# INVESTIGATING THE GENETIC DIVERSITY OF IMMUNE GENES IN NON-NATIVE POPULATIONS OF AMERICAN BULLFROGS (*RANA CATESBEIANA*)

By

Luisa Nereyda Segovia

A Thesis Presented to

The Faculty of California State Polytechnic University, Humboldt

In Partial Fulfillment of the Requirements for the Degree

Master of Science in Biology

Committee Membership

Dr. Karen Kiemnec-Tyburczy, Committee Chair

Dr. Catalina Cuellar-Gempeler, Committee Member

Dr. Amy Sprowles, Committee Member

Dr. John Reiss, Committee Member

Dr. Paul Bourdeau, Program Graduate Coordinator

December 2022

# ABSTRACT

## INVESTIGATING THE GENETIC DIVERSITY OF IMMUNE GENES IN NON-NATIVE POPULATIONS OF AMERICAN BULLFROGS (*RANA CATESBEIANA*)

Luisa Nereyda Segovia

The American Bullfrog (*Rana catesbeiana*) is a highly invasive species that has successfully colonized habitats around the world. The genetic variability of immune genes within invasive populations, like that of bullfrogs, may contribute to how resistant a population is to pathogens. The objective of this project was to characterize the genetic diversity of an immune gene in invasive bullfrog populations in California to better understand how persistent these populations might be over time. To characterize immune gene variability, I isolated exon 2 of the Major Histocompatibility Complex (MHC) Class II Beta chain gene and a neutral mitochondrial marker, *cytochrome b* (*cyt b*), from five bullfrog populations. I used standard population genetic metrics to compare the genetic diversity between these two loci across populations. I found high diversity in MHC II exon relative to *cyt b*, however these populations harbored differing levels of diversity at both loci. I also reconstructed a phylogeny of the isolated alleles with those from other ranids and found that the majority of my MHC alleles clustered with other *R. catesbeiana* alleles instead of with those of other ranids, except two alleles may exhibit transspecies polymorphism. The majority of the selection tests detected significant positive selection acting on MHC alleles, while there was little evidence on *cyt b*. Overall, these non-native

populations have similar genetic diversity to other native amphibian populations. I conclude that they likely have sufficient genetic diversity to persist in the face of novel pathogens they may encounter in non-native habitats.

# ACKNOWLEDGEMENTS

I want to acknowledge all the people that have helped me over the past few years. Without the guidance and mentorship of my advisor, Dr. Karen Kiemnec-Tyburczy, I would have never gotten the opportunity to work on this project as an undergraduate or have the opportunity to continue working on this project as a graduate student. She saw potential in me and has offered lots of support over the course of my undergraduate degree as well as during my adventure as a graduate student. I would also like to acknowledge all the help I received from my lab assistant Angelica Klawiter, who worked on many colony PCRs as well as on *cyt b* isolation. The members of my thesis committee committed a lot of time and effort throughout this project. I would also like to thank all the different sources of funding I have received over the years. I received funding from LSAMP-INRSEP during my undergraduate years and received funding from the Department of Biological Sciences Master's award during my graduate years. Last, but not least, I would like to thank my friends and family, who have offered endless amounts of support during my endeavors at Cal Poly Humboldt.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

# INTRODUCTION

The introduction of invasive species to non-native ranges has caused many deleterious effects to native populations. Many of these introduced species are difficult to control as their populations outcompete native populations for resources and may lead to the extinction of native species (Blackburn et al. 2014). Regardless of the origin of the introduction, invasive species are a major cause of biodiversity loss and extinction of many native fauna (Ficetola et al. 2008). It is important to understand why these species are successful in novel environments, and thus how to predict and possibly prevent future invasions (Colautti and Lau 2015).

The introduction of a non-native species to a new environment may lead to a founder effect–a decrease in genetic diversity within the founding population relative to the source population (Uller and Leimu 2011). Reduction in genetic diversity may cause issues like inbreeding depression and reduction of fitness, making it difficult for these founding populations to establish in their new environment. Yet there are many invasive species that adapt to new environments even with extremely low genetic diversity (Schrieber and Lachmuth 2017). This scenario is referred to as the ‘genetic paradox of invasion’ and it illustrates that there are other mechanisms involved in the success of a novel population (Schriber and Lachmuth 2017). A possible explanation for the success of an invasive population in new environments is that they maintain high levels of diversity at ecologically relevant loci even if there is low genetic diversity at the neutral loci typically used for population genetic analysis (Estoup et al. 2016).

# Invasion Ecology of Focal Species

A prime example of an invasive species is the American Bullfrog (*Rana catesbeiana*), which has been introduced globally and successfully colonized a variety of habitats (Hammon et al. 2017). *Rana catesbeiana* is a member of the family Ranidae and is native to Eastern North America. It has been introduced to over 40 countries across four continents (Funk et al. 2011, Schloegel et al. 2012). Bullfrogs are opportunistic predators that can consume a variety of different prey ranging from invertebrates to fish, reptiles, amphibians, and even small mammals (Snow and Witmer 2010). *Rana catesbeiana* has been introduced by humans to areas outside its native range for a variety of reasons. In the U.S., this species has been used commercially in the pet trade and as a food source (Gervasi et al. 2013, Schloegel et al. 2009, Yap et al. 2018).

There are many examples of *R. catesbeiana* introductions leading to declines in the populations of native fauna, predominately other amphibian taxa. The introduction of this amphibian has caused the spread of disease and reduction in available food resources for native amphibians (Laufer et al. 2018). Laufer et al. (2018) surveyed ponds in Uruguay where *R. catesbeiana* were introduced due to bullfrog farming. He discovered that in ponds with feral bullfrogs, there was a significant decrease in native amphibians and other fauna compared to that documented before the introduction of bullfrogs. Groffen et al. (2019) described the various reasons for the introduction of bullfrogs (e.g., use as an alternative food source and for cultural practices) in South Korea and the effects that these bullfrogs had on the native fauna once they were released (after the demand for

bullfrogs declined). For example, he showed that these established bullfrog populations directly impacted three anurans (two species of *Pelophylax* and a hylid) and indirectly affected local frog-eating rat snakes, *Oocatochus rufodorsatus* (Groffen et al. 2019).

The success of these invasive frogs has driven researchers to study the ecological impacts of *R. catesbeiana* in native and non-native regions, but few studies have used genetics to understand the persistence and diversity of *R. catesbeiana*. Sanz et al. (2022) used environmental DNA (eDNA) sampling to detect the presence of the American bullfrog on the Iberian Peninsula. They detected bullfrogs in several locations around the peninsula and determined that these individuals may be in the early stages of population establishment. Environmental DNA is a useful early detection tool especially in amphibian studies as it can detect larval stages in aquatic habitats (Sanz et al. 2022). However, to better understand the genetic composition of these invading species, in-depth studies of genomes are needed. Several molecular studies have examined the divergence of neutral markers throughout the native and non-ranges of *R. catesbeiana* (e.g., Austin et al. 2004a, b, Funk et al. 2011, Kamath et al. 2016). Assumedly neutral loci, like mitochondrial loci or microsatellites, are often used as a metric to quantify genetic diversity of a population and identify demographic patterns but are not as helpful for predicting a population's evolutionary potential (Estoup et al. 2016, Uller and Leimu 2011). For this sort of information, protein-coding loci that are likely to have direct effects on fitness are more useful.

# Functional Loci of the Vertebrate Immune System

My study focuses on a key gene of the acquired immune system's Major Histocompatibility Complex (MHC). The MHC is a complex of genes that are involved in the acquired immune response of jawed vertebrates and thus have multiple functions in this taxonomic group. MHC genes influence many traits within vertebrate populations such as immune recognition, kin recognition, cooperation, and pregnancy outcomes (Milinski et al. 2006). Since MHC plays an influential role in the success of a population, it is often used in wildlife studies to better understand how populations will persist in the long term.

The MHC is made up of two gene classes; class I genes encodes a transmembrane proteins that recognize peptides derived from *intracellular* pathogens while class II genes encode a transmembrane proteins that recognize *extracellular* pathogens (Teacher et al. 2009). MHC class I and class II proteins form functional complexes that differ structurally and are expressed in different cell types. MHC class I genes encode a glycoprotein that is found on the surfaces of nearly all nucleated cells while MHC class II genes encode glycoproteins that are located on the surface of antigen presenting cells (APCs) (Martin et al. 2017a). It has been shown that MHC class I proteins use an endogenous pathway while MHC class II proteins use an exogenous pathway in order to create the antigen-MHC complex that will be recognized by the immune system. In MHC class II expressing cells, the MHC class II glycoprotein is synthesized in the endoplasmic reticulum (ER) and will leave the ER once fully formed. The MHC class II will travel to

an endosome, where an engulfed antigen has been fragmented and will form a complex with these fragmented antigens. Once these complex is formed it will travel to the cell membrane to be recognized by an immune cell (Martin et al. 2017a). Historically, the two MHC classes were thought to work in separate pathways; but recent work has shown communication between MHC classes, suggesting that both MHC classes can respond to microbial pathogens regardless of the pathogen's origin (Dengjel et al. 2005).

The structure of the MHC class II protein is made up of two homogenous peptides: the alpha chain and the beta chain (Martin et al. 2017a). Each chain has two extracellular domains (α1and α2 or β1 and β2) and each domain is encoded by a separate exon. In the mature protein, the subunits α1 and β1 form pockets where peptides derived from pathogens are bound (Fig. 1). These pockets are formed by the coupling of the N-terminal domains of both these subunits and these pockets are influenced by the amino acid residues that are found in these regions. The amino acid residues that make up these pockets are characterized to better understand whether there is evidence of positive selection (Zhang et al. 2022). This is important to note since the ability to recognize an antigen is important in the survival of individuals and the long-term persistence of populations (Savage et al. 2019).

The most well-understood function of MHC genes in amphibians is their role in disease susceptibility and resistance. A diverse class I and class II MHC repertoire may help individuals be resistant to diseases, as well as increase population persistence in the presence of pathogens (Savage et al. 2019). In amphibians, particular PBRs have demonstrated increased protection or susceptibility towards diseases like

chytridiomycosis, caused by infection with a fungal pathogen *Batrachochytrium dendrobatidis* (Bd) (Bataille et al. 2015, Teacher et al. 2009, Savage et al. 2019, Lau et al. 2017). Specific PBRs in MHC class II β1 domain, like those in the P4 and P6 binding pockets have been correlated with chytridiomycosis resistance before infection occurs (Bataille et al. 2015). These chytridiomycosis-resistant alleles are not found in high frequency in natural populations, but there has been a consistent pattern of chytridiomycosis-induced selection upon codons in MHC class II (Bataille et al. 2015, Kosch et al. 2016). These studies illustrate the usefulness of characterizing functional immune genes in vertebrate populations.

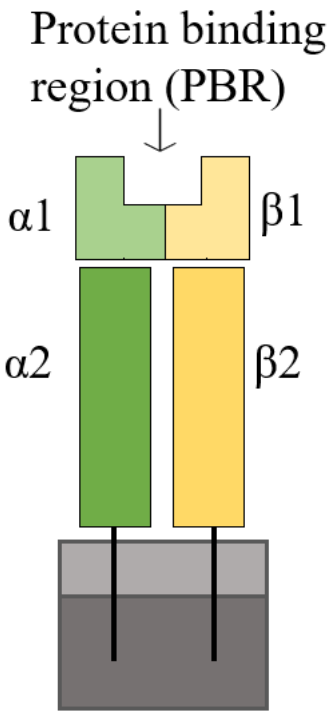


Figure 1: MHC Class II protein structure with pathogen-derived peptide binding regions that are formed by the complexing of multiple proteins. The two beta and two alpha domains on MHC class II are shown in different colors. Illustrated structure adapted from Martin et al. 2017a)

The number of MHC genes found in the genome varies across vertebrates, but these loci are generally some of the most polymorphic protein-coding loci in the genome (Zeisset and Beebee 2010). For example, humans have three MHC class I genes and four MHC class II genes (Janeway et al. 2001). Each of these genes contain multiple exons that encode different domains of the protein. The MHC class II beta chain gene–particularly its second exon that encodes the β1 domain–is frequently used in wildlife disease studies because of its high polymorphism and because genetic variability at this locus results in pathogen resistance at the population level (Sommer et al. 2005). High genetic variability can lead to populations having a better response to environmental stressors like pathogens, parasites, environmental changes while low genetic variability can lead to populations being unable to respond to environmental stressors (Ujvari and Belov 2011). Thus, this locus can be used as a proxy for predicting the potential an invasive population has for pathogen resistance.

Little is known, however, about the diversity of MHC genes in invasive populations of amphibians because of limited sampling. To date, only two studies have investigated MHC evolution in non-native populations. In a study of the invasion of Cane toads (*Rhinella marina*) in Australia, Lillie et al. (2016) found low MHC class I and class II allelic diversity in these invasive populations. This pattern—combined with microsatellite data—suggested that the Cane toad populations had undergone bottlenecks. A study of MHC class II alleles in both native and non-populations of *Rana catesbeiana* throughout the U.S. (LaFond et al. 2021) found widespread MHC allele sharing across native and non-ranges, but also demonstrated limited trans-species polymorphism,

indicating a complex MHC evolutionary history in *R. catesbeiana*. LaFond (2021) also determined that the unique evolutionary history of MHC alleles might play an important role in *R. catesbeiana* resistance to Bd infection.

In this study, I characterized the genetic diversity and evolutionary history of the second exon of the MHC class II beta chain gene in five invasive bullfrog populations in California, in geographic areas not targeted in previous efforts. I investigated the evolutionary history of the alleles by reconstructing the phylogeny of my alleles combined with those from other ranids. I also compared the genetic diversity of the MHC locus to a neutrally evolving mitochondrial marker, *cytochrome b* (hereafter *cyt b*), using demography, population structure and phylogenetic analyses. This comparison is useful when trying to better understand the effects of functional loci, like those of the MHC complex, on invasive species. Along with mitochondrial markers, previous studies have used other putatively neutral markers such as microsatellites to understand historical and contemporary influences on population genetic structure (Austin et al. 2004b) and estimate bullfrog larval fitness (Zeisset and Beebee 2010). I chose *cyt b* for comparison because the use of this protein-coding gene allowed me to compare the rates of synonymous and non-synonymous substitutions between the two genes. Finally, I compared the genetic diversity found in my focal populations to that of other native and invasive species. Since MHC is a polymorphic gene that can be critical for the successful adaption of a new population, I predicted that MHC will maintain high diversity across the five populations and will be undergoing positive selection, while *cyt b* will maintain

its neutrality. Alternatively, these invasive populations may not maintain their high MHC diversity due to the founder effect.

# MATERIALS AND METHODS

## Tissue Collection and DNA Isolation

*Rana catesbeiana* were collected from different National Wildlife Refuges (Table 1) by James Bettaso, U.S. Forest Service, as part of a survey for abnormal frog development conducted in 2008-2009 (Reeves et al. 2013). I chose to sample four localities from this study that were the most geographically distant. I added one additional locality (Mad River, Humboldt County, CA) because of its coastal location and proximity to Cal Poly Humboldt (Fig. 2). The whole specimens were either stored individually or multiple in jars containing 70% ethanol until a small tissue sample was collected from each.

