Comparative immunogenetics, microbiomics, and wildlife disease ecology across sympatric imperiled and invasive cottontails

Kimberly B. Neil

Advisor: David M. Rand

A dissertation submitted to the Graduate School of Brown University in fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Ecology and Evolutionary Biology

Providence, RI
October 2020

This dissertation by Kimberly B. Neil is accepted in its present form
by the Department of Ecology and Evolutionary Biology as satisfying the
dissertation requirement for the degree of Doctor of Philosophy

Date_____ _____
David M. Rand, Advisor

Recommended to the Graduate Council

Date_____ _____
Thomas J. McGreevy, Jr., Reader

Date_____ _____
Sohini Ramachandran, Reader

Date_____ _____
Daniel M. Weinreich, Reader

Approved by the Graduate Council

Date_____ _____
Andrew G. Campbell, Dean of the Graduate School

Kimberly B. Neil (née Cohen) kimberly_cohen@brown.edu

EDUCATION

Brown University, Providence, RI 2020
Ph.D., Department of Ecology and Evolutionary Biology
Primary Supervisor: Dr. David Rand

Colby College, Waterville, ME B.A. 2009
Major: Biology, Minor: Classical Civilization,
Summa cum laude, Distinction in Biology, Honors in Biology

AWARDS, GRANTS, AND FELLOWSHIPS

Grant-in-Aid of Research, American Society of Mammalogists	2019
Bushnell Dissertation Improvement Grant	2018
NSF Graduate Research Fellow	2015-2018
IBES Research and Training Award	2016, 2015
NSF IGERT Fellow	2013-2015
Bushnell Research and Education Award	2013
Phi Beta Kappa, Colby College	2009
Webster Chester Biology Prize, Colby College	2009
Students Special Projects Grant, Colby College	2008

RESEARCH EXPERIENCE

Ph.D. candidate, Ecology and Evolutionary Biology 2013-2020
Brown University, Providence, RI
Laboratory of Dr. David Rand

Investigating functionally significant variation and its drivers in an imperiled and invasive mammal pair, the New England cottontail (*Sylvilagus transitionalis*) and the eastern cottontail (*S. floridanus*). Research projects encompass:

- Neutral population structure, immunogenetic diversity, and selection regimes at the major histocompatibility complex and toll-like receptors
- Host-associated ear microbiota, genetic and immunogenetic drivers, and potential pathogens
- Fecal microbiome across native, invasive, and reintroduced cottontail populations

Fellow, NSF IGERT Program 2013-2015
Brown University, Providence, RI

Worked within an interdisciplinary peer team to investigate host-parasite dynamics in an amphipod and its phenotype-altering trematode parasite using a combination of whole genome sequencing and transcriptomics

Ph.D. student, **Ecology, Behavior, and Evolution** 2011-2013
Boston University, Boston, MA
Laboratory of Dr. Tom Kunz (lab closed in 2013 due to medical leave)

Investigated immune system genes and immune transcriptomics in little brown bats (*Myotis lucifugus*) infected with the epizootic white-nose syndrome

Research Assistant, **Center for Brain Science** 2009-2011
Harvard University, Boston, MA
Laboratory of Dr. Joshua Sanes
Generated novel transgenic mice and virus constructs engineered for neuron-specific fluorescence ("Brainbow"), assisted in genotyping mouse colony, and trained new research technicians

Senior Honors Researcher, **Biology Department** 2008-2009
Colby College, Waterville, ME
Laboratory of Dr. Cathy Bevier
Examined effect of larval diet on flight endurance and lipid content in the painted lady butterfly (*Vanessa cardui*)

Wildlife Medicine Intern, **New England Wildlife Center** Summer 2008
South Weymouth, MA
Rotated through wards of the wildlife hospital and rehabilitation center performing duties related to veterinary care

Research Intern, Ocean Mammal Institute Jan. 2008
Maui, HI
Researched the effects of underwater noise pollution on behaviors of the North Pacific humpback whale (*Megaptera novaeangliae*) as part of a highly diverse, multi-institutional team

PUBLICATIONS

Nunez JCB, Flight PA, **Neil KB**, Rong S, Eriksson LA, Ferranti DA, Rosenblad MA, Blomberg A, Rand DM (2020) Footprints of natural selection at the mannose-6-phosphate isomerase locus in barnacles. Proc Natl Acad Sci USA 117(10):5376-5385

Cai D, **Cohen KB**, Luo T, Lichtman JW, Sanes JR (2013) Improved Tools for the Brainbow Toolbox. Nature Methods 10(6):540-547

RESEARCH TALKS

Immunogenetics and potential pathogens of the New England cottontail (*Sylvilagus transitionalis*) June 2018
American Society of Mammalogists • Manhattan, KS
Authors: Neil KB, McGreevy TJ, Rand DM

Potential pathogens of New England's cottontails May 2018
Department of Ecology and Evolutionary Biology Seminar
Brown University • Providence, RI

Immunogenetics and potential pathogens of the New England cottontail (*Sylvilagus transitionalis*) April 2018
The Northeast Association of Fish and Wildlife Agencies • Burlington, VT • Invited Talk
Authors: Neil KB, McGreevy TJ, Rand DM

Monitoring for pathogens and parasites in the New England cottontail Jan. 2018

New England Cottontail Technical Committee Meeting • Dover, NH
Authors: Neil KB, McGreevy TJ, Rand DM

Immunogenetic variation of the New England cottontail Jan. 2017
New England Cottontail Technical Committee Meeting • Charlestown, RI
Authors: Neil KB, McGreevy TJ, Rand DM

A tale of two genomes: metatranscriptomics of a host-parasite system Nov. 2016
Department of Ecology and Evolutionary Biology Seminar
Brown University • Providence, RI

From bats to bunnies: white-nose syndrome and disease ecology of New England's cottontails Nov. 2013
Department of Ecology and Evolutionary Biology Seminar
Brown University • Providence, RI

Diet manipulation in *Vanessa cardui*: effect on larval and adult morphology, lipid content, and flight capability April 2009
Colby College Undergraduate Research Symposium • Waterville, ME
Authors: Cohen KB and Bevier CR

POSTER PRESENTATIONS

Immunogenetic variation across New England's invasive and imperiled cottontails June 2019
Evolution Meeting • Providence, RI
Authors: Neil KB, McGreevy TJ, Rand DM

Parasite manipulation of host gene expression in a trematode-amphipod system June 2019
Evolution Meeting • Providence, RI
Authors: Ferranti D, Nunez CBJ, Williams S, Neil KB, Spierer A, Morgan D, Brown B, Greenhill M, Okami N, Lyons A, Johnson D, Rand DM

Immunogenetic variation in a declining native rabbit and its introduced congener June 2017
American Society of Mammalogists Meeting • Moscow, ID
Authors: Neil KB, McGreevy TJ, Rand DM

Metaparasitomics: A proposal to investigate amphipod and trematode genomes at Plum Island Ecosystems LTER Nov. 2013
Marine Biological Laboratory Symposium • Cape Cod, MA
Authors: Cohen KB, Kinney K, McKerrow W, Osborne B, Spierer A, Smith KF, Giblin A, Cardon Z, Rand DM

Characterizing major histocompatibility complex class II diversity in *Myotis lucifugus* Oct. 2012
North American Society for Bat Research Symposium • San Juan, Puerto Rico
Authors: Cohen KB, Kunz TH, Sorenson M

Diet manipulation in *Vanessa cardui*: effect on larval and adult morphology, lipid content, and flight capability April 2009

Maine Biological and Medical Sciences Symposium • Salisbury Cove, ME
Authors: Cohen KB and Bevier CR

TEACHING EXPERIENCE

Guest Lecturer: Ecology and Evolution of Infectious Disease (BI 0380) Oct. 2019, 2018, 2017, 2016
Brown University • Providence, RI
Invited to deliver a lecture titled 'Disease and immunology of New England's cottontails'

Guest Lecturer: Ecology and Evolution of Infectious Disease (BI 0380) Oct. 2019, 2018, 2017, 2016
Brown University • Providence, RI
Invited to design and deliver a lecture titled 'Disease ecology and Lyme disease'

Graduate Mentor, **Leadership Alliance** Summer 2019
Brown University, Providence, RI
Co-designed and led a weekly support section designed to enhance professional development of undergraduate mentees in preparation for summer research, science communication, and applying to graduate research programs

Teaching Assistant, **Ecology and Evolution of Infectious Disease** Fall 2017; 2016
Brown University, Providence, RI
Co-designed and led weekly discussion sections, conducted office hours, and graded problem sets

Guest Lecturer: Conservation Medicine (BI 0140) Oct. 2015
Brown University • Providence, RI
Invited to design and deliver a lecture titled 'Land use change and disease'

Guest Lecturer: Evolutionary Ecology (BI 303) May 2013
Boston University • Boston, MA
Invited to design and deliver a guest lecture titled 'The ecology of wildlife disease: white-nose syndrome'

Teaching Fellow, **Evolutionary Ecology Laboratory** Spring 2013
Boston University, Boston, MA
Autonomously developed weekly laboratory exercises, modeled proper laboratory techniques and safety, designed and evaluated assignments, and held weekly office hours

Teaching Fellow, **Introductory Biology Laboratory** Fall 2011
Boston University, Boston, MA
Instructed two laboratory sections designed to introduce students to foundational "wet lab" skills and the scientific method, attended weekly preparatory meetings, and evaluated student assignments

TEACHING LEADERSHIP

Teaching Fellow, Sheridan Center for Teaching and Learning Spring 2020; 2019
Brown University, Providence, RI
Conducted research on undergraduate perceptions of teaching and learning in biology discussion and laboratory sections, with special emphasis on best practices for retention of historically underrepresented groups

Senior Teaching Consultant, **Sheridan Center for Teaching and Learning** 2015-2020
Brown University, Providence, RI
Observe Brown community members in the classroom and consult on their teaching practices and classroom environment

Workshop Leader, **Seminar in Reflective Teaching** Fall 2017; 2018; 2019
The Sheridan Center for Teaching and Learning, Brown University, Providence, RI
Led a four-part workshop on evidence-based teaching practices relating to critical reflection, inclusive classrooms, classroom communication, learning design, and engaged learning

Ivy Plus Teaching Transformations Graduate Student Summit May 3, 2019
Columbia University, New York, NY
Representative for Brown's Sheridan Center for Teaching and Learning

Workshop Leader, Feedback and Grading in STEM Oct. 2, 2018; Sept. 5, 2017
The Sheridan Center for Teaching and Learning, Brown University, Providence, RI
Led workshop for STEM graduate students and postdocs, with special emphasis on maintaining an inclusive classroom, implementing rubrics, and providing both formative and summative feedback

Course, **Course Design** Jan.- May 2018
The Sheridan Center for Teaching and Learning, Brown University, Providence, RI
Interactive seminar to individually develop two semester-long, college-level courses that employ active learning techniques and principles of universal design

Course, Seminar in Reflective Teaching Sept. 2013-May 2014
The Sheridan Center for Teaching and Learning, Brown University, Providence, RI
Year-long certificate program that instructs graduate students in evidence-based pedagogical practices that foster excellent communication, inclusive teaching, and active reflection

MENTORING

Rob Blattner – Developed research project on cottontail immunogenetics and secured competitive Voss Fellowship to fund laboratory work. Trained in PCR, Sanger sequencing, data analysis, and foundational R skills. 2017-2019

Emily Longman – Met through Brown University's New Scientist Program. Trained to assist with PCR, gel electrophoresis, and preparing amplicons for Sanger sequencing. Secured Royce Fellowship. 2014-2015

Paul Rizzo – Trained to assist with molecular "wet lab" work, including DNA extractions, PCR, and gel electrophoresis. Secured Undergraduate Teaching and Research Award (UTRA) funding. 2013-2014

DIVERSITY AND ACADEMIC SERVICE

Member, American Society of Mammalogists Human Diversity Committee	2018-Present
Graduate Student Liaison, Sheridan Center for Teaching and Learning, Brown University	2019-2020
Learning Community Fellow: Community Engaged Scholarship, Swearer Center, Brown University	2019-2020
Panel Organizer, Leadership Alliance Graduate Student Q&A	June-Aug. 2019
Graduate student coordinator, Brown University Department of Ecology and Evolutionary Biology internal seminar series	Sept.-Dec. 2016
Mentor, New Scientist Program, Brown University	Jan.-May 2014
President, Biology Club, Colby College	2007-2009

SCIENCE COMMUNICATION AND OUTREACH

Mount Hope Child Care Center STEM Volunteer. Designed and donated a science workstation for a local preschool classroom, and led hands-on science activities for preschool and pre-K students. — 2015-2020

Rhode Island Science and Engineering Fair. Judge for Junior and Senior Division in Plant Sciences and Cellular and Molecular Biology. — March 2019

Research Spotlight. Interviewed for the Institute at Brown for Environment and Society (IBES) Magazine. Link: https://www.brown.edu/academics/institute-environment-society/spotlights/kimberly-neil — Aug. 2018

'University doctoral candidate investigates dwindling rabbit species native to campus'. Interviewed for news story in the 'News, Science & Research' feature of the Brown Daily Herald. Link: https://www.browndailyherald.com/2018/04/23/university-doctoral-candidate-investigates-dwindling-rabbit-species-native-campus/ — April 2018

Vartan Elementary School STEM Volunteer. Entered a Providence, RI first grade classroom to lead hands-on STEM activities relevant to geology and environmental science. — 2015-2016

Day in the Life of a Scientist. Hosted a group of young women interested in biology from a local Rhode Island middle school. Designed a half day of lab activities centered around my research. — Nov. 2015

Biology Outreach and Inquiry with Boston University Graduate Students (BIOBUGS). Program that brings Boston area public high school students to campus for hands-on laboratory activities. Independently designed curricula for an Urban Ecology module and volunteered within the laboratory classroom for both ecology and microbiology-themed units. — 2011-2013

Imprinting: implications for wildlife rehabilitation programs. Public talk given at the New England Wildlife Center to discuss behavioral imprinting and how this can impact reintroduced wildlife. — Aug. 2008

PROFESSIONAL AFFILIATIONS

American Society of Mammalogists
Graduate Women in Science and Engineering
500 Women Scientists – Providence
New England Cottontail Population Monitoring Working Group
New England Cottontail Research Working Group

PREFACE

Knowledge of genetic diversity in natural populations, the maintenance of this variation, and its underlying selective drivers are central to both biological conservation and evolutionary ecology. Understanding species' genetic architecture and demography can inform both theory-driven science and management decisions of imperiled taxa (Allendorf *et al.* 2010; Ouborg *et al.* 2010; Steiner *et al.* 2013). Traditionally, conservation genetics has employed neutral markers for these purposes (Primmer 2009; Allendorf *et al.* 2010). However, a solely neutral focus may not capture selective pressures or ecologically important processes relevant to species' long-term survival (Bekessy *et al.* 2003). With rapid advances in metagenomics, the definition of functionally and ecologically significant diversity has expanded to also include the natural microflora associated with individuals. The critical role of the microbiome in host health, development, and physiology is well-accepted, yet research encompassing the roles of host genetics, species identity, and environment in governing the microbiomes of wildlife is lacking.

The study of New England's imperiled and invasive cottontails is an attractive system for investigating these areas. The New England cottontail (*Sylvilagus transitionalis*; NEC) is a threatened mammal spanning multiple geographically isolated populations, whereas the eastern cottontail (*S. floridanus*; EC) is an introduced species that has successfully expanded throughout much of the northeastern United States. Since the 1960s, NEC have declined in abundance and are presently found in just five remnant groups that occupy less than 16 percent of their historic range (Litvaitis *et al.* 2003). While NEC population decay is largely attributed to habitat loss, it is also coincident with EC introduction (Litvaitis *et al.* 2003, 2007). Beginning in the early 1900s, ECs were intentionally introduced into New England to supplement small game hunting, until concerns of the zoonotic disease tularemia halted their import in the 1940s (Ayres &

Feemster 1948). It is estimated that an excess of 200,000 EC comprising at least six subspecies were introduced (Johnston 1972; Chapman & Morgan 1973). Several key demographic and ecological differences exist between the two species that may have contributed to their dissimilar trajectories. While NECs specialize on densely vegetated shrubland and early successional habitat, EC are more "ecologically flexible" and can effectively utilize human-modified landscapes (Smith & Litvaitis 2000; Litvaitis *et al.* 2007; Cheeseman *et al.* 2018).

The goal of this dissertation is to apply a comparative lens to comprehensively investigate functionally and ecologically significant variation within an invasive and imperiled mammal system of immense conservation value. I apply an integrative approach uniting conservation genetics, immunogenetics, microbial ecology, and evolutionary ecology to simultaneously advance biological theory and translate this knowledge to applied species management. I outline my chapters in greater detail below.

In Chapter 1, I focus on understanding neutral and adaptive variation across New England's EC and NEC. To do so, I utilize a combination of sampling genome-wide neutral single nucleotide polymorphisms (SNPs) and targeted gene screening at functional loci of both the adaptive and the innate branches of the immune system. I show that both species exhibit neutral population structure, with NEC clustering according to geographic isolation. I also find that despite population decline, NEC retain immunogenetic diversity at all targeted immune genes, and that EC display exceptionally high immunogenetic diversity, particularly at major histocompatibility complex (MHC) genes. I demonstrate that types of selection are different across immune genes and species. On an evolutionary timescale, historical balancing selection has operated on MHC loci, and this continues to be detected for NEC. However, EC MHC genes are highly differentiated relative to neutral SNPs, indicating that diversifying selection and local

adaptation may be driving observed patterns. For both species, genes of the innate immune system are primarily structured by purifying selection, although codon-specific positive selection is detected. This work will be submitted for publication in fall/winter of 2020.

In Chapter 2, I expand beyond considering only genetic variation to also host-associated microbiota and the interplay with species identity, immune genes, and host-genetics. I focus on one geographic location where both NEC and EC are found in sympatry, Cape Cod, Massachusetts. I develop a novel modified 16s-rRNA sequencing approach to apply traditional 16s-rRNA sequencing for bacterial capture while excluding host mitochondrial reads, thereby making this methodology ideal for 16s-rRNA sequencing of tissues that are rich in host DNA. With my approach, I investigate microbiota associated with EC and NEC ear biopsies. I find that no species signatures are evident in community composition, but that NEC host a greater number of amplicon sequence variants (ASVs). This finding is among the first to explore whether invasive species are "missing" microbiota relative to ecologically similar natives, which is an expansion of the missing parasites hypothesis (Torchin *et al.* 2003). I also show that host genetics influence microbiota composition. I find that particular genomic SNPs associate with presence/absence of certain microbial families, and that crucial immune genes are among those SNPs. I also demonstrate that specific MHC alleles associate with both presence and absence of microbial families, indicating that MHC may have evolved to confer both resistance to harmful pathogens and tolerance to beneficial commensals. We also show that for EC, MHC heterozygosity impacts the collective diversity of ear microbiota. I plan to submit this work for publication in winter of 2020.

In Chapter 3, I investigate how fecal microbiome is structured across multiple sympatric populations of invasive EC and imperiled NEC. I show that recovered bacterial taxa are

consistent with those reported for other herbivores and folivores, including the European hare (*Lepus europaeus*) and European rabbit (*Oryctolagus cuniculus*). I demonstrate that species signature strongly dominates in structuring fecal microbiome, and that microbiomes of sympatric populations of conspecifics do not converge. Compositional dissimilarities between species are primarily driven by differences in presence/absence of "core microbiome" members, and this corresponds with inferred functional differences in generation of metabolites/energy, biosynthesis, and degradation/utilization of compounds. I also demonstrate that within species, population differences in microbiome are detected and primarily driven by differential abundance of rare taxa. I plan to submit this chapter for publication in winter/spring of 2020/21.

Overall, this dissertation addresses fundamental questions in evolutionary ecology, microbial ecology, and conservation genetics while also informing ongoing species management efforts. This collective body of work contributes to our understanding of functional, ecologically significant variation in imperiled and invasive species, and also utilizes novel approaches. In particular, I 1) utilize a unique system characterized by multiple populations of a closely-related imperiled species and its successful invader, 2) develop a modified 16s-rRNA sequencing methodology that is appropriate for non-fecal tissues, 3) am among few studies to wed double-digest restriction-associated DNA (ddRAD) sequencing with microbiota analyses in non-model systems, and 4) am among the first studies that have applied either immunogenetics or genome-wide association studies to understanding microbiota of natural systems. This dissertation therefore represents a significant advancement to our understanding of functionally and ecologically significant variation and its potential drivers in imperiled and invasive systems.

References

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature*, **11**, 697–709.

Ayres JC, Feemster RF (1948) Epidemiology of tularemia in Massachusetts with a review of the literature. *New England Journal of Medicine Journal of Medicine*, **238**, 187–194.

Bekessy SA, Ennos RA, Burgman MA, Newton AC, Ades PK (2003) Neutral DNA markers fail to detect genetic divergence in an ecologically important trait. *Biological Conservation*, **110**, 267–275.

Chapman TR, Morgan RP (1973) Allen Press. *Wildlife Monographs*, **36**, 3–54.

Cheeseman AE, Ryan SJ, Whipps CM, Cohen JB (2018) Competition alters seasonal resource selection and promotes use of invasive shrubs by an imperiled native cottontail. *Ecology and Evolution*, **8**, 11122–11133.

Johnston JE (1972) Identification and distribution of cottontial rabbits in southern New England. University of Connecticut.

Litvaitis JA, Barbour MS, Brown AL *et al.* (2007) Testing Multiple Hypotheses to Identify Causes of the Decline of a Lagomorph Species: The New England Cottontail as a Case Study Natural History of New England Cottontails. In: *Lagomorph Biology: Evolution, Ecology, and Conservation* (eds Alves P, Ferrand N, Hacklander K), pp. 167–185. Berling Heidelberg.

Litvaitis JA, Johnson B, Jakubas W, Morris K (2003) Distribution and habitat features associated with remnant populations of New England cottontails in Maine. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **81**, 877–887.

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW (2010) Conservation genetics in

transition to conservation genomics. *Trends in Genetics*, **26**, 177–187.

Primmer CR (2009) From conservation genetics to conservation genomics. *Ann NY Acad Sci*, **1162**, 357–368.

Smith DF, Litvaitis JA (2000) Foraging strategies of sympatric lagomorphs: implications for differential success in fragmented landscapes. *Canadian Journal of Zoology*, **78**, 2134–2141.

Steiner CC, Putnam AS, Hoeck PEA, Ryder OA (2013) Conservation Genomics of Threatened Animal Species. *Annual Review of Animal Biosciences*, **1**, 261–281.

Torchin ME, Lafferty KD, Dobson AP, McKenzie VJ, Kuris AM (2003) Introduced species and their missing parasites. *Nature*, **421**, 628–630.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my advisor, David Rand, for his unwavering support, guidance, and patience over the last several years. You adopted me into your lab, and you continuously encouraged me to both pursue research that I am passionate about and opportunities that would develop my proficiencies and confidence. I also would like to extend my gratitude to my co-advisor, T.J. McGreevy, Jr. You introduced me to the world of applied wildlife biology and helped me to cultivate skillsets of value both to theory-driven and applied science. I am extremely appreciative for your continued feedback, assistance, and advice. I am extremely thankful for my other committee members, Sohini Ramachandran and Dan Weinreich, for their continued contributions, encouragement, and support during my years at Brown.

Thank you to past and present members of the Rand lab, who made me feel included in the lab right away and provided valuable feedback on my research, especially Jim Mossman, Joaquin Nunez, Adam Spierer, Shawn Williams, and Faye Lemieux. Thank you to McGreevy lab members for sharing samples, data, and helping me navigate your halls of freezers. Thank you to Tyler Kartzinel and his lab members for graciously including me in their meetings and sharing laboratory space. Thank you to Becky Kartzinel for providing valuable feedback on my work and allowing me to regularly work out of your laboratory. Thank you to Martha Cooper for very generously helping me equip my first lab space.

I am extremely grateful to past and present EEB staff for their immense support over the years. Shannon Silva, Lianne Mendonca, Jesse Marsh, Alexandra Collins, Danielle Leone-Camara – you helped me to navigate all of the logistical and administrative aspects of my PhD, and I am deeply thankful.

This research would not have been possible without the help of numerous collaborators outside of Brown. I want to thank the many wildlife agencies that provided samples for my work: New Hampshire Fish and Game (Heidi Holman), Massachusetts Department of Fish and Game (David Scarpitti), Connecticut Department of Energy and Environmental Protection (Howard Kilpatrick), and Rhode Island Department of Environmental Management (Brian Tefft). Thank you to the Roger Williams Park Zoo, especially Lou Perrotti, Kim Wojick, and Mike McBride, for collecting numerous samples from your captive cottontail population, and to the Queens Zoo for also sharing specimens. Thank you to the New England Cottontail Population Working Group and the New England Cottontail Research Working Group for including me in your efforts and meetings. Thank you to Christopher Reid of Bryant University for allowing me use of your biosafety cabinet.

I also want to thank Tom Kunz, my first graduate school PI, who introduced me to the exciting world of wildlife disease and immunogenetics. My research interests and drive to succeed were heavily influenced by your early guidance and perpetual kindness, and I will always be grateful for having known you.

I am so immensely appreciative to my rich support network of friends and family. First, I want to thank my EEB cohort and friends from Brown, who always made me feel supported, capable, and included. You were often my biggest cheerleaders, and I am forever grateful. Thank you to Nikole Bonacorsi, KC Cushman, Bianca Brown, Courtney Reed, Priya Nakka, Brooke Osborne, and Patrick Freeman. Thank you to my best friends from outside of Brown, for always lending an ear and a helping hand when most needed, especially Talia Hitt and Jamie Tyrol.

I extend my warmest thanks to my loving family – without you this wouldn't be possible. Thank you foremost to my mom and dad, Steve and Karen Cohen. You have provided me with

every opportunity in life, always encouraged me to work hard and follow my passions, and taught me to go after my professional aspirations while also taking care of a family. Thank you for your steadfast support, love, and encouragement. Thank you to my sister Allie Cohen and soon-to-be brother-in-law Zach Pearlstein, and to my brother Dave Cohen, sister-in-law Kristie Cohen, and their children Kobe, Eva, and Olivia. Thank you to my amazing in-laws, Joan and Bob Neil, and the extended Neil family that embraced me – Jon, Aki, Ella, and Solana Cone, Jesse Neil, and Anthony Neil. Thank you to my grandparents, and particularly my grandpa, Merrill Cohen, who passed away this summer. You were the first scientist role model in my life and an avid lover of wildlife; even though you aren't here for my defense, you played a big role in me making it this far.

Lastly, thank you to my beloved husband, Chris, and my son Griffin. One of the joys of finishing my thesis is now looking back at how much our family has grown (and is currently growing) since I began. Chris, you always believed in me with unwavering confidence. You thoughtfully and ardently support me as a scientist, student, spouse, and mother, and this journey would not have been possible with anyone else but you. Griffin, your curiosity, warmth, and carefree love of life motivate me more than anything. Thank you both for continuously reminding me how joyful and vibrant life is.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

Contrasting population structure and patterns of immunogenetic diversity across an imperiled cottontail and its sympatric invader

Kimberly B. Neil, Thomas J. McGreevy, Jr., Joaquin C. B. Nunez, and David M. Rand

Introduction

Knowledge of genetic variation in natural populations, the maintenance of this diversity, and underlying selective drivers are central tenets of both evolutionary ecology and biological conservation. Understanding species' genetic architecture and demographic trends can inform knowledge of gene flow, population size fluctuations, and also management decisions for threatened taxa (Allendorf *et al.* 2010; Ouborg *et al.* 2010; Steiner *et al.* 2013). Traditionally, conservation genetics has employed neutral markers for these purposes (Primmer 2009; Allendorf *et al.* 2010; Steiner *et al.* 2013). However, an exclusively neutral focus on understanding species' genetic diversity may not capture ecologically and functionally significant genetic variation (Bekessy *et al.* 2003). Understanding adaptive variation within and across species is crucial for assessing local adaptation, long-term population viability, and adaptive potential in the face of rapid anthropogenic change (Allendorf *et al.* 2010; Steiner *et al.* 2013). Consequently, the application of immunogenetics to conservation biology and evolutionary ecology can lend insights into the maintenance of adaptive variation across challenges of global relevance, including population decline, species' invasion, and habitat fragmentation.

