach terenowych, udziale w projektowaniu eksperymentów, udziale w analizowaniu danych oraz przygotowaniu rycin i tabel, oraz redagowaniu treści manuskryptu.

Swój wkład ilościowy w powstanie pracy oceniam na 40%.

Jako promotor potwierdzam, że mgr inż. Ewa Pikus jest wiodącym autorem niniejszej pracy (60% udziału ilościowego), a jej wkład polegał na udziale w projektowaniu hipotez badawczych, udziale w badaniach terenowych, udziale w projektowaniu eksperymentów, wykonaniu prac laboratoryjnych, analizie statystycznej i bioinformatycznej uzyskanych danych, zebraniu danych literaturowych, przygotowaniu tekstu manuskryptu oraz powiązanych z nim materiałów dodatkowych, przygotowaniu rycin i tabel.

2. Pikus, E., Dunn, P., Minias, P. High MHC diversity does not confer fitness advantage in a wild bird. Journal of Animal Ecology (praca zgłoszona do druku).

Niniejszym oświadczam, że jestem współautorem powyższej pracy, a mój wkład w jej powstanie polegał na udziale w projektowaniu hipotez badawczych, udziale w badaniach terenowych, udziale w projektowaniu eksperymentów, udziale w analizowaniu danych oraz przygotowaniu rycin i tabel, oraz redagowaniu treści manuskryptu.

Swój wkład ilościowy w powstanie pracy oceniam na 40%.

Jako promotor potwierdzam, że mgr inż. Ewa Pikus jest wiodącym autorem niniejszej pracy (50% udziału ilościowego), a jej wkład polegał na udziale w projektowaniu hipotez badawczych, udziale w badaniach terenowych, udziale w projektowaniu eksperymentów, wykonaniu prac laboratoryjnych, analizie statystycznej i bioinformatycznej uzyskanych danych, zebraniu danych literaturowych, przygotowaniu tekstu manuskryptu oraz powiązanych z nim materiałów dodatkowych, przygotowaniu rycin i tabel.

3. Pikus, E., Włodarczyk, R., Jedlikowski, J. and Minias, P., 2021. Urbanization processes drive divergence at the major histocompatibility complex in a common waterbird. PeerJ, *9*, p.e12264.

Niniejszym oświadczam, że jestem współautorem powyższej pracy, a mój wkład w jej powstanie polegał na udziale w projektowaniu hipotez badawczych, udziale w badaniach terenowych, udziale w projektowaniu eksperymentów, udziale w analizowaniu danych oraz przygotowaniu rycin i tabel, oraz redagowaniu treści manuskryptu.

Swój wkład ilościowy w powstanie pracy oceniam na 30%.

Jako promotor potwierdzam, że mgr inż. Ewa Pikus jest wiodącym autorem niniejszej pracy (50% udziału ilościowego), a jej wkład polegał na udziale w projektowaniu hipotez badawczych, udziale w badaniach terenowych, udziale w projektowaniu eksperymentów, wykonaniu prac laboratoryjnych, analizie statystycznej i bioinformatycznej uzyskanych danych, zebraniu danych literaturowych, przygotowaniu tekstu manuskryptu oraz powiązanych z nim materiałów dodatkowych, przygotowaniu rycin i tabel, przygotowaniu odpowiedzi na uwagi recenzentów.

Podpis

Milwaukee, 30.11.2021

Prof. Peter O. Dunn

Behavioral and Molecular Ecology Group,

Department of Biological Sciences

University of Wisconsin-Milwaukee

DECLARATION OF CO-AUTORSHIP

Pikus, E., Dunn, P., Minias, P. High MHC diversity does not confer fitness advantage in a wild bird. Journal of Animal Ecology (submitted).

I confirm that I am co-author of the above manuscript and my contribution included bioinformatic analyses (ddRAD sequencing) and revision of the manuscript.

I estimate my quantitative contribution to manuscript preparation at 10%.

Signature

Łódź, 30.11.2021

dr hab. Piotr Minias, prof. UŁ

Katedra Badania Różnorodności Biologicznej,
Dydaktyki i Bioedukacji
Uniwersytet Łódzki

OŚWIADCZENIE

Niniejszym oświadczam, że z przyczyn niezależnych uzyskanie oryginału oświadczenia o wkładzie prof. Petera O. Dunna w powstanie pracy (Rozdział 2) nie było możliwe. Poświadczam również, że treść zawarta w załączonym skanie jest zgodna ze stanem faktycznym.

Podpis

Łódź, 30.11.2021

dr hab. Radosław Włodarczyk

Katedra Badania Różnorodności Biologicznej,
Dydaktyki i Bioedukacji
Uniwersytet Łódzki

OŚWIADCZENIE O WKŁADZIE W POWSTANIE PRACY

Pikus, E., Włodarczyk, R., Jedlikowski, J. and Minias, P., 2021. Urbanization processes drive divergence at the major histocompatibility complex in a common waterbird. PeerJ, *9*, p.e12264.

Niniejszym oświadczam, że jestem współautorem powyższej pracy, a mój wkład w jej powstanie polegał na udziale w badaniach trenowych oraz udziale w redagowaniu treści manuskryptu.

Swój wkład ilościowy w powstanie pracy oceniam na 10 %.

Podpis

Warszawa, 30.11.2021

dr Jan Jedlikowski

Zakład Ekologii i Ochrony Środowiska

Uniwersytet Warszawski

OŚWIADCZENIE O WKŁADZIE W POWSTANIE PRACY

Pikus, E., Włodarczyk, R., Jedlikowski, J. and Minias, P., 2021. Urbanization processes drive divergence at the major histocompatibility complex in a common waterbird. PeerJ, *9*, p.e12264.

Niniejszym oświadczam, że jestem współautorem powyższej pracy, a mój wkład w jej powstanie polegał na udziale w badaniach trenowych oraz udziale w redagowaniu tekstu manuskryptu.

Swój wkład ilościowy w powstanie pracy oceniam na 10 %.

Podpis