Table 1: Collection site names, abbreviations, GPS coordinates, date of collection, and sample size for all localities in California where American bullfrog samples were collected. National Wildlife Refuges are abbreviated as NWR.

| Collection Site | Abbreviation | Latitude | Longitude | Date of collection | Sample size |
|---|---|---|---|---|---|
| Lower Klamath Lake NWR | LKL | 41.99756 | -121.83445 | 11-Aug-2008 | 20 |
| Mad River, Humboldt County | MAD | 41.443521 | -120.496557 | 30-Aug-2018 | 20 |
| Sutter NWR | SUT | 39.14871 | -122.04114 | 24-Jun-2008 | 20 |
| Merced NWR | MER | 37.17504 | -120.62276 | 25-Jun-2008 | 20 |
| Kern NWR | KRN | 35.71841 | -119.58896 | 12-Jul-2018 | 20 |

Genomic DNA was extracted from the 100 *R. catesbeiana* liver or skin tissue samples using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. The concentrations of genomic DNA were determined by spectrophotometry using a Nanodrop 1000 (Fisher Scientific, Waltham, MA). The handling and processing of these tissues was approved by the Cal Poly Humboldt Institutional Animal Care and Use Committee (protocol 16/17.B.102-E).

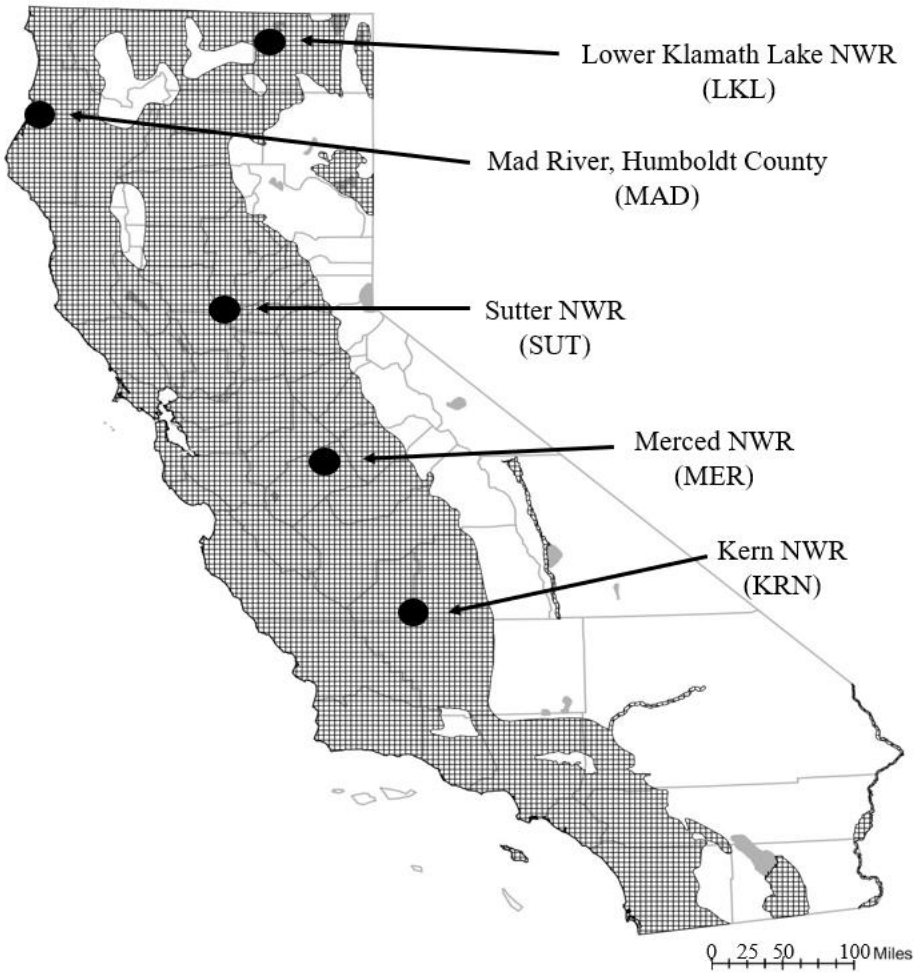


Figure 2: Introduced bullfrog range in California. The grey shaded regions indicate bullfrog presence. The black circles represent the five localities sampled for this

project (map adapted from Zeiner et al. [1988-1990], gridlines indicate yearlong range).

## Amplification of MHC Alleles

Polymerase Chain Reaction (PCR) to amplify MHC class II exon 2 alleles was performed using the ForN and RevA primers developed by Mulder (2017). These primers amplify the single beta chain gene from this species. The conditions for the PCR were as follows: 2 µl of bullfrog genomic DNA, 1 µM of each primer, and GoTaq Master Mix (Promega, Madison, WI) at 1X. This amplification was carried out at 95°C for 3 minutes and 30 seconds for the initial denaturation step; followed by 34 cycles of 95°C for 30 seconds of denaturation, annealing at 52°C for 30 seconds, and extension at 72°C for 30 seconds; with a final extension at 72°C for 5 minutes.

The PCR results were visualized using standard gel electrophoresis with ethidium bromide staining. PCRs containing products of the predicted size were purified using the Wizard SV PCR Purification kit (Promega).

## Amplification of *Cytochrome b*

I amplified *cyt b* using the primers MVZ15-L (Moritz et al. 1992) and CytbAR-H (Goebel et al. 1999), which amplified a 1047-base pair segment of the coding region. The conditions of this PCR were as follows, 0.5 µl of DNA, 2 µl of 10 µM of each primer and GoTaq Master Mix (Promega) at 1X. This amplification was carried out at 94°C for 6 minutes for the initial denaturation step; followed by 35 cycles of 94°C for 20 seconds of denaturation, annealing at 52°C for 30 seconds, and extension at 72°C for 30 seconds;

with a final extension at 72°C for 5 minutes. The amplified PCR products were visualized using gel electrophoresis and purified using the Wizard SV PCR Purification kit (Promega).

## Cloning and Sequencing

All purified MHC PCR products were cloned using the pGEM T-easy Vector System (Promega). Eight unique clones per individual were screened using standard M13 primers and clones containing PCR inserts of the intended size were sent out for sequencing at Eurofins Genomics (Louisville, KY) or MC Lab (San Francisco, CA). The purified *cyt b* PCR products were sequencing directly without cloning by either Eurofins Genomics or MC Lab using the forward and reverse primers used in PCR.

I cleaned all my sequences by scanning the chromatograms by eye and removed the vector sequence from MHC alleles using MEGA v11 (Kumar et al. 2018). I then aligned my sequences from each individual using ClustalW and assessed the homozygosity or heterozygosity of each individual based on whether I recovered one or two alleles from eight unique clones.

## Analysis of Genetic Diversity

After sequences were cleaned and edited, genetic diversity was assessed using several different programs. DnaSP v5 (Librado and Rozas 2009) was used to calculate gene diversity metrics (expected heterozygosity [$H_e$] for MHC and haplotype diversity [$H_d$] for *cyt b*), number of MHC alleles and *cyt b* haplotypes per population, nucleotide

diversity ($\pi$), and Tajima's D. Arlequin v3.5.22 (Excoffier and Lischer 2010) was used to calculate pairwise $F_{ST}$ and significance of $F_{ST}$ values for both loci. Microsoft Excel was used to create allele frequency histograms using proportions of each allele.

## Phylogenetic Reconstruction of MHC β1 Alleles and *Cyt b*

I isolated MHC class II exon 2 by removing the introns from the isolated MHC sequences. The exon 1/ exon 2 boundary was determined by Mulder et al. (2017) and I used this boundary to trim approximately 9bp of intron 1 and 9bp of intron 2 from my sequences. I combined my MHC class II exon 2 sequences with representative sequences from all other ranid species available in GenBank alignment to make an alignment used to generate phylogenetic trees. This analysis was conducted in order to investigate the evolutionary history of the *R. catesbeiana* alleles relative to other native and non-native species' alleles in the family Ranidae (to which bullfrogs belong). I aligned the sequences using the ClustalW algorithm implemented in MEGA. To determine which evolutionary model of nucleotide substitution best fit my data, I used FindModel (http://www.hiv.lanl.gov/). Evolutionary relationships were inferred using maximum likelihood with the Hasegawa-Kishino-Yano (HKY) plus gamma model with 10,000 bootstrap replicates implemented in the IQ-tree web software (http://iqtree.cibiv.univie.ac.at/; Trifinopoulos et al. 2016). The rendering of the phylogenetic reconstruction of the evolutionary history of ranid MHC class II alleles was done in MEGA.

I aligned my *cyt b* haplotypes with three reference sequences (Austin et al. 2004a, Kamath et al. 2016) to create my phylogenetic tree. I aligned my sequences using ClustalW algorithm implemented in MEGA. I then used FindModel to determine the best evolutionary model that fit with my *cyt b* data. Evolutionary relationships were inferred using maximum likelihood based on the Hasegawa-Kishino-Yano (HKY) plus gamma model with 10,000 bootstrap replicates implemented in IQ-tree. Phylogenetic reconstruction of *cyt b* was visualized using MEGA.

## Haplotype Networks

Population Analysis with Reticulate Trees (PopART) v1.7 (Bandelt et al. 1999) was used to generate a haplotype network for MHC and *cyt b* haplotypes using the TCS algorithm (Leigh and Bryant 2015). I aligned my MHC alleles and *cyt b* haplotypes using MEGA and used my alignments to I created my haplotype network files. I then used MEGA to convert my fasta files into nexus files. I then added a frequency matrix into my nexus files. I used this modified nexus file to create my haplotype networks using the available dropdown menu. The haplotype networks depict the proportion of alleles or haplotypes found at each locality.

## Analysis of Codon-by-Codon Selection

I used the Hypothesis testing using Phylogeny (HyPhy) Datamonkey 2.0 server (www.datamonkey.org) to test for positive selection (Weaver et al. 2018). Genetic Algorithm for Recombination Detection (GARD) was used to identify

intragenic interference that can bias selection measurements (Kosakovsky Pond et al. 2005). If any intragenic interference was found, GARD created a partitioned data set that was used for the following codon-by-codon tests for positive selection: MEME (Mixed Effects Model of Evolution), FEL (Fixed Effects Likelihood), SLAC (Single Likelihood Ancestry Counting) and FUBAR (Fast, Unconstrained Bayesian AppRoximation). The program MEME uses mixed-effects maximum likelihood to test episodic positive selection on individual amino acid sites (Murrell et al. 2012). In contrast, FEL uses a maximum-likelihood approach to infer nonsynonymous ($d_N$) and synonymous ($d_S$) substitution rates on a per-site basis for a given alignment and corresponding phylogeny (Pond et al. 2005). The SLAC program uses a combination of maximum likelihood and counting approaches to infer $d_N/d_S$ substitution rates on a per-site basis for an alignment and corresponding phylogeny (Pond et al. 2005). The FUBAR program uses Bayesian approaches to infer $d_N/d_S$ substitution rates on a per-site basis for a given phylogeny, and it assumes there is constant selection pressures on each site along the entire phylogeny (Murrel et al.  2013). I predicted that sites under positive (also described as diversifying) selection may be involved with binding foreign peptides. Selection on putative binding sites provides evidence that this MHC gene is rapidly changing in these invasive populations as it is in many other native and non-native taxa. To determine if sites under selection were putative PBRs, I aligned my sequences with human MHC sequences (there are no crystal structures for any amphibian MHC proteins). I noted which positions in my bullfrog amino acid alignment matched known PBRs in human MHC proteins (determined by crystallography studies, e.g., Tong et al. 2006).

I ran a series of tests for positive selection on *cyt b*, although I did not predict it would be experiencing this type of selection based on the assumption of neutrality made in previous studies (e.g., Kamath et al. 2016). First, I used GARD to identify any intragenic interference that can bias selection measurements in my *cyt b* data. I also conducted the same four selection analyses done on MHC (MEME, SLAC, FEL and FUBAR) on the *cyt b* haplotypes.

## Analysis of Branch Selection

I used the one branch selection test that is available for partioned datasets on the Hypothesis testing using Phylogeny (HyPhy) datamonkey server: Branch-site Unrestricted Statistical Test for Episodic Diversification (BUSTED). BUSTED is a gene-wide test that will identify whether an entire gene has experienced positive selection (Murrell et al. 2015). After running my MHC data through GARD, I used that portioned dataset as the input file for BUSTED.

# RESULTS

## Analysis of Genetic Diversity

After analyzing a total of 100 *L. catesbeianus* across my five localities, I found 18 unique MHC class II beta chain exon 2 alleles (272 bp in length). None of these unique alleles had stop codons (Appendix A) suggesting that all these alleles encode functional proteins. These alleles were named with the prefix "RACA" with numbers one through eighteen to differentiate them (RACA 1-18). Out of these 18 alleles, seven alleles were found in KRN, seven alleles were found in LKL, three alleles were found in MAD, four alleles were found in MER, and nine alleles were found in SUT. Of the seven alleles found in KRN, one allele (RACA 10) was unique to that locality. In LKL, three alleles (RACA 04, 05, 18) out seven were unique to this location. In SUT, five (RACA 02, 11, 12, 13, 14) out of nine alleles were unique to this location. However, MAD contained no alleles unique to this location.

Allele frequencies varied widely across the five localities. For example, allele RACA 16 was the most common allele in MAD, but not detected in KRN or SUT (Fig. 3). Allele RACA 1 was at approximately equal frequency in KRN, LKL and SUT, but not found in MAD or MER. No single allele was most frequent in all populations. Alleles RACA 01, 08 and 16 had the highest frequencies in particular populations while others (e.g., RACA 05, 06 and 11) were in low frequency (and often private to that locality).

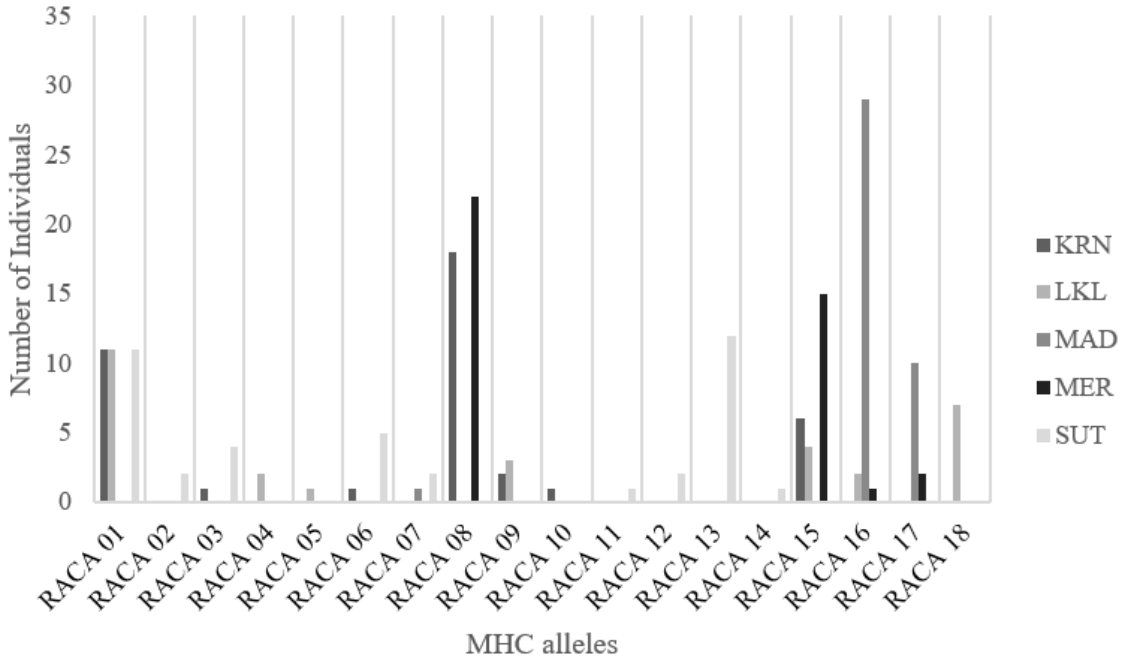


Figure 3: MHC allele frequency histogram. Bars represent the frequency of each allele across all the focal localities and each color represents a single locality.