Genes of the major histocompatibility complex (MHC) are traditionally targets for understanding adaptive variation in wildlife (Sommer 2005; Ujvari & Belov 2011). MHC genes encode glycoproteins that present foreign peptide for recognition by the immune system and are critical in initiating the adaptive branch of the vertebrate immune response. MHC genes are also exceptionally diverse (Bernatchez & Landry 2003); indeed, they are among the most polymorphic vertebrate genes, and sequence diversity at MHC regions is typically elevated compared to genomic averages (Garrigan & Hedrick 2003). This high diversity is thought to be

maintained by several nonexclusive mechanisms, including pathogen-driven selection for 1) rare MHC variants that parasites/pathogens are less likely to adapt to compared to common host alleles, 2) MHC heterozygotes, who should be able to detect a larger repertoire of parasites/pathogens, or 3) fluctuating selective pressures over differing spatial and temporal scales (Bernatchez & Landry 2003; Piertney & Oliver 2006; Spurgin & Richardson 2010). There is a rich body of literature characterizing MHC architecture in non-model wildlife to inform species' evolutionary histories, guide conservation decisions, or test for associations between MHC genotype and pathogens (Hughes 1991; Sommer 2005; Ujvari & Belov 2011).

However, immunity is complex, and there is a need to move beyond using only MHC genes as indicators of adaptive immunogenetic variation and to additionally incorporate other ecologically relevant markers (Acevedo-Whitehouse & Cunningham 2006). Recently, studies of wildlife immunogenetics have targeted toll-like receptors (TLRs) due to their functional importance (Tschirren *et al.* 2011, 2012, 2013; Fornůsková *et al.* 2013; Morger *et al.* 2014; Cui *et al.* 2015; Quéméré *et al.* 2015). In contrast to MHC, TLRs belong to the innate branch of the immune system and form an ancient gene family found across plants, invertebrates, and vertebrates (Beutler 2004; Zhou *et al.* 2007). Additionally, while the adaptive branch initiates a fast, highly specific immune response with lasting immunological memory, the innate system is typically slower and non-specific, recognizing broad pathogen-associated molecular patterns (PAMPs) shared across potentially pathogenic taxa (*e.g.,* peptidoglycan, double-stranded RNA) (Takeda *et al.* 2003; Beutler 2004). TLRs have classically been purveyed as under purifying selection, but recent work supports that diversity and selection regimes are variable across species, TLR loci, and intragenic functional regions (Tschirren *et al.* 2012, 2013; Fornůsková *et al.* 2013; Morger *et al.* 2014; Cui *et al.* 2015; Quéméré *et al.* 2015).

We investigate immune genes in a sympatric imperiled and invasive cottontail, that of the native New England cottontail (*Sylvilagus transitionalis*; NEC) and nonnative eastern cottontail (*S. floridanus*; EC). Evolutionary relationships among *Sylvilagus* sp. are difficult to discern due to both a lack of comprehensive studies as well as rapid diversification of the genus (Silva *et al.* 2019), but it is likely that EC and NEC are relatively recently diverged (Matthee *et al.* 2004; Ge *et al.* 2013; Silva *et al.* 2019). NEC were historically native throughout much of New England and New York but have severely declined in abundance since the 1960s (Litvaitis 2003; Litvaitis *et al.* 2007). Contemporary NEC occupy less than 16 percent of their historical range and are found on five geographically disjunct remnant groups that cluster correspondingly at microsatellite loci (Litvaitis *et al.* 2004; Fenderson *et al.* 2011). However, it is unknown if these population identities are recapitulated at nuclear markers surveyed over a genome-wide scale, or if NEC remnant groups diverge at functionally significant loci. NEC are also the current target of a reintroduction program, whereby wild-borne individuals are captured from throughout their native range, bred in captivity, and released into managed habitats in Rhode Island, New Hampshire, and Maine (Fuller & Tur 2012, 2015).

NEC population decay is largely attributed to land-use changes (Litvaitis *et al.* 2003, 2007) but is also coincident with the introduction of EC. Beginning in the early 1900s, over 200,000 EC from at least six subspecies were introduced into the northeastern United States to supplement small game hunting (Johnston 1972; Chapman & Morgan 1973), until concerns of the zoonosis tularemia halted their import in the 1940s (Ayres & Feemster 1948; Johnston 1972). EC have since expanded throughout most of New England, limiting the areas where NEC can exist in allopatry (Tash & Litvaitis 2007). While invaders often experience population bottlenecks (Tsutsui *et al.* 2000), large and genetically diverse founder populations coupled with

multiple invasion waves and high migration between introduction sites can mitigate these effects (Kolbe *et al.* 2004, 2007; Genton *et al.* 2005; Dlugosch & Parker 2008). Mitochondrial DNA sequencing of invasive EC failed to detect population structure (Litvaitis *et al.* 1997; Sullivan 2013), but whether neutral markers support a similar pattern is unknown.

This system provides a rare opportunity to investigate population structure and adaptive variation across an imperiled species and its closely related, sympatric invader. Species such as NEC with small, fragmented populations are subject to potential inbreeding depression, the accumulation of deleterious alleles, and loss of adaptive variants (Reed 2005; O'Grady *et al.* 2006; Edmands 2007). Because of immune genes' crucial roles in processes such as disease resistance, lifespan, and reproductive output, conservation biologists often consider MHC when evaluating endangered species' recovery plans and trajectories (Bernatchez & Landry 2003; Sommer 2005; Ujvari & Belov 2011). However, whether MHC diversity is lost following population decline is highly variable. If pathogen-driven selection is weak, population size is small, or genetic drift is strong, then MHC alleles are predicted to be lost (Mainguy *et al.* 2007; Radwan *et al.* 2010a). However, there are cases where MHC diversity persists despite population bottlenecks (Aguilar *et al.* 2004; Oliver & Piertney 2012). Additionally, spatially explicit pathogen regimes coupled with low population gene flow may accordingly select for biogeographic signal of MHC alleles and/or genotypes. Because NEC are part of an ongoing conservation program, it is especially imperative to investigate immunogenetic diversity in reintroduced populations. Admixture of founders may increase diversity in reintroduced groups, as private alleles may combine into novel genotype combinations. Conversely, limited founder individuals may cause overrepresentation of certain alleles in captive-bred offspring (Marsden *et al.* 2013) and therefore decrease diversity relative to wild conspecifics or introduce potentially

maladaptive genotypes (Čížková *et al.* 2012).

Understanding the role of immunogenetics in invasive species is of particular value for informing both invasion ecology theory and practical species management. Immune genes may modulate the ability of invaders to adapt to novel environments (Lee & Klasing 2004; White & Perkins 2012; Cornet *et al.* 2016). For example, invasion success is linked to immune system functioning across select case studies of birds, fish, amphibians, and mammals (Cornet *et al.* 2016). However, investigations of invaders' immunogenetic architecture and associated selective regimes are lacking. Multiple waves of EC admixture could support high immunogenetic diversity and novel genotype combinations useful for pathogen resistance. Alternatively, invaders may experience relaxed selection on immune genes due to lack of pathogens in their novel range or tradeoffs related to dispersal (Sakai *et al.* 2001; Shine *et al.* 2011). Studying immune genes in invasive species provides valuable information both about the evolutionary processes that contribute to invasions and management strategies that may mitigate their success.

We investigate the following novel research questions:

1) How are populations structured across New England's native and invasive cottontails?
2) Is immunogenetic diversity evident in imperiled NECs?
3) Are invasive EC more diverse at functional loci than are NEC?
4) Is there evidence of selection acting on immune genes in NEC and EC?
5) Can knowledge of EC and NEC immunogenetics inform ongoing species management?

We expect to detect population structure across NEC using a survey of neutral, genome-wide single nucleotide polymorphisms (SNPs), as has been reported previously using microsatellite markers (Fenderson *et al.* 2011). Studies of EC mitochondrial markers fail to denote population structure (Litvaitis *et al.* 1997; Sullivan 2013), but whether this is

recapitulated at nuclear markers is unknown. Multiple, widespread admixture events at introduction coupled with high gene flow could obscure population signal. However, local populations may form due to differential regional founder histories, local adaptation, and/or low gene flow.

Whether imperiled species retain and maintain immunogenetic diversity following population decline is highly variable (Aguilar *et al.* 2004; Mainguy *et al.* 2007; Radwan *et al.* 2010b; Oliver & Piertney 2012). However, due to historically pervasive balancing selection at MHC genes (Bernatchez & Landry 2003; Piertney & Oliver 2006), we expect to detect species-level diversity at immune loci. Due to admixture of multiple EC subspecies during introduction, we expect to detect high immunogenetic diversity both within and across EC populations relative to NEC. On an evolutionary timescale, MHC genes are typically under balancing selection, whereas TLR loci are canonically thought to be subject to purifying selection (Takeda *et al.* 2003; Beutler 2004; Piertney & Oliver 2006). We expect to detect these historical signatures, and to detect contemporary selection patterns consistent with the role of pathogen-driven selection in natural systems. Since knowledge of population structure and adaptive variation are widely used to inform conservation programs, we expect our findings to be of applied relevance to ongoing NEC reintroduction efforts and EC species management.

Methods

Sample collection and DNA extraction

EC and NEC tissue samples originated from multiple sources: 1) DNA specimens from the Wildlife Genetics and Ecology Laboratory (WGEL) at the University of Rhode Island (URI) (IACUC protocol AN11-12-011), 2) ear biopsy, blood, and/or cadaver tissue donated by New England state wildlife biologists, and 3) the Roger Williams Zoo and Queens Zoo NEC captive

breeding programs. Tissue samples were extracted either at WGEL or in a class-II biosafety cabinet at Bryant University using the Qiagen DNeasy kit (Cat. No. 69506).

ddRAD library preparation and sequencing

Double-digest restriction-associated DNA (ddRAD) sequence was undertaken by WGEL using 275 samples for EC and 218 samples for NEC, spanning geographically throughout New England and New York (**Figure 1**). Briefly, libraries were prepared according to (Peterson *et al.* 2012) using the restriction enzymes SphI-HF (New England Biolabs, Ipswich, MA) and MluCI (New England Biolabs, Ipswich, MA). Samples were standardized to 400ng of DNA by quantifying using the high-sensitivity Quant-iT^{TM} dsDNA Assay Kit (Invitrogen). Sets of 12 samples were uniquely barcoded, pooled at equal concentrations, and polymerase chain reaction (PCR) amplified for 12 cycles to increase the concentration and incorporate adapters. Four sets of 12 samples were pooled at equal concentration and size selected for 333 bp to 414 bp using a Pippin Blue (Sage Science). The size selected pools were cleaned using Dynabeads (Thermofisher Scientific) to remove amplicons that did not have the correct pair of forward and reverse adapters. Fifty base pair single-reads were sequenced on the Illumina HiSeq 2000 in six lanes by Harvard University's Bauer Core Facility.

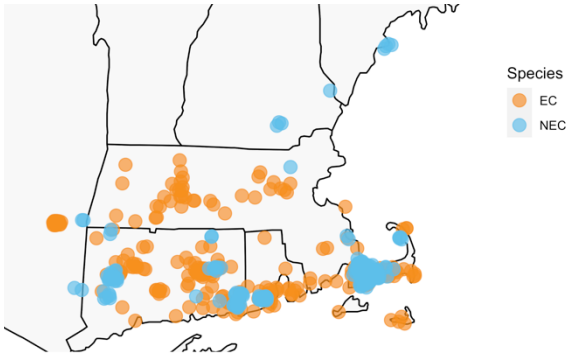

Figure 1. Sampling of EC and NEC for ddRAD. Circles represent individual samples across New York, Connecticut, Rhode Island, Massachusetts, New Hampshire, and Maine. Color corresponds to species.

ddRAD parameterization, SNP filtering, and variant calling

Raw sequences from each ddRAD library were processed using STACKS v.2.2 (Catchen

et al. 2013). Libraries for each species were processed separately using the *de novo* pipeline. Parameter selection within STACKS is critical and can impact downstream analyses (Paris *et al.* 2017; Rochette & Catchen 2017). Therefore, we performed a series of optimization tests as described in Paris *et al.* 2017 and Rochette & Catchen 2017. Separately for EC and NEC, we systematically tested values of *M* (the minimum number of mismatches allowed between stacks to merge into a putative locus) ranging from 0 to 6, and values of *n* (the number of mismatches allowed between putative loci during catalog construction) ranging from 0 to 5. The impact of changing *M* or *n* across four parameters was evaluated: 1) the number of assembled loci, 2) the percent of polymorphic loci, 3) the total number of SNPs recovered, and 4) the distribution of SNPs within and across loci (**Figures S1 & S2**). We selected optimal *M* and *n* values that either maximized the parameter of interest or that reflected the point after which the degree of change was minimal: EC *M*=3, *n*=2, NEC: *M*=2, *n*=3.

Following completion of STACKS, SNP data was further filtered using an iterative scheme modified from O'Leary *et al.* 2018. An iterative filtering scheme employs repeating and alternately increasing filtering thresholds to remove individuals with missing data and loci with low call rates across individuals. This approach can retain greater data relative to applying hard thresholds uniformly across a dataset (O'Leary *et al.* 2018). vcftools v.0.0.16 (Danecek *et al.* 2011) was used to remove loci with a mean depth per locus of less than 10 or a minor allele count less than 3. Subsequently, an iterative scheme was applied to retain individuals or loci with a genotype call rate > 50%, individual missing data < 90%, genotype call rate > 60%, individual missing data <70%, genotype call rate > 70%, individual missing data < 50%, genotype call rate > 80%, and individual missing data < 40%. Divergence from Hardy-Weinberg equilibrium was calculated at each locus using the 'Populations' module of STACKS, and loci that significantly

departed from Hardy-Weinberg proportions following Benjamini-Hochberg correction were removed. With the selected parameters, a separate catalog of ddRAD loci was constructed per species. For EC, the final catalog comprised 10,688 loci across 249 individuals, and the final NEC catalog consisted of 4,194 loci across 192 individuals.

Immune gene primer design, library preparation, and sequencing

We targeted exon 2 of the MHC class II genes DRB and DQA, which encodes the hypervariable peptide-binding region (PBR) that directly interacts with antigenic peptide (Hughes & Yeager 1998). MHC class II glycoproteins interact with extracellular antigens, including bacteria, parasites, and helminths (Piertney & Oliver 2006). Multiple DRB and DQA primers used in other small mammals (Smulders *et al.* 2003; Schad *et al.* 2004; Koutsogiannouli *et al.* 2014) and designed specifically for this study were tested to assess amplification fidelity and allelic dropout.

Final DRB PCR amplification used primers KBN.F (5'-GAG TGY CAT TTC YHC AAY GGG AC-3') and KBN.R (5'-RRY CYS GTA GTT GTR TCK GCA-3'), which are modified versions of the mammalian DRB primers JS1 and JS2 (Schad *et al.* 2004). PCR reactions contained between 5 and 20 ng of DNA, 0.5µM of each primer, 200µM dNTPs, 1X Phusion HF Buffer, 0.75 units Phusion DNA polymerase, and molecular-grade water up to 30 µl final volume. Cycling parameters consisted of an initial denaturation at 98°C for 30s, followed by 30 cycles of 98°C for 10 sec, 58°C for 30 sec, and 72°C for 15 sec, and a final elongation at 72°C for 7 min.

DQA PCR amplification initially used leporid DQA primers B and J (Fain *et al.* 2001) to amplify the full PBR and flanking introns. Recovered sequences were used to design and test multiple primers that capture the majority of the PBR. Final PCR reactions used primers B and KBN.DQA.R (5'-GCA GCA GTR GWR TTG GAG YVT TT-3') and reagents as described above. Cycling parameters consisted of

an initial denaturation at 98°C for 60 sec, followed by 10 cycles of 98°C for 30 sec, 58°C for 30 sec, and 72°C for 15 sec; 20 cycles of 98°C for 30 sec, 68°C for 30 sec and 72°C for 15 sec; and a final elongation at 72°C for 7 min.

147 EC and 213 NEC were sequenced (**Tables S1 & S2**). An initial pilot run of 82 samples and 12 replicates was completed, followed by subsequent sequencing of remaining individuals in a single run. We utilized an internal barcoding strategy coupled with multiplexing: DQA and DRB were separately amplified using one of eight possible internal barcodes, and eight uniquely barcoded PCRs were pooled together at equimolar ratio per well across all samples, thereby producing a single 96-well plate for subsequent library preparation. Cleaned amplicons were sent to the University of Rode Island Genomics and Sequencing Center for Illumina MiSeq library preparation and sequencing. An Illumina v2 500 cycle kit was used for 2x200 base pair (bp) paired-end sequencing. Briefly, the Nextera XT Index Kit was used to attach indices and adapters, and products were purified using an AMPure bead cleanup. Samples were analyzed by the Agilent BioAnalyzer DNA 1000 chip and quantified using Qubit fluorometry. Samples were normalized to equimolar ratio before pooling, and the final pooled library was quantified using qPCR prior to sequencing.

For TLRs, I targeted the leucine-rich repeat (LRR) regions of TLR-2 and TLR-4, both of which are involved in bacterial PAMP recognition. The LRR region is encompassed within the TLR ectodomain that interacts with pathogen, and motifs are variable across vertebrates (Matsushima *et al.* 2007). Custom primers were designed using alignments from the European rabbit (*Oryctolagus cuniculus*) and multiple rodent species. For TLR-2, forward primer 197F (5'-CTT GAC CTG TCY AAC AAY R-3') and reverse primer 1415R (5'-GAG ATT GTT RTT RCT AAC RTC-3') were utilized to amplify a 1092 bp region encompassing 14/20 LRRs. For

TLR-4 forward primer 35F (5'-AGG TGT RAA ATT CAD ACA AT-3') and reverse primer 1205R (5'-ATT KCC AGC CAT TTT TAA-3') were used to amplify a 912 bp amplicon spanning 9/20 LRRs.

TLR PCR reactions utilized EconoTaq Plus Green 2x Master Mix (Lucigen), 1µM of each primer, between 5 and 20 ng of DNA, and molecular-grade water up to a 40 µl final volume. Cycling parameters consisted of an initial denaturation at 94°C for 2 min, followed by 35 cycles of 94°C for 30 sec, 54°C for 30 sec, and 72°C for 2.5 min, and a final elongation at 72°C for 10 min. Product was checked on a 1.5 percent agarose gel. Amplicons were sent to Genewiz (Boston) for PCR cleanup and bi-directional Sanger sequencing on an ABI370xl sequencer (Applied Biosystems).

Immune gene genotyping and classification of alleles

Non-model MHC gene sequencing requires specialized data-processing considerations for several reasons (Babik *et al.* 2009; Babik 2010; Sommer *et al.* 2013; Lighten *et al.* 2014; Biedrzycka *et al.* 2016): 1) MHC genes commonly undergo duplication events, and therefore the number of expected loci recovered per individual is typically unknown and variable, 2) The use of degenerate primers coupled with GC-rich regions may lead to substantial biases in allele frequencies per amplicon, and 3) Differentiating between PCR artefacts and low-frequency "true" alleles is challenging.

These considerations have manifested in various approaches to calling non-model MHC genotypes. We utilized AmpliSAT (Sebastian *et al.* 2016), a highly customizable "adjustable clustering" method that was identified in a recent test of MHC genotyping strategies as the preferred approach (Biedrzycka *et al.* 2016). The AmpliSAT pipeline was used to merge, clean, and retain sequences with a minimum Phred score of 30. AmpliCHECK was run using default

Illumina parameters and a minimum per amplicon frequency of 0.001 to preliminarily explore the data. Within the AmpliSAT suite, AmpliSAS was run utilizing default Illumina parameters, but a threshold for "minimum dominant frequency" during clustering was adjusted to 15% for DQA and 20% for DRB. AmpliSAS filtering was constrained to remove chimeras and low-frequency variants by stipulating a "minimum amplicon frequency" value of 10% for DQA and 15% for DRB. Clustering and filtering thresholds were informed by multiple approaches: 1) preliminary data exploration using AmpliCHECK, 2) ordering clusters of variants per amplicon by frequency and frequency with respect to the dominant variant, and identifying the value for which variants are unlikely to be artefacts (**Figures S3 & S4**), and 3) calculating the maximum per amplicon frequency of each variant and identifying a value below which no artefacts are found (**Figures S5 & S6**). These analyses were conducted separately for each barcoded set of amplicons.

Following variant recovery, MHC alleles were further classified as "True" or "Unclassified" according to a decision tree adapted from Grogan *et al.* 2016 (**Figure S7**). "True" alleles were utilized in all downstream analyses, and any individuals containing "Unclassified" alleles were excluded from further processing.

TLR forward and reverse trace sequences were inspected manually, checked for SNPs and indels using novoSNP (Weckx *et al.* 2005), and aligned using MEGA v.7.0.26 (Kumar *et al.* 2016). Haplotypes were phased using PHASE in DnaSP v.5.10.01 (Librado & Rozas 2009). Haplotypes that could not be reconstructed with a 90 percent probability or higher were excluded from further analysis.

For both MHC and TLR genes, recovered amplicons were aligned in BioEdit v.7.2.6 (Hall 2011) and checked for stop codons or reading frame shifts. Amplicons were also BLAST

searched to confirm gene identity and match to Lagomorpha, or if unavailable, Rodentia.

Identifying genetic structure

Population structure was assessed separately for EC and NEC by four primary means: 1) a principle component analysis (PCA) and discriminant analysis of principle components (DAPC), 2) phylogenetic inference, 3) ADMIXTURE, and 4) Estimating Effective Migration Surfaces (EEMS).

DAPC was implemented in the R package 'adegenet' (Jombart 2016). Uneven population sampling can bias both DAPC and admixture analysis, and therefore, for NEC, we subsampled the MACC population. While PCA summarizes the overall variability among individuals by accounting for both within- and among-group divergence, DAPC specifically seeks to maximize the between-group component and thereby discriminate individuals into evolutionary clusters (Jombart 2016). Populations were defined with no prior assumptions about groups. A *k*-means clustering algorithm was performed, and Bayesian Information Criterion (BIC) scores were used to compare models for k = 1-30. Because this is useful to inform a set of possible *k*-values that can summarize the data rather than a "true" *k*, we evaluated 3, 4, and 5 clusters per species. A cross-validation analysis was performed to identify the optimal number of principle components (PCs) to retain per *k*. This procedure utilizes a training data set in which it performs DAPC across varying PC numbers, and then predicts the accuracy of cluster assignment using a validation data set.

Unrooted phylogenies were inferred for each species using SVDquartets (Chifman & Kubatko 2014) in PAUP* v4.0a166 (Wilgenbusch & Swofford 2003). For each possible set of individuals, SVDquartets infers genealogies using a coalescent framework, then assembles a species phylogeny using a quartet assembly method that satisfies the maximum number of

quartet trees. Critically, SVDquartets utilizes the full dataset and is computationally tractable even for large SNP datasets. A random subsample of 30,000,000 quartets was evaluated, and support of inferred relationships was determined using 100 bootstrap replicates. Unrooted phylogenies were generated in R using the packages 'phytools', 'treeio', and 'ggtree' (Revell 2012; Yu *et al.* 2017; Wang *et al.* 2020).

Structure was inferred by ADMIXTURE with cross-validation (CV) (Alexander *et al.* 2009; Alexander & Lange 2011). The number of genetic clusters *k* was pre-defined from 2 to 10, and optimal *k* values to consider were chosen based on the lowest average CV scores across 10 randomly-seeded replicates. The R package 'pophelper' (Francis 2017) was used to visualize ADMIXTURE plots, and populations were labeled by DAPC cluster identity.

Lastly, Estimating Effective Migration Surfaces (EEMS) (Petkova *et al.* 2015) was used to estimate effective migration rates separately for NEC and EC. EEMS infers expected genetic dissimilarities between individuals based on a stepping-stone model and circuit theory's concept of "resistance distance". Briefly, a dense grid is overlaid across the sampling landscape, and migrants are exchanged between neighboring demes. Both migration rates and habitat partitioning of the grid into "cells" are varied such that observed genetic dissimilarities match those modeled under EEMS. These estimates are then visually summarized across the sampling landscape, permitting identification of effective migration that deviates from isolation by distance expectations. A dissimilarity matrix was generated using the EEMS tool 'bed2diffs_v2', which imputes missing individual genotypes using observed mean allele frequencies. Each EEMS run was randomly seeded, and we used 10 M iterations, 1 M burn-in iterations, and a thinning interval of 9,999. Grids that ranged from 100 to 400 demes were evaluated, and a final deme size of 200 was used for both species. Data were analyzed and visualized in R using the

packages ‘rEEMSplots’ and ‘reemsplots2’, which are distributed with EEMS.

Statistical analyses

For each immune gene and species, we calculated basic metrics of sequence diversity per population using DnaSP v.5.10.01 and MEGA v.7.0.26 (Librado & Rozas 2009; Kumar *et al.* 2016): the average number of nucleotide differences (*k*), haplotype diversity (Hd), the number of haplotypes (H), the number of segregating sites (S), amino-acid p-distance, and nucleotide diversity (pi, π). Observed and expected levels of heterozygosity were calculated and compared in R using the packages ‘adegenet’ and ‘pegas’ (Paradis 2010; Jombart 2016).

To estimate population differentiation for neutral SNPs and immune genes, population differentiation was calculated using Hedrick’s G_{ST} (G'_{ST}) (Hedrick 2005). G'_{ST} is an F-statistic that corrects levels of population divergence against the maximum possible level, and therefore it is appropriate when including comparisons of highly diverse, multi-allelic genes such as immune loci (Hedrick 2005; Tschirren *et al.* 2012; Meyer-Lucht *et al.* 2016). The R package ‘diveRsity’ was used to estimate G'_{ST} using a bias-correcting bootstrapping method and to generate 95 percent confidence intervals around the mean (Keenan *et al.* 2013). Populations were defined according to genetic clusters identified by previously described DAPC analysis, phylogenetic inference, ADMIXTURE, and EEMS. To assess whether population differentiation at immune loci is comparable to neutral expectations, population divergence was compared between immune loci and genome-wide SNP estimates (Tschirren *et al.* 2012). Differences in differentiation between neutral and immune loci were assessed by two means: 1) visually assessing whether 95% confidence intervals overlapped with neutral expectations, and 2) evaluating the probability of observing differentiation patterns at least as extreme as those in the data. This was done separately for each gene and in R using the base function ‘dbinom’, a group

size of nine, and a 50 percent probability of falling on either side of neutrality (e.g., more differentiated relative to neutral SNPs or less differentiated).

Standardized heterozygosity (sh) was calculated independently per species and gene using the R package 'Rhh' (Alho *et al.* 2010) and the formula of Coltman *et al.* 1999. Differences in sh between populations was evaluated using a Kruskal-Wallis test, followed by a pairwise Wilcoxon rank sum test with Bonferroni correction for multiple testing. Correlation between genome-wide sh and immune loci was performed independently per species and immune gene using the base R function 'cor.test' and Kendall's *tau*. Isolation by distance was evaluated by testing for correlations between genetic distance ($G'_{ST}/(1-G'_{ST})$ and geographic distance (log of Euclidean distance), and performed separately per species and locus. The R packages 'vegan' and 'dartR' were used to perform Mantel tests of isolation by distance with 10,000 permutations (Dixon 2003; Gruber *et al.* 2018).

To visualize genetic relationships between immune alleles, haplotype networks were constructed per species and immune gene using a minimum-spanning network and the program PopArt (Leigh & Bryant 2015). Haplotype redundancy is defined as two or more unique nucleic acid sequences encoding the same amino acid sequence, and all recovered immune variants were checked for haplotype redundancy using BioEdit v.7.2.6 (Hall 2011). Phylogenetic trees were constructed for each immune gene using RAXML v.8.2.10, 1,000 bootstraps, and the general time-reversible (GTR) model of nucleotide substitution with a Gamma distribution to account for variable rates of evolution across sites (Stamatakis 2014). Additional leporid sequences were incorporated when available from Surridge *et al.* 2008, Oppelt *et al.* 2010, Smith *et al.* 2011, and Koutsogiannouli *et al.* 2014.

To infer selection acting on immune genes, the substitution rates of non-synonymous

substitutions at non-synonymous sites (d_N) and synonymous substitutions at synonymous sites (d_S) were calculated using the program MEGA v.7.0.26 (Kumar *et al.* 2016). Codon-based selection was assessed for MHC using Phylogenetic Analysis by Maximum Likelihood (PAML) in CODEML (Yang 2007). Under neutrality, the expected ratio of d_N to d_S is 1, and significant departures can indicate positive or purifying selection. We compared observed values of d_N/d_S (ω) to four model distributions: M0 (one-ratio), which assumes the same ω across all codons; M1A, which assumes sites are either $\omega = 0$ or $\omega = 1$, allowing for codon conservation or neutrality; M2A, which adds a third class of sites for $\omega > 1$ and therefore permits positive selection; M7, which allows for $0 < \omega < 1$ and varies ω according to the beta distribution; and M8, which additionally permits for $\omega > 1$. Pairs of nested models were compared in R using a likelihood-ratio test: M0 was compared to M3, M1a to M2a, and M7 to M8.

Codon-based selection in TLRs was investigated using two approaches: Mixed Effects Model of Evolution (MEME) to test for episodic selection (Murrell *et al.* 2012) and Fast Unconstrained Bayesian AppRoximation (FUBAR) to test for pervasive selection (Murrell *et al.* 2013). A significance threshold of $p < 0.1$ was used for MEME, and a Bayesian posterior probability of > 0.9 was set for FUBAR. We checked whether codons under selection were located in LRR regions by annotating TLR gene domains based on Matsushima *et al.* 2007.

Results

Population structure

For EC, DAPC clearly defined four population clusters that were largely supported by admixture analysis, phylogenetic inference, and EEMS (**Figure 2**). Surprisingly, the geographic area of Cape Cod, Massachusetts partitioned into two clusters: one that spanned throughout Cape Cod, Massachusetts, eastern Massachusetts, and Rhode Island, and one that was only found on a

subsection of Cape Cod. EEMS analysis also revealed that genetic clusters tended to follow patterns of increased genetic similarity in vertical bands, and recapitulated an internal barrier within Cape Cod. Significant isolation by distance was not observed (Mantel's $r = 0.870$, $p = 0.080$), but genome-wide standardized heterozygosity differed between all populations ($p < 1.6 \times 10^{-11}$), except for Cluster3 vs. Cluster4 ($p = 0.10$).

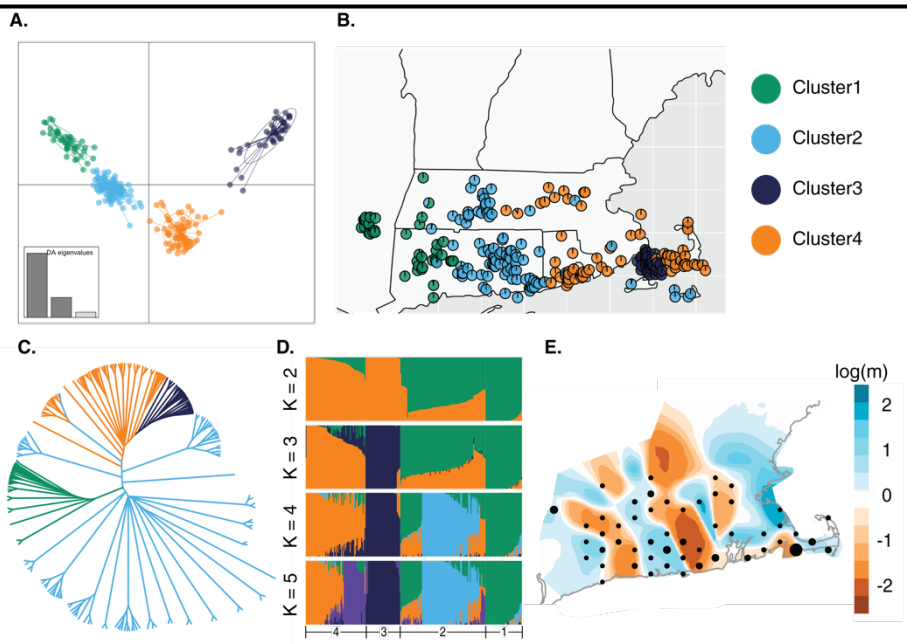

Figure 2. EC population structure. Population color assignment is consistent across all panels. **A)** Discriminant analysis of principal components (DAPC) identified 4 genetic clusters. **B**) Genetic cluster membership probability per individual, as inferred from DAPC. **C)** Unrooted phylogeny inferred using SVDquartets in PAUP and 1,000 bootstraps. **D)** Admixture analysis with genetic clusters corresponding to color and evaluating optimal clusters of $k = 2$-5. Genetic cluster membership probability per individual, as inferred from DAPC. **E)** Estimated effective migration surfaces (EEMS). Migration rates (m) are reported on a log scale, whereby positive values indicate a greater effective migration rate than expected and negative values indicate a lower effective migration rate than expected. Points represent sampling locations and are proportional to sample size.

Wild NEC population identity was largely consistent with the microsatellite analysis of Fenderson *et al.* 2011 (**Figure 3**). NEC located in the managed habitat of Great Swamp Management Area, Rhode Island (RIGS) also clustered separately, although closely with ECT. In contrast to EC, NEC displayed significant isolation by distance for all populations (Mantel's $r = 0.736$, $p = 0.042$) as well as significantly divergent standardized heterozygosity between all

populations (p = 2.2 x10^{-16}) except for 1) ECT vs. RIGS, and 2) MENH vs. MACC.

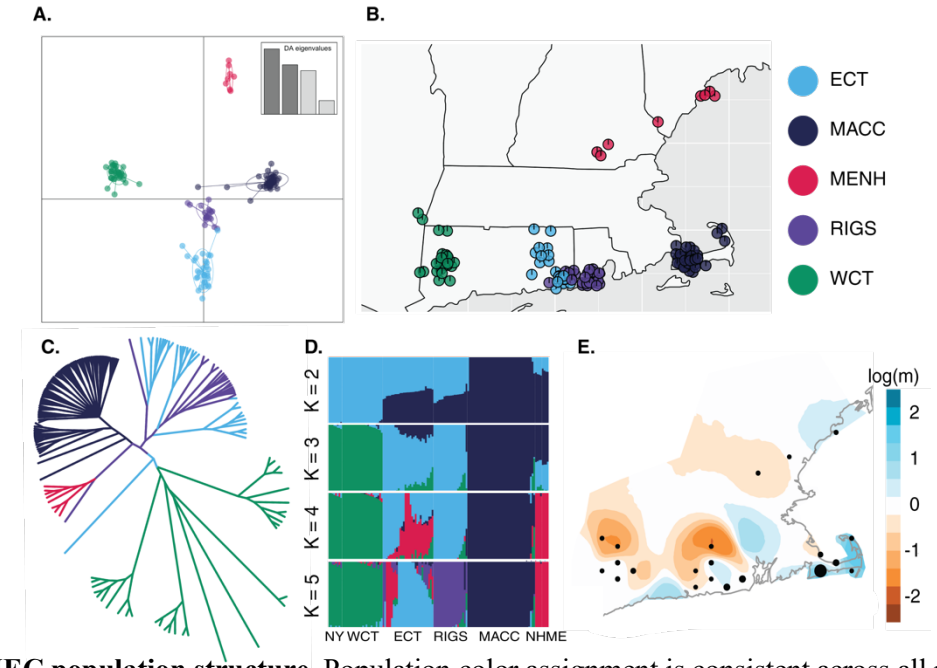

Figure 3. NEC population structure. Population color assignment is consistent across all panels. **A)** Discriminant analysis of principal components (DAPC) identified 5 genetic clusters. **B**) Genetic cluster membership probability per individual, as inferred from DAPC. **C)** Unrooted phylogeny inferred using SVDquartets in PAUP and 1,000 bootstraps. **D)** Admixture analysis with genetic clusters corresponding to color and evaluating optimal clusters of k = 2-5. Genetic cluster membership probability per individual, as inferred from DAPC. **E)** Estimated effective migration surfaces (EEMS). Migration rates (m) are reported on a log scale, whereby positive values indicate a greater effective migration rate than expected and negative values indicate a lower effective migration rate than expected. Points represent sampling locations and are proportional to sample size.

Immune gene diversity

MHC genes were consistently more diverse than TLRs across both species, with DRB displaying the greatest diversity (**Tables S3-S6; Figure 4**). Multiple putative alleles among EC (DQA: 17, DRB: 35, TLR-2: 26, and TLR-4: 13) and NEC (DQA: 4, DRB: 10, TLR-2: 8, TLR-4: 6) were detected, and EC had significantly greater haplotype diversity (p_{DRB} = 0.014, p_{DQA} = 0.013, p_{TLR2} = 0.034, p_{TLR4} =0.032) and amino acid p-distance (p_{DRB} = 0.007, p_{DQA} = 0.0002, p_{TLR2} = 2.2e-16, p_{TLR4} = 3.425e-10) across all immune genes relative to NEC.

No correlations were evident between immunogenetic p-distance and genome-wide standardized heterozygosity for EC (tau_{DRB} = -0.013, p_{DRB} = 0.844; tau_{DQA} = -0.126, p_{DQA} =

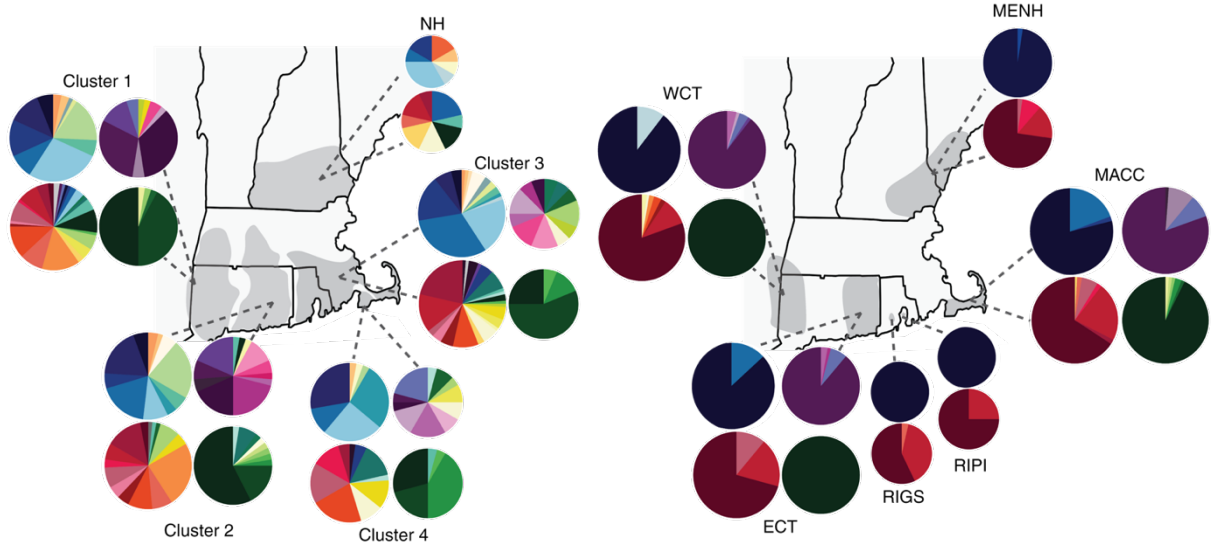

Figure 4. Immunogenetic diversity across invasive EC and imperiled NEC populations. Pie charts depict allele frequencies for a given immune gene. Allele identity indicated by color. Gray shading refers to range from which samples are obtained and correspond with genetic clusters. Top left = DQA, bottom left = DRB, top right = TLR-2, bottom right = TLR-4. Note, MHC only depicted for MENH and NH populations.

0.092; tau_{TLR2} = 0.173, p_{TLR2} = 0.095; tau_{TLR4}= -0.024, p_{TLR4} = 0.817). For NEC, no relationship was evident for MHC genes, but both TLR genes were negatively correlated with genome-wide standardized heterozygosity (tau_{DRB} = -0.014, p_{DRB} = 0.841; tau_{DQA} = -0.048, p_{DQA} = 0.626; tau_{TLR2} = -0.175, p_{TLR2} = 0.046; tau_{TLR4}= -0.180, p_{TLR4} = 0.043).

Selection at immune genes

Recovered immune patterns were highly complex. Haplotype diversity was exceptionally high for all EC immune genes across populations and consistently exceeded genome-wide neutral estimates (**Figure 5**). For NEC, DRB haplotype diversity was consistently high relative to genome-wide neutral estimates, but DQA and TLR-2 were variable across populations. TLR-4 haplotype diversity was consistently low. No isolation by distance patterns were observed for any EC or NEC immunogenetic loci (all $p > 0.05$).

Of the recovered haplotypes, EC haplotype networks revealed complicated networks spanning across populations (**Figures 6-8**). MHC networks contained numerous haplotypes at

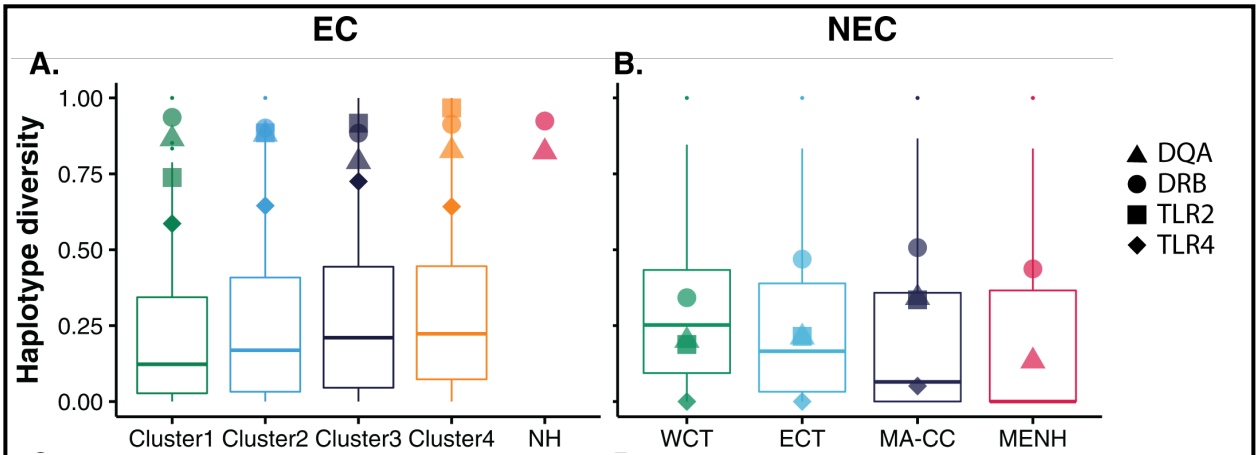

Figure 5. Haplotype diversity at genome-wide neutral SNPs and immune genes. A) EC. **B)** NEC. Populations are abbreviated according to genetic population identity. Boxplots show interquartile range and median, and symbol point estimates correspond to haplotype diversity for immune genes by shape.

intermediate and rare frequencies across populations and were separated by long branches, indicating the presence of multiple distinct lineages and balancing selection at an evolutionary timescale (**Figures 6 & 7**). For TLRs, an abundance of low frequency haplotypes was evident, all of which were closely related and marked by considerable haplotype redundancy, meaning that multiple alleles encoded the same amino acid sequence (**Figures 8 & 9**). This pattern is consistent with elevated pi values and amino acid p-distances at MHC loci relative to TLRs (**Tables S3-S6**). For NEC, MHC haplotype networks were also consistent with balancing selection, although fewer variants were recovered relative to EC. TLR loci were each marked by one high frequency haplotype, surrounded by rare and closely-related variants that displayed considerable haplotype redundancy. This pattern was more evident for TLR-4 than TLR-2.

Gene trees for MHC genes did not cluster by species or genera, but rather formed distinct lineages retained across speciation boundaries (**Figure 10**). In contrast, TLR-2 sequences clearly formed EC- and NEC-specific clades (**Figure S8**). TLR-4 also did not primarily cluster by species, but recovered relationships were poorly supported (**Figure S9**).

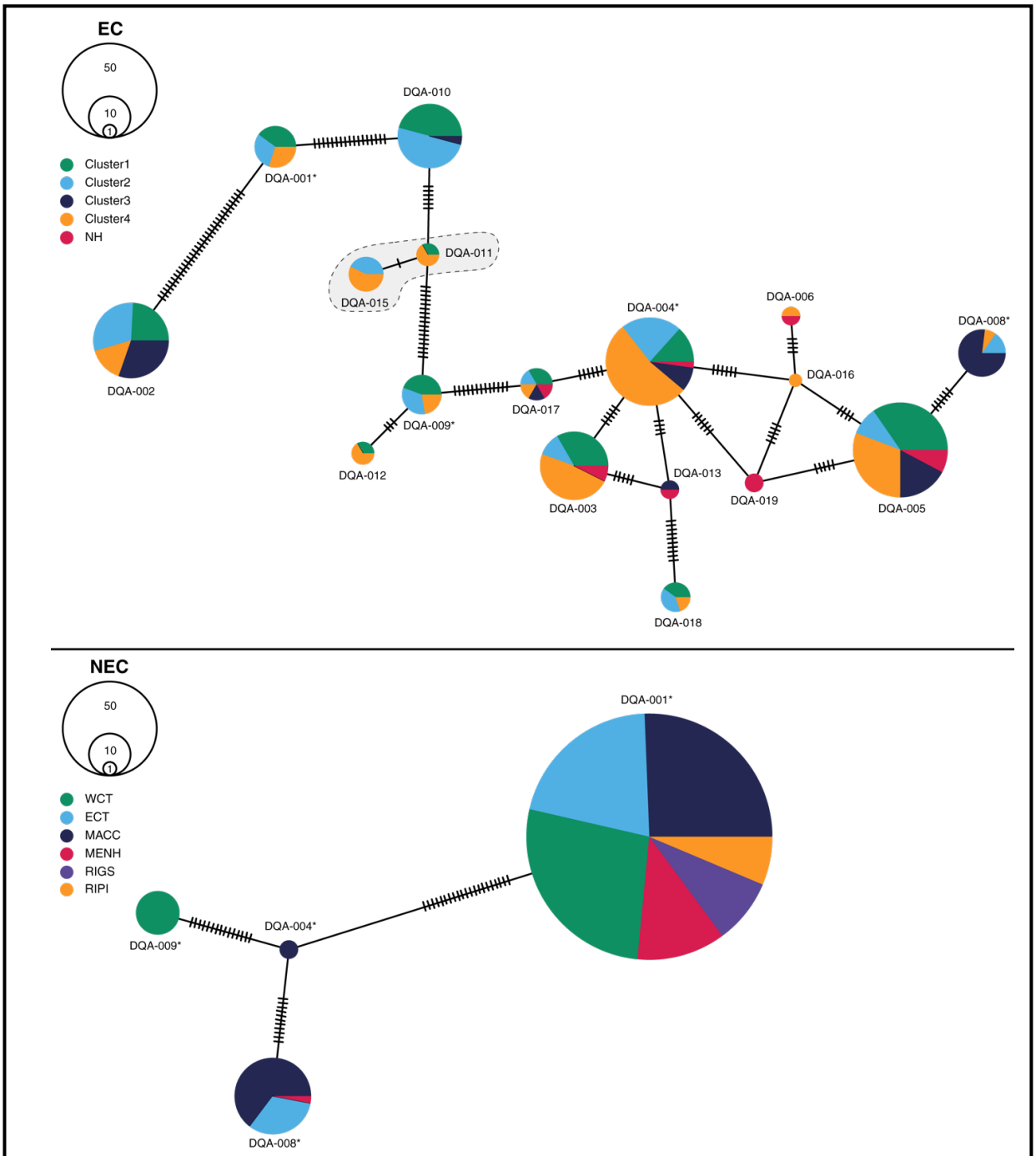

Figure 6. Haplotype network for DQA. EC depicted in top panel, NEC depicted in bottom panel. Pie chart area proportional to allele frequency, and colors correspond to population identity. Nucleotide differences indicated by hatchmarks. Asterisk by allele name indicates variant is shared between species. Redundant haplotypes encircled and shaded.

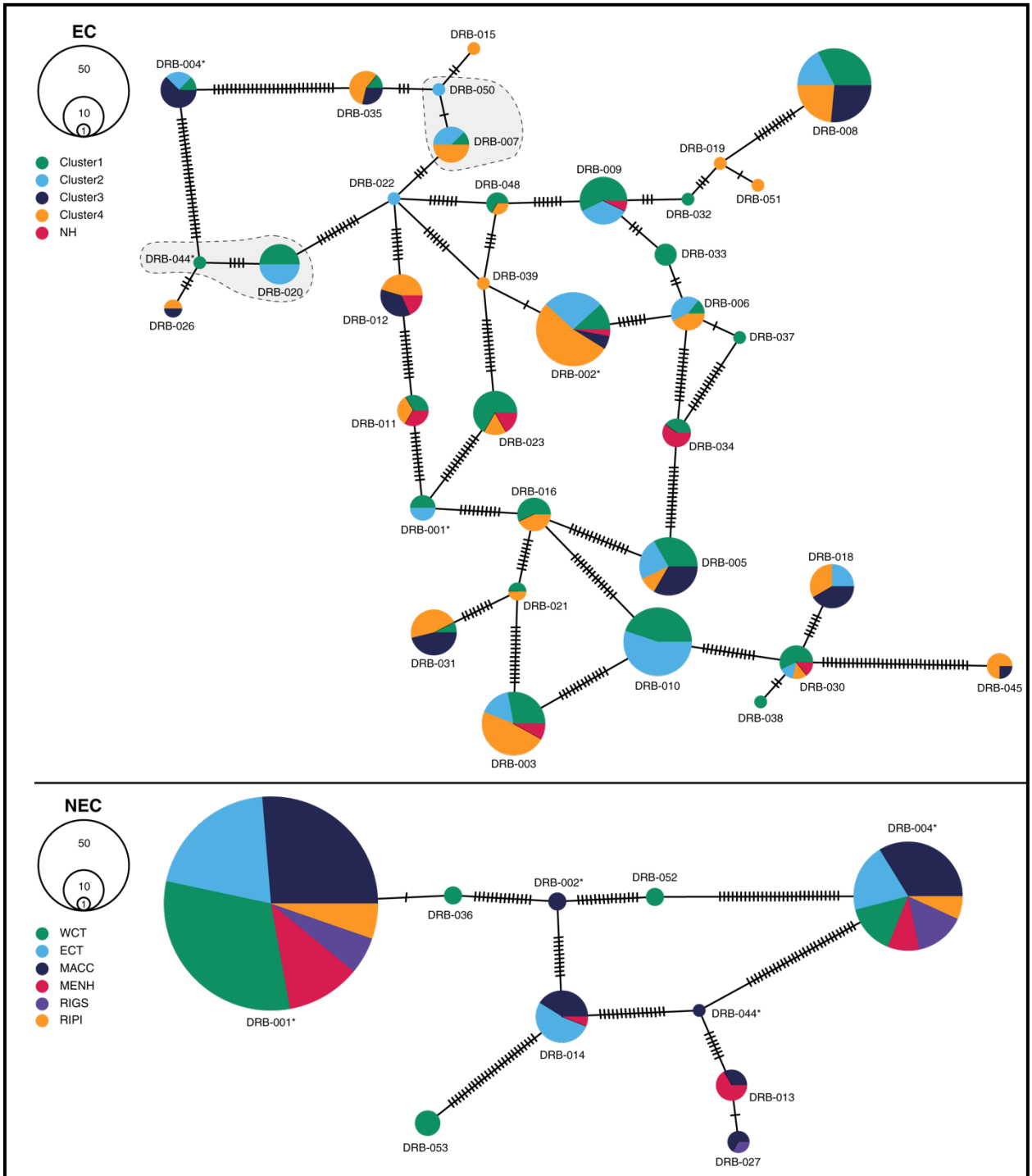

Figure 7. Haplotype network for DRB. EC depicted in top panel, NEC depicted in bottom panel. Pie chart area proportional to allele frequency, and colors correspond to population identity. Nucleotide differences indicated by hatchmarks. Asterisk by allele name indicates variant is shared between species. Redundant haplotypes encircled and shaded.

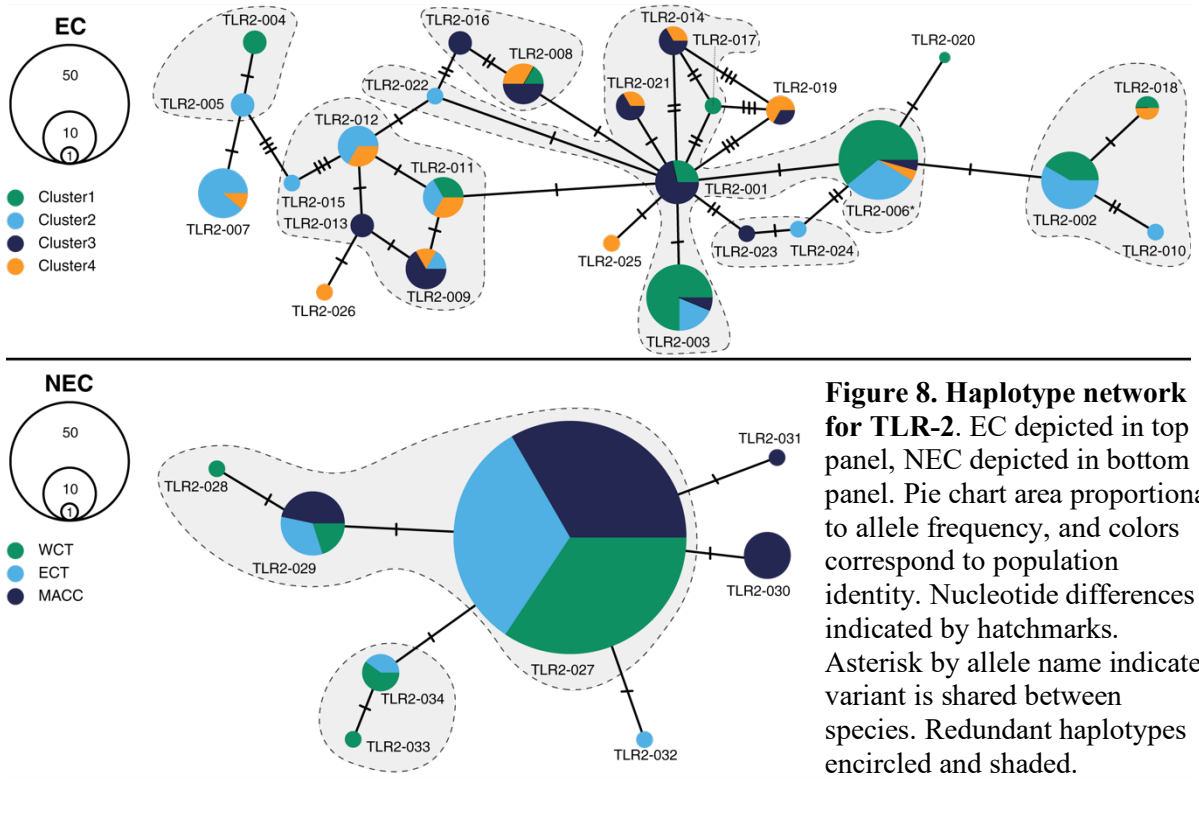

Figure 8. Haplotype network for TLR-2. EC depicted in top panel, NEC depicted in bottom panel. Pie chart area proportional to allele frequency, and colors correspond to population identity. Nucleotide differences indicated by hatchmarks. Asterisk by allele name indicates variant is shared between species. Redundant haplotypes encircled and shaded.