MHC haplotype diversity ($H_d$), expected heterozygosity ($H_e$) and nucleotide diversity ($\pi$) was also variable across the five localities (Table 2). Expected heterozygosity ($H_e$) for all 100 samples was 0.094 and ranged from 0.499-0.716. Haplotype diversity for all samples was 0.019, while haplotype diversity ranged from 0.052-0.272 across individual localities. Nucleotide diversity for all 100 samples was 0.01427, while that for individual localities ranged from 0.0120-0.16149. The number of polymorphic sites (S) per locality ranged from 76-122. The mean Tajima's D value for MHC was 0.59852 across all localities, with three localities having negative values that ranged from -0.50841- -0.84568 and two localities having positive Tajima's D values that ranged from 0.12324-0.17393. MAD had the highest $H_d$ while the lowest $H_d$ was found in

SUT. LKL and SUT had the same $H_d$. MAD had the highest π value while KRN had the lowest π value. I found the highest $H_e$ in MER while SUT had the lowest $H_e$ value of the five localities.

Table 2: Genetic diversity of MHC in American bullfrogs in five localities in California. Diversity indices reported are the number of polymorphic sites (S), haplotype diversity ($H_d$), expected heterozygosity ($H_e$) and nucleotide diversity (π). Locality abbreviations and sample size (n) are shown. Significant Tajima's D values are shown in bold.

| Locality | n | S | $H_d$ | $H_e$ | π | Tajima's D |
|---|---|---|---|---|---|---|
| LKL | 20 | 113 | 0.076 | 0.23852 | 0.03018 | 0.12325 |
| MAD | 20 | 76 | 0.272 | 0.2211 | 0.6692 | -0.84568 |
| SUT | 20 | 89 | 0.052 | 0.01219 | 0.16149 | 0.17393 |
| MER | 20 | 105 | 0.177 | 0.24034 | 0.03166 | -0.32146 |
| KRN | 20 | 122 | 0.076 | 0.16382 | 0.012 | -0.01619 |

I identified ten different 925 bp *cyt b* haplotypes in 100 *R. catesbeiana* across the five localities (Appendix B). Out of these ten haplotypes, five of these haplotypes were found in KRN, four of these haplotypes were found in LKL, two of these haplotypes were found in MAD, three haplotypes were found in MER and three haplotypes were found in SUT. Two of the five haplotypes found in KRN (CA8 and CA10) were unique to KRN and two of the four haplotypes found in LKL (CA1 and CA5) were unique to

LKL. SUT (CA9) had one private haplotype, while MAD and MER had no private haplotypes. One haplotype (MT1) that was found in four out of the five localities. MT1 and H2 were found in high frequency in four out of the five localities, although CA2 was the highest in LKL (Fig. 4). Four *cyt b* haplotypes (MT1, H2, C9 and C10) were found high frequency (present in more than five individuals) across one or more localities. Four other *cyt b* haplotypes (H7, H11, CA1 and CA8) were found in low frequencies (less than four individuals) across one or more localities. *Cyt b* haplotype CA2 was found to have high frequency in LKL but had low frequency in MAD and MER.

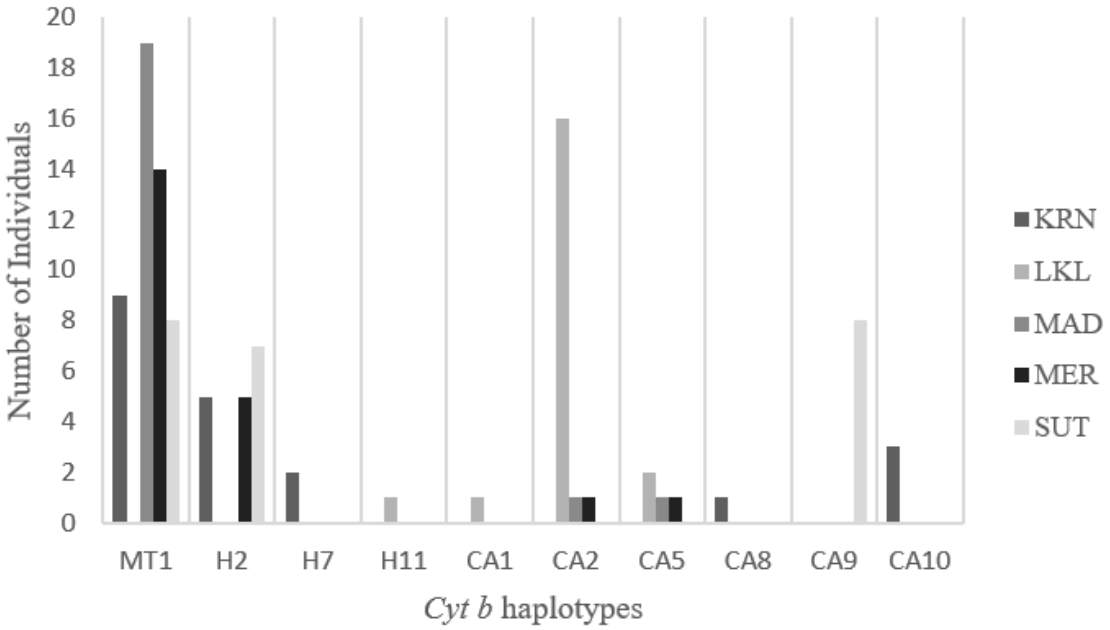


Figure 4: *Cyt b* haplotype frequency histogram. The bars represent frequency of each haplotype found and the color represents each locality.

*Cytochrome b* haplotype diversity ($H_d$) and nucleotide diversity ($\pi$) varied across the localities (Table 3). Haplotype diversity for all 100 samples was 0.002, while haplotype diversity ranged from 0.043-0.131 across individual localities. Nucleotide

diversity for all 100 samples was 0.00207, while nucleotide diversity ranged from 0.00010-0.00197. The number of polymorphic sites (S) per locality ranged from 1-20. The mean Tajima's D value for *cyt b* was -0.73124 across all localities with three negative values and two positive value that ranged from -1.16439-2.36792.

Table 3: Genetic diversity of *cyt b* in American bullfrogs in five localities. Diversity indices reported are the number of polymorphic sites (S), haplotype diversity ($H_d$), and nucleotide diversity ($\pi$). Significant values for Tajima's D are shown in bold ($P< 0.05$). Locality abbreviations and sample size (n) are shown in the table below.

| Population | n | S | $H_d$ | $\pi$ | Tajima's D |
|---|---|---|---|---|---|
| LKL | 20 | 20 | 0.131 | 0.03018 | -0.76309 |
| MAD | 20 | 1 | 0.088 | 0.06959 | -0.28699 |
| SUT | 20 | 5 | 0.043 | 0.02881 | 0.35181 |
| MER | 20 | 2 | 0.104 | 0.03125 | **2.36792*** |
| KRN | 20 | 7 | 0.072 | 0.03669 | -1.16439 |

## Evidence of Population Structure

The pairwise $F_{ST}$ analysis only revealed two significant MHC pairwise values between KRN and MER and between MER and LKL (Table 4). In contrast, there were

seven significant *cyt b* pairwise $F_{ST}$. While both loci showed significant differentiation between KRN and MER, the other tests were only significant for one locus or the other.

Table 4: Pairwise $F_{ST}$ values for both MHC and *cyt b*. Pairwise MHC $F_{ST}$ values are in the lower diagonal and *cyt b* $F_{ST}$ values are in the upper diagonal. Bolded values are significant. Significant $F_{st}$ values are in bold ($P<0.05$).

| | LKL | MAD | SUT | MER | KRN |
|---|---|---|---|---|---|
| LKL | | 0.24176 | **0.24136** | 0.23960 | 0.25524 |
| MAD | 0.38881 | | **0.35000** | **0.13806** | 0.68776 |
| SUT | 0.30115 | 0.67186 | | 0.26342 | 0.20780 |
| MER | **0.07313** | 0.51197 | **0.27838** | | **0.10449** |
| KRN | 0.20296 | **0.19368** | **0.11507** | **0.12344** | |

## Phylogenetic Reconstruction of MHC Alleles and *Cyt b*

The phylogenetic analysis that included all MHC alleles found in other *Rana* species recovered two major clades (noted by blue and purple lines in Figs. 5 and 6). The majority of the alleles isolated in my study clustered together in the first clade, but three of my alleles were found in the second clade. The majority of the alleles grouped with MHC class II alleles isolated from bullfrogs by previous authors (e.g., LaFond et al. 2022, Kiemnec-Tyburczy et al. 2010) within the two major clades. I documented two possible instances of trans-species polymorphism (where different species have maintained similar alleles through speciation due to selective pressures). Allele RACA 15

grouped with an allele from *R. sylvatica* with moderate support and RACA 05, 17, and 18 were in a clade that included an allele from *R. clamitans*.

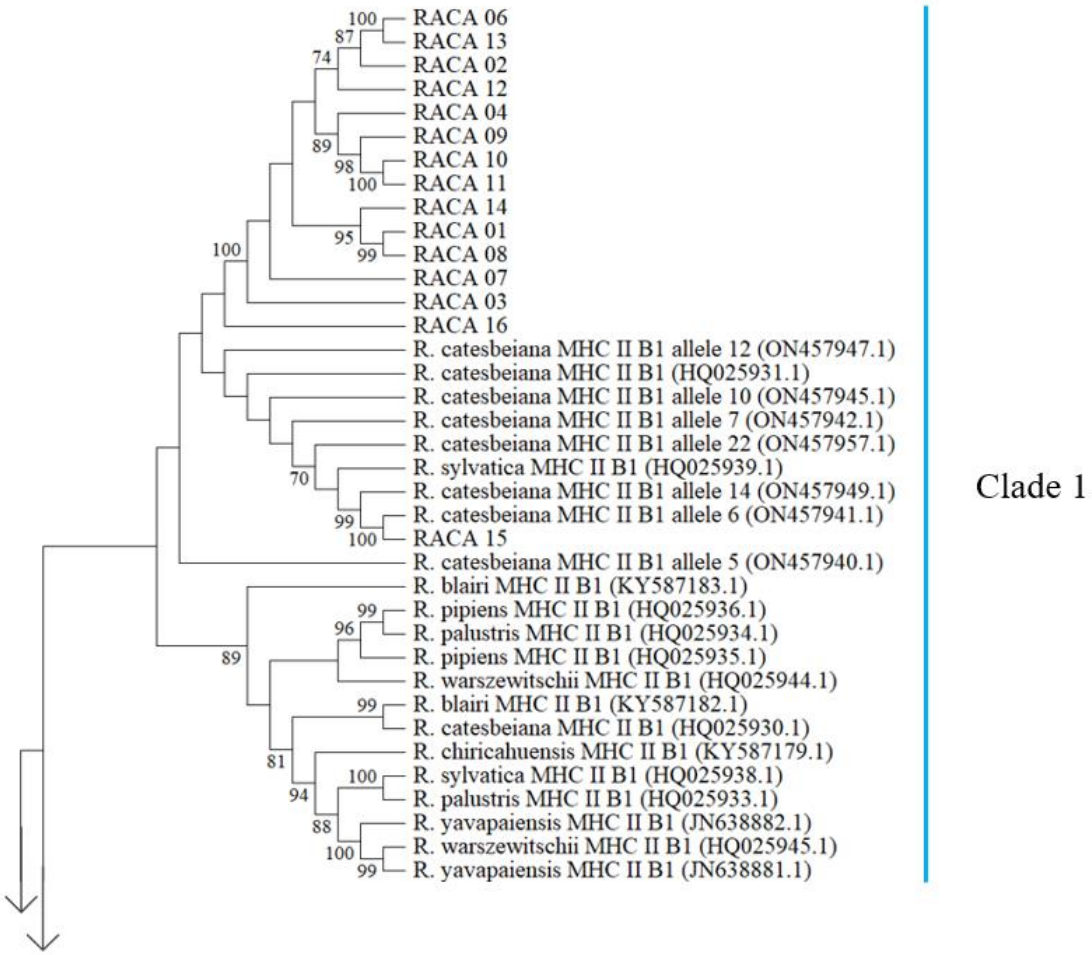


Figure 5: MHC II exon 2 maximum likelihood phylogenetic reconstruction with 10,000 bootstrap replicates rooted to *Xenopus* species. This tree was split into two images, where the top of the tree is seen in this image. The two-colored lines depict the two clades that are form in this tree. The light blue line indicates the first clade seen in this image.

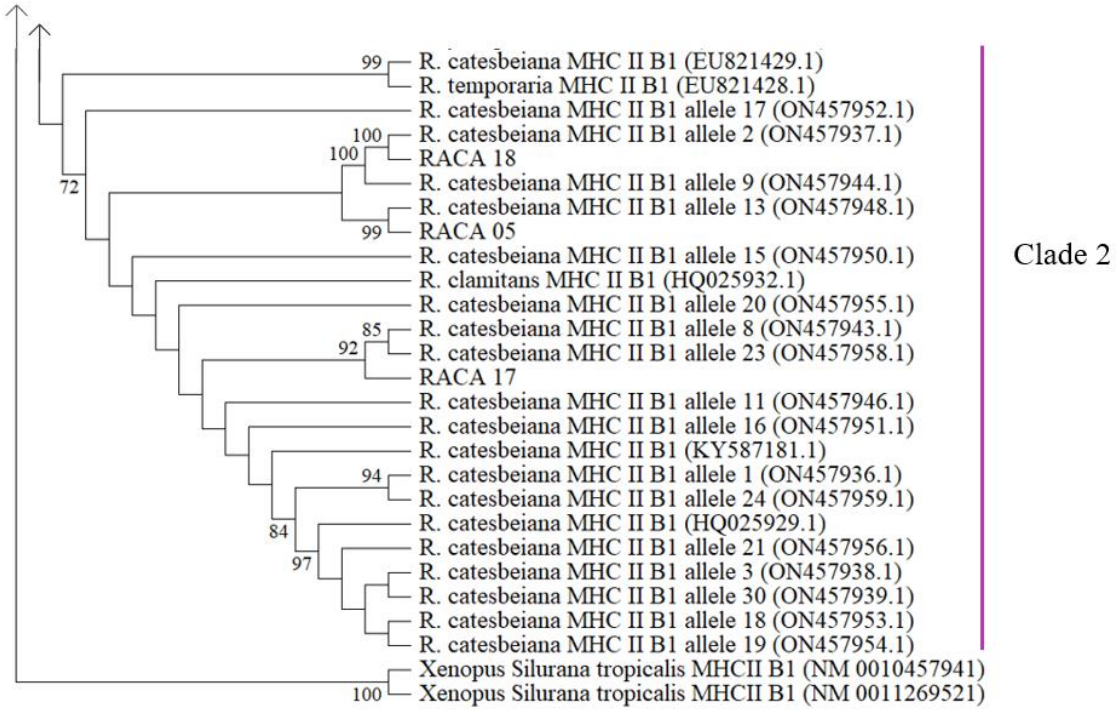


Figure 6: MHC II exon 2 maximum likelihood phylogenetic reconstruction with 10,000 bootstrap replicates rooted to *Xenopus* species. This tree was split into two images, where the bottom of the tree is seen in this image. The second clade is shown by purple color on this image. Three of my alleles (RACA 05, 17 and 18) are found in the second of two clades that form in this tree. The other 15 alleles are found in the top half of the tree seen in Figure 5. The outgroup is seen in this image.

In the *cyt b* phylogeny, all of the haplotypes found in this study were nested in a single, highly supported clade with other *R. catesbeiana* haplotypes that have been isolated from frog populations throughout the U.S. (Fig. 7). Haplotype MT1 was identified in introduced populations in Montana by Kamath et al. (2016) while haplotypes H2 and 11 were first identified in sampling within the native range (east of Mississippi river) by Austin et al. (2004a).

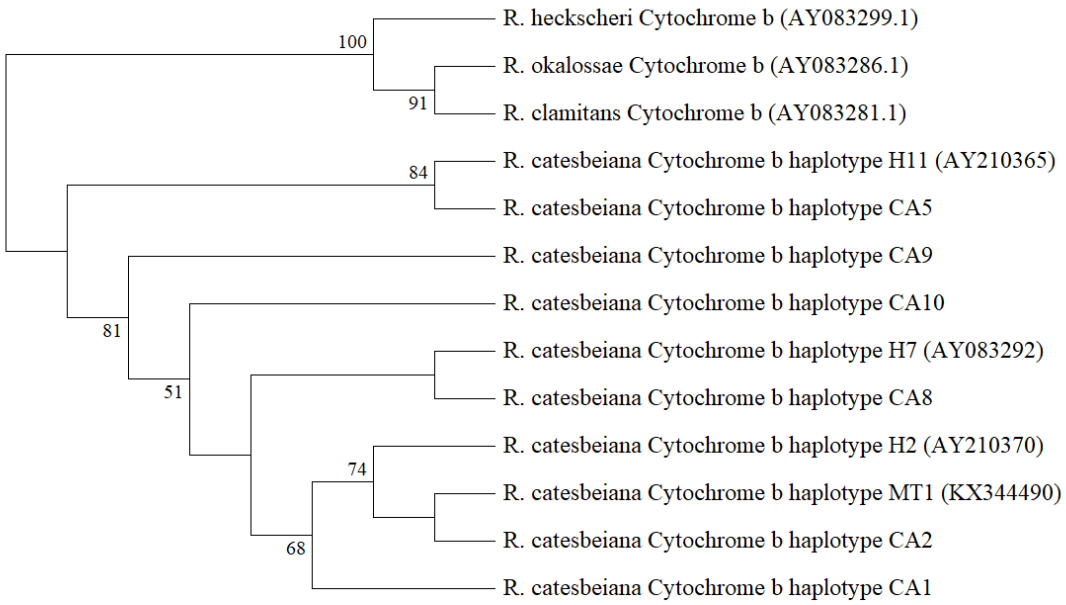


Figure 7: *Cyt b* maximum likelihood phylogenetic reconstruction with 10,000 bootstrap replicates. The haplotypes found in this study (CA1, 2, 5, 8, and 10) and haplotypes that have been previously identified (starting with 'MT' and 'H') are found with other *R. catesbieana cyt b* haplotypes (Austin et al. 2004a, Kamath et al. 2016). *R. heckscheri*, *R. okaloossae*, and *R. clamitans* were used as outgroups for this tree.

## Haplotype Networks

TCS haplotype networks for MHC and *cyt b* depict proportions of each haplotype found at each locality (Figs. 8 and 9) and the number of changes between the alleles. The networks are another way to visualize the notable differences in allele frequencies among populations. For example, RACA 16 was predominately found in MAD, while RACA 08 was found only in KER and MER (Fig. 8). In the MHC network, four alleles (RACA 01, 15 and 16) had a high frequency in three localities. Another five alleles (RACA 03, 06,

07, 08, 09 and 17) had a high frequency in two localities. Nine alleles (RACA 02, 04, 05, 10, 11, 13, 14 and 18) were only isolated from one locality.

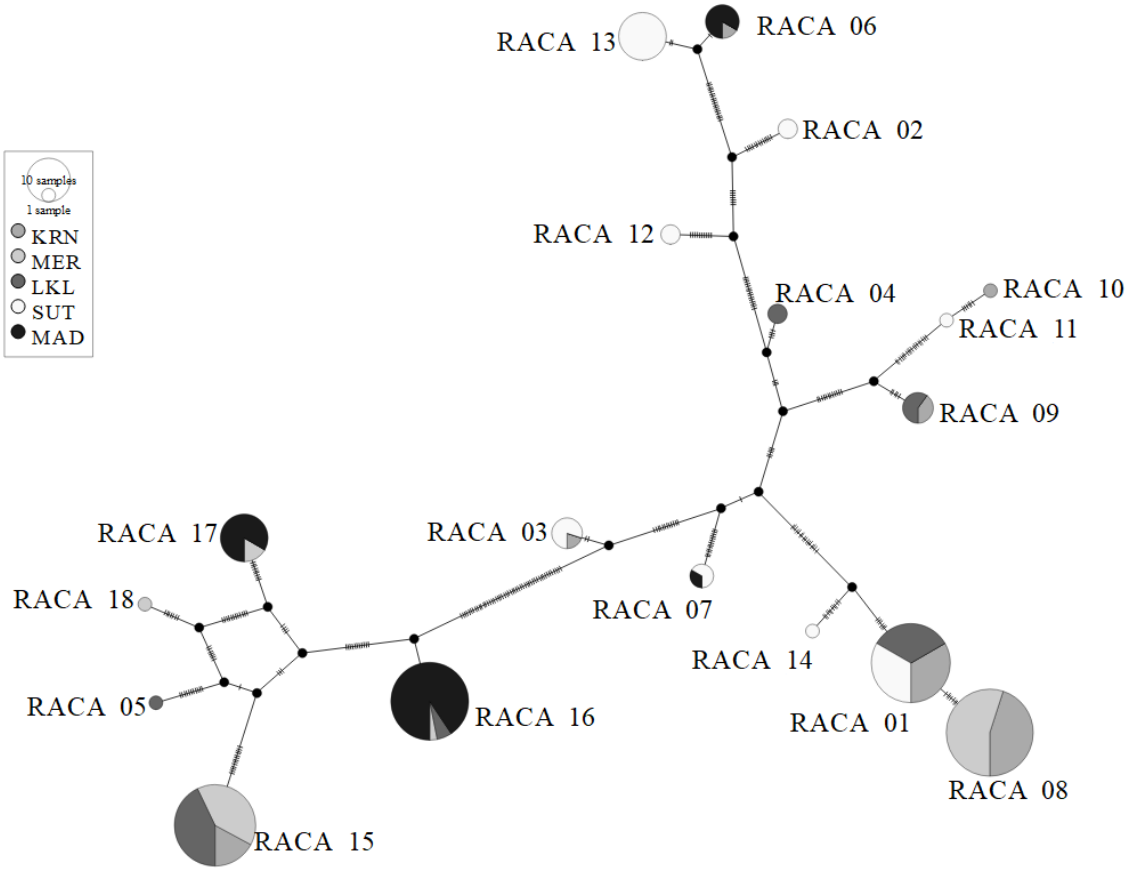


Figure 8: TCS haplotype network depicting proportions of each MHC class II exon 2 haplotype found at each locality. The number of nucleotide substitutions between each haplotype is represented by individual hash marks along the line connecting them. Localities are abbreviated and haplotype names are shown beside each circle in the network.

In the *cyt b* haplotype network, three haplotypes (MT1, H2, and CA2) were found in three localities and seven haplotypes are found in only one locality at low frequency. The networks visually depict the differences between loci seen in the indices of genetic

diversity; there were typically fewer nucleotide substitutions between *cyt b* haplotypes than there were between MHC alleles.

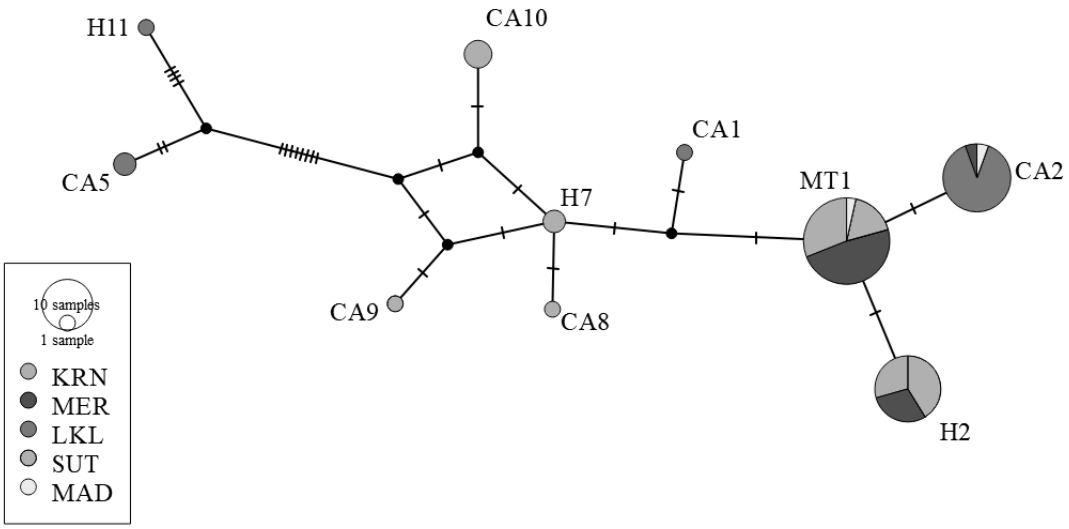


Figure 9: TCS haplotype network depicting proportions of each *cyt b* haplotype found at each locality. The number of nucleotide substitutions between each haplotype is represented by individual hash marks along the line connecting them. Haplotypes with names that start with 'CA' were identified in this study and haplotypes that begin with 'MT' or 'H' were previously described by Austin (2004a) or Kamath (2016). Localities are abbreviated and haplotype names are shown beside each circle in the network.

## Analysis of Codon-by-Codon Selection

GARD found evidence of three recombination breakpoints and created a new partitioned dataset with three partitions. This new dataset was used for all analyses of codon-by-codon selection. In my MHC alignment of exon 2, one amino acid codon showed significant evidence of positive selection in all three site-specific tests and two

amino acid positions showed significant evidence of positive selection in two site-specific tests (Table 6). Out of the 14 sites that had evidence of positive selection in at least one test, three of these sites aligned to known PBRs in human MHC (Tong et al. 2006). Specifically, codons 11 and 66 were associated with P6/P7 binding pockets and codon 67 was associated with P4/P7.

In contrast, only one site in my *cyt b* alignment showed evidence of positive selection and this site was detected only with a single method: FUBAR (Table 5). The three other site-specific tests showed no evidence of positive selection

Table 5: Summary of results of six tests for positive selection on *cyt b*. Corresponding statistical cutoffs for each test and amino acid positions found under selection for each method is shown.

| Method | Statistical cut off | | |
|---|---|---|---|
| | | 80 | 153 |
| MEME | $p \leq 0.05$ | | |
| SLAC | $p \leq 0.05$ | | |
| FEL | $p \leq 0.05$ | | |
| FUBAR | Posterior probability $\leq 0.95$ | X | X |
| BUSTED | $p \leq 0.05$ | Found no evidence of episodic diversifying selection | |

Analysis of Branch Selection

BUSTED found evidence of episodic positive selection occurring in the partitioned MHC data set (Table 6). Since BUSTED is a gene-level test, it will not give

any specific branch sites that are undergoing selection in a portioned dataset with more than one evolutionary history for the whole gene, however it did support that it was appropriate to investigate the more specific codon-by-codon analyses using MEME, FEL, SLAC and FUBAR. BUSTED found no significant signature of positive selection occurring on any branch of the *cyt b* phylogeny (Table 6).

Table 6: Summary of results of six tests for positive selection on MHC class II exon 2. Corresponding statistical cutoffs for each test and amino acid positions found under selection for each method are shown. The amino acid positions (11 and 66) that are bolded and highlighted in light blue corresponds to the P6/P7 pocket residues (Bataille et al. 2015). The amino acid (67) that is bolded and highlighted in purple corresponds to the P4/P7 pocket residue (Bataille et al. 2015).

| Method | Statistical cut off | 3 | **11** | 18 | 39 | 40 | 43 | 46 | 47 | 50 | 64 | **66** | **67** | 76 | 79 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MEME | $p \leq 0.05$ | X | **X** | X | X | X | X | X | X | X | X | **X** | **X** | X | X |
| SLAC | $p \leq 0.95$ | | | | X | | | | | | | | | | |
| FEL | $p \leq 0.05$ | | | | | | | X | | | | | | | X |
| FUBAR | Posterior probability $\leq 0.95$ | | | | | | | X | | | | | | | |
| BUSTED | $p \leq 0.05$ | | | | | There | was | evidence | found | in | exon | 2. | | | |

# DISCUSSION

## Genetic Diversity was Highly Variable Across Populations and Loci

The aim of this study was to characterize a functionally important region of the MHC class II beta gene from five invasive *R. catesbeiana* populations in different areas of California. To understand how other processes aside from selection may be affecting MHC evolution, I compared my MHC data to a neutral maker, *cyt b*. I predicted that the MHC exon would be more diverse than *cyt b* and—because it is a polymorphic gene in many other vertebrates—that it would be under positive selection. I isolated a total of 18 unique MHC alleles compared to ten *cyt b* haplotypes from 100 individuals. This value is on par with other *R. catesbeiana* studies as well as those found in other ranids. LaFond (2022) found 28-30 alleles across 277 individuals (140 frogs from the native range and 137 from invasive populations). In a study of 272 *R. pipiens* individuals from 14 populations within their native range, Trujillo (2021) identified 37 unique MHC alleles. Savage and Zamudio isolated 84 alleles from eight populations of *R. yavapaiensis* in Arizona (Savage & Zamudio. 2016). In another study, Savage (2019) found 20 MHC alleles from 381 wood frogs (*Rana sylvatica*). Although all these studies had a larger sample sizes compared to my study, they all had similarly high MHC diversity.