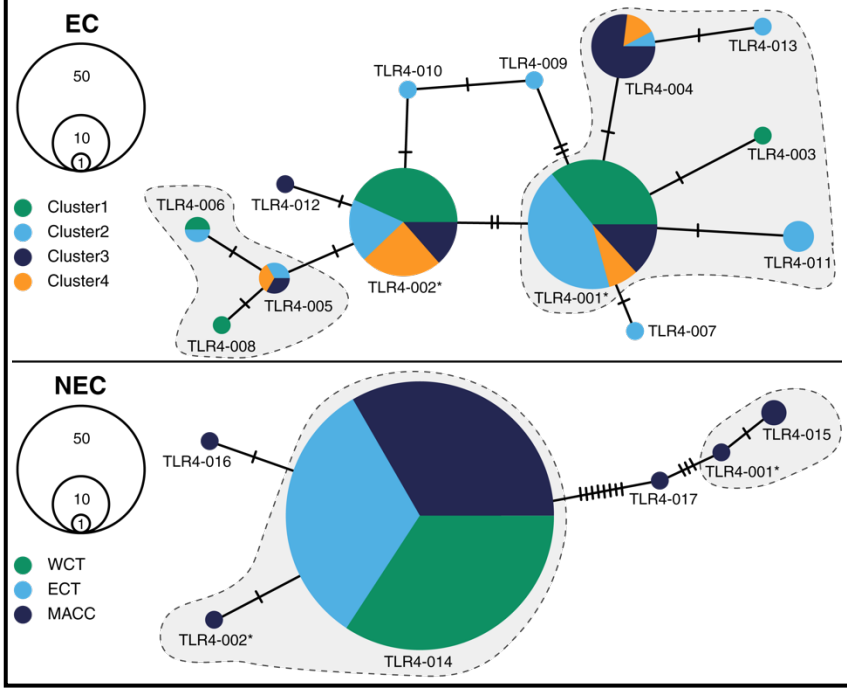

Figure 9. Haplotype network for TLR-4. EC depicted in top panel, NEC depicted in bottom panel. Pie chart area proportional to allele frequency, and colors correspond to population identity. Nucleotide differences indicated by hatchmarks. Asterisk by allele name indicates variant is shared between species. Redundant haplotypes encircled and shaded.

PAML analyses revealed that for both DQA and DRB, models that permitted positive evolution (M2a, M8) were significantly better fits over those that did not (M1a, M7) (all *p*-values < 0.0001). At the amino-acid level, specific codons were under selection for both MHC and TLR genes (**Figure 10**). Positive selection is evident in both DQA and DRB, and residues with high d_N/d_S values are often located within the PBR. The DRB locus overall had more codons under positive selection and greater rates of non-synonymous substitutions per non-synonymous site

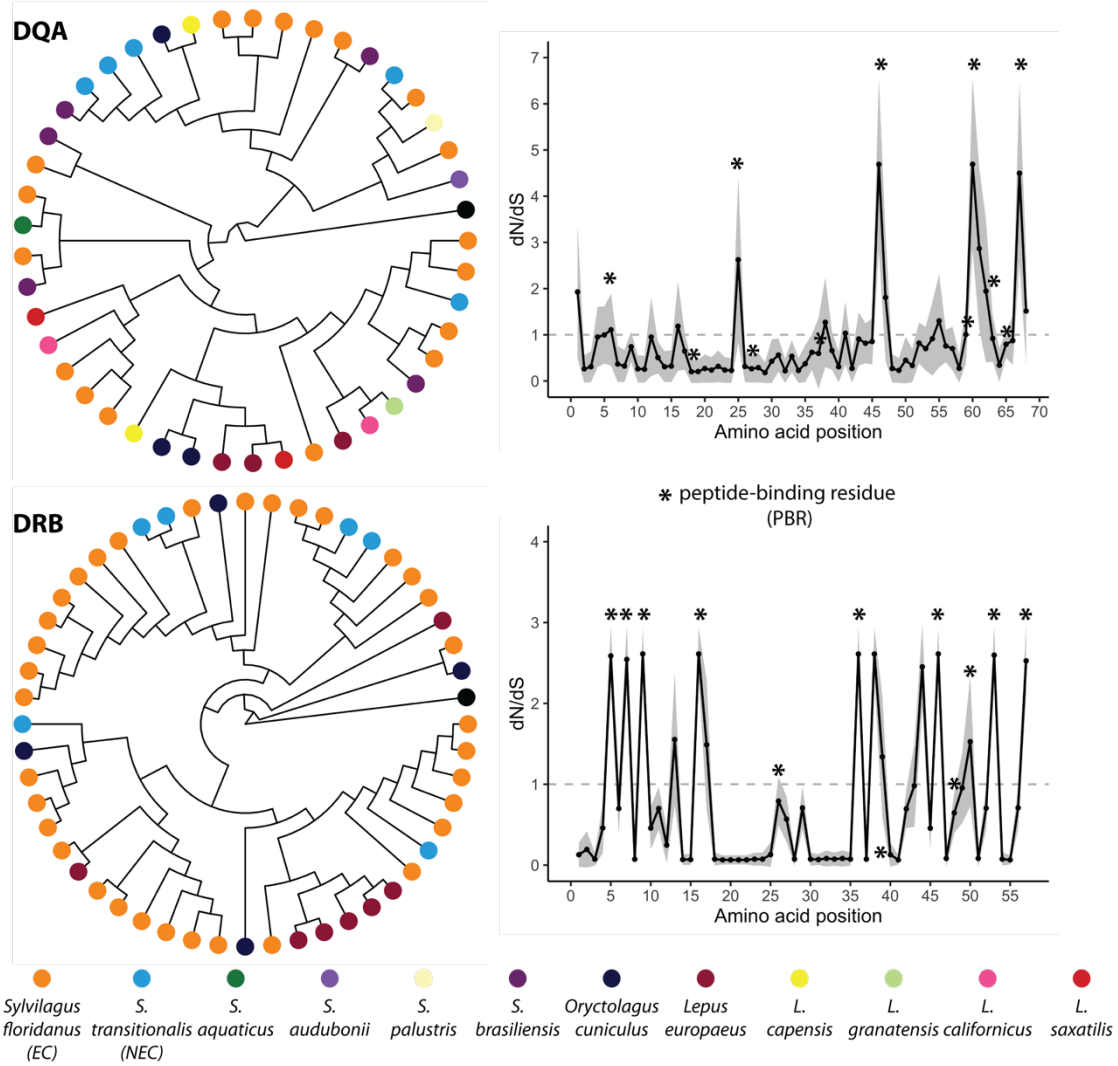

Figure 10. Historical selection acting on DQA (top) and DRB (bottom). Left: Phylogeny of targeted loci across Leporidae. Taxa are colored by species. NEC and EC sequences were generated from this study, and additional lagomorph sequences were retrieved from NCBI. *Mus musculus* used as outgroup (black). Right: Codons under positive selection. d_N/d_S estimated using PAML and a model permitting positive selection (M8). * indicates residues that contribute to the peptide-binding region (PBR).

than DQA. For both MHC loci, PBR regions had greater d_N than d_S for both species, and this pattern was reversed for non-PBR regions, meaning that d_S was greater than d_N (**Figure 11**). Elevated d_N/d_S at PBR sites also corresponded with elevated p-distance at PBR sites relative to non-PBR sites (**Figure 11**).

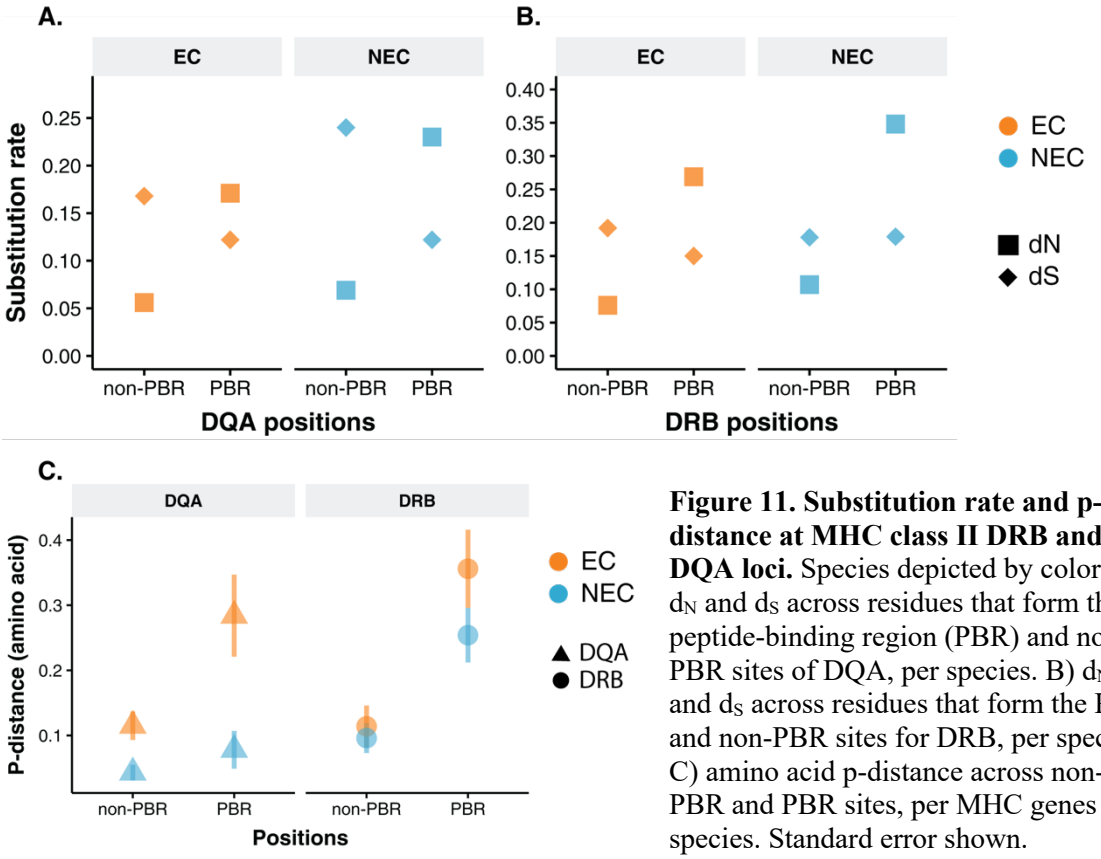

Figure 11. Substitution rate and p-distance at MHC class II DRB and DQA loci. Species depicted by color. A) d_N and d_S across residues that form the peptide-binding region (PBR) and non-PBR sites of DQA, per species. B) d_N and d_S across residues that form the PBR and non-PBR sites for DRB, per species. C) amino acid p-distance across non-PBR and PBR sites, per MHC genes and species. Standard error shown.

For each TLR locus, codons were identified as under pervasive selection for both species (**Table 1**). Both TLR-2 and TLR-4 included codons under positive and purifying selection, as well as codons that

Table 1. Selection on TLR loci by codon.

Locus	Selection	EC	NEC	Both
TLR-2	Positive	115*†		103, 115†
	Negative	35, 47†, 168, 246†, 256	35, 246†	35, 47†, 246†, 256
TLR-4	Positive	247†	39, 286	39, 247†, 286
	Negative	135, 147†, 163, 233†, 276†	135, 163	20, 135, 147†, 163, 233†, 276†

* Also detected by Mixed Effect Model of Evolution (MEME)
† Codon located in leucine-rich repeat (LRR) region

are located within LRRs. For both loci, more codons under selection were identified in EC than in NEC, and a subset of codons were under selection in both EC and NEC.

When comparing population differentiation (G'_{ST}) at neutral vs. immune loci, clear patterns emerged at a species level (**Figure 12**). For NEC, genome-wide patterns revealed strong population differentiation relative to lack of discernable structure at immune loci. In contrast, EC generally displayed greater biogeographic signal at immune loci relative to genome-wide SNPs. Due to high intraspecific variability in immune genes for EC, the 95% confidence intervals are substantial, and occasionally overlap with neutral expectations. Probabilities of observing the evident species-level divisions in differentiation patterns (i.e., NEC above the 1:1 line and EC below) were significant globally and per gene ($p_{global} = 9.2 \times 10^{-9}$, $p_{DQA} = 0.002$, $p_{DRB} = 0.018$, $p_{TLR2} = 0.002$, and $p_{TLR4} = 0.018$).

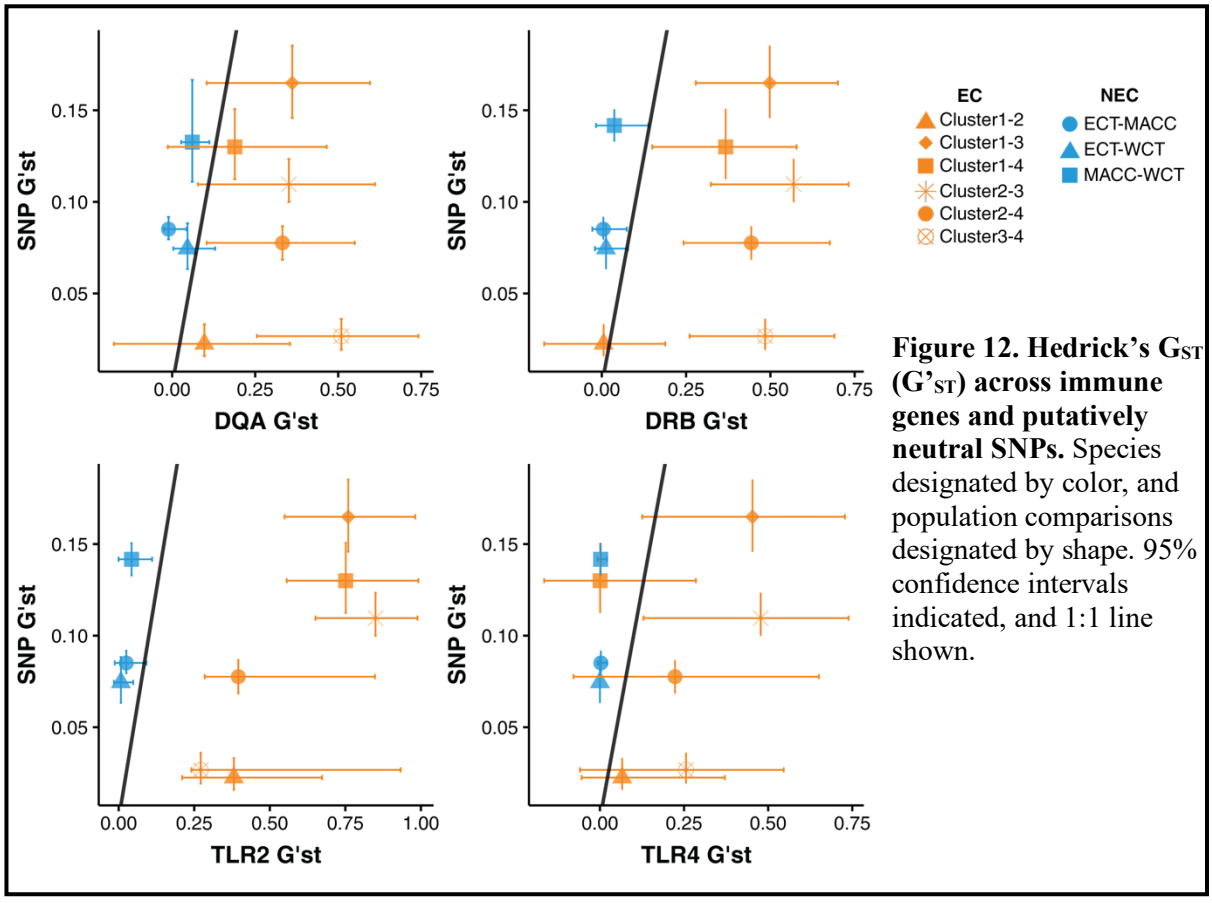

Figure 12. Hedrick's G_{ST} (G'_{ST}) across immune genes and putatively neutral SNPs. Species designated by color, and population comparisons designated by shape. 95% confidence intervals indicated, and 1:1 line shown.

Discussion

Population structure

We investigate population structure and immunogenetic variation in two ecologically similar species with vastly different demographic histories and anthropogenic threats. For the first time, we report population structure at EC based on neutral markers (**Figure 2**). Early studies using mitochondrial DNA (mtDNA) reported a lack of biogeographic signal (Litvaitis *et al.* 1997; Sullivan 2013). However, a lack of mtDNA population structure coupled with strong neutral marker structuring has been reported for other systems and could be due to different selection regimes between the two marker types or sex-biased dispersal patterns (Kim *et al.* 1998; Worthington Wilmer *et al.* 1999; Piertney *et al.* 2000; Wirth & Bernatchez 2001). Invasion theory predicts that multiple waves of introductions and high gene flow between sites may produce weak population structure (Dlugosch & Parker 2008; Zalewski *et al.* 2010). However, following the expansion phase, distinct genetic populations may form due to combinations of genetic drift and local adaptation (Biedrzycka *et al.* 2020). While no significant isolation by distance was detected, population differentiation, as measured by G'_{ST}, was greatest when measured between the most geographically distant populations (e.g., Cluster1 vs. 3, and Cluster 1 vs, 4; **Figure 12**). Similarly, the populations closest in proximity shared the lowest G'_{ST} values (Cluster1 vs. 2, and Cluster3 vs. 4; **Figure 12**).

NEC population structure is consistent with that described by Fenderson *et al.* 2011 (**Figure 3**). Geographic populations generally cluster genetically, with the exception of Maine and New Hampshire samples forming a single genetic population. The RIGS population formed its own cluster, but was close in DAPC space to ECT, and there was some phylogenetic overlap between the two populations (**Figure 3**). A majority of founding NEC breeders in the

reintroduction program were obtained from ECT, particularly in the early years of the program, and therefore the genetic makeup of RIGS appears to reflect this shared ancestry (T.J. McGreevy, Jr., personal communication). In contrast to EC, significant isolation by distance was detected, and population differentiation was greatest between the most distant WCT and MACC comparisons.

MHC diversity

EC MHC displayed high polymorphism, both in terms of sequence diversity and number of haplotypes (**Figures 4-7; Tables S3 & S4**). For DQA, we detected 17 alleles among 123 individuals, yielding a mean amino acid p-distance of 0.150 ± 0.024. In comparison, a study of 267 European brown hares spanning across Belgium and Austria detected 10 DQA alleles (Goüy de Bellocq *et al.* 2009) and reported a mean amino acid p-distance for the European rabbit of nearly half the value that we report for EC. Indeed, Surridge *et al.* 2008 surveyed leporid DQA across 53 individuals spanning 16 species and four genera and reports that pairwise nucleotide comparisons between EC sequences exceeded all other pairwise comparisons. High polymorphism was also recovered for DRB. We report 35 alleles among 198 individuals and species-wide amino acid p-distance of 0.186 ± 0.032. A 2010 study of 518 Lepus europaeus reports 18 alleles, and a 2014 survey spanning 801 hares across 16 populations and two continents uncovered 61 alleles. Although a multitude of EC DRB alleles were recovered, amino acid p-distance values were similar to those reported for *L. europaeus* (0.197 ± 0.042, 0.180 ± 0.034; (Smith *et al.* 2011; Koutsogiannouli *et al.* 2014).

Because six subspecies of EC were co-introduced to New England, it is possible that MHC diversity is high due to extensive population admixture. However, haplotype diversity at immune genotypes substantially exceeds genome-wide neutral estimates (**Figure 5**), markedly

different differentiation patterns emerge at immune loci relative to neutral SNPs (**Figure 11**), and there is no correlation between genome-wide standardized heterozygosity and immunogenetic diversity. These patterns indicate that MHC variability is not simply a reflection of relatively recent widespread admixture. Invasion theory posits that immunological function may modulate invaders' success (Lee & Klasing 2004; White & Perkins 2012; Cornet *et al.* 2016), and other studies have reported that invaders display highly diverse MHC genotypes (Biedrzycka *et al.* 2020). Indeed, Schwensow *et al.* 2017 report that invasive European rabbits in Australia retained high MHC diversity following biocontrol efforts and population bottlenecks and speculate that this may play a role in the species' increasing resistance to rabbit hemorrhagic disease virus. Therefore, it is possible that EC's high immunogenetic diversity may have facilitated its establishment and expansion throughout New England.

Despite population reductions and significant habitat fragmentation, NEC have retained immunogenetic diversity at all four loci targeted (**Figures 4-7; Tables S3-S6)**. Fewer alleles were recovered at all loci relative to EC and other studies of leporids, and amino acid p-distance was lower as well (Goüy de Bellocq *et al.* 2009; Smith *et al.* 2011; Koutsogiannouli *et al.* 2014). Haplotype diversity and observed heterozygosity was also lower compared to EC (**Figure 5**). Whether species retain immunogenetic diversity following population decline is highly variable (Bernatchez & Landry 2003; Aguilar *et al.* 2004; Mainguy *et al.* 2007; Oliver & Piertney 2012), and in extreme cases imperiled species may become fixed for a single allele at historically polymorphic loci (Smulders *et al.* 2003). It is possible that ancestral NEC polymorphism waned following population decline, but it is also possible that current immunogenetic diversity reflects maintenance of historically moderate-to-low levels. Analysis of historical museum specimens obtained prior to population decline could help elucidate this question.

Selection at immune genes

Both DQA and DRB display strong evidence of balancing selection on an evolutionary timescale, which is highly typical for MHC genes (Bernatchez & Landry 2003; Piertney & Oliver 2006) and supported by multiple lines of evidence. We report high sequence polymorphism and haplotype diversity relative to neutral expectations (**Figure 5**), and elevated rates of non-synonymous substitutions at PBR sites (**Figures 10 & 11**). Haplotype networks for both species are characteristic with balancing selection, as evidenced by complex networks of intermediate-frequency clusters with long branches, and allele sharing between the two species (**Figures 6 & 7**). A greater number of alleles were shared among populations for DRB than DQA, and for EC than NEC, suggesting more lineages were retained in DRB relative to DQA and in EC relative to NEC. Although NEC retained fewer immunogenetic variants relative to EC, segregating alleles are marked by long branches and generally cluster in distinct transspecies lineages. Lenz 2011 proposed that one mechanism driving balancing selection at MHC loci may stem from maintaining few but relatively divergent alleles, as more divergent PBRs theoretically provide greater immune surveillance relative to more similar PBRs. The pattern of retention of divergent MHC alleles specifically following population reductions or bottlenecks has been described in other species (van Oosterhout *et al.* 2006; Vlček *et al.* 2016; Marmesat *et al.* 2017).

When a phylogenetic approach was adopted, transspecies polymorphism was clearly evident and leporid phylogeny was not reflected in observed MHC evolutionary patterns (**Figure 10**). For both MHC loci, we observe multiple ancient and distinct lineages retained across species boundaries. Transspecies polymorphism is a hallmark of MHC evolution and a signal of historical balancing selection (Hughes & Yeager 1998; Piertney & Oliver 2006). Footprints of historical positive selection at both MHC loci were also detected at specific codons, primarily

those involved in antigen-binding within the PBR region (**Figure 10**). We report greater ratios of non-synonymous substitutions vs. synonymous substitutions at PBR sites relative to residues not involved in antigen-binding ("non-PBR" sites) and the reversed pattern at non-PBR sites (**Figure 11**). This supports prevailing immunogenetic theory of positive selection residues that interact directly with antigen and purifying selection at non-PBR sites (Hughes & Nei 1988).

No isolation by distance or geographic signal based on G'_{ST} was evident for any NEC immune genes (**Figure 12**). This pattern of biogeographic structuring at neutral loci coupled with low signal at immune loci is another signature of balancing selection and regional homogenization. Interestingly, EC immune genes generally were more highly diverged between populations relative to neutral estimates (**Figure 12**), which could indicate contemporary diversifying selection and local adaptation. Local pathogen regimes and divergent population dynamics could impose selection pressures that lead to population differences in immunogenetic diversity. Similar patterns have been uncovered for other species (Ekblom *et al.* 2007; Loiseau *et al.* 2009; Kyle *et al.* 2014; Meyer-Lucht *et al.* 2016) and were attributed to contemporary spatial variation in pathogen-driven selection. However, 95% confidence intervals for G'_{ST} estimates are quite large due to high intrapopulation and intraindividual variability. More extensive sampling could provide further clarity.

For DQA and DRB, multiple populations across both EC and NEC had less observed than expected heterozygosity (**Tables S3-S6**). Heterozygote advantage is one proposed mechanism underlying high levels of polymorphism typical of MHC genes, and this opposing trend was initially surprising. It is possible that allelic dropout could explain this pattern; however, we tested multiple unique primer pairs specifically to assess this occurrence, and this was never observed. It is also possible that particular MHC alleles confer increased resistance to

local pathogens, and therefore homozygote individuals as well as heterozygotes would benefit. Koutsogiannouli *et al.* 2014 observe the same trend at DQA and DRB loci of the European hare and similarly posit that allele-driven selection may explain this trend. Indeed, a 2010 study of the European rabbit found that a single specific DRB variant conferred resistance to hepatic coccidia infection in young rabbits (Oliver & Piertney 2010).

TLR genes

For each TLR locus, far fewer alleles were recovered relative to MHC loci, along with multiple private and low-frequency alleles (**Tables S3-S6; Figures 4, 6-9**). Kloch & Biedrzycka 2020 posit that this pattern may arise in TLRs due to pathogen-driven selection retaining older pathogen lineages associated with common haplotypes, and numerous private alleles arise due to more recent spatially-driven selection. In our study, recovered haplotypes are highly genetically similar, with TLRs displaying low nucleotide diversity, low amino acid p-distance, and recurrent haplotype redundancy (**Tables S5-S6; Figures 8 & 9**). These patterns are consistent with purifying selection removing deleterious mutations.

While TLR studies in lagomorphs are still in their infancy, Stefanović *et al.* 2020 detected purifying selection in TLR-2 of the European hare, and Awadi *et al.* 2018 similarly reports patterns of both purifying selection within TLR-2 domains and low biogeographic variability. For both TLR-2 and TLR-4, the majority of codons under selection were identified as subject to purifying selection, although positive selection was also detected (**Table 1**). Positively selected codons have been identified in the extracellular domain and LRRs of TLR-2 in rodents (Tschirren *et al.* 2011, 2012; Kloch & Biedrzycka 2020) and TLR-3 in the European rabbit (Abrantes *et al.* 2013). Tschirren *et al.* 2012 posits that historical TLR evolution may be shaped by episodic selective sweeps coupled with subsequent purifying selection, but that species are

frequently at different stages within this process due to varying pathogen pressures or intrinsic host immunological capabilities. Indeed, the authors speculate that this is the case for sympatric yellow-necked mice and bank voles displaying markedly different patterns at the TLR-2 locus. A similar scenario could be occurring with EC and NEC, whereby NEC are at the end stage of this process, and EC are diversifying and adapting locally (Tschirren *et al.* 2012). Less variation in TLRs compared to MHC, particularly as seen within NEC, could also be partially driven by stochastic processes associated with population decline rather than purifying selection alone. However, largely monomorphic TLRs have been reported for other mammals with varying demographic histories (Tschirren *et al.* 2011).

Conservation Implications

Understanding whether MHC diversity differs in wild, reintroduced, and translocated populations can be factored into species management decisions. Here, we investigate MHC diversity in two managed Rhode Island populations, those of Patience Island (RIPI) and those of Great Swamp Management Area (RIGS), which stems from Patience Island-derived reintroduced individuals. While sample sizes for both sites are low, DRB diversity is evident at both locales (**Figures 4 & 7; Table S4**). DRB haplotype diversity estimates do not significantly differ from wild populations (**Figure 5**), and observed heterozygosity and p-distance are high. In contrast, only one DQA haplotype is recovered for each population, which is shared at high frequency among all wild populations (**Figures 4 & 6; Table S3**). This is likely due to undersampling, but it is also possible that founder effects and/or genetic drift have eliminated DQA diversity at these sites. Limited DQA diversity may impact NEC reproduction, and further investigation is warranted. In a study of European brown hare, Smith *et al.* 2010 found that while the DRB locus had little impact on hare reproduction, DQA heterozygosity was associated with reproductive

ability and litter size. Whether reduced DQA diversity in managed populations could impact reproduction in NEC remains unknown.

Local adaptation at immune genes has important implications for conservation programs due to tradeoffs between increasing gene flow and preserving local adaptation. For NEC, we detect little evidence for pervasive local adaptation at immune genes. Rather, we find signatures of balancing selection observed at both evolutionary and contemporary timescales, resulting in little biogeographic signal and regional homogenization relative to neutral SNPs. However, private alleles were detected for four immunogenetic loci, and less heterozygosity than expected could indicate that particular alleles are locally adaptive.