As I predicted, genetic diversity of the MHC locus was higher than that of *cyt b* based on most—but not all—of the metrics I calculated. This is important to note when trying to understand the significant role that MHC class II has on the persistence of an

invasive population. When I compared the MHC alleles to *cyt b* haplotypes, I found an overall difference in polymorphic sites; MHC alleles had more polymorphic sites than *cyt b*. Expected heterozygosity in MHC had a higher range of values compared *cyt b* haplotype diversity. Haplotype diversity, however, was slightly lower across all MHC alleles lower than it was for *cyt b*. MHC nucleotide diversity was similar in value to that of *cyt b.* There were no significant Tajima's D values in MHC and there was one significant value with *cyt b*. The one population with significant *cyt b* Tajima's D positive value had a nonsignificant negative MHC Tajima's D value and may suggest that this locus is undergoing balancing selection, but it is difficult to say as this pattern could also be caused by a recent decrease in population size.

There was high variability in most of genetic diversity metrics across localities. Comparisons of the two loci showed some congruence; *cyt b* $H_d$ in LKL was twice that of most other localities while nucleotide diversity was almost twice as large in MAD than any other and similarly, MHC $H_d$ and $\pi$ were highest in this locality. These patterns suggest that demographic processes, rather than just selection, may have contributed to the high diversity in MHC in some bullfrog populations. A founder effect may have caused an initial decrease in genetic diversity if only a few frogs were originally introduced at a site. This scenario may explain the low diversity in some of the feral populations in China, for example. Because bullfrogs were introduced to some areas in California multiple times since 1896 (Heard 1904), some populations may have been "infused" with additional alleles from the native range. It is clear from my analysis that some contemporary populations have persisted with relatively low genetic diversity.

LaFond et al.'s (2022) study on MHC in invasive and native populations of bullfrogs demonstrated a unique evolutionary history of MHC. Some my alleles cluster with LaFond et al.'s (2022) alleles in my phylogenetic reconstruction. Allele RACA 17 clusters with their allele 23, which was identified in the native bullfrog range and was found only in one population in Georgia. Allele RACA 05 clusters with another native allele from a population from Virginia. The fact that I had two alleles cluster with two native alleles supports the historical accounts that show bullfrogs were introduced to California multiple times and from different source populations. Even more interesting is that allele 2 from LaFond et al.'s (2022) study was found in both native and invasive populations. Allele 2 was found in three Southern California populations, one Arizona population and one Oregon population (LaFond 2022). Allele RACA 18 actually clusters with Allele 2 from LaFond's (2021) study with high bootstrap support. Allele RACA 15 clusters with allele 6 from LaFond's (2021) study and it was only identified in invasive populations in Southern California. It is possible that certain MHC alleles are more prevalent in either the native or invasive populations and that they may provide some sort of fitness benefit to these bullfrogs allowing them to successfully establish in novel locations.

During an alien species invasion, it has been suggested that the lack of multiple introductions and small starting populations may lead to low genetic diversity in invasive populations (Bai et al. 2012, Dlugosch and Parker 2008). In a recent unpublished study from China, Zhang et al. (2022) found that invasive bullfrog populations in Mainland

China had undergone a population bottleneck caused by a single introduction that led to a decrease in MHC diversity. However, Zhang et al. (2022) suggested that there is potential for successful invasions even with low diversity, as rapid evolution may occur by natural selection in the novel environment.

## Positive Selection has Shaped MHC, But Not *Cyt b*, in Bullfrogs

The results of this study revealed that MHC II B1 domain encoded by exon 2 has experienced positive selection. In contrast, codon-by-codon and branch selection tests of the *cyt b* sequences showed little to no evidence of positive selection. This pattern is consistent with *cyt b* evolving neutrally in these invasive populations. In total, my MHC alignment had evidence of positive selection acting on a total 14 amino acid sites. Out of those 14 amino acid sites, three of those sites aligned are putative PBRs. This is important because these PBRs have shown to increase susceptibility or resistance to diseases like chytridiomycosis (Bataille et al. 2016). The success of an introduced invader, like bullfrogs, may be related to whether the colonizing individuals have particular alleles, and it is likely that these populations are undergoing disease-orientated selection pressures that may help these populations persist long term.

More generally, it was recently shown that MHC alleles are linked to relative risk of being infected with Bd in bullfrogs (La Fond 2022). In this study, specific alleles seemed to be associated with risk, but that relationship varied across populations (e.g., an allele might increase relative risk in native populations, but not invasive ones). Although

I did not isolate the same alleles as in LaFond (2022), the high levels of sequence similarity between my alleles and those isolated by LaFond (2022) suggest a possibility that these alleles are affecting infection status in the populations I sampled. These results illustrate that further research is needed to test these effects directly by either increasing sampling in future studies or by comparing the available LaFond (2022) sequences to the sequences characterized in this study and running further tests.

## Contrasting Evolutionary History of MHC and *Cyt b*

All bullfrog *cyt b* haplotypes formed a single well-supported clade regardless of the location of the individual sampled. This topology was different from that seen in the phylogenetic history of the MHC class II alleles from multiple ranid species. Fifteen MHC class II alleles identified in this study cluster together and three alleles (RACA 5, 7 and 15) clustered with previously identified alleles (LaFond et al. 2022). This is important to note as LaFond et al. (2022) identified MHC alleles in both the native and non-native ranges of bullfrogs. The clustering of some of my alleles with LaFond et al.'s (2022) alleles suggest that regardless of origin, they may confer immunological support to invading bullfrog populations and may help them establish populations even with decrease genetic diversity. This also supports evidence of limited trans-species polymorphism in *R. catesbeiana* as they generally cluster together in a phylogenetic reconstruction regardless of their origin (Figs. 5 and 6). Other ranid species demonstrate extensive trans-species polymorphism with their MHC II exon 2 alleles as they form multiple clades in the MHC evolutionary tree, which indicates that these species retain

ancestral MHC alleles through balancing selection (Kiemnec-Tyburczy et al. 2010, Savage et al. 2018, Trujilo et al. 2017). Some studies that have shown MHC II B1 experiencing directional selection due to local pathogen pressures driving one allele to high frequency, while others have shown long-term patterns of balancing selection that maintain trans-species polymorphism (Cortazar-Chinarro et al. 2018). The lack of significant $F_{ST}$ values for the MHC locus among most of my localities suggests that balancing selection may be maintaining similar alleles across sites in bullfrogs. Thus, the MHC class II exon 2 is shaped by complex selection pressures and historical demographic events (Cortazar-Chinarro et al. 2018).

## Implications for Persistence of Invasive Populations

My study provides insight into the genetic characteristics of established, non-native populations of *R. catesbeiana* in California. I found that the isolated *R. catesbeiana* MHC II exon 2 alleles had evidence of positive selection and are mostly more diverse than the *cyt b* haplotypes. This is important to note as immune genes like those of the MHC can play an important role in the success of invading species. Several studies have proven the important function of immune genes in the process of alien species invasion. Martin (2017b) studied the influence of TLR-4 (a specific Toll-Like Receptor involved in initiating an inflammatory response when microbial elements are identified) in the range expansion of house sparrows (*Passer domesticus*) in Kenya. He determined that a higher expression of TLR-4 helped control parasite loads in these novel house sparrow populations (Martin et al. 2017b). Inflammation is a costly immune

response that can be self-damaging, but it benefitted the house sparrows when establishing new populations because they were able to fight off new parasites (Martin et al. 2017). This may have helped the house sparrows bypass the negative effects of the founder effect and successfully established new populations. Biedrzycka (2020) identified low MHC diversity in invasive raccoon (*Procyon lotor*) populations compared to native populations, yet these invasive populations maintained a high frequency of MHC alleles that showed rabies resistance (Biedrzycaka et al. 2020). Monzon-Arguello (2014) did a similar study on the MHC class II β1 locus on two invasive trout species (*Oncorhynchus mykiss* and *Salmo trutta*). These two trout species were introduced to the Chilean Patagonia for recreational use and have founded self-sustaining populations (Monzon-Arguello et al. 2014). He found that both trout species had similar MHC diversity to native range trout, however their MHC functional diversity was limited, which may decrease immune responses and affect the long-term survival of these salmonid populations (Monzon-Arguello et al. 2014). These studies illustrate the need for further studies on invasive populations that focus on examining a direct link between MHC alleles and disease susceptibility. Combining these two aspects of invasion ecology will provide a better understanding on how invasive species like *R. catesbeiana* are able to adapt to new environments and how their functional loci play important roles in their success.

# LITERATURE CITED

Aguilar, A., Roemer, G., Debenham, S., Binns, M., Garcelon, D., Wayne, R. K. (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. Proceedings of the National Academy of Sciences of the United States of America 10: 3490–3494

Austin, J. D., Lougheed, S. C., Boag, P. T. (2004a) Discordant temporal and geographic patterns in maternal lineages of eastern north American frogs, *Rana catesbeiana* (Ranidae) and *Pseudacris crucifer* (Hylidae). Molecular Phylogenetics and Evolution 32: 799–816

Austin, J. D., Lougheed, S. C., Boag, P. T. (2004b) Controlling for the effects of history and nonequilibrium conditions in gene flow estimates in northern bullfrog (*Rana catesbeiana*) populations. Genetics 168: 1491–1506

Bataille, A., Cashins, S. D., Grogan, L., Skerratt, L. F., Hunter, D., McFadden, M., Scheele, B., Brannelly, L. A., Macris, A., Harlow, P. S., Bell, S., Berger, L., Waldman, B. (2015) Susceptibility of amphibians to chytridiomycosis is associated with MHC class II conformation. Proceedings of the Royal Society B: Biological Sciences 282: 20143127

Biedrzycka, A., Konopinski, M., Hoffman, E., Trujillo, A., Zalewski, A. (2020) Comparing raccoon major histocompatibility complex diversity in native and introduced ranges: Evidence for the importance of functional immune diversity for adaptation and survival in novel environments. Evolutionary Applications 13: 752–767

Blackburn, T. M., Essl, F., Evans, T., Hulme, P. E., Jeschke, J. M., Kuhn, I., Kumschick, S., Markova, Z., Mrugala, A., Nentwig, W., Pergl, J. Pysek, P., Rabitsch, W., Ricciardi, A., Richardson, D. M., Sendek, A., Villa, M., Wilson, J. R. U., Winter, M., Genovesi, P., Bacher, S. (2014) A unified classification of alien species based on the magnitude of their environmental impacts. PLOS Biology 12: e1001850

Colautti, R. I, Lau, J. (2015) Contemporary evolution during invasion: Evidence for differentiation, natural selection, and local adaptation. Molecular Ecology 24: 1999–2017

Dengjel, J., Schoor, O., Fischer, R., Reich, M., Kraus, M., Muller, M., Kreymborg, K., Altenberend, F., Brandenburg, J., Kalbacher, H., Brock, R., Driessen, C., Rammensee, H. G., Stevanovic, S. (2005) Autophagy promotes MHC class II

presentation of peptides from intracellular source proteins. Proceedings of the National Academy of Sciences of the USA 102: 7922–7927

Estoup, A., Ravigne, V., Hufbauer, R., Vitalis, R., Gautier, M., Facon, B. (2016) Is there a genetic paradox of biological invasion? Annual Review of Ecology, Evolution, and Systematics 47: 51–72

Excoffier, L., Lischer, H. E. L. (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics under Linux and Windows. Molecular Ecology Resources 10: 564–567

Ficetola, G. F., Bonin, A., Miaud, C. (2008) Population genetics reveals origin and number of founders in a biological invasion. Molecular Ecology 17: 773–782

Funk, W. C., Garcia, T. S., Cortina, G. A., Hill, R. H. (2011) Population genetics of introduced bullfrogs, *Rana (Lithobates) catesbeianus*, in the Willamette Valley, Oregon, USA. Biological Invasions 13: 651–658

Hammond, S. A., Warren, R. L., Vandervalk, B. P., Kucuk, E., Khan, H., Gibb, E. A., Pandoh, P., Kirk, H., Zhao, Y., Jones, M., Mungall, A. J., Coope, R., Pleasance, S., Moore, R. A., Holt, R. A., Round, J. M., Ohora, S., Walle, B. V., Veldhoen, N., Helbing, C. C., Birol, I. (2017) The North American bullfrog draft genome provides insight into hormonal regulation of long noncoding RNA. Nature Communications 8: 1433

Heard, M. (1904) A California frog ranch. Out West 21: 20-27

Gervasi, S. S., Urbina, J., Chestnut, T., Relyea, R., Blaustein, A. (2013) Experimental evidence for American bullfrog (*Lithobates catesbeianus*) susceptibility to chytrid fungus (*Batrachochytium dendrobatidis*). EcoHealth 10: 166-171

Goebel, A. M., Donnelly, J. M., Atz, M. E. (1999) PCR primers and amplification methods for 125 ribosomal DNA, the control region, cytochrome oxidase I, and cytochrome b in bufonids and other frogs, and an overview of PCR primers which have amplified DNA in amphibians successfully. Molecular Phylogenetics and Evolution 11: 163–199

Janeway CA Jr, Travers P, Walport M, Shlomchik, M. J. Immunobiology: The Immune System in Health and Disease. 5th edition. New York: Garland Science; 2001. The major histocompatibility complex and its functions. Available from: https://www.ncbi.nlm.nih.gov/books/NBK27156/

Kamath, P.L., Sepulveda, A.J., Layhee, M. (2016) Genetic reconstruction of a bullfrog invasion to elucidate vectors of introduction and secondary spread. Ecology and Evolution 6: 5221–5233

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35: 1547–1549

Kiemnec-Tyburczy, K. M., Richmond, J. Q., Savage, A. E., Zamudio, K. R. (2010) Selection, trans-species polymorphism and locus identification of major histocompatibility complex class IIB alleles of New World ranid frogs. Immunogenetics 62: 741–751

Kosakovsky Pond, S. L., Frost, S. D. W., Muse, S. V. (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics, 21: 676–679

Kosakovsky Pond, S. L. Posada, D., Gravenor, M. B., Woelk, C. H., Frost, S. D. W. (2006) Automated phylogenetic detection of recombination using a genetic algorithm. Molecular Biology and Evolution 23: 1891–1901

Kosch, T. A., Bataille, A., Didinger, C., Eimes, J.A., Rodriguez-Brenes, S., Ryan, M. J., Waldman, B. (2016) Major histocompatibility complex selection dynamics in pathogen-infected tungara frog (*Physalaemus pustulosus)* populations. Biology Letters 12: 20160345

LaFond, J. C. (2021) Invasive bullfrogs maintain high levels of immune gene diversity despite elevated BD infections relative to native populations. Master's thesis, University of Central Florida

LaFond, J., Martin, K. R., Dahn, H., Richmond, J. Q., Murphy, R. W., Rollinson, N., Savage, A. E. (2022) Invasive bullfrogs maintain MHC polymorphism including alleles associated with chytrid fungal infection. Integrative and Comparative Biology 62: 262–274

Laufer, G., Gobel, N., Borteiro, C., Soutullo, A., Martinez-Debat, C., O de Sa, R. (2018) Current status of American bullfrog, *Lithobates catesbeianus*, invasion in Uruguay and exploration of chytrid infection. Biological Invasions 20: 285–291

Lillie, M., Dubey, S., Shine, R., Belov, K. (2016) Variation in Major Histocompatibility Complex diversity in invasive cane toad populations. Wildlife Research 44: 565–572

Martin, S. J., Burton, D. R., Roitt, I. M., Delves, P. J. (2017a) Roitt's Essential Immunology. 13th edition. Wiley-Blackwell, New Jersey.

Martin, L. B, Kilvitis, H. J., Brace, A. J., Cooper, L., Haussmann, M. F., Mutati, A., FAsanello, V., O'Brien, S., Ardia, D. R. (2017b) Costs of immunity and their role in the range expansion of the house sparrow in Kenya. Journal of Experimental Biology 220: 2228–2235

Monzon-Arguello, C., Garcia de Leaniz, C., Gajardo, G., Consuegra, S. (2014) Eco-immunology of fish invasions: the role of MHC variation. Immunogenetics 66: 393–402

Moritz, C., Schneider C. J., Wake, D. B. (1992) Evolutionary relationships within the *Ensatina eschscholtzii* complex confirm the ring species interpretation. Systematic Biology 41: 273–291

Mulder, K. P., Cortazar-Chinarro, M., Harris D. J., Crottini, A., Campbell Grant E. H., Fleischer, R. C., Savage, A. E. (2017) Evolutionary dynamics of an expressed MHC class IIβ locus in the Ranidae (Anura) uncovered by genome walking and high-throughput amplicon sequencing. Developmental and Comparative Immunology 76: 177–188

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., Kosakovsky Pond, S. L. (2012) Detecting individual sites subject to episodic diversifying selection. PLoS Genetics 8: e1002764

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., Scheffler, K. (2013) FUBAR: A fast, unconstrained Bayesian approximation for inferring selection. Molecular Biology and Evolution 20: 1196–1205

Savage, A. E., Muletz-Wolz, C. R., Grant, E. H. G, Fleischer, R. C., Mulder, K. P. (2019) Functional variation at an expressed MHC class IIB locus associates with ranavirus infection intensity in larval anuran populations. Immunogenetics 71: 335–346

Savage, A. E., Zamudio, K. R. (2011) MHC genotypes associate with resistance to a frog-killing fungus. Proceedings of the National Academy of Sciences USA 40: 16705–16710

Schriber, K., Lachmuth, S. (2017) The genetic paradox of invasions revisited: the potential role of inbreeding x environment interactions in invasion success. Biology Reviews 92: 939–952

Schloegel, L. M., Picco, A. M., Kilpatrick, A. M., Davies, A. J., Hyatt, A. D., Daszak, P. (2009) Magnitude of the US trade in amphibians and presence of *Batrachochytrium dendrobatidis* and ranavirus infection in imported North American bullfrogs (*Rana catesbeiana*). Biological Conservation 142: 1420–1426

Snow N.P., Witmer, G.W. (2010) American bullfrogs as invasive species: a review of the introduction, subsequent problems, management options, and future directions. Proceedings of the Vertebrate Pest Conference 24: 86–89

Reeves MK, Medley KA, Pinkney AE, Holyoak M, Johnson PTJ, et al. (2013) Localized Hotspots Drive Continental Geography of Abnormal Amphibians on U.S. Wildlife Refuges. PLOS ONE 8: e77467

Teacher, A. G. F., Garner, T. W. J., Nichols, R. A. (2009) Evidence for directional selection at a Novel Major Histocompatibility class I marker in wild common frogs (*Rana temporaria*) exposed to a viral pathogen (Ranavirus). PLOS ONE e4616

Trifinopoulos, Nguyen, L. T., von Haeseler, A., Minh, B. Q. (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research, 44: W232–W235

Ujvari, B., Belov, K. (2011) Major Histocompatibility Complex (MHC) markers in conservation biology. International Journal of Molecular Sciences 12: 5168–5186

Uller, T., Leimu, R. (2011) Founder events predict changes in genetic diversity during human-mediated range expansions. Global Change Biology 17: 3478–3485

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. Molecular Biology and Evolution 35: 773–777

Yap, T. A., Koo, M. S., Ambrose, R. F., Vredenburg, V. T. (2018) Introduced bullfrog facilitates pathogen invasion in the western United States. PLOS ONE 13: e0188384

Zhang, J., Wang, P., Xu, C., Liu, X., Wang, S., Yang, X., Hou, X., Li, X. (2022) Adaptive selection and reshaping of the genetic structure of an immune gene in invasive *Lithobates catesbeianus*. Authorea. May 23, 2022

Zeiner, D.C., W.F. Laudenslayer, Jr., K.E. Mayer, and M. White, eds. (1988-1990) California's Wildlife. Vol. I-III. California Depart. of Fish and Game, Sacramento, California.

# APPENDICES

Appendix A: MHC Class II exon 2 alleles isolated in this study

>RACA_01
TTTAAAGGCCAGTGCTATTATCGTAACGGCACCGAAGATATTTGGAGCGTGA
GCCGTTATATGTATAACCAGGAAGAATATCTGTATTTTGATAGCGATAAAGG
CTTTTTTATTCCGGTGACCGAACTGGGCCGTCTGGATGCGGAAAGCTATAACA
AAAACCCGGATATTCTGGAACGTAACCGTGCGGAAGTGGAAACCATTTGCAA
ACATAACTATCAGCTGTATAAAGCGCTGACCGTGGATCTGAAA
>RACA_02
TGCAAATGCCAGTGCTATTATCGTAACGGCACCGAAGATATTCGTTATCTGAA
ACGTTGGATGTATAACCAGGAAGTGTATGTGTATTTTGATAGCGATAAAGGC
TTTTATATTCCGGTGACCGAATATGGCCGTGTGAGCGCGGATTATTGGAACAA
AGATCCGGATCGTCTGGGCCAGGTGCGTGCGGAAGTGGAAACCGTGTGCAAA
CATAACTATCAGATTTATAAACCGGCGGCGATTGATCTGAAA
>RACA_03
TATAAAGGCCAGTGCTATTTTCGTAACGGCACCGAAGATGTGCGTCTGCTGG
AACGTTATATTTATAACCGTGAAGTGTATGTGTATTTTGATAGCGATAAAGGC
TTTTATATTCCGGTGACCGAATATGGCCGTCCGGATGCGGATTATTGGAACAA
CAACCCGGATATTCTGGGCCAGGCGCGTGCGGTGGTGGAAACCGTGTGCAAA
CATAACTATCAGATTCTGAAACCGATTACCGTGGATCTGAAA
>RACA_04
TATAAAGGCCAGTGCTATTATCGTAACGGCACCGAAGATATTCGTCTGCTGC
AGCGTTATATTTATAACCAGGAAGAATTTCTGTATTTTGATAGCGATAAAGGC
TTTTTTATTCCGGTGACCGAACTGGGCCGTCCGAGCGCGGAAAGCTGGAACA
AAGATCCGGATATTCTGGGCCAGGTGCGTGCGAGCGTGGAAACCATTTGCAA
ACATAACTATCAGATTGATAAACCGCTGACCGTGGATCTGAAA
>RACA_05
TTTAAGGGTCAGTGTTATTACCGGAACGGGACGGAGGATGTCAGGTTTCTGG
CACGTTGGATGTACAATCAGGAGGAGTATCTGTACTTCGACAGTGATAAAGG
ATTCTTCATCCCCGTAACCGAGCTAGGCAGACCGAGCGCTGAGTCCTGGAAC
AAAGATCCGGATATATTAGGACAAGCACGGGCTGAGGTGGAGAGAGTCTGC
AAACCCAACTATCAGATCTACAAACCTGCAGCCATAGACCTGAAA
>RACA_06

TATAAAGCGCAGTGCTATTATCGTAACGGCACCGAAGATATTCGTCTGCTGG
CGCGTTGGATGTATAACCAGGAAGAATATGTGTATTTTGATAGCGATAAAGG
CTTTTTTATTCCGAAAACCGAATATGGCCGTCTGGATGCGGATTATTGGAACA
GCAACCCGGATATTCTGGAAAGCGAACGTGCGGCGGTGGAAACCGTGTGCAA
ACATAACTATCAGATTTATAAACCGGCGGCGATTGATCTGAAA
>RACA_07
TGCAAAGGCCAGTGCTATTTTCGTAACGGCACCGAAGATGTGCGTCTGCTGG
CGCGTTATATTTATAACCAGGAAGTGTATGTGTATTTTGATAGCGATAAAGGC
TTTTATATTCCGAAAACCGAACTGGGCCGTCCGGATGCGGAAAGCTATAACA
AAAACCCGGATATTCTGGGCCAGGCGCGTGCGGCGGTGGAAACCATTTGCAA
ACATAACTATCAGATTGATAAACCGCTGACCGTGGATCTGAAA
>RACA_08
TTTAAATGGCAGTGCTATTATCGTAACGGCACCGAAGATATTTGGAGCGTGA
GCCGTTATATGTATAACCAGGAAGAATATCTGTATTTTGATAGCGATAAAGG
CTTTTTTATTCCGGTGACCGAACTGGGCCGTCTGGATGCGGAAAGCTATAACA
AAAACCCGGATATTCTGGAACGTAACCGTGCGGAAGGCGAAACCATTTGCAA
ACATAACTATCATCTGTATAAAGCGCTGACCGTGAACCTGAAA
>RACA_09
TTTAAAGCGCAGTGCTATTTTCGTAACGGCACCGAAGATATTAAATATCTGGT
GTGCTATATTTATAACCGTGAAGAATTTCTGTATTTTAACAGCGATAAAGGCT
TTTATATTCCGAAAACCGAACTGGGCCGTCCGAGCGCGGAAAGCTGGAACAA
AGATCCGGATATTCTGGGCCAGGTGCGTGCGAGCGTGGTGACCATTTGCAAA
CATAACTATCAGATTGATAAACCGCTGACCGTGGATCTGAAA
>RACA_10
TTTAAAGCGCAGTGCTATTATCGTAACGGCATGGAAGATGTGAAATATCTGG
TGCGTTATATGTATAACCGTGAAGAATATCTGTATTTTGATAGCGATAAAGGC
TTTTATATTCCGAAAACCGAACTGGGCCGTCTGGATGCGGAAAGCTATAACA
AAAACCCGGATATTCTGGGCCAGATTCGTGCGAGCGTGGAACGTATTTGCAA
ACATAACTATCCGATTTTTAAACCGCTGACCGTGGATCTGAAA
>RACA_11
TTTAAAGCGCAGTGCTATTATCGTAACGGCATGGAAGATATTAAATATCTGGT
GCGTTATATGTATAACCGTGAAGAATATCTGTATTTTAACAGCGATAAAGGCT
TTTATATTCCGAAAACCGAACTGGGCCGTCTGGATGCGGAAAGCTATAACAA
AAACCCGGATATTCTGGGCCAGGTGCGTGCGAGCGTGGAACGTATTTGCAAA
CATAACTATCCGATTTTTAAACCGCTGACCGTGGATCTGAAA
>RACA_12

TTTAAAGGCCAGTGCTATTATCGTAACGGCACCGAAGATGTGCGTTTTCTGGC
GCGTTGGATGTATAACCAGGAAGAATATCTGTATTTTGATAGCGATAAAGGC
TTTTTTATTCCGGTGACCGAACTGGGCCGTCCGAGCGCGGAAAGCTGGAACA
AAGATCCGGATATTCTGGGCCAGGCGCGTGCGGAAGTGGAACGTGTGTGCAA
ACCGAACTATCAGATTTATAAACCGGCGGCGATTGATCTGAAA
>RACA_13
TTTAAATGCCAGTGCTATTATCGTAACGGCACCGAAGATATTCGTCTGCTGGC
GCGTTGGATGTATAACCAGGAAGAATATGTGTATTTTGATAGCGATAAAGGC
TTTTTTATTCCGAAAACCGAATATGGCCGTCTGGATGCGGATTATTGGAACAG
CAACCCGGATATTCTGGAAAGCGAACGTGCGGCGGTGGAAACCGTGTGCAAA
CATAACTATCAGATTTATAAACCGGCGGCGATTGATCTGAAA
>RACA_14
TTTAAAGGCCAGTGCTATTATCGTAACGGCACCGAAGATATTCGTTATCTGAA
ACGTTGGACCTATAACCGTGAAGAATTTCTGTATTTTGATAGCGATAAAGGCT
TTTTTATTCCGGTGACCGAACTGGGCCGTCTGGATGCGGAAAGCTATAACAA
AAACCCGGATATTCTGGAACGTAACCGTGCGGAAGTGGAAACCATTTGCAAA
CATAACTATCAGCCGTATAAAGCGCTGACCGTGGATCTGAAA
>RACA_15
TGTAAGTGTCAGTGTTATTACCGGAACGGGACGGAGGATATCAGGTATCTGA
AACGTTGGATGTACAATCAGGAGGTGTATGTGTACTTCGACAGTGATAAGGG
ATTCTACATCCCCGTAACCGAGTATGGGAGAGTGTCCGCTGATTACTGGAAC
AAAGATCCGGATAGATTAGGACAAGTGCGGGCTGAGGTGGAGACAGTCTGC
AAACACAACTATCAGATCTACAAACCTGCAGCCATAGACCTGAAA
>RACA_16
TATAAGGGTCAGTGTTATTTCCGGAACGGGACGGAGGATGTCAGGCTTCTGG
AACGTTACATTTACAATCGGGAGGTGTATGTGTACTTCGACAGTGATAAAGG
ATTCTACATCCCCGTAACCGAGTATGGGAGACCGGACGCTGATTACTGGAAC
AACAATCCGGATATATTAGGACAAGCACGGGCTGTGGTGGAGACAGTCTGCA
AACACAACTATCAGATCCTAAAACCTATAACCGTAGACCTGAAA
>RACA_17
CGTAAGGGTCAGTGTTATTTCCGGAACGGGACGGAGGATGTCAGGCTTCTGG
CACGTTACATTTGCAATCAGGAGGAGTTTCTGTACTTCGACAGTGATAAAGG
ATTCTACATCCCCAAAACCGAGCTGGGCAGACCGGACGCCGAGTCCTATAAC
AAAAATCCGGATATATTAGGACAAGCACGGGCTGCGGTGGAGACAATCTGCA
AACACAACTATCAGATCGACAAACCTCTAACCGTAGACCTGAAA
>RACA_18

TTTAAGTGTCAGTGTTATTACCGGAACGGGACGGAGGATATCAGGTATCTGA
AACGTTGGACGTACAATCGGGAGGAGTTTTTGTACTTCGATAGTGATAAAGG
ATTCTACATCCCCAAAACCGAGCTGGGCAGACCGATCGCTGAGTCCTGGAAC
AAAGATCCGGATAGATTAGGACAAGTACGGGCTGAGGTGGAGACAATCTGC
AAACACAACTATCAGCTCTACAAACCTCTAACCGTAGACCTGAAA