Invasive species are of huge economic and environmental impacts, and understanding any genetic bases for successful establishment is crucial. High diversity at immune loci may increase survival during epizootics and provide greater material for adaptation to disease threats (Osborne *et al.* 2015; Schwensow *et al.* 2017). Should disease emerge within New England's cottontails, it is possible that EC will be better equipped to develop adaptive resistance relative to NEC. This is of particular and urgent concern given that the deadly rabbit hemorrhagic disease virus serotype 2 (RHDV2) was reported in New York in 2020 (United States Department of Agriculture 2020). While our immunogenetic study focuses on loci primarily involved in bacterial detection, immunogenetic patterns at other loci involved in viral detection may be similar, meaning that EC potentially possess a greater pool of adaptive variation to draw from.

This study also highlights the importance of undertaking comparative immunogenetic studies among non-model systems of conservation concern. Even among closely related and ecologically similar species, it is difficult to apply findings between species. Rather, management decisions may be better guided by system-specific investigations.

Acknowledgements

We would like to thank collaborators who helped with EC and NEC specimen procurement: New Hampshire Fish and Game, Massachusetts Department of Fish and Game, Connecticut Department of Energy and Environmental Protection, Rhode Island Department of Environmental Management, Roger Williams Park Zoo, and the Queens Zoo. Thank you to Adrienne Kovach of the University of New Hampshire for sending New Hampshire EC and NEC samples for MHC screening. We thank Tom Husband and Mary Sullivan of the University of Rhode Island for their assistance with ddRAD funding and library preparation as part of an ongoing WGEL project. We additionally thank Daniel Weinreich of Brown University for assistance with statistical analyses of population differentiation. We also thank Brown University undergraduate researchers Robert Blattner, Emily Longman, and Paul Rizzo for their assistance with molecular wet lab work. Lastly, we thank Janet Atoyan of the URI Genomics Core for her assistance with Illumina MiSeq library preparation and sequencing. This research was supported by computational resources at Brown University's Center for Computation and Visualization. K.B.N. and J.C.B.N. were supported by the NSF Graduate Research Fellowship Program (GRFP) and NSF IGERT Grant.

References

Abrantes J, Areal H, Esteves PJ (2013) Insights into the European rabbit (Oryctolagus cuniculus) innate immune system: Genetic diversity of the toll-like receptor 3 (TLR3) in wild populations and domestic breeds. *BMC Genetics*, **14**.

Acevedo-Whitehouse K, Cunningham AA (2006) Is MHC enough for understanding wildlife immunogenetics? *Trends in Ecology and Evolution*, **21**, 433–438.

Agriculture USD of (2020) *Factsheet: Rabbit hemorrhagic disease*.

Aguilar A, Roemer G, Debenham S *et al.* (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*, **101**, 3490–4.

Alexander DH, Lange K (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, **12**.

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.

Alho JS, Välimäki K, Merilä J (2010) Rhh: An R extension for estimating multilocus heterozygosity and heterozygosity-heterozygosity correlation. *Molecular Ecology Resources*, **10**, 720–722.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature*, **11**, 697–709.

Awadi A, Ben Slimen H, Smith S *et al.* (2018) Genetic diversity of the toll-like receptor 2 (TLR2) in hare (Lepus capensis) populations from Tunisia. *Comptes Rendus - Biologies*, **341**, 315–324.

Ayres JC, Feemster RF (1948) Epidemiology of tularemia in Massachusetts with a review of the literature. *New England Journal of Medicine Journal of Medicine*, **238**, 187–194.

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

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

Bekessy SA, Ennos RA, Burgman MA, Newton AC, Ades PK (2003) Neutral DNA markers fail

to detect genetic divergence in an ecologically important trait. *Biological Conservation*, **110**, 267–275.

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**, 363–377.

Beutler B (2004) Inferences, questions and possibilities in Toll-like receptor signalling. *Nature*, **430**, 257–263.

Biedrzycka A, Konopiński 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.

Biedrzycka A, Sebastian A, Migalska M, Westerdahl H, Radwan J (2016) Testing genotyping strategies for ultra-deep sequencing of a co-amplifying gene family : MHC class I in a passerine bird. *Molecular Ecology Resources*.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: An analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Chapman TR, Morgan RP (1973) Allen Press. *Wildlife Monographs*, **36**, 3–54.

Chifman J, Kubatko L (2014) Quartet inference from SNP data under the coalescent model. *Bioinformatics*, **30**, 3317–3324.

Čížková D, Javůrková V, Champagnon J, Kreisinger J (2012) Duck's not dead: Does restocking with captive bred individuals affect the genetic integrity of wild mallard (Anas platyrhynchos) population? *Biological Conservation*, **152**, 231–240.

Coltman DW, Pilkington JG, Smith JA, Pemberton JM (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution*, **53**, 1259–1267.

Cornet S, Brouat C, Diagne C, Charbonnel N (2016) immunology and bioinvasion : revisiting the evolution of increased competitive ability hypotheses. *Evolutionary Applications*, **9**, 952–962.

Cui J, Cheng Y, Belov K (2015) Diversity in the Toll-like receptor genes of the Tasmanian devil (Sarcophilus harrisii). *Immunogenetics*, **67**, 195–201.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Dixon P (2003) VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, **14**, 927–930.

Dlugosch KM, Parker IM (2008) Founding events in species invasions: Genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, **17**, 431–449.

Edmands S (2007) Between a rock and a hard place: Evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology*, **16**, 463–475.

Ekblom R, Sæther SA, Jacobsson P *et al.* (2007) Spatial pattern of MHC class II variation in the great snipe (Gallinago media). *Molecular Ecology*, **16**, 1439–1451.

Fain M, Zhao T, Kindt T (2001) Improved typing procedure for the polymorphic single-copy RLA-DQA gene of the rabbit reveals a new allele. *Tissue Antigens*, **57**, 332–338.

Fenderson LE, Kovach AI, Litvaitis JA, Litvaitis MK (2011) Population genetic structure and history of fragmented remnant populations of the New England cottontail (Sylvilagus transitionalis). *Conservation Genetics*, **12**, 943–958.

Fornůsková A, Vinkler M, Pagès M *et al.* (2013) Contrasted evolutionary histories of two Toll-like receptors (Tlr4 and Tlr7) in wild rodents (MURINAE). *BMC evolutionary biology*, **13**, 194.

Francis RM (2017) pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, **17**, 27–32.

Fuller S, Tur A (2012) Conservation Strategy for the New England Cottontail (Sylvilagus transitionalis). , 143.

Fuller S, Tur A (2015) New England Cottontail Conservation Progress. , 303–312.

Garrigan D, Hedrick PW (2003) Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution*, **57**, 1707–1722.

Ge D, Wen Z, Xia L *et al.* (2013) Evolutionary History of Lagomorphs in Response to Global Environmental Change. *PLoS ONE*, **8**.

Genton BJ, Shykoff JA, Giraud T (2005) High genetic diversity in French invasive populations of common ragweed, Ambrosia artemisiifolia, as a result of multiple sources of introduction. *Molecular Ecology*, **14**, 4275–4285.

Goüy de Bellocq J, Suchentrunk F, Baird SJE, Schaschl H (2009) Evolutionary history of an MHC gene in two leporid species: characterisation of Mhc-DQA in the European brown hare and comparison with the European rabbit. *Immunogenetics*, **61**, 131–44.

Grogan KE, McGinnis GJ, Sauther ML, Cuozzo FP, Drea CM (2016) Next-generation genotyping of hypervariable loci in many individuals of a non-model species: Technical and theoretical implications. *BMC Genomics*, **17**, 1–16.

Gruber B, Unmack PJ, Berry OF, Georges A (2018) dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources*, **18**, 691–699.

Hall T (2011) BioEdit: An important software for molecular biology. *GERF Bulletin of Biosciences*, **2**, 60–61.

Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.

Hughes AL (1991) MHC Polymorphism and the Design of Captive Breeding Programs. *Conservation Biology*, **5**, 249–251.

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, **336**, 403–405.

Hughes AL, Yeager M (1998) Natural selection and major histocompatibility complex loci of vertebrates. *Annual Review of Genetics*, **32**, 415–435.

Johnston JE (1972) Identification and distribution of cottontial rabbits in southern New England. University of Connecticut.

Jombart T (2016) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

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**, 782–788.

Kim I, Phillips CJ, Monjeau JA *et al.* (1998) Habitat islands, genetic diversity, and gene flow in a Patagonian rodent. *Molecular Ecology*, **7**, 667–678.

Kloch A, Biedrzycka A (2020) Post-glacial phylogeography and variation in innate immunity loci in a sylvatic rodent, bank vole Myodes glareolus. *Mammalian Biology*, **100**, 141–154.

Kolbe JJ, Glor RE, Rodríguez Schettino L *et al.* (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 177–81.

Kolbe JJ, Larson A, Losos JB (2007) Differential admixture shapes morphological variation among invasive populations of the lizard Anolis sagrei. *Molecular Ecology*, **16**, 1579–1591.

Koutsogiannouli EA, Moutou KA, Stamatis C, Walter L, Mamuris Z (2014) Genetic variation in

the major histocompatibility complex of the European brown hare (Lepus europaeus) across distinct phylogeographic areas. *Immunogenetics*, **66**, 379–392.

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution*, **33**, 1870–1874.

Kyle CJ, Rico Y, Castillo S *et al.* (2014) Spatial patterns of neutral and functional genetic variations reveal patterns of local adaptation in raccoon (Procyon lotor) populations exposed to raccoon rabies. *Molecular Ecology*, **23**, 2287–2298.

Lee KA, Klasing KC (2004) A role for immunology in invasion biology. *Trends in Ecology and Evolution*, **19**, 523–529.

Leigh JW, Bryant D (2015) POPART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, **6**, 1110–1116.

Lenz TL (2011) Computational prediction of MHC II-antigen binding supports divergent allele advantage and explains trans-species polymorphism. *Evolution*, **65**, 2380–2390.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Lighten J, van Oosterhout C, Bentzen P (2014) Critical review of NGS analyses for de novo genotyping multigene families. *Molecular Ecology*, **23**, 3957–3972.

Litvaitis JA (2003) Are pre-Columbian conditions relevant baselines for managed forests in the northeastern United States? *Forest Ecology and Management*, **185**, 113–126.

Litvaitis JA, Barbour MS, Brown AL *et al.* (2007) Testing Multiple Hypotheses to Identify Causes of the Decline of a Lagomorph Species : The New England Cottontail as a Case Study Natural History of New England Cottontails. In: *Lagomorph Biology: Evolution, Ecology, and Conservation* (eds Alves P, Ferrand N, Hacklander K), pp. 167–185. Berling

Heidelberg.

Litvaitis JA, Johnson B, Jakubas W, Morris K (2003) Distribution and habitat features associated with remnant populations of New England cottontails in Maine. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **81**, 877–887.

Litvaitis MK, Litvaitis JA, Lee W-J, Kocher TD (1997) Variation in the mitochondrial DNA of the Sylvilagus complex occupying the northeastern United States. *Canadian Journal of Zoology*, **75**, 595–605.

Litvaitis JA, Tash JP, Litvaitis MK *et al.* (2004) A range-wide survey to determine the current distribution of New England cottontails. *Wildlife Society Bulletin*, **34**, 1190–1197.

Loiseau C, Richard M, Garnier S *et al.* (2009) Diversifying selection on MHC class i in the house sparrow (Passer domesticus). *Molecular Ecology*, **18**, 1331–1340.

Mainguy J, Worley K, Côté SD, Coltman DW (2007) Low MHC DRB class II diversity in the mountain goat: past bottlenecks and possible role of pathogens and parasites. *Conservation Genetics*, **8**, 885–891.

Marmesat E, Schmidt K, Saveljev AP, Seryodkin I V., Godoy JA (2017) Retention of functional variation despite extreme genomic erosion: MHC allelic repertoires in the Lynx genus. *BMC Evolutionary Biology*, **17**, 1–16.

Marsden CD, Verberkmoes H, Thomas R, Wayne RK, Mable BK (2013) Pedigrees, MHC and microsatellites: An integrated approach for genetic management of captive African wild dogs (Lycaon pictus). *Conservation Genetics*, **14**, 171–183.

Matsushima N, Tanaka T, Enkhbayar P *et al.* (2007) Comparative sequence analysis of leucine-rich repeats (LRRs) within vertebrate toll-like receptors. *BMC Genomics*, **8**, 1–20.

Matthee CA, Jansen Van Vuuren B, Bell D, Robinson TJ (2004) A molecular supermatrix of the

rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the Miocene. *Systematic Biology*, **53**, 433–447.

Meyer-Lucht Y, Mulder KP, James MC *et al.* (2016) Adaptive and neutral genetic differentiation among Scottish and endangered Irish red grouse (Lagopus lagopus scotica). *Conservation Genetics*, **17**, 615–630.

Morger J, Bajnok J, Boyce K *et al.* (2014) Naturally occurring Toll-like receptor 11 (TLR11) and Toll-like receptor 12 (TLR12) polymorphisms are not associated with Toxoplasma gondii infection in wild wood mice. *Infection, Genetics and Evolution*, **26**, 180–184.

Murrell B, Moola S, Mabona A *et al.* (2013) FUBAR: A fast, unconstrained bayesian AppRoximation for inferring selection. *Molecular Biology and Evolution*, **30**, 1196–1205.

Murrell B, Wertheim JO, Moola S *et al.* (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, **8**.

O'Grady JJ, Brook BW, Reed DH *et al.* (2006) Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biological Conservation*, **133**, 42–51.

O'Leary SJ, Puritz JB, Willis SC, Hollenbeck CM, Portnoy DS (2018) These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, **27**, 3193–3206.

Oliver M, Piertney S (2010) Beyond splitting hares and rabbiting on about major histocompatibility complex complexity. *Molecular Ecology*, **19**, 4099–4101.

Oliver MK, Piertney SB (2012) Selection Maintains MHC Diversity through a Natural Population Bottleneck. *Molecular Biology and Evolution*, **29**, 1713–1720.

van Oosterhout C, Joyce DA, Cummings SM *et al.* (2006) Balancing Selection, Random Genetic Drift, and Genetic Variation At the Major Histocompatibility Complex in Two Wild

Populations of Guppies (Poecilia Reticulata). *Evolution*, **60**, 2562.

Oppelt C, Starkloff A, Rausch P, Von Holst D, Rödel HG (2010) Major histocompatibility complex variation and age-specific endoparasite load in subadult European rabbits. *Molecular ecology*, **19**, 4155–67.

Osborne AJ, Pearson J, Negro SS *et al.* (2015) Heterozygote advantage at MHC DRB may influence response to infectious disease epizootics. *Molecular Ecology*, **24**, 1419–1432.

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW (2010) Conservation genetics in transition to conservation genomics. *Trends in Genetics*, **26**, 177–187.

Paradis E (2010) Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, **26**, 419–420.

Paris JR, Stevens JR, Catchen JM (2017) Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*, **8**, 1360–1373.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, **7**.

Petkova D, Novembre J, Stephens M (2015) Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, **48**, 94–100.

Piertney SB, Maccoll ADC, Bacon PJ *et al.* (2000) Matrilineal genetic structure and female-mediated gene flow in red grouse (Lagopus lagopus scoticus): An analysis using mitochondrial DNA. *Evolution*, **54**, 279–289.

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

Primmer CR (2009) From conservation genetics to conservation genomics. *Ann NY Acad Sci*,

1162, 357–368.

Quéméré E, Galan M, Cosson JF *et al.* (2015) Immunogenetic heterogeneity in a widespread ungulate: The European roe deer (Capreolus capreolus). *Molecular Ecology*, **24**, 3873–3887.

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

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

Reed DH (2005) Relationship between population size and fitness. *Conservation Biology*, **19**, 563–568.

Revell LJ (2012) phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.

Rochette NC, Catchen JM (2017) Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, **12**, 2640–2659.

Sakai AK, Allendorf FW, Holt JS *et al.* (2001) The population biology of invasive species. *Annu. Rev. Ecol. Syst.*, **32**, 305–32.

Schad J, Sommer S, Ganzhorn JU (2004) MHC variability of a small lemur in the littoral forest fragments of southeastern Madagascar. *Conservation Genetics*, **5**, 299–309.

Schwensow N, Mazzoni CJ, Marmesat E *et al.* (2017) High adaptive variability and virus-driven selection on major histocompatibility complex (MHC) genes in invasive wild rabbits in Australia. *Biological Invasions*, **19**, 1255–1271.

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**, 498–510.

Shine R, Brown GP, Phillips BL (2011) An evolutionary process that assembles phenotypes through space rather than through time. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 5708–5711.

Silva SM, Ruedas LA, Santos LH, E Silva JDS, Aleixo A (2019) Illuminating the obscured phylogenetic radiation of South American Sylvilagus Gray, 1867 (Lagomorpha: Leporidae). *Journal of Mammalogy*, **100**, 31–44.

Smith S, De Bellocq JG, Suchentrunk F, Schaschl H (2011) Evolutionary genetics of MHC class II beta genes in the brown hare, Lepus europaeus. *Immunogenetics*, **63**, 743–751.

Smith S, Mang T, De Bellocq JG *et al.* (2010) Homozygosity at a class II MHC locus depresses female reproductive ability in European brown hares. *Molecular Ecology*, **19**, 4131–4143.

Smulders MJM, Snoek LB, Booy G, Vosman B (2003) Complete loss of MHC genetic diversity in the Common Hamster (Cricetus cricetus) population in The Netherlands. Consequences for conservation strategies. *Conservation Genetics*, **4**, 441–451.

Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in zoology*, **2**, 16.

Sommer S, Courtiol A, Mazzoni CJ (2013) MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. *BMC Genomics*, **14**, 1.

Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 979–988.

Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of

large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Stefanović M, Djan M, Veličković N *et al.* (2020) Purifying selection shaping the evolution of the Toll-like receptor 2 TIR domain in brown hares (Lepus europaeus) from Europe and the Middle East. *Molecular Biology Reports*, **47**, 2975–2984.

Steiner CC, Putnam AS, Hoeck PEA, Ryder OA (2013) Conservation Genomics of Threatened Animal Species. *Annual Review of Animal Biosciences*, **1**, 261–281.

Sullivan ME (2013) Comparative Phylogeography and a Mitochondrial DNA Barcode for Identifying Three Sympatric Lagomorphs in the Northeastern United States. University of Rhode Island.

Surridge AK, van der Loo W, Abrantes J *et al.* (2008) Diversity and evolutionary history of the MHC DQA gene in leporids. *Immunogenetics*, **60**, 515–525.

Takeda K, Kaisho T, Akira S (2003) Toll-like Receptors. *Annual Review of Immunology*, **21**, 335–376.

Tash JP, Litvaitis JA (2007) Characteristics of occupied habitats and identification of sites for restoration and translocation of New England cottontail populations. *Biological Conservation*, **137**, 584–598.

Tschirren B, Andersson M, Scherman K *et al.* (2013) Polymorphisms at the innate immune receptor TLR2 are associated with Borrelia infection in a wild rodent population. *Proc Biol Sci*, **280**, 20130364.

Tschirren B, Andersson M, Scherman K, Westerdahl H, Råberg L (2012) Contrasting Patterns of Diversity and Population Differentiation At the Innate Immunity Gene Toll-Like Receptor 2 (Tlr2) in Two Sympatric Rodent Species. *Evolution*, **66**, 720–731.

Tschirren B, Råberg L, Westerdahl H (2011) Signatures of selection acting on the innate

immunity gene Toll-like receptor 2 (TLR2) during the evolutionary history of rodents. *Journal of Evolutionary Biology*, **24**, 1232–1240.

Tsutsui ND, Suarez A V., Holway DA, Case TJ (2000) Reduced genetic variation and the success of an invasive species. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 5948–5953.

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

Vlček J, Hoeck PEA, Keller LF *et al.* (2016) Balancing selection and genetic drift create unusual patterns of MHCIIβ variation in Galápagos mockingbirds. *Molecular ecology*, **25**, 4757–4772.

Wang LG, Lam TTY, Xu S *et al.* (2020) Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. *Molecular Biology and Evolution*, **37**, 599–603.

Weckx S, Del-Favero J, Rademakers R *et al.* (2005) novoSNP, a novel computational tool for sequence variation discovery. *Genome Research*, **15**, 436–442.

White TA, Perkins SE (2012) The ecoimmunology of invasive species. *Functional Ecology*, **3**, 1313–1323.

Wilgenbusch JC, Swofford D (2003) Inferring Evolutionary Trees with PAUP*. *Current Protocols in Bioinformatics*, **00**, 6.4.1-6.4.28.

Wirth T, Bernatchez L (2001) Genetic evidence against panmixia in the European eel. *Nature*, **409**, 1037–1040.

Worthington Wilmer J, Hall L, Barratt E, Moritz C (1999) Genetic structure and male-mediated gene flow in the ghost bat (Macroderma gigas). *Evolution*, **53**, 1582–1591.

Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Yu G, Smith DK, Zhu H, Guan Y, Lam TTY (2017) Ggtree: an R Package for Visualization and Annotation of Phylogenetic Trees With Their Covariates and Other Associated Data. *Methods in Ecology and Evolution*, **8**, 28–36.

Zalewski A, Michalska-Parda A, Bartoszewicz M, Kozakiewicz M, Brzeziński M (2010) Multiple introductions determine the genetic structure of an invasive species population: American mink Neovison vison in Poland. *Biological Conservation*, **143**, 1355–1363.

Zhou H, Gu J, Lamont SJ, Gu X (2007) Evolutionary analysis for functional divergence of the toll-like receptor gene family and altered functional constraints. *Journal of Molecular Evolution*, **65**, 119–123.

Supplemental Figures

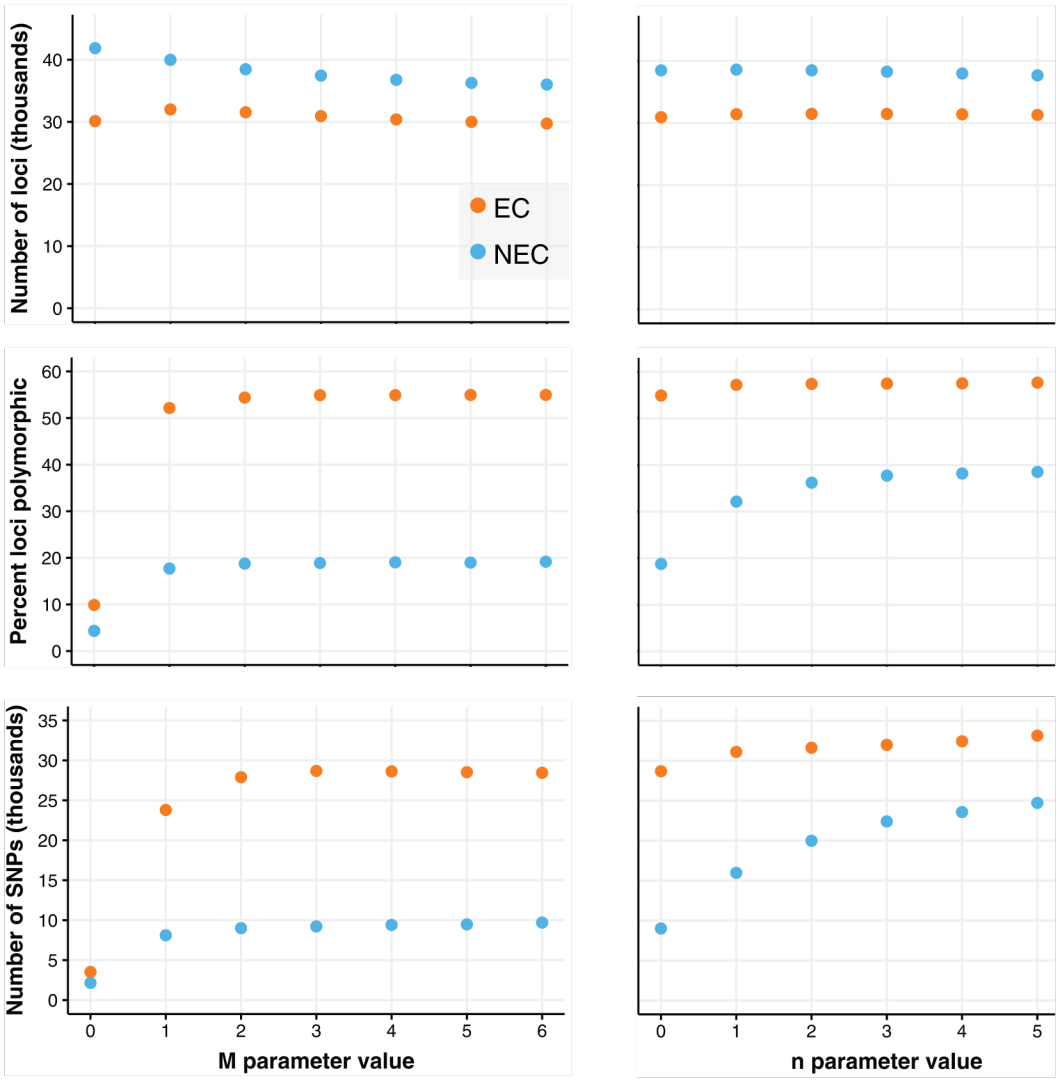

Figure S1. STACKS parameter testing of *M* and *n* values across EC and NEC. Metrics are (top) the number of assembled loci, (middle) the percent of assembled loci that are polymorphic, and (bottom) the number of SNPs recovered. EC are shown in orange, NEC shown in blue.

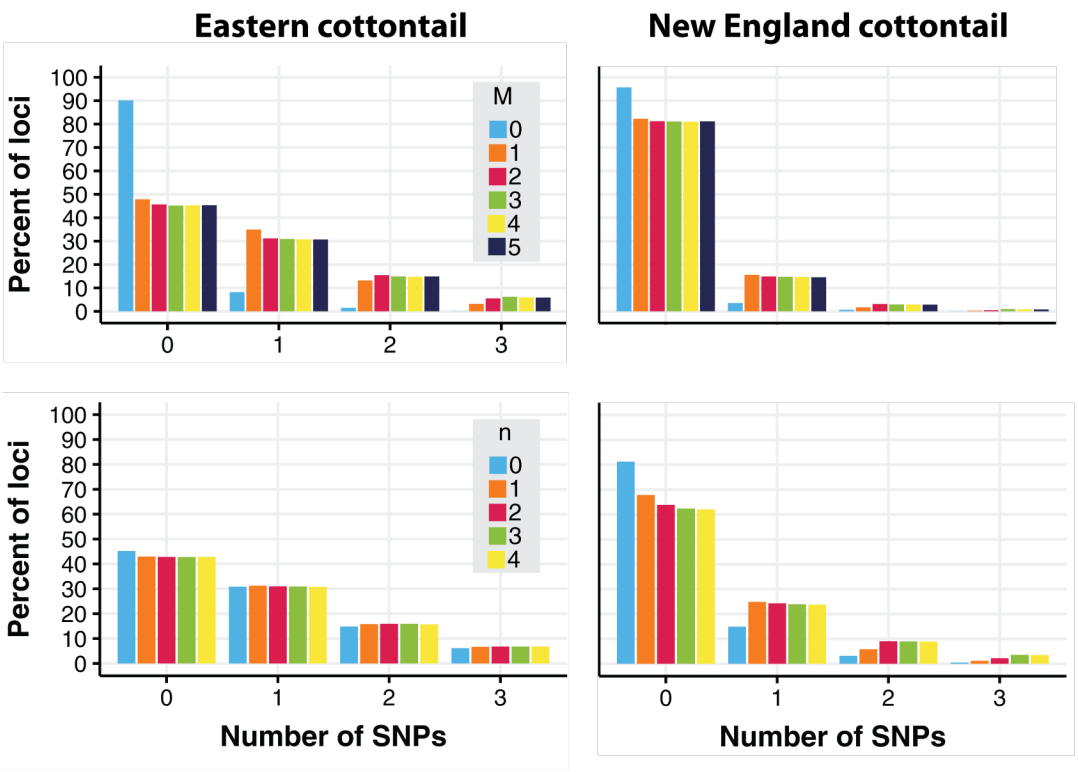

Figure S2. SNP distribution across varying parameter values of *M* and *n*. The distribution of the percent of loci with 0, 1, 2, or 3 SNPs across values of M=0-5 or n=0-4.