Appendix B: *cytochrome b* haplotypes isolated in this study

>MT1
TGACCTGCCCTCCCCCGCCAATATCTCATCCTGATGAAACTTTGGCTCACTCC
TCGGAGTTTGCCTTGTTGCCCAAATCGTCACTGGCTTATTCCTAGCCATACAC
TATACAGCCGACACTTCCCTTGCATTCTCATCTATCGCTCACATCTGCCGTGA
TGTTAATAACGGCTGACTCCTACGAAATCTCCATGCCAACGGAGCATCATTCT
TTTTTATCTGTATTTATTTCCACATCGGCCGAGGCCTCTACTACGGCTCCTACC
TTTATAAAGAAACATGAAATATCGGTGTAATTCTACTGTTTTTAGTAATAGCT
ACAGCTTTTGTAGGTTACGTCCTACCATGGGGCCAAATATCATTCTGAGGCG
CCACAGTAATTACTAACCTTCTTTCAGCCGCCCCATACATCGGTTCCGACCTT
GTTCAATGAATTTGAGGGGGTTTTTCAGTAGACAACGCCACCCTTACCCGATT
CTTTACTTTTCACTTCATCCTCCCGTTTATTATCGCAGCAGCAAGTATGATCCA
CCTGCTTTTCCTCCATCAAACTGGCTCATCCAACCCAACAGGCCTCAATTCAA
ACTTAGATAAGGTCTCCTTTCATCCATACTTCTCCTACAAAGACTTATTCGGC
TTCACTATTATACTTGGAGCCTTAGCAGCCCTATCAACCTTTGCCCCCAACCT
TCTAGGTGACCCAGACAATTTCACACCAGCTAATCCCCTAGTTACACCCCCT
CACATTAAACCAGAGTGGTACTTCCTATTCGCTTACGCTATTCTCCGCTCCAT
CCCTAATAAACTAGGCGGCGTTCTCGCCCTATTGTTCTCGATCTTAATCCTCTT
CCTAATGCCTATCATCCACACCTCCAAGCTCCGTTCACTCATATTCCGCCCAA
TTGCTAAAATCTTCTTCTGAACC
>H2
TGACCTGCCCTCCCCCGCCAATATCTCATCCTGATGAAACTTTGGCTCACTCC
TCGGAGTTTGCCTTGTTGCCCAAATCGTCACTGGCTTATTCCTAGCCATACAC
TATACAGCCGACACTTCCCTTGCATTCTCATCTATCGCTCACATCTGCCGTGA
TGTTAATAACGGCTGACTCCTACGAAATCTCCATGCCAACGGAGCATCATTCT
TTTTTATCTGTATTTATTTCCACATCGGCCGAGGCCTCTACTACGGCTCCTACC
TTTATAAAGAAACATGAAATATCGGTGTAATTCTACTGTTTTTAGTAATAGCT
ACAGCTTTTGTAGGTTACGTCCTACCATGGGGCCAAATATCGTTCTGAGGCG
CCACAGTAATTACTAACCTTCTTTCAGCCGCCCCATACATCGGTTCCGACCTT
GTTCAATGAATTTGAGGGGGTTTTTCAGTAGACAACGCCACCCTTACCCGATT
CTTTACTTTTCACTTCATCCTCCCGTTTATTATCGCAGCAGCAAGTATGATCCA
CCTGCTTTTCCTCCATCAAACTGGCTCATCCAACCCAACAGGCCTCAATTCAA
ACTTAGATAAGGTCTCCTTTCATCCATACTTCTCCTACAAAGACTTATTCGGC
TTCACTATTATACTTGGAGCCTTAGCAGCCCTATCAACCTTTGCCCCCAACCT
TCTAGGTGACCCAGACAATTTCACACCAGCTAATCCCCTAGTTACACCCCCT
CACATTAAACCAGAGTGGTACTTCCTATTCGCTTACGCTATTCTCCGCTCCAT
CCCTAATAAACTAGGCGGCGTTCTCGCCCTATTGTTCTCGATCTTAATCCTCTT

CCTAATGCCTATCATCCACACCTCCAAGCTCCGTTCACTCATATTCCGCCCAA
TTGCTAAAATCTTCTTCTGAACC
>H7
TGACCTGCCCTCCCCCGCCAATATCTCATCCTGATGAAACTTTGGCTCACTCC
TCGGAGTTTGCCTTGTTGCCCAAATCGTCACTGGCTTATTCCTAGCCATACAC
TATACAGCCGACACTTCCCTTGCATTCTCATCTATCGCTCACATCTGCCGTGA
TGTTAATAACGGCTGACTCCTACGAAATCTCCATGCCAACGGAGCATCATTCT
TTTTTATCTGTATTTATTTCCACATCGGCCGAGGCCTCTACTACGGCTCCTACC
TTTATAAAGAAACATGAAATATCGGTGTAATTCTACTGTTTTTAGTAATAGCT
ACAGCTTTTGTAGGTTACGTTCTACCATGGGGCCAAATATCATTCTGAGGCG
CCACAGTAATTACTAACCTTCTTTCAGCCGCCCCATACATCGGTTCCGACCTT
GTTCAATGAATTTGAGGGGGTTTTTCAGTAGACAACGCCACCCTTACCCGATT
CTTTACTTTTCACTTCATCCTCCCGTTTATTATCGCAGCAGCAAGTATGATCCA
CCTGCTTTTCCTCCATCAAACTGGCTCATCCAACCCAACAGGCCTCAATTCAA
ACTTAGATAAGGTCTCCTTTCATCCATACTTCTCCTACAAAGACTTATTCGGC
TTCACTATTATACTTGGAGCCTTAGCAGCCCTATCAACCTTTGCCCCCAACCT
TCTAGGTGACCCAGACAATTTCACACCAGCTAATCCCCTAGTTACACCCCCT
CACATTAAACCAGAGTGGTACTTCCTATTCGCTTACGCCATTCTCCGCTCCAT
CCCTAATAAACTAGGCGGCGTTCTCGCCCTATTGTTCTCGATCTTAATCCTCTT
CCTAATGCCTATCATCCACACCTCCAAGCTCCGTTCACTCATATTCCGCCCAA
TTGCTAAAATCTTCTTCTGAACC
>H11
TGACCTCCCCTCCCCCGCCAATATCTCATCCTGATGAAACTTTGGCTCACTCC
TCGGAGTATGCCTTGTTACCCAAATCGTCACTGGCTTATTCCTAGCCATACAC
TATACAGCTGACACTTCCCTTGCATTCTCATCTATCGCTCACATCTGCCGTGAT
GTTAATAACGGCTGACTCCTACGAAATCTCCATGCCAACGGAGCATCATTCTT
TTTTATCTGTATTTATTTTCACATCGGCCGAGGTCTCTACTACGGCTCCTACCT
TTATAAAGAAACATGAAATATCGGTGTAATTCTACTGTTTTTAGTAATAGCTA
CAGCTTTTGTAGGTTACGTTCTACCATGGGGCCAAATATCATTCTGAGGCG
CCACAGTAATTACTAACCTTCTCTCAGCCGCCCCATACATCGGTTCTGACCTT
GTTCAATGGATTTGAGGGGGTTTTTCAGTAGACAACGCCACCCTTACCCGATT
CTTCACTTTTCACTTCATCCTCCCGTTTATTATCGCAGCAGCAAGTATGATCCA
CCTGCTTTTCCTCCATCAAACTGGCTCATCCAACCCAACAGGCCTCAATTCAA
ACTTAGATAAGGTCTCCTTTCATCCATACTTCTCCTACAAAGACTTATTTGGCT
TCACTATTATACTTGGAGCCTTAGCAGCCCTATCAACCTTTGCCCCCAACCTT
CTAGGTGACCCAGACAACTTCACACCAGCTAATCCCCTAGTTACACCCCCT

CACATTAAACCAGAGTGGTATTTCCTATTCGCCTACGCCATTCTCCGCTCCAT
CCCTAATAAACTAGGCGGTGTTCTCGCCCTATTGTTCTCGATCTTAATCCTCTT
CCTAATGCCTATCATCCACACCTCCAAACTCCGTTCACTCATATTCCGCCCAA
TTGCTAAAATCTTCTTCTGAACC
>CA1
TGACCTNCCCTCCCNCGCCAATATNTNATCNTGATGAAACTTTNGCTCACTCC
TCGGAGTNTGCNTTGTTGCCCAAATCGTCACTGGCTTATTCCTAGCCATACAC
TATACAGCCAACACTTCCCTTGCATTCTCATCTATCGCTCACATCTGCCGTGA
TGTTAATAACGGCTGACTCCTACGAAATCTCCATGCCAACGGAGCATCATTCT
TTTTTATCTGTATTTATTTCCACATCGGCCGAGGCCTCTACTACGGCTCCTACC
TTTATAAAGAAACATGAAATATCGGTGTAATTCTACTGTTTTTAGTAATAGCT
ACAGCTTTTGTAGGTTACGTTCTACCATGGGGCCAAATATCATTCTGAGGCGC
CACAGTAATTACTAACCTTCTTTCAGCCGCCCCATACATCGGTTCCGACCTTG
TTCAATGAATTTGAGGGGGTTTTTCAGTAGACAACGCCACCCTTACCCGATTC
TTTACTTTTCACTTCATCCTCCCGTTTATTATCGCAGCAGCAAGTATGATCCAC
CTGCTTTTCCTCCATCAAACTGGCTCATCCAACCCAACAGGCCTCAATTCAAA
CTTAGATAAGGTCTCCTTTCATCCATACTTCTCCTACAAAGACTTATTCGGCTT
CACTATTATACTTGGAGCCTTAGCAGCCCTATCAACCTTTGCCCCCAACCTTC
TAGGTGACCCAGACAATTTCACACCAGCTAATCCCCTAGTTACACCCCCTCAC
ATTAAACCAGAGTGGTACTTCCTATTCGCTTACGCTATTCTCCGCTCCATCCCT
AATAAACTAGGCGGCGTTCTCGCCCTATTGTTCTCGATCTTAATCCTCTTCCTA
ATGCCTATCATCCACACCTCCAAGCTCCGTTCACTCATATTCCGCCCAATTGC
TAAAATCTTCTTCTGAACC
>CA2
TGACCTGCCCTCCCCCGCCAATATCTCATCCTGATGAAACTTTGGCTCACTCC
TCGGAGTTTGCCTTGTTGCCCAAATCGTCACTGGCTTATTCCTAGCCATACAC
TATACAGCCGACACTTCCCTTGCATTCTCATCTATCGCTCACATCTGCCGTGA
TGTTAATAACGGCTGACTCCTACGAAATCTCCATGCCAACGGAGCATCATTCT
TTTTTATCTGTATTTATTTCCACATCGGCCGAGGCCTCTACTACGGCTCCTACC
TTTATAAAGAAACATGAAATATCGGTGTAATTCTACTGTTTTTAGTAATAGCT
ACAGCTTTTGTAGGTTACGTCCTACCATGGGGCCAAATATCATTCTGAGGCGC
CACAGTAATTACTAACCTTCTTTCAGCCGCCCCATACATCGGTTCCGACCTTG
TTCAATGAATTTGAGGGGGTTTTTCAGTAGACAACGCCACCCTTACCCGATTC
TTTACTTTTCACTTCATCCTCCCGTTTATTATCGCAGCAGCAAGTATGATCCAC
CTGCTTTTCCTCCATCAAACTGGCTCATCCAACCCAACAGGCCTCAACTCAAA
CTTAGATAAGGTCTCCTTTCATCCATACTTCTCCTACAAAGACTTATTCGGCTT
CACTATTATACTTGGAGCCTTAGCAGCCCTATCAACCTTTGCCCCCAACCTTC

TAGGTGACCCAGACAATTTCACACCAGCTAATCCCCTAGTTACACCCCCTCAC
ATTAAACCAGAGTGGTACTTCCTATTCGCTTACGCTATTCTCCGCTCCATCCCT
AATAAACTAGGCGGCGTTCTCGCCCTATTGTTCTCGATCTTAATCCTCTTCCTA
ATGCCTATCATCCACACCTCCAAGCTCCGTTCACTCATATTCCGCCCAATTGC
TAAAATCTTCTTCTGAACC
>CA5
TGACCTCCCNTCCCCCGCCAATATCTCATCCTGATGAAACTTTGGCTCACTCC
TCGGAGTATGCCTTGTTGCCCAAATCGTCACTGGCTTATTCCTAGCCATACAC
TATACAGCTGACACTTCCCTTGCATTCTCATCTATGGCTCACATCTGCCGTGA
TGTTAATAACGGCTGACTCCTACGAAATCTCCATGCCAACGGAGCATCATTCT
TTTTTATCTGTATTTATTTTCACATCGGCCGAGGTCTCTACTACGGCTCCTACC
TTTATAAAGAAACATGAAATATCGGTGTAATTCTACTGTTTTTAGTAATAGCT
ACAGCTTTTGTAGGTTACGTTCTACCATGGGGCCAAATATCATTCTGAGGCGC
CACAGTAATTACTAACCTTCTCTCAGCCGCCCCATACATCGGTTCTGACCTTG
TTCAATGAATTTGAGGGGGTTTTTCAGTAGACAACGCCACCCTCACCCGATTC
TTTACTTTTCACTTCATCCTCCCGTTTATTATCGCAGCAGCAAGTATGATCCAC
CTGCTTTTCCTCCATCAAACTGGCTCATCCAACCCAACAGGCCTCAATTCAAA
CTTAGATAAGGTCTCCTTTCATCCATACTTCTCCTACAAAGACTTATTTGGCTT
CACTATTATACTTGGAGCCTTAGCAGCCCTATCAACCTTTGCCCCCAACCTTC
TAGGTGACCCAGACAATTTCACACCAGCTAATCCCCTAGTTACACCCCCTCAC
ATTAAACCAGAGTGGTATTTCCTATTCGCCTACGCCATTCTCCGCTCCATCCC
TAATAAACTAGGCGGTGTTCTCGCCCTATTGTTCTCGATCTTAATCCTCTTCCT
AATGCCTATCATCCACACCTCCAAACTCCGTTCACTCATATTCCGCCCAATTG
CTAAAATCTTCTTCTGAACC
>CA8
TGACCTGCCCTCCCCCGCCAATATCTCATCCTGATGAAACTTTGGCTCACTCC
TCGGAGTTTGCCTTGTTGCCCAAATCGTCACTGGCTTATTCCTAGCCATACAC
TATACAGCCGACACTTCCCTTGCATTCTCATCTATCGCTCACATCTGCCGTGA
TGTTAATAACGGCTGACTCCTACGAAATCTCCATGCCAACGGAGCATCATTCT
TTTTTATCTGTATTTATTTCCACATCGGCCGAGGCCTCTACTACGGCTCCTACC
TTTATAAAGAAACATGAAATATCGGTGTAATTCTACTGTTTTTAGTAATAGCT
ACAGCTTTTGTAGGTTACGTTCTACCATGGGGCCAAATATCATTCTGAGGCGC
CACAGTAATTACTAACCTTCTTTCAGCCGCCCCATACATCGGTTCCGACCTTG
TTCAATGAATTTGAGGGGGTTTTTCAGTAGACAACGCCACCCTTACCCGATTC
TTTACTTTTCACTTCATCCTCCCGTTTATTATCGCAGCAGCAAGTATGATCCAC
CTGCTTTTCCTCCATCAAACTGGCTCATCCAACCCAACAGGCCTCAATTCAAA
CTTAGATAAGGTCTCCTTTCATCCATACTTCTCCTACAAAGACTTATTCGGCTT

CACTATTATACTTGGAGCCTTAGCAGCCCTATCAACCTTTGCCCCCAACCTTC
TAGGTGACCCAGACAATTTCACACCAGCTAATCCCCTAGTCACACCCCCTCAC
ATTAAACCAGAGTGGTACTTCCTATTCGCTTACGCCATTCTCCGCTCCATCCC
TAATAAACTAGGCGGCGTTCTCGCCCTATTGTTCTCGATCTTAATCCTCTTCCT
AATGCCTATCATCCACACCTCCAAGCTCCGTTCACTCATATTCCGCCCAATTG
CTAAAATCTTCTTCTGAACC
>CA9
TGACCTGCCCTCCCCCGCCAATATCTCATCCTGATGAAACTTTGGCTCACTCC
TCGGAGTTTGCCTTGTTGCCCAAATCGTCACTGGCTTATTCCTAGCCATACAC
TATACAGCCGACACTTCCCTTGCATTCTCATCTATCGCTCACATCTGCCGTGA
TGTTAATAACGGCTGACTCCTACGAAATCTCCATGCCAACGGAGCATCATTCT
TTTTTATCTGTATTTATTTCCACATCGGCCGAGGCCTCTACTACGGCTCCTACC
TTTATAAAGAAACATGAAATATCGGTGTAATTCTACTGTTTTTAGTAATAGCT
ACAGCTTTTGTAGGTTACGTTCTACCATGGGGCCAAATATCATTCTGAGGCGC
CACAGTAATTACTAACCTTCTCTCAGCCGCCCCATACATCGGTTCCGACCTTG
TTCAATGAATTTGAGGGGGTTTTTCAGTAGACAACGCCACCCTTACCCGATTC
TTTACTTTTCACTTCATCCTCCCGTTTATTATCGCAGCAGCAAGTATGATCCAC
CTGCTTTTCCTCCATCAAACTGGCTCATCCAACCCAACAGGCCTCAATTCAAA
CTTAGATAAGGTCTCCTTTCATCCATACTTCTCCTACAAAGACTTATTCGGCTT
CACTATTATACTTGGAGCCTTAGCAGCCCTATCAACCTTTGCCCCCAACCTTC
TAGGTGACCCAGACAATTTCACACCAGCTAATCCCCTAGTTACACCCCCTCAC
ATTAAACCAGAGTGGTACTTCCTATTTGCTTACGCCATTCTCCGCTCCATCCCT
AATAAACTAGGCGGCGTTCTCGCCCTATTGTTCTCGATCTTAATCCTCTTCCTA
ATGCCTATCATCCACACCTCCAAGCTCCGTTCACTCATATTCCGCCCAATTGC
TAAAATCTTCTTCTGAACC
>CA10
TGACCTGCCCTCCCCCGCCAATATCTCATCCTGATGAAACTTTGGCTCACTCC
TCGGAGTTTGCCTTGTTGCCCAAATCGTCACTGGCTTATTCCTAGCCATACAC
TATACAGCCGACACTTCCCTTGCATTCTCATCTATCGCTCACATCTGCCGTGA
TGTTAATAACGGCTGACTCCTACGAAATCTCCATGCCAACGGAGCATCATTCT
TTTTTATCTGTATTTATTTCCACATCGGCCGAGGTCTCTACTACGGCTCCTACC
TTTATAAAGAAACATGAAATATCGGTGTAATTCTACTGTTTTTAGTAATAGCT
ACAGCTTTTGTAGGTTACGTTCTACCATGAGGCCAAATATCATTCTGAGGCGC
CACAGTAATTACTAACCTTCTTTCAGCCGCCCCATACATCGGTTCCGACCTTG
TTCAATGAATTTGAGGGGGGTTTTCAGTAGACAACGCCACCCTTACCCGATTC
TTTACTTTTCACTTCATCCTCCCGTTTATTATCGCAGCAGCAAGTATGATCCAC
CTGCTTTTCCTCCATCAAACTGGCTCATCCAACCCAACAGGCCTCAATTCAAA

CTTAGATAAGGTCTCCTTTCATCCATACTTCTCCTACAAAGACTTATTCGGCTT
CACTATTATACTTGGAGCCTTAGCAGCCCTATCAACCTTTGCCCCCAACCTTC
TAGGTGACCCAGACAATTTCACACCAGCTAATCCCCTAGTTACACCCCCTCAC
ATTAAACCAGAGTGGTACTTCCTATTCGCTTACGCCATTCTCCGCTCCATCCC
TAATAAACTAGGCGGCGTTCTCGCCCTATTGTTCTCGATCTTAATCCTCTTCCT
AATGCCTATCATCCACACCTCCAAGCTCCGTTCACTCATATTCCGCCCAATTG
CTAAAATCTTCTTCTGAACC

Appendix C: MHC Class II Pairwise Distance Matrix

| | R01 | R02 | R03 | R04 | R05 | R06 | R07 | R08 | R09 | R10 | R11 | R12 | R13 | R14 | R15 | R16 | R17 | R18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| R01 | | | | | | | | | | | | | | | | | | |
| R02 | 0.2063 | | | | | | | | | | | | | | | | | |
| R03 | 0.1984 | 0.1508 | | | | | | | | | | | | | | | | |
| R04 | 0.1349 | 0.1548 | 0.1429 | | | | | | | | | | | | | | | |
| R05 | 0.3214 | 0.2817 | 0.3175 | 0.2937 | | | | | | | | | | | | | | |
| R06 | 0.1825 | 0.1429 | 0.1508 | 0.1825 | 0.2976 | | | | | | | | | | | | | |
| R07 | 0.1429 | 0.1825 | 0.1071 | 0.1032 | 0.3135 | 0.1627 | | | | | | | | | | | | |
| R08 | 0.0278 | 0.2262 | 0.2262 | 0.1627 | 0.3373 | 0.2024 | 0.1706 | | | | | | | | | | | |
| R09 | 0.1944 | 0.2103 | 0.1984 | 0.0952 | 0.3135 | 0.2262 | 0.1429 | 0.2143 | | | | | | | | | | |
| R10 | 0.1548 | 0.2381 | 0.1905 | 0.1548 | 0.3294 | 0.2103 | 0.1310 | 0.1746 | 0.1151 | | | | | | | | | |
| R11 | 0.1587 | 0.2302 | 0.1984 | 0.1468 | 0.3294 | 0.2063 | 0.1389 | 0.1786 | 0.0913 | 0.0238 | | | | | | | | |
| R12 | 0.1746 | 0.1190 | 0.1786 | 0.1151 | 0.2183 | 0.1587 | 0.1508 | 0.2024 | 0.1746 | 0.1706 | 0.1786 | | | | | | | |
| R13 | 0.1746 | 0.1310 | 0.1508 | 0.1825 | 0.2937 | 0.0159 | 0.1587 | 0.1944 | 0.2341 | 0.2183 | 0.2143 | 0.1508 | | | | | | |
| R14 | 0.0635 | 0.1905 | 0.1905 | 0.1389 | 0.3254 | 0.1865 | 0.1587 | 0.0913 | 0.1825 | 0.1548 | 0.1587 | 0.1706 | 0.1786 | | | | | |
| R15 | 0.3452 | 0.2143 | 0.3095 | 0.3254 | 0.0992 | 0.3016 | 0.3373 | 0.3532 | 0.3413 | 0.3611 | 0.3492 | 0.2976 | 0.2937 | 0.3413 | | | | |
| R16 | 0.3413 | 0.3095 | 0.2103 | 0.3056 | 0.1230 | 0.3016 | 0.2817 | 0.3571 | 0.3214 | 0.3333 | 0.3333 | 0.3254 | 0.3056 | 0.3452 | 0.1151 | | | |
| R17 | 0.3095 | 0.3532 | 0.2976 | 0.2857 | 0.1111 | 0.3254 | 0.2341 | 0.3254 | 0.2857 | 0.3016 | 0.3016 | 0.3175 | 0.3254 | 0.3175 | 0.1667 | 0.1032 | | |
| R18 | 0.3095 | 0.2857 | 0.3294 | 0.2857 | 0.0992 | 0.3333 | 0.3095 | 0.3175 | 0.2817 | 0.3095 | 0.3095 | 0.3016 | 0.3214 | 0.2817 | 0.1071 | 0.1429 | 0.1111 | |

*The alleles names have been shortened; R stands for RACA.

Appendix D: *cytochrome b* Pairwise Distance Matrix

| | CA1 | CA2 | CA3 | CA4 | CA5 | CA6 | CA7 | CA8 | CA9 | CA10 | MT1 | H11 | H2 | H7 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CA1 | | | | | | | | | | | | | | |
| CA2 | 0.00325 | | | | | | | | | | | | | |
| CA3 | 0.00433 | 0.00542 | | | | | | | | | | | | |
| CA4 | 0.00975 | 0.01083 | 0.00758 | | | | | | | | | | | |
| CA5 | 0.01517 | 0.01838 | 0.01300 | 0.00542 | | | | | | | | | | |
| CA6 | 0.00433 | 0.00324 | 0.00650 | 0.01192 | 0.01946 | | | | | | | | | |
| CA7 | 0.00542 | 0.00432 | 0.00758 | 0.01300 | 0.02054 | 0.00108 | | | | | | | | |
| CA8 | 0.00325 | 0.00432 | 0.00108 | 0.00867 | 0.01622 | 0.00541 | 0.00649 | | | | | | | |
| CA9 | 0.00433 | 0.00541 | 0.00433 | 0.00975 | 0.01514 | 0.00649 | 0.00757 | 0.00324 | | | | | | |
| CA10 | 0.00433 | 0.00541 | 0.00433 | 0.00975 | 0.01514 | 0.00649 | 0.00757 | 0.00324 | 0.00432 | | | | | |
| MT1 | 0.00217 | 0.00108 | 0.00433 | 0.00975 | 0.01730 | 0.00216 | 0.00324 | 0.00324 | 0.00432 | 0.00432 | | | | |
| H11 | 0.01733 | 0.02054 | 0.01517 | 0.00975 | 0.00649 | 0.02162 | 0.02270 | 0.01838 | 0.01730 | 0.01730 | 0.01946 | | | |
| H2 | 0.00325 | 0.00216 | 0.00542 | 0.01083 | 0.01838 | 0.00108 | 0.00216 | 0.00432 | 0.00541 | 0.00541 | 0.00108 | 0.02054 | | |
| H7 | 0.00217 | 0.00324 | 0.00217 | 0.00758 | 0.01514 | 0.00432 | 0.00541 | 0.00108 | 0.00216 | 0.00216 | 0.00216 | 0.01730 | 0.00324 | |

Appendix E: MHC Class II amino acid alignment

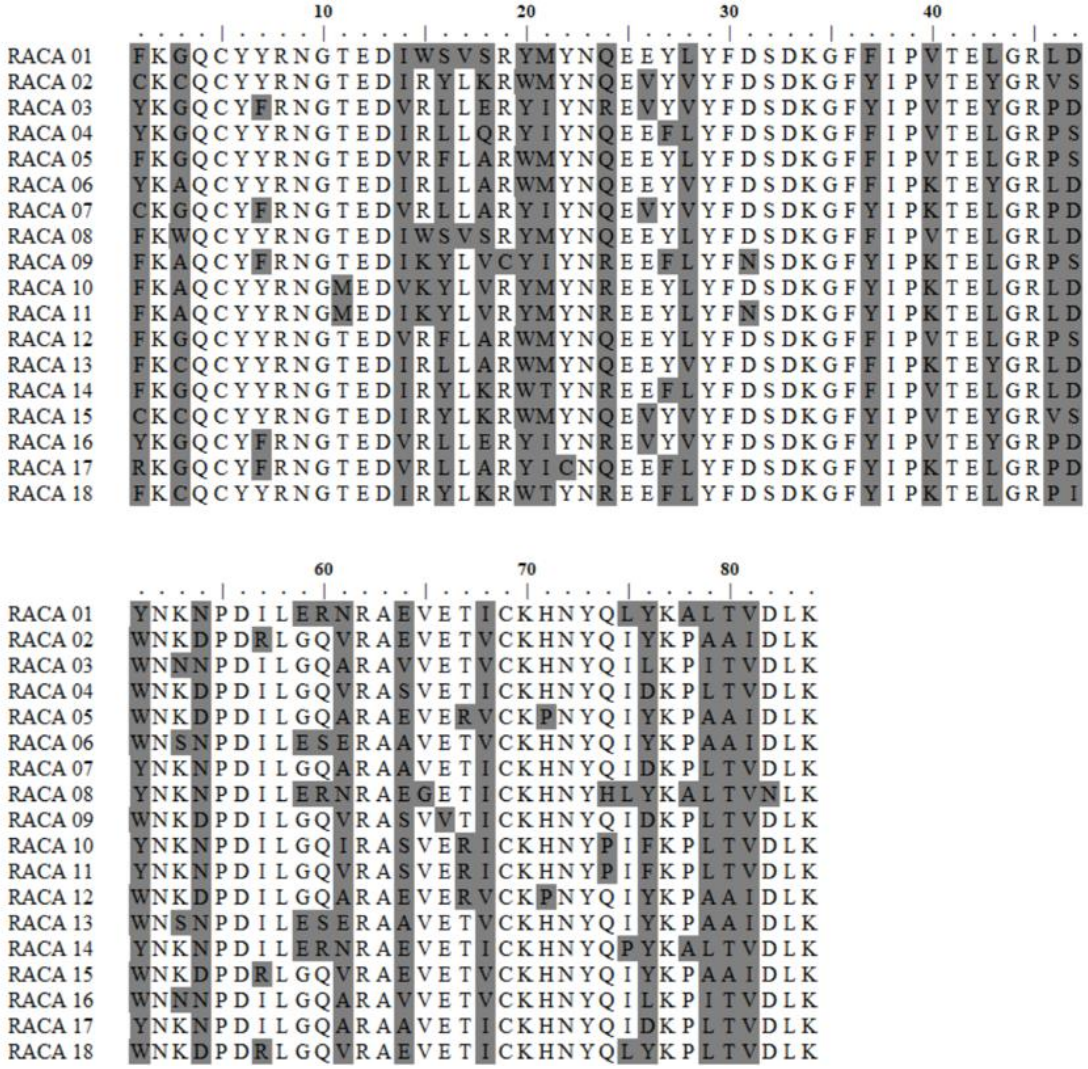


Figure A. The shading represents amino acid residues that differ in between the alleles.