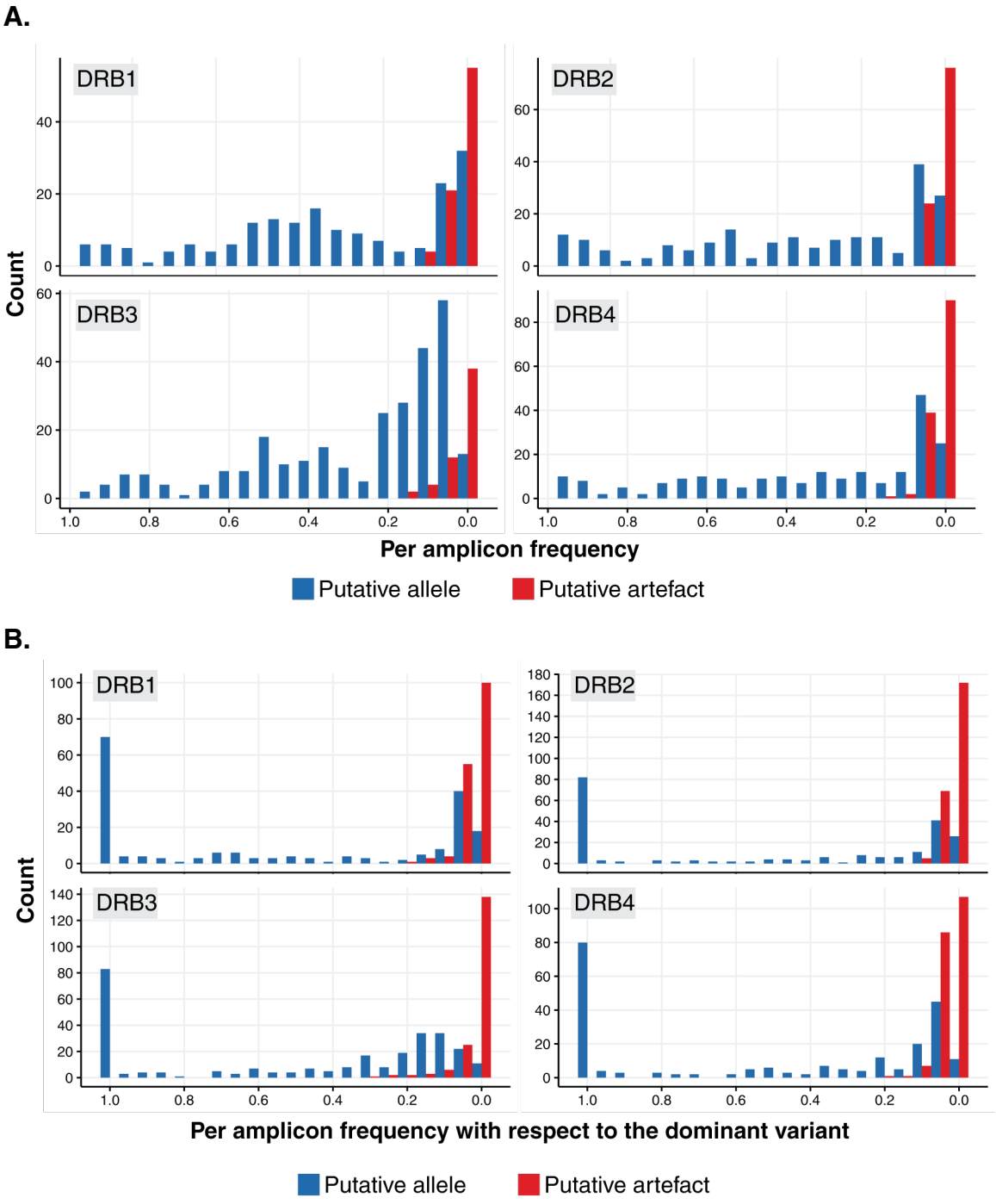

Figure S3. AmpliSAT parameter optimization for DRB thresholds. Four batches of sequencing were individually evaluated (DRB1-DRB4). Putative alleles and artefacts were identified by AmpliSAT. **A)** Counts of putative alleles and artefacts based on their frequency within a given sample amplicon. **B)** Counts of putative alleles and artefacts based on their frequency with respect to the dominant (most abundant) variant within a given sample amplicon.

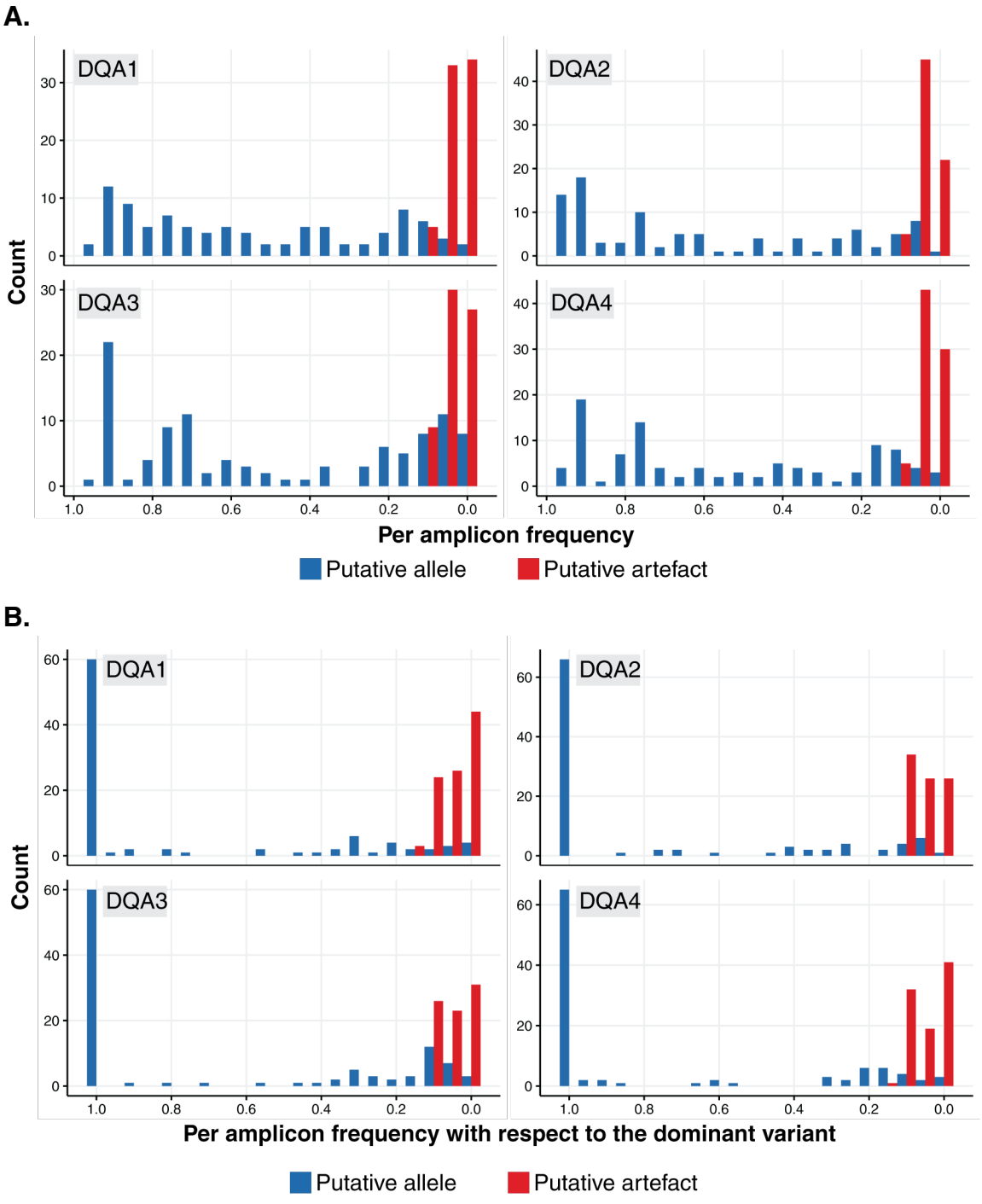

Figure S4. AmpliSAT parameter optimization for DQA thresholds. Four batches of sequencing were individually evaluated (DQA1-DQA4). Putative alleles and artefacts were identified by AmpliSAT. **A)** Counts of putative alleles and artefacts based on their frequency within a given sample amplicon. **B)** Counts of putative alleles and artefacts based on their frequency with respect to the dominant (most abundant) variant within a given sample amplicon.

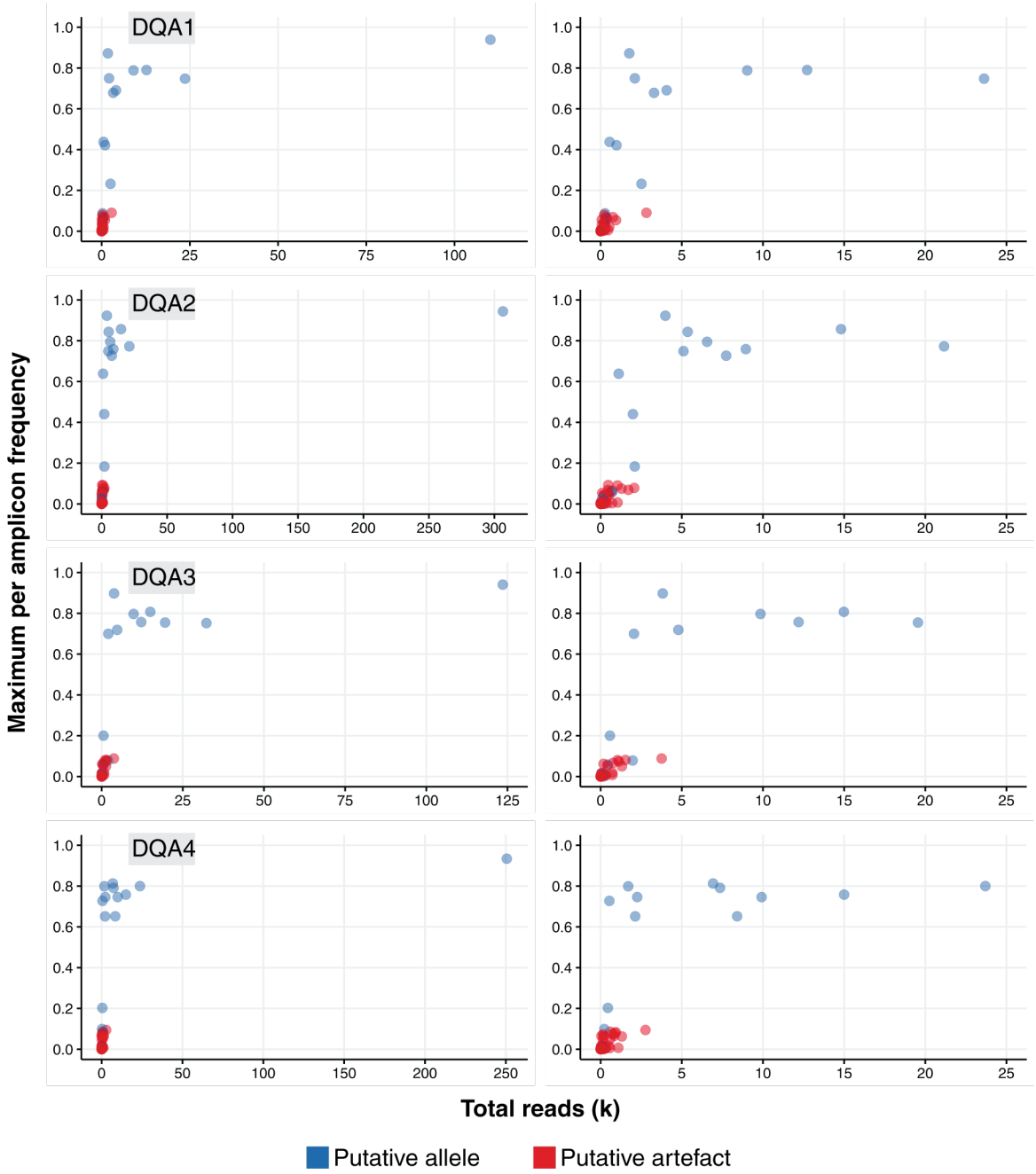

Figure S5. AmpliSAT parameter optimization for DQA thresholds. Four batches of sequencing were individually evaluated (DQA1-DQA4). Putative alleles and artefacts were identified by AmpliSAT. (Left) Panels show the total number of reads for a given sequence (i.e., putative allele or putative artefact) vs. the maximum frequency at which that sequence is detected within a single sample amplicon. (Right) Panels depict a zoomed-in view.

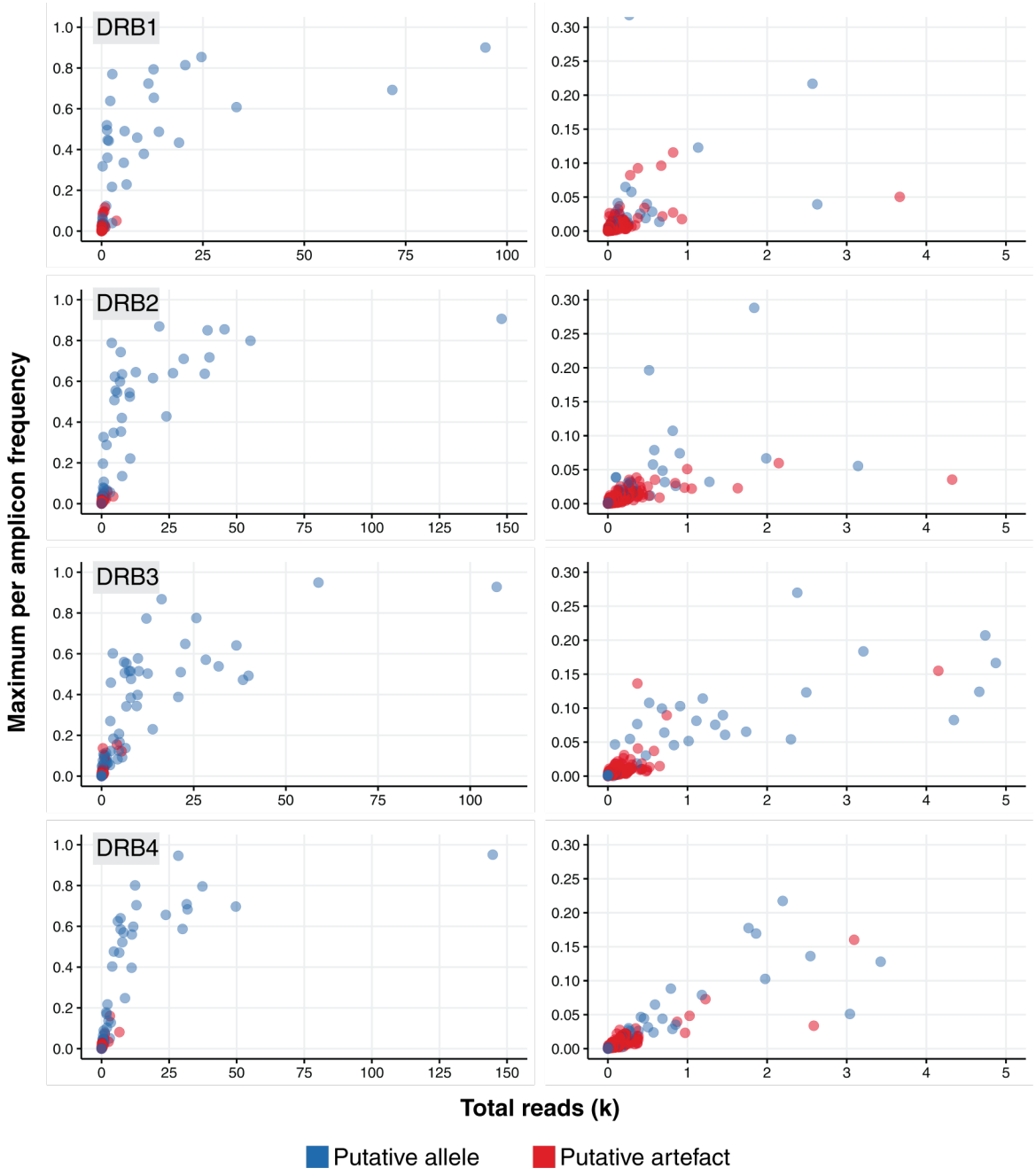

Figure S6. AmpliSAT parameter optimization for DRB thresholds. Four batches of sequencing were individually evaluated (DRB1-DRB4). Putative alleles and artefacts were identified by AmpliSAT. (Left) Panels show the total number of reads for a given sequence (i.e., putative allele or putative artefact) vs. the maximum frequency at which that sequence is detected within a single sample amplicon. (Right) Panels depict a zoomed-in view.

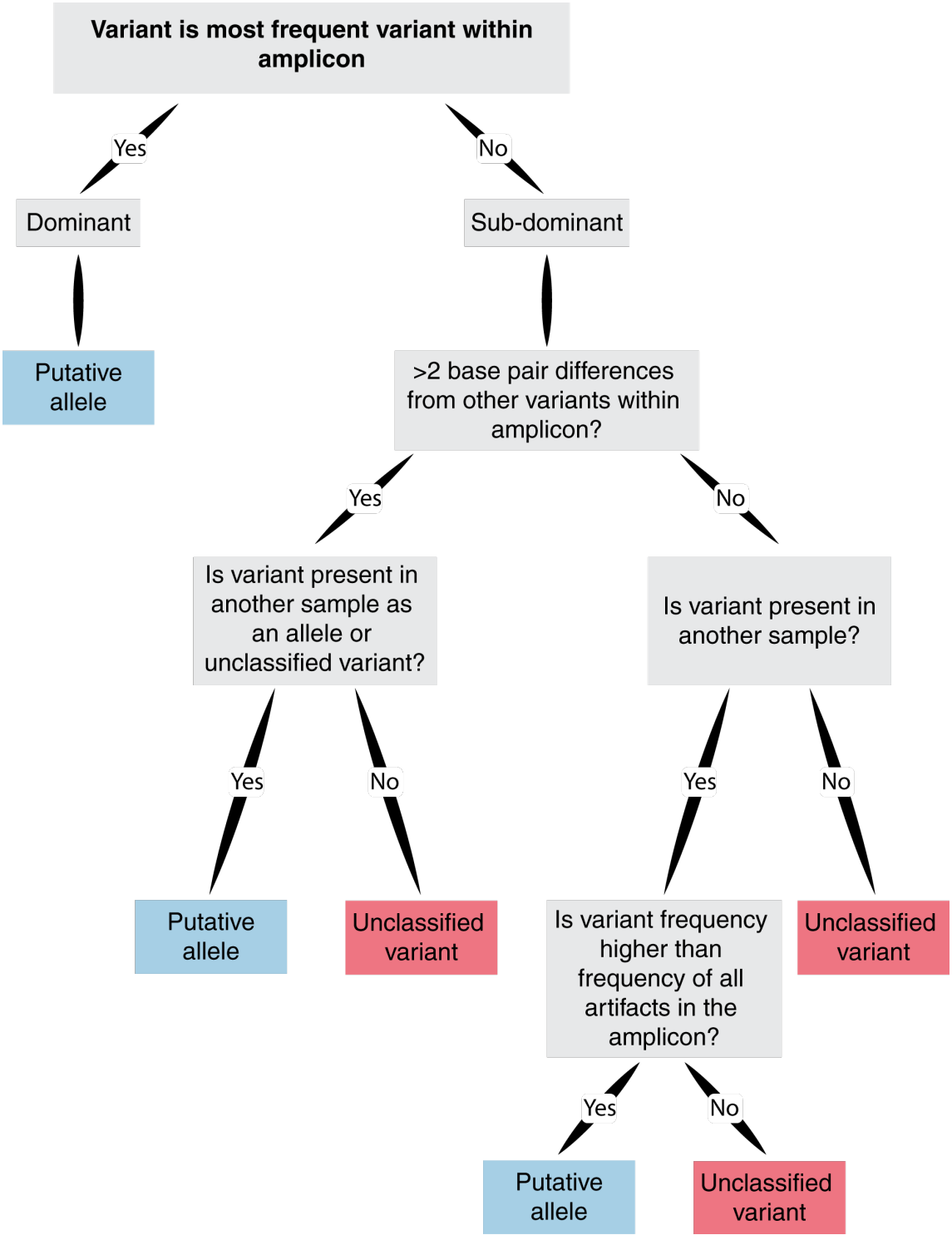

Figure S7. Decision tree for classifying MHC variants as putative alleles or unclassified variants. Adapted from Grogan *et al.* 2016 and applied separately to DQA and DRB.

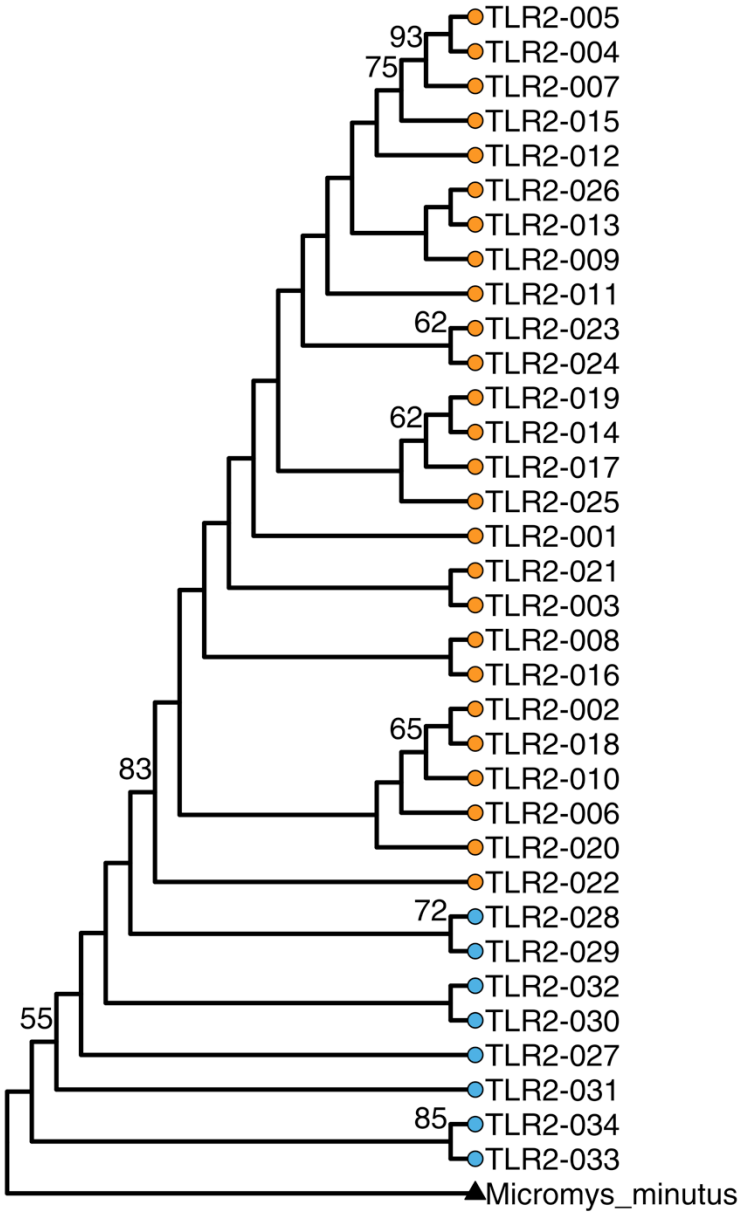

Figure S8. Maximum-likelihood tree of TLR-2 sequences. Branch support above 50 percent labeled and obtained from 1,000 bootstrap replicates. Tree was rooted using the rodent outgroup *Micromys minutus* (accession number HM215603.1). EC indicated by orange circles, and NEC are indicated by blue circles.

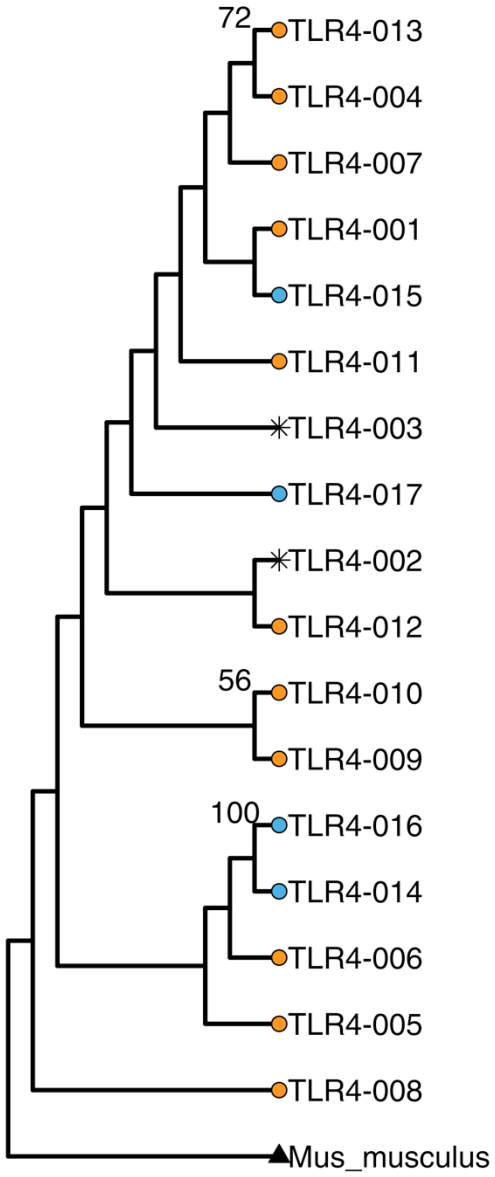

Figure S9. Maximum-likelihood tree of TLR-4 sequences. Branch support above 50 percent labeled and obtained from 1,000 bootstrap replicates. Tree was rooted using the rodent outgroup *Mus musculus* (accession number KF696967.1). EC indicated by orange circles, NEC are indicated by blue circles, and asterisk indicate a sequence found in both species.

Supplementary Tables

Table S1. Sample size per immune gene across EC populations.

Gene	Cluster1	Cluster2	Cluster3	Cluster4	NH
DQA	33	27	18	38	6
DRB	46	32	21	42	6
TLR-2	15	19	12	10	0
TLR-4	19	20	12	10	0

Table S2. Sample size per immune gene across NEC populations.

Gene	ECT	MACC	MENH	RIGS	RIPI	WCT
DQA	42	57	22	16	11	54
DRB	42	57	23	14	10	64
TLR-2	35	38	0	0	0	35
TLR-4	41	37	0	0	0	39

Table S3. Diversity statistics for DQA across populations of EC and NEC.

Species	Population	2n	S	h	Hd	k	pi	p-distance	Ho	He
EC	Cluster1	66	64	11	0.865 (0.003)	18.8	0.109 (0.016)	0.152 (0.024)	0.344*	0.844*
	Cluster2	54	65	11	0.88 (0.003)	21.7	0.126 (0.018)	0.168 (0.026)	0.711*	0.860*
	Cluster3	36	51	7	0.79 (0.005)	17.8	0.111 (0.02)	0.141 (0.025)	0.611*	0.769*
	Cluster4	76	67	14	0.826 (0.002)	13.5	0.094 (0.014)	0.136 (0.022)	0.711	0.815
	NH	12	17	5	0.822 (0.031)	6.4	0.033 (0.009)	0.073 (0.021)	0.500	0.806
	All	244	68	17	0.871 (0.009)	17.8	0.106 (0.015)	0.150 (0.024)	0.558*	0.873*
NEC	ECT	84	29	2	0.212 (0.006)	6.2	0.041 (0.009)	0.058 (0.012)	0.214	0.228
	MACC	114	30	3	0.342 (0.004)	9.6	0.056 (0.012)	0.08 (0.016)	0.351*	0.339*
	MENH	44	34	4	0.133 (0.010)	3.4	0.009 (0.002)	0.013 (0.003)	0.047	0.046
	RIGS	32	0	1	0	0	0	0	0	0
	RIPI	22	0	1	0	0	0	0	0	0
	WCT	108	39	3	0.201 (0.001)	6.4	0.048 (0.01)	0.067 (0.011)	0.170	0.186
	All	404	43	4	0.224 (0.001)	6.4	0.037 (0.007)	0.052 (0.01)	0.196*	0.214*

Table S4. Diversity statistics for DRB across populations of EC and NEC.

Species	Population	2n	S	h	Hd	k	pi	p-distance	Ho	He
EC	Cluster1	92	66	25	0.936 (0.001)	19.9	0.143 (0.023)	0.180 (0.032)	0.810*	0.930*
	Cluster2	64	63	15	0.901 (0.003)	20.0	0.145 (0.023)	0.174 (0.031)	0.774*	0.885*
	Cluster3	42	76	10	0.884 (0.003)	24.2	0.185 (0.027)	0.222 (0.036)	0.524*	0.862*
	Cluster4	84	75	22	0.913 (0.002)	19.9	0.147 (0.022)	0.172 (0.031)	0.976	0.902
	NH	12	45	8	0.924 (0.016)	18.5	0.130 (0.022)	0.183 (0.035)	0.857	0.857
	All	294	81	35	0.940 (0.001)	20.9	0.153 (0.023)	0.186 (0.032)	0.811*	0.938*
NEC	ECT	84	46	4	0.469 (0.006)	13.7	0.107 (0.02)	0.145 (0.023)	0.390*	0.454*
	MACC	114	49	7	0.507 (0.004)	15.5	0.123 (0.022)	0.163 (0.026)	0.439*	0.503*
	MENH	46	48	6	0.437 (0.012)	12.8	0.099 (0.019)	0.130 (0.021)	0.318	0.375
	RIGS	28	47	3	0.537 (0.010)	19.3	0.158 (0.031)	0.200 (0.033)	0.643	0.518
	RIPI	20	37	2	0.395 (0.023)	14.6	0.120 (0.025)	0.152 (0.025)	0.500	0.375
	WCT	128	63	8	0.342 (0.004)	10.3	0.082 (0.015)	0.101 (0.016)	0.185*	0.309*
	All	420	68	10	0.409 (0.230)	13.8	0.110 (0.020)	0.143 (0.023)	0.364*	0.434*

Table S5. Diversity statistics for TLR-2 across populations of EC and NEC.

Species	Population	2n	S	h	Hd	k	pi	p-distance	Ho	He
EC	Cluster1	30	12	12	0.738 (0.010)	2.2	0.002 (0.001)	0.001 (0.001)	0.944*	0.739*
	Cluster2	38	15	12	0.885 (0.004)	4.5	0.004 (0.001)	0.004 (0.002)	0.842*	0.861*
	Cluster3	24	14	11	0.917 (0.006)	2.6	0.003 (0.001)	0.003 (0.002)	0.818	0.876
	Cluster4	20	20	12	0.967 (0.001)	3.9	0.004 (0.001)	0.005 (0.002)	1.00	0.906
	All	112	28	26	0.917 (0.001)	3.5	0.003 (0.001)	0.003 (0.001)	0.893*	0.908*
NEC	ECT	70	3	4	0.215 (0.008)	0.2	0.0002 (0)	0	0.179	0.165
	MACC	76	3	4	0.335 (0.007)	0.4	0.0004 (0)	0	0.342	0.338
	WCT	70	3	4	0.188 (0.007)	0.2	0.000 (0)	0	0.042*	0.155*
	All	216	6	8	0.254 (0.001)	0.3	0.0002 (0)	0	0.211*	0.243*

Table S6. Diversity statistics for TLR-4 across populations of EC and NEC.

Species	Population	2n	S	h	Hd	k	pi	p-distance	Ho	He
EC	Cluster1	40	6	5	0.586 (0.007)	1.3	0.001 (0.001)	0.004 (0.002)	0.529*	0.578*
	Cluster2	38	9	10	0.645 (0.012)	1.4	0.001 (0.001)	0.003 (0.002)	0.333	0.623
	Cluster3	24	5	5	0.725 (0.011)	1.5	0.002 (0.001)	0.003 (0.002)	0.818	0.686
	Cluster4	20	4	4	0.642 (0.026)	1.4	0.002 (0.001)	0.004 (0.002)	0.600	0.629
	All	122	12	13	0.692 (0.001)	1.5	0.002 (0.001)	0.004 (0.002)	0.579*	0.690*
NEC	ECT	74	0	0	0	0	0	0	0	0
	MACC	82	13	6	0.147 (0.003)	0.9	0.001 (0.000)	0.002 (0.001)	0.053	0.052
	WCT	78	0	0	0	0	0	0	0	0
	All	234	13	6	0.051 (0.020)	0.3	0 0	0.001 (0.000)	0.020	0.020

CHAPTER 2

Host-associated microbiota across imperiled and invasive cottontails: genetic and immunogenetic associations and potential pathogens

Kimberly B. Neil, Thomas J. McGreevy, Jr., and David M. Rand

Introduction

Host-associated microbes strongly impact species' health, immunity, physiology, and development (Kinross *et al.* 2011; Costello *et al.* 2012; Amato 2013). The most well-studied host-microbiome relationships encompass gut- or fecal-associated microbiota. A combination of host-associated constraints (e.g., immune system, genetics, physiological needs) and environmental elements (e.g., local exposures) shape these communities (Zoetendal *et al.* 2001; Mueller *et al.* 2006; Ley *et al.* 2008; Fallani *et al.* 2010). However, less is known about the microbial communities of other tissues. Recently, increasing attention has been turned to understanding skin microbiota (Byrd *et al.* 2018; Ross *et al.* 2019). Skin microbes act as crucial barriers to invading potential pathogens; indeed, the presence of some commensal bacteria species may provide "colonization resistance" against harmful microorganisms by occupying resources (Buffie & Pamer 2013; Byrd *et al.* 2018). Skin-associated microorganisms are also vital for immune system development, particularly in educating immune cells and promoting tolerance to beneficial commensal microbiota (Naik *et al.* 2012; Byrd *et al.* 2018). To-date, the majority of research on skin microbiome is conducted in humans, model species, domestic animals, and amphibians (Ross *et al.* 2019). Very few studies have been reported for wild mammalian systems.

Relatedly, understanding genetic influence on microbial communities is an active area of research, but a substantial knowledge gap exists for non-model, wild systems. In humans, genome-wide quantitative trait locus (QTL) studies have identified correlations between numerous genomic elements and gut and/or fecal microbiota (Org *et al.* 2015; Goodrich *et al.* 2016; Kurilshikov *et al.* 2017). Given the important interplay between host-associated microbes and the immune system, recent work has also investigated the influence of functional immune

loci on microbiomes, particularly by investigating genes of the major histocompatibility complex (MHC). MHC genes belong to the adaptive branch of the vertebrate immune system and recognize foreign peptides, thereby initiating an immune response (Piertney & Oliver 2006). Hypotheses on MHC evolution primarily focus on the role of immune genes in regard to pathogens, but MHC may also occupy a key role in microbiome regulation through tolerance of beneficial species (Bolnick *et al.* 2014). Additionally, MHC genes are highly polymorphic and may contribute substantially to inter-individual variation in microbial communities (Bolnick *et al.* 2014). Recent work in mice identified that MHC polymorphisms impact the mouse gut microbiome (Kubinak *et al.* 2015; Khan *et al.* 2019), and Bolnick *et al.* 2014 similarly found that specific MHC binding motifs and heterozygosity were associated with gut microbial abundance in three-spine sticklebacks.

Investigating host-associated microbiome and its genetic underpinnings is particularly needed for wildlife of conservation concern due to the potential impact on disease susceptibility and resistance. Indeed, monitoring reintroduced, imperiled, and translocated species for disease threats is widely urged for effective conservation planning and wildlife management (Munson & Cook 1993; Viggers *et al.* 1993; Cunningham 1996; Mathews *et al.* 2006). Baseline pathogen data can serve as a reference for identifying emerging disease trends (Mathews *et al.* 2006), as well as prevent incidental co-introduction of exotic diseases to immunologically naïve populations during translocations (Viggers *et al.* 1993; Boyce *et al.* 2011). The use of metagenomics is exceptionally valuable for this task, as microbiome studies allow a comprehensive approach to surveillance rather than the more commonly employed "one-host-one-pathogen" approach (Pedersen *et al.* 2007). Additionally, non-model taxa typically have

poorly characterized pathogen communities, and therefore *a priori* pathogen-specific targeting skews heavily towards zoonoses or known epizootic agents of charismatic fauna.

Here, I adopt a comprehensive approach to simultaneously investigate microbiome, potential pathogens, and potential genetic drivers of microbial patterns within a naturally-occurring system of native and invasive cottontails: the imperiled New England cottontail (*Sylvilagus transitionalis*; NEC) and the non-native eastern cottontail (*S. floridanus*; EC). Historically, NEC were native to much of New England and New York but have declined in abundance since the 1960s (Litvaitis 2003; Litvaitis *et al.* 2007). Currently, NEC occupy five disjunct remnant groups occupying less than 16 percent of their historical range (Litvaitis *et al.* 2004; Fenderson *et al.* 2011) and are the target of a reintroduction program. NEC population decay is largely attributed to land use change (Litvaitis *et al.* 2003, 2007) but is also coincident with the intentional introduction of EC. In the early 1900s, EC were released into the northeastern United States to supplement small game hunting (Chapman & Morgan 1973), until concerns of co-introduction of the zoonosis tularemia halted their import in the 1940s (Ayres & Feemster 1948; Johnston 1972). Given that the bacterial agent of tularemia, *Francisella tularensis*, exists enzootically in some native EC populations (Woolf *et al.* 1993) and that the first cases in New England emerged in 1978 following EC introduction (Feldman *et al.* 2001), it is possible that EC introduced this pathogenic bacterium to the region.

The role of invasive species in pathogen/parasite dynamics is well-studied, and ecological theory predicts several possible scenarios following species' invasion (Dunn 2009; Dunn & Hatcher 2015). Invaders may have low levels of pathogens/parasites due to a combination of escape from native colonizers ('enemy release hypothesis'; Torchin *et al.* 2003) and incompetence for pathogens/parasites in their nonnative range. Conversely, invaders could retain

native microorganisms and additionally accumulate generalist pathogens/parasites in their introduced range (Poulin *et al.* 2003). It is also possible for nonnative species to introduce exotic pathogens/parasites to native species with deleterious effects ('spillover'; Strauss *et al.* 2012). However, how this may translate more broadly beyond pathogens and to also encompass host-associated microbiota is unknown.

In this study I report new findings on the following research objectives:

1) Evaluate differences in ear microbiota between invasive EC and imperiled NEC,
2) Assess compositional differences in microbiota associated with host genetic traits at both genome-wide loci and targeted immune loci, and
3) Identify any potential disease risks to EC, NEC, and/or humans.

Methods

Sample Collection

DNA samples extracted from approximately two-mm ear biopsy punches were obtained from archives at the Wildlife Genetics and Ecology Laboratory (WGEL) at the University of Rhode Island (URI). Ear biopsies were chosen for several reasons: 1) It is a non-lethal sampling method that yields high quality DNA for genomic analyses. Using fecal-derived DNA for host genomic analyses is often infeasible due to both the low yield and low quality of recovered host DNA, yet working with imperiled species necessitates minimally-invasive tissue sampling that does not require animal sacrifice, 2) Ear biopsy provides a snapshot of host-associated microbiota across skin, blood, and surrounding tissue matrix, providing a comprehensive understanding of host-associated microbial symbionts and, 3) ticks often feed on rabbit ears, making this tissue an attractive point for identifying vector-transmitted potential pathogens.

Biopsies were originally collected in combination by state wildlife agencies and/or WGEL research staff (IACUC protocol # AN111-12-011) and extracted into DNA using Qiagen DNeasy kits (Cat. No. 69504) according to manufacturer's instructions. Negative controls of water were included with all extraction batches. I targeted samples from Cape Cod, Massachusetts, which supports both an NEC remnant population and co-occurring invasive ECs. A total of 43 NEC and 38 EC were sequenced (**Figure 1**).

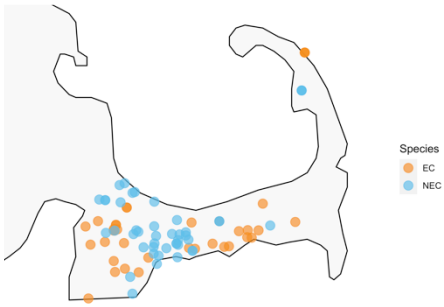

Figure 1. Sampling for EC and NEC. Samples encompass the Cape Cod, Massachusetts populations of NEC and EC.

ddRAD library preparation, sequencing, and SNP calling

The collective 81 samples used are part of a larger double-digest restriction-associated DNA sequencing (ddRAD) project at WGEL. Briefly, libraries were prepared according to Peterson *et al.* 2012, using the restriction enzymes SphI-HF (New England Biolabs, Ipswich, MA) and MluC1 (New England Biolabs, Ipswich, MA) and size selected for 338 to 418 bp using a Pippin Blue. Fifty base pair (bp) single-end reads were sequenced on the Illumina HiSeq 2000 by Harvard University's Bauer Core Facility.

Raw sequences were separately processed for each species using STACKS v.2.2 and its *de novo* pipeline (Catchen *et al.* 2013). Parameter selection can significantly impact downstream analyses, and therefore I performed a series of optimization tests according to Paris *et al.* 2017 and Rochette & Catchen 2017 to identify the best parameters for *M* (the minimum number of mismatches allowed between stacks to merge them into a putative locus) and *n* (the number of mismatches allowed between putative loci during catalog construction). A separate catalog of ddRAD loci was constructed for each species using selected parameters.

Following completion of STACKS, single nucleotide polymorphism (SNP) data was further filtered to remove low call individuals and loci using an iterative filtering scheme modified from O'Leary *et al.* 2018 and implemented in vcftools v.0.1.16 (Danecek *et al.* 2011). An iterative scheme was applied to retain individuals or loci with a genotype call rate > 50%, individual missing data < 90%, genotype call rate > 60%, individual missing data <70%, genotype call rate > 70%, individual missing data < 50%, genotype call rate > 80%, and individual missing data < 40%. Divergence from Hardy-Weinberg equilibrium was calculated for each locus using the Populations module of STACKS. Loci that significantly departed from Hardy-Weinberg proportions following Benjamini-Hochberg correction were removed.

MHC primer design, gene sequencing, and amplicon calling

We targeted exon 2 of the MHC class II genes DRB and DQA, which encodes the hypervariable peptide-binding site (PBR) that directly interacts with antigenic peptide (Hughes & Yeager 1998). MHC class II proteins interact with extracellular antigens, including bacteria (Piertney & Oliver 2006), and therefore are an excellent *a priori* target for microbiome association studies. Multiple DRB and DQA primers used in other small mammals (Smulders *et al.* 2003; Schad *et al.* 2004; Koutsogiannouli *et al.* 2014) and designed specifically for this study were tested and assessed for amplification fidelity and dropout.

Final DRB amplification used primers KBN.F and KBN.R, which are modified versions of the mammalian DRB primers JS1 and JS2 (Schad *et al.* 2005). DQA PCR amplification used forward primer B (Fain *et al.* 2001) and KBN.DQA.R. DQA and DRB were separately amplified and pooled together at equimolar ratio for a given sample. Amplicons were cleaned using AmPure beads and sent to the URI Genomics and Sequencing Center for Illumina MiSeq library preparation and sequencing. An Illumina v2 500 cycle kit was used for 2x200 paired-end

sequencing. Briefly, the Nextera XT Index Kit was used to attach indices and adapters, and products were again purified by AMPure bead cleanup. Samples were analyzed by the Agilent BioAnalyzer DNA 1000 chip and quantified using Qubit fluorometry. Samples were normalized to equimolar ratio before pooling, and the final pooled library was quantified prior to sequencing using qPCR.

The program suite AmpliSAT (Sebastian *et al.* 2016) was used to call MHC genotypes. Non-model MHC gene sequencing requires specialized data-processing considerations due to the possibility of unknown duplication events, the use of degenerate primers coupled with GC-rich regions leading to amplification biases, and the difficulty of differentiating between "true" alleles and high-frequency chimeras or PCR artefacts (Babik 2010; Sommer *et al.* 2013; Lighten *et al.* 2014). AmpliSAT is highly customizable and was identified in a study of MHC genotyping strategies as the preferred approach (Biedrzycka *et al.* 2016). AmpliCHECK was run using default exploratory recommendations, and AmpliSAS utilized suggested Illumina parameters with the "minimum dominant frequency" adjusted to 15% for DQA and 20% for DRB. Chimeras and low-frequency variants were removed. Following variant recovery, alleles were further classified as "True" or "Unclassified" according to a decision tree adapted from Grogan *et al.* 2016. "True" alleles were utilized in all downstream analyses, whereas any individuals containing "Unclassified" alleles were excluded from further processing.

16s-rRNA microbiome sequencing, quality filtering, and read processing

16s-rRNA primer sites are targeted to anneal to genomic sequences shared broadly across the kingdom of bacteria. However, these sites are often also homologous to mitochondrial DNA. When using fecal samples or blood, host DNA is minimal, but when using other tissue types, the high prevalence of mitochondrial DNA can be problematic in swamping bacterial signal. To

mitigate this issue, I applied a nested PCR approach that first amplifies a 470 bp region using a forward primer with high amplification fidelity across Bacteria but low sequence homology with mammalian mitochondria, and a canonically used V4 reverse primer, 806R. This is then followed by a second round of amplification using the primers 16s-V4F and 16s-V4R from Razzauti *et al.* 2015, which captures an approximately 200 bp hypervariable V4 region and is also sufficient for identifying a variety of bacterial pathogens (Razzauti *et al.* 2015; Galan *et al.* 2016). A pilot sequencing run was conducted using cadaver tissue-derived cottontail samples and a mock community of known composition, where three cycle profiles were tested for impact on taxonomic abundance, alpha diversity, and beta diversity: (primary:secondary) 20:15, 15:20, and 10:25 and two different amounts of primary amplicon were used: 5 or 10 uL. 15:20 cycles and 5 uL of primary amplicon were chosen as appropriate parameters. This was also consistent with Yu *et al.* 2015 who found that nested PCR was appropriate for 16s-RNA gene sequencing if the primary amplification round did not exceed 20 cycles.

All PCR was conducted using Phusion High-Fidelity DNA Polymerase, and negative controls were included that consisted of molecular-grade water in place of DNA. Laboratory work was conducted in a PCR workstation and sterilized with ultraviolet light, and in separate pre- and post-PCR work areas. Amplified products were purified using AMPure beads (Beckman Coulter) and shipped to the URI Genomics Core for further library preparation then sequencing on the Illumina MiSeq platform (2 x 250 bp paired-end) as described above for MHC.

Forward and reverse reads were imported into Qiime2 (v.2020.2) for standard quality control and assembly (Bolyen *et al.* 2019). Within Qiime2, the Dada2 plug-in was used for denoising, to remove chimeric sequences, and to identify amplicon sequence variants (ASVs). Taxonomy was assigned using the Silva 132 reference database, and sequences identified as

chloroplasts or mitochondria were pruned from the dataset. A phylogenetic tree was generated for beta diversity estimates.

Statistical analyses

Statistical analyses were primarily conducted using RStudio v.1.1.463. vcftools v.0.1.16 was used to convert STACKS vcf files into pedfiles and imported into R using "trio" (Danecek *et al.* 2011; Schwender *et al.* 2014). Genome-wide standardized heterozygosity was calculated separately for each individual in species batches using the package "Rhh" (Alho *et al.* 2010) and the formula of Coltman *et al.* 1999.

Genome-wide association study (GWAS) was conducted to identify candidate SNPs associated with microbial presence/absence or abundance. Taxa were grouped at the family level due to 1) modest sample size precluding using genus or amplicon sequence variant (ASV)-level testing, and 2) another study which identified the family level as appropriate for detecting microbiome and host-genetic associations (Bolnick *et al.* 2014). Taxonomic presence/absence was utilized for all families present at a species-wide frequency of 30 percent or greater: Bacillaceae, Burkholderiaceae, Carnobacteriaceae, Enterobacteriaceae, Enterococcaceae, Micrococcaceae, Moraxellaceae, Neisseriaceae, Parvibaculaceae, Pasteurellaceae, Planococcaceae, Pseudomonadaceae, Rhizobiaceae, Rhizobiales Incertae Sedes, Ruminococcaceae, Sphingobacteriaceae, Staphylococcaceae, Stappiaceae, Streptococcaceae, Weeksellaceae, Xanthobacteraceae, and Xanthomonadaceae. The software package GEMMA v0.98.1 (Zhou & Stephens 2012) was used to generate univariate linear mixed models (LMM) with a centered relatedness matrix to also statistically control for any subpopulation structure within the Cape Cod populations. A significance threshold of $p < 1.0 \times 10^{-5}$ was used; however, because Bonferroni correction can be overly conservative (Johnson *et al.* 2010), we additionally

and cautiously report SNPs that exceed $p < 1.0x10^{-4}$. In parallel, we also applied a Benjamini-Hochberg correction, which identified a subset of the aforementioned SNPs as statistically significant. To identify genes associated with potential outlier SNPs, the encompassing 50 bp scaffold was BLAST searched against the European rabbit (*Oryctolagus cuniculus*) genome. Gene functions were inferred using the databases Genecard and UniProtKB. The proportion of genotypes associated with the presence or absence of families with outlier SNPs is additionally reported.

MHC heterozygosity and allele frequency distributions were calculated separately for each locus and individual, and individual amino acid p-distance was calculated using MEGA v.7.0.26 (Kumar *et al.* 2016). Correlation between genome-wide standardized heterozygosity and MHC amino acid p-distance was evaluated separately for each species and MHC locus using Spearman's *rho*. Generalized linear models (GLM) were utilized to test whether MHC affects taxon presence/absence using the microbial families listed above and the core R function "glm". Separately for each host species, GLMs were generated for each taxonomic family, with presence/absence as the response variable and MHC alleles as explanatory variables. All MHC alleles were included in the null model, and backwards elimination was used to identify any MHC allele(s) that were statistically significant at $p < 0.05$. Separate models were also generated with explanatory variables of MHC heterozygosity, amino acid p-distance, and genome-wide standardized heterozygosity estimates.

One family, Pseudomonacae, was present in greater than 95 percent of individuals. Here, we also evaluated whether relative abundance associated with specific MHC alleles, MHC heterozygosity, MHC amino acid distance, and standardized heterozygosity. We used

quasibinomial GLMs separately for each species, as this type of model effectively accounts for uncertainty in relative abundance and is robust to high variance and skew (Bolnick *et al.* 2014).

Microbial data were imported into R using the package 'phyloseq' (McMurdie & Holmes 2013). Compositional differences among species were visualized by principal coordinates analyses (PCoA). To assess whether explanatory variables (e.g., species, standardized heterozygosity, MHC heterozygosity) impacted ordinated patterns, samples were sub-sampled to a depth of 5,000 ASVs and non-parametric multivariate analysis of variance (permanova) was performed using the package 'vegan' (Dixon 2003). Bray-Curtis dissimilarity, and weighted and unweighted Unifrac distance matrices were calculated using the package 'phyloseq' (McMurdie & Holmes 2013). To evaluate whether compositional diversity was associated with MHC diversity, weighted Unifrac distances were calculated separately per species and for MHC homozygotes and heterozygotes at each locus. Differences between species and heterozygotes/homozygotes were tested for statistical significance using a Kruskal-Wallis test with Bonferroni correction. To assess whether geographic distance between individuals impacts beta-diversity, geographic comparison of individual microbiome dissimilarities/distances and Euclidean geographic distances were performed per species, using the Mantel function in the package 'vegan' and 10,000 permutations (Dixon 2003).

Alpha diversity estimates included the number of observed ASVs, Shannon index, and Simpson diversity. For each sample and alpha diversity estimate, the mean of 20 iterations at a sub-sampling depth of 5,000 ASVs was calculated. A non-parametric two-sample t-test was used to assess differences in alpha diversity between species and within species, between MHC heterozygotes and homozygotes. To test for a genetic impact on individual ASV count, a negative binomial GLM was separately generated for each species and also included the

explanatory variables standardized heterozygosity, DRB mean amino acid distance, and DQA mean amino acid distance.

Potential bacterial pathogens were identified at the genus level as in Razzauti *et al.* 2015, Galan *et al.* 2016, and extensive literature searches. First, sequences were rarified to a depth of 5,000 to prevent sequencing depth biases in recovering rare taxa. Differences in infection prevalence between EC and NEC were conducted at the genus level using a Fisher's exact test. We attempt species-level identification using a phylogenetic approach. Each recovered ASV sequence belonging to a potentially pathogenic genus was BLAST searched to identify closest species-level matches. Sequences were downloaded from GENBANK and aligned and trimmed using BioEdit v7.2.6. (Hall 2011). Maximum-likelihood trees were generated using RAxML v8.2.10 and general time reversible (GTR) model with gamma distributed rate variation among sites (Stamatakis 2014). Potential species-level matches and/or species-level clustering are recorded for any clade encompassing a recovered ASV with greater than 50 percent support.

Results

Microbiome analyses

Microbiome for both EC and NEC was dominated by the phylum Proteobacteria, followed by Firmicutes and to a lesser extent Bacteroidetes (**Figure 2**). At the ASV level, 24.0% of EC ASVs were shared with NEC, and for NEC 16.9% of ASVs were shared with EC (**Figure 3**); at deeper taxonomic levels, the shared percentage increased until it converged for both species at the phylum level to 79%. Simpson and Shannon diversity were similar between each species, but NEC individuals had a greater number of observed ASVs relative to ECs (mean ± SD: EC, 34.7 ± 49.5; NEC, 47.4 ± 56.1; $p = 0.026$) (**Figure 4**). Species did not visibly cluster in PCoA space (**Figure 4**) but did significantly differ by permanova testing of Bray-Curtis

dissimilarity ($R^2 = 0.059$, $p = 0.003$); however, the species' variances were significantly different ($p = 0.006$).

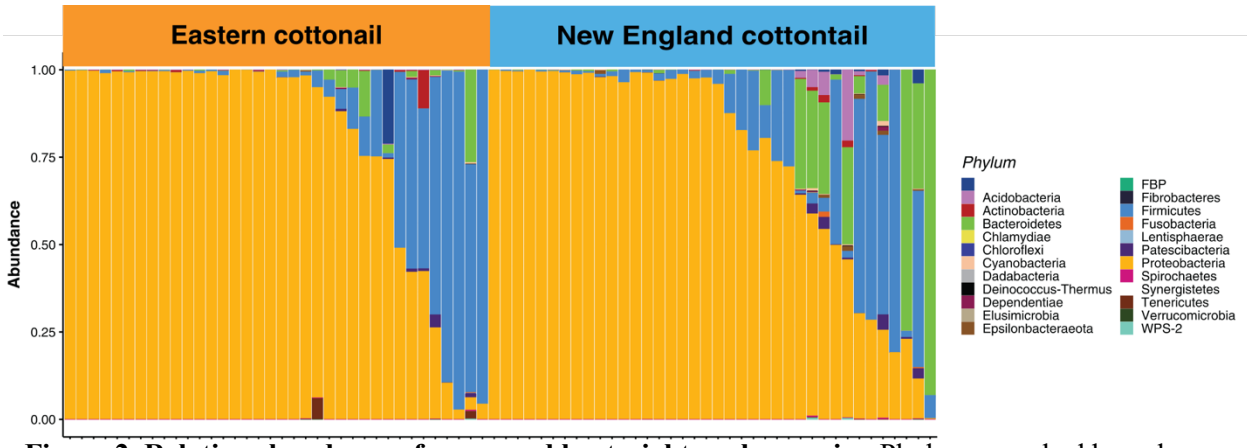

Figure 2. Relative abundance of recovered bacterial taxa by species. Phyla are marked by color.

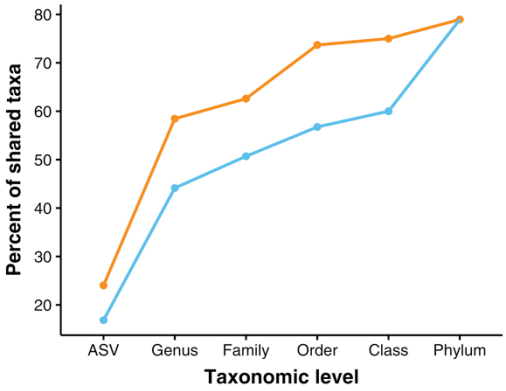

Figure 3. Microbial taxa shared between EC and NEC at varying taxonomic levels. EC represented by orange line, NEC are represented by blue.

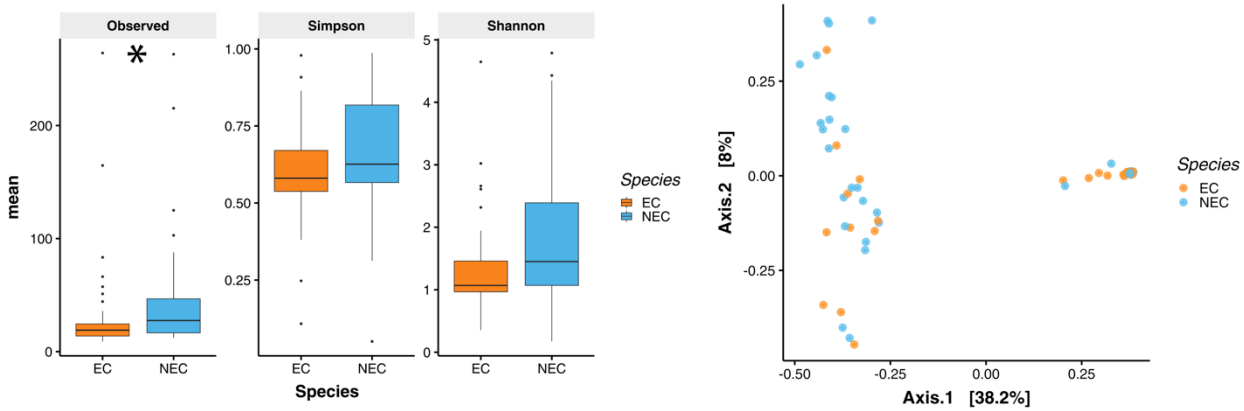

Figure 4. Microbial diversity by species. (Left) Alpha diversity estimates between species. Data were rarefied to 5,000 ASVs per sample. Boxplot depicts the median and interquartile range. Asterisks indicates statistical significance at $p < 0.05$. (Right) PCoA by species. Circles indicate the location of individuals within microbiome space and are colored by species.

Of the 22 taxonomic families prevalent in greater than 30 percent of EC and NEC, six significantly differed between the species (**Figure 5**). EC had greater mean relative abundances for Enterobacteriaceae, Parvibaculaceae, Rhizobiales Incertae Sedes, Xanthobacteraceae, and Xanthomonadaceae, whereas NEC had greater mean relative abundance of Moraxellaceae.

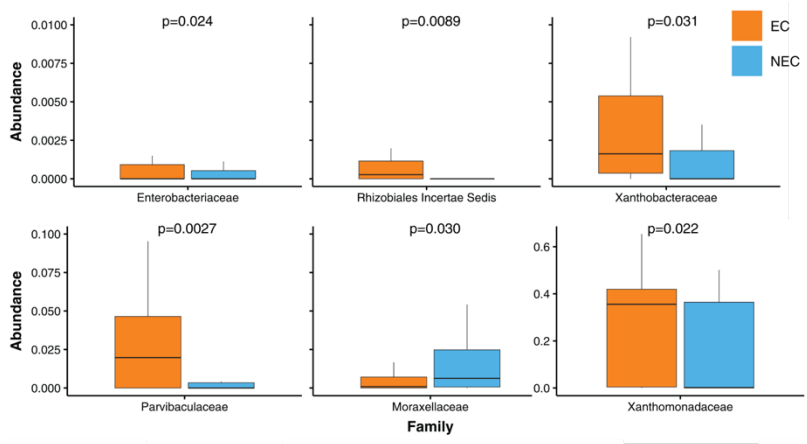

Figure 5. Microbial abundance differs between EC and NEC. Six families differ in microbial relative abundance between the species at $p < 0.05$. Boxplots depict the median and interquartile range.

Genome-wide host-microbe association analyses

A total of 8,786 SNPs were recovered for EC, and 3,054 for NEC. Even with a modest number of markers, we were able to identify associations between multiple taxonomic families and specific SNPs in both NEC and EC (**Figure 6**). For EC, nine families had potential outlier SNPs associated with presence/absence at $p < 1.0\text{x}10^{-4}$ (**Table 1**): Moraxellaceae, Parvibaculaceae, Pasteurellaceae, Rhizobiaceae, Staphylococcaceae, Stappiaceae, Weeksellaceae, Xanthobacteraceae, and Xanthomonadaceae. The majority of SNPs were mapped to specific *O. cuniculus* genes with functions including immune system regulation and epithelial dynamics, and several with known expression patterns in the skin (**Table 1**). For NEC, three families had potential outlier SNPs associated with presence/absence at $p < 1.0\text{x}10^{-4}$: Burkholderiaceae, Moraxellaceae, and Planococcaceae, and one family, Ruminococcaceae, had

an outlier SNP at $p < 1.0x10^{-5}$. All NEC SNPs were mapped to specific *O. cuniculus* protein-coding genes, with functions including immune system regulation, metabolism, and epithelial dynamics (**Table 1**).

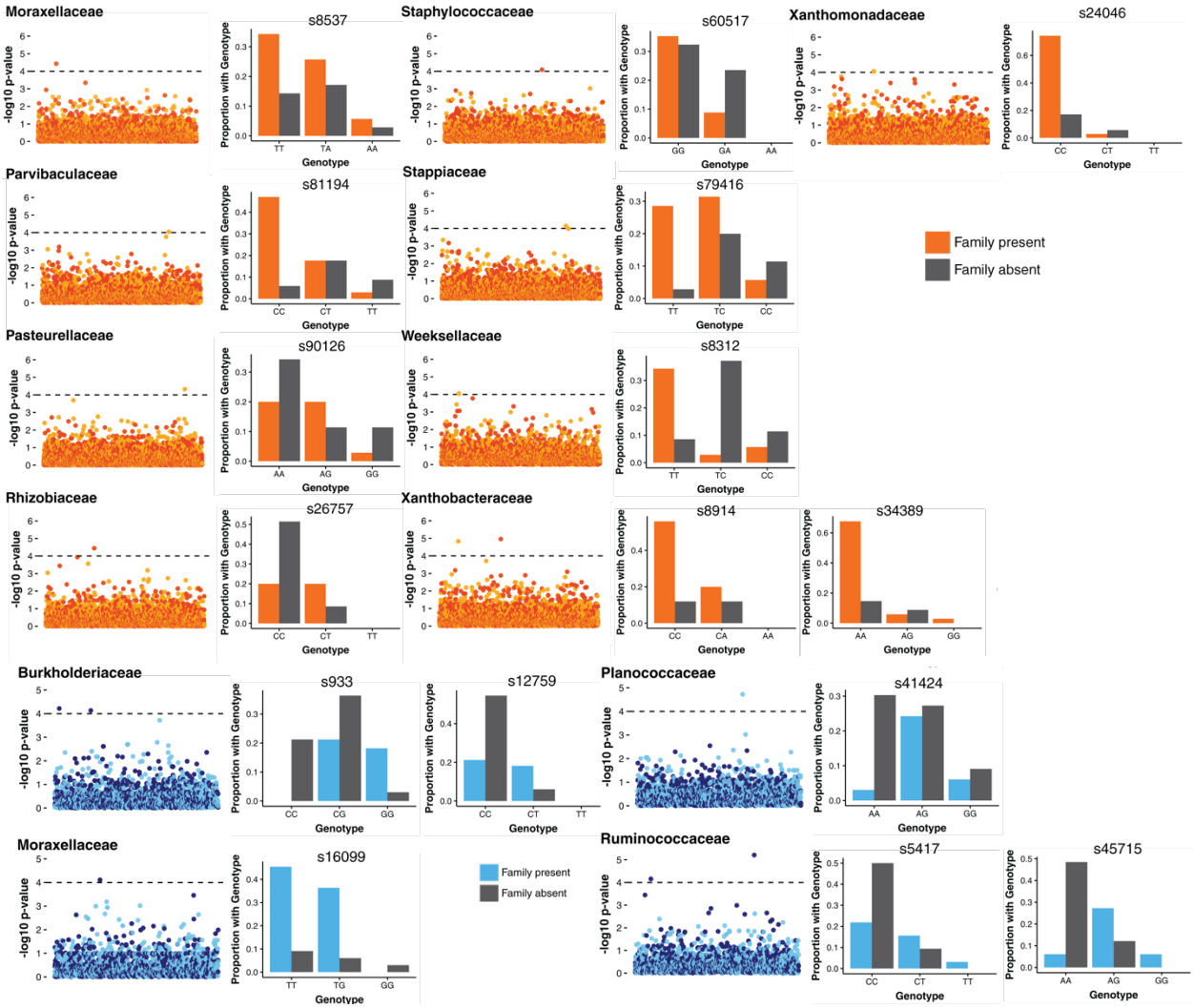

Figure 6. Genetic association with microbial family presence/absence. Dashed line indicates SNP threshold for -log10 p-value of 4. Barplots show the proportion of individuals with presence/absence of a particular SNP genotype.
(Top) = EC, (Bottom) = NEC.

Table 1. Gene matches and inferred functions from outlier SNPs. Protein functions inferred from Genecard and UniProtKB. * indicates SNP is significant following Benjamini-Hochberg correction.

Taxon	SNP	p-value (-log10)	*Orcytolagus cuniculus* blast match	Protein function
			NEC	
Burkholderiaceae	s933	4.23	T-cell receptor beta (TCRB) genes	T cell receptors recognize foreign antigens bound to major histocompatibility complex (MHC) proteins
Burkholderiaceae	s12759	4.13	solute carrier family 27 member 3	Family of membrane proteins involved in lipid metabolism; expressed in multitude of human tissues, including skin, lymphoid organs, blood vessels, mucosal epithelia
Moraxellaceae	s16099	4.11	FLJ44290 gene	Synonymous with DNAH12 gene; encodes dynein axonemal heavy chain 12; expressed in multitude of human tissues, including mucosal epithelia, skin, and lymphoid organs
Planococcaceae	s41424	4.73*	uroplakin-3b	Component of mammalian epithelium; expressed in human esophagus, lung, ovary, skin, and urinary bladder
Ruminococcaceae	s5417	4.16	RAB3 GTPase activating protein catalytic subunit 1	Catalytic subunit of Rab GTPase activating protein; expressed in multiple human tissues, including skin, lymphoid organs, and salivary gland
Ruminococcaceae	s45715	5.21*	kappa 1 light chain immunoglobulin gene region	Kappa 1 light chain forms the small subunit of immunoglobulin (antibody)

		EC		
Moraxellaceae	s8537	4.43	Fre5 prolactin precursor (PRL) gene, promoter region	Prolactin is an anterior pituitary hormone involved in growth regulation of multiple tissues, including cells of the immune system; may suppress apoptosis; crucial for lactation
Parvibaculaceae	s81194	4.05	butyrophilin like 2 (BTNL2)	Encodes a major histocompatibility complex (MHC) associated type-I transmembrane protein; involved in immune surveillance; expressed in European rabbit skin, kidney, liver, testes, and prefrontal cortex
Rhizobiaceae	s26757	4.44	BTB domain containing 7 (BTBD7)	Protein coding gene that facilitates epithelial dynamics, alters cell morphology, and decreases cell-cell adhesion
Staphylococcaceae	s60517	4.09	*Oryctolagus cuniculus* clone 16788057A1	Unknown
Stappiaceae	s79416	4.14	Galectin like (LGALSL)	Protein coding gene involved with carbohydrate binding; expressed in multiple European rabbit tissues, including skin and blood
Weeksellaceae	s8312	4.04	*Oryctolagus cuniculus* uncharacterized LOC108177289	Unknown
Xanthobacteraceae	s8914	4.83*	CDK5 regulatory subunit associated protein 2 (CDK5RAP2)	Protein coding gene involved in cyclin-dependent kinase 5 activity (CDK5); expressed in multitude of European rabbit

				tissues, including skin, kidney, blood, smooth muscle, and liver
Xanthobacteraceae	s34389	4.96*	CFTR gene	Encodes member of the ATP-binding cassette (ABS) transported superfamily; involved in ion and water secretion/absorption in epithelial cells
Xanthomonadaceae	s24046	4.06	nectin cell adhesion molecule 2 (NECTIN2)	Protein is a membrane component of adherens junctions; involved in viral entry; modulator of T-cell signaling
Pasteurellaceae	s90126	4.33	afamin (AFM)	Member of the albumin gene family; involved in activity of Wnt family members and binding of Vitamin E

MHC-specific association analyses

A greater number of MHC alleles was recovered for EC than NEC for both DQA and DRB loci, and more DRB alleles were segregating than DQA alleles (**Figure 7**). MHC heterozygosity was also higher for EC than NEC, although individual amino acid p-distance was not significantly different between the species (**Figure 7**). No correlation was observed for either species between MHC gene amino acid p-distance and genome-wide standardized heterozygosity (all $p > 0.05$).

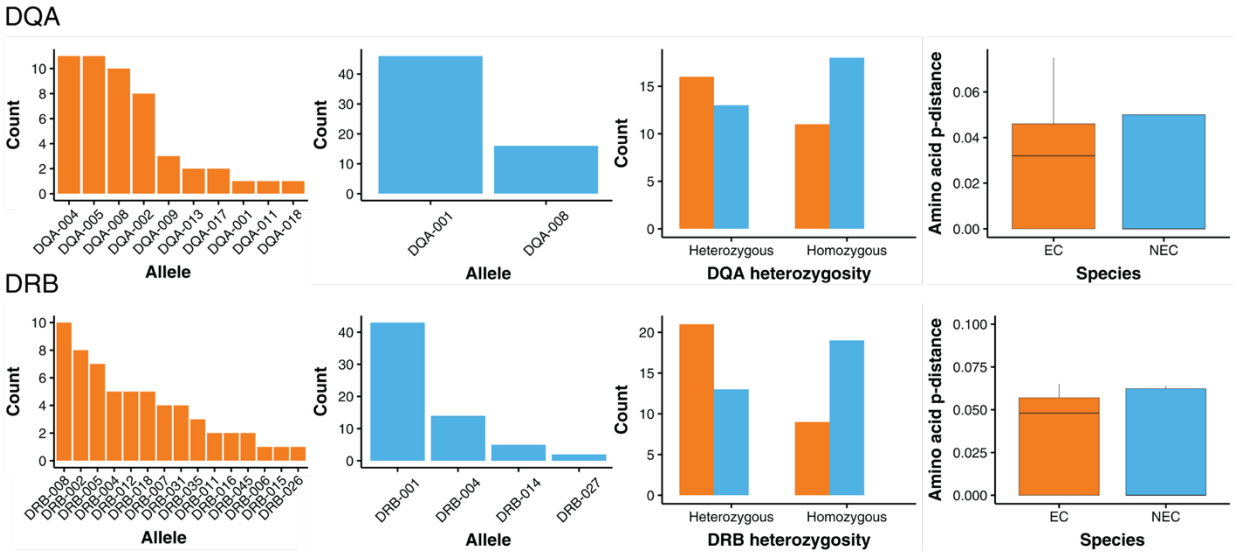

Figure 7. Characterization of MHC loci DQA and DRB. EC represented by orange, NEC represented in blue. Top panels depict DQA, bottom panels DRB. Left-most panels depict allele distribution for EC and NEC at DRB and DQA. Third panel depicts heterozygosity at DRB and DQA per species. Right-most panel depicts amino acid p-distances per species. Boxplot displays median and interquartile range.

For EC, DQA heterozygotes had a greater number of observed ASVs relative to homozygotes (mean ± SD: heterozygotes, 41.3 ± 68.6; homozygotes, 13.3 ± 4.27; $p = 0.003$). No significant difference existed for EC DRB heterozygotes (mean ± SD: homozygotes, 33.2 ± 37.1; heterozygotes, 46.3 ± 81.5; p = 0.538), and no significant difference was found for NEC at either the DQA (mean ± SD: homozygotes, 35.9 ± 31.0; heterozygotes, 52.6 ± 61.8; p = 0.934) or DRB locus (mean ± SD: homozygotes, 37.3 ± 30.4; heterozygotes, 52.8 ± 61.4; p = 0.815). Generalized linear models revealed no association between EC or NEC alpha diversity and either

genome-wide standardized heterozygosity or DQA or DRB amino acid distance (all $p > 0.05$). When comparing compositional diversity between species dependent on heterozygosity, EC weighted Unifrac distances were significantly greater in heterozygotes at DRB and DQA relative to homozygotes (**Figure 8**). Compositional diversity within NEC was equally high for both heterozygotes and homozygotes regardless of heterozygosity at each locus (**Figure 8**).

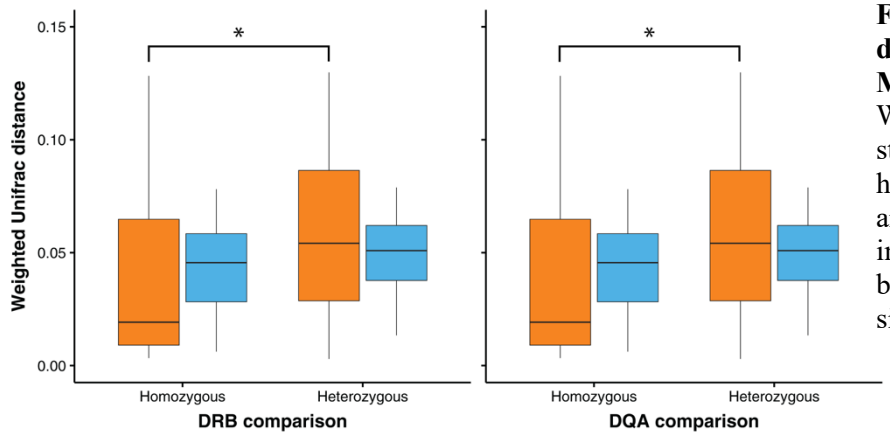

Figure 8. Microbiome distance as a function of MHC heterozygosity. Weighted Unifrac distances stratified according to heterozygosity at DRB (left) and DQA (right). EC shown in orange, NEC shown in blue. Asterisk indicates significance at $p < 0.05$.

We detected associations between microbial family-level presence/absence and specific MHC variants. For EC, Streptococcaceae, Xanthobacteraceae, and Rhizobiales Incertae Sedis were significantly associated with different DRB alleles, and Rhizobiaceae was associated with a DQA allele (**Table 2**). For NEC, the families Ruminococcaceae, Carnobacteriaceae, Xanthobacteraceae, Xanthomonadaceae, Rhizobiaceae, and Planococcaceae were associated with DRB alleles (**Table 2**).

Table 2. Generalized linear models that indicate significant associations between presence/absence of microbial families and particular MHC alleles.

Family	Parameter	Estimate	Std. Error	P-value
	NEC			
Ruminococcaceae	DRB-004	1.65	0.619	0.008
Xanthomonadaceae	DRB-004	-2.33	0.862	0.007
	DRB-014	-3.29	1.10	0.003
Carnobacteriaceae	DRB-014	2.32	0.892	0.009
Xanthobacteraceae	DRB-004	-2.17	0.677	0.001
Rhizobiaceae	DRB-004	-1.64	0.663	0.014
Planococcaceae	DRB-001	-2.94	0.927	0.001
	EC			
Streptococcaceae	DRB-008	1.89	0.824	0.022
Xanthobacteraceae	DRB-007	-1.95	0.831	0.019
Rhizobiaceae	DQA-005	-2.30	0.866	0.008
Rhizobiales Incertae Sedis	DRB-002	-2.34	0.874	0.007

Twenty-seven potentially pathogenic bacterial genera were detected (**Figure 9**; **Table S1**), and there were no differences in genus-level prevalence between EC and NEC (all *p*-values > 0.05). A large number of detected genera are commonly found commensals or environmental bacteria that have the potential to become opportunistic pathogens (**Table S1**; Razzauti *et al.* 2015; Galan *et al.* 2016). For most genera, multiple ASVs were recovered at low prevalence and potential species-level matches were not possible using the 16s-rRNA V4 region amplified. Some genera and/or species of documented pathogenicity to rabbits were recovered, including *Acinetobacter baumannii* (Pereira *et al.* 2020), *Campylobacter* spp. (Schoeb & Fox 1990), *Mycobacterium* spp. (Judge *et al.* 2006; Ludwig *et al.* 2009), *Pasteurella multocida* (Suckow *et al.* 2002), *Pseudomonas aeruginosa*, *Staphylococcus aureus* (Osebold & Gray 1960; McCoy &

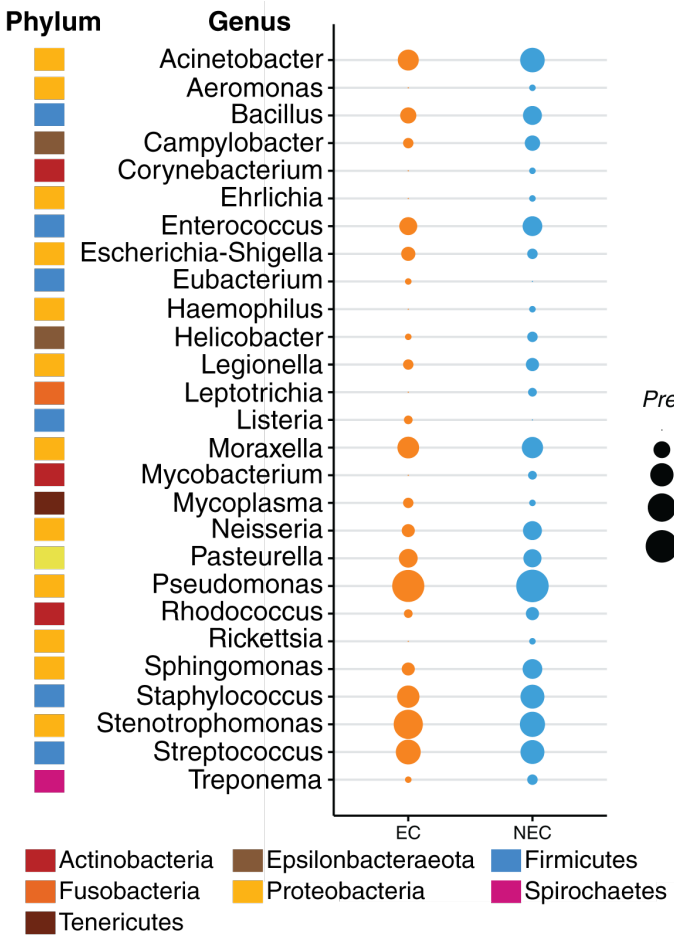

Figure 9. Prevalence of potentially pathogenic genera and host species. Bubble plot circle area depicts prevalence, with EC shown in orange and NEC in blue. Microbial phyla specified by color according to scheme in Figure 2. All pairwise comparisons between species not significant (Fisher's exact test, $p > 0.05$).

Steenbergen 1969), *Streptococcus* spp. (Ren *et al.* 2014), and *Treponema* spp. (Suckow *et al.* 2002). Additionally, two genera of serious tickborne zoonoses were detected: *Ehrlichia*, the etiological agent of ehrlichiosis, and *Rickettsia*, which can cause typhus and various spotted fevers, including rickettsial pox and Rocky Mountain spotted fever.

Discussion

This study is among the first to investigate host-associated microbiome and its genetic underpinnings across a wild system of conservation concern. Ear microbiome between native NEC and invasive EC was similar (**Figures 2 & 4**), with the phyla Proteobacteria and Firmicutes appearing most abundant. This is similar to findings of Ross *et al.* 2019, who report that Proteobacteria tends to dominate in non-human mammal skin. Relatedly, lack of discernable host-species signatures has been reported previously for microbiome of multi-species studies on amphibian skin (Ellison *et al.* 2019). Although Bray-Curtis dissimilarity did differ between EC and NEC, the variance was also significantly different, indicating that any species-level beta diversity differences may be due to dispersion. Alpha diversity estimates were similar between species, with the exception of NEC harboring a greater number of observed ASVs than EC (**Figure 4**). Ecological theory predicts that nonnative species may lose pathogens as they invade new regions (Torchin *et al.* 2003), resulting in greater parasitism of ecologically similar native species relative to invaders (Sellers *et al.* 2015). How this applies more broadly to microbiota symbionts is unknown, but our study could suggest lower levels of microbiota colonization in invaders relative to native congeners.

Within species, microbiota were variable, as evidenced by moderate-to-high intraspecific beta diversity values, particularly for NEC (Bray-Curtis dissimilarity, mean ± SD; EC: 0.626 ± 0.401; NEC: 0.872 ± 0.252). Microbiome composition is due to both stochastic and deterministic

events (Stegen et al. 2012), and therefore a highly variable microbiome may suggest that stochastic dispersal, colonization and extinction processes, such as ecological drift, may occupy an influential role (Rosindell *et al.* 2012; Li *et al.* 2018). However, we also detected notable influence of host genetics on microbial symbionts. For both EC and NEC, multiple SNPs were recovered that associate with microbial families. Due to modest sample size and a *p*-value threshold of 1.0×10^{-4}, we cautiously suggest these SNPs and associated genes as worthy of further investigation as to their role of microbiota. Additionally, because we lack positional information for identified genes and our SNP data represented a genome-wide sampling, it is possible that identified genes are in linkage disequilibrium with true genetic drivers of microbial patterns and can be used to identify regions of interest for further study.

Interestingly, of the 16 candidate genes identified in both NEC and EC, four genes with known immune function were identified (**Table 1**): 1) T-cell receptor beta (TCRB) genes encode part of the T-cell receptor peptide structure, which is a crucial receptor for recognizing foreign antigens bound to MHC proteins and initiating the adaptive immune response, 2) kappa 1 light chain immunoglobulin gene region, which forms the small subunit of immunoglobulin (antibody), 3) butyrophilin like 2 (BTNL2) encodes MHC-associated transmembrane proteins, is involved in immune surveillance, and is expressed in European rabbit skin, and 4) nectin cell adhesion molecule 2 (NECTIN2) encodes a protein component of adherens junctions and is involved in both viral entry and T-cell signaling. While functional validation is not possible at this stage, these hits illustrate the promise of our genome-wide approach in identifying genes implicating immune function in host-associated microbial patterns.

We also found that MHC polymorphism may impact microbiota in natural populations through heterozygosity and specific allele-microbe associations. EC DQA heterozygotes had a

greater number of observed ASVs than homozygotes, and EC heterozygotes also had greater weighted Unifrac distances relative to homozygotes (**Figure 8**). Similar results were reported by Khan *et al.* 2019 who found that MHC heterozygote mice had a more variable fecal taxonomic composition relative to MHC homozygotes. This increased among-individual variation in MHC heterozygotes may be due to heterozygotes supporting a greater number of ASVs. Because no correlation was found between genome-wide standardized heterozygosity and either MHC heterozygosity or MHC amino acid distance, it is unlikely that these findings are due to genome-wide effects of heterozygosity.

Associations of specific MHC alleles and microbial families were detected for both EC and NEC (**Table 2**). Both negative and positive associations were recovered, suggesting that MHC-microbiota interactions may span both tolerance to and defense against particular microbes. Therefore, MHC associations may reflect alleles that have evolved under pathogen-driven selection to recognize specific bacterial pathogens or pathogenic families, or that tolerate functionally important mutualists, as suggested by Bolnick *et al.* 2014. Within both NEC and EC, MHC alleles that did form associations were segregating at moderate to high frequencies (**Figure 7**). Overall, NEC had fewer segregating MHC alleles than EC, and DRB alleles were associated with multiple taxa. In contrast, for EC, fewer associations between specific MHC alleles and families were discovered, and no MHC allele associated with more than one family. Due to the correlational aspect of this MHC study, it is also crucial to acknowledge that associations with MHC may arrive due to linkage disequilibrium with other immune genes not investigated.

Multiple potential pathogens were identified for both EC and NEC (**Table S1**; **Figure 9**). Many of these belong to genera containing known pathogenic species and rare, opportunistic

pathogens, but for most taxa, species-level identification was not possible (**Table S1**). While genus-level prevalence was similar for EC and NEC for most potential pathogens (**Figure 9**), multiple ASVs contributed to each genus, and many ASVs were found only in a single individual or a single host species (**Table S1**).

Several identified pathogens are of particular concerns in rabbits. Pasteurellosis is a serious rabbit disease caused by the bacterium *Pasteurella multocida*. Multiple ASVs belonging to the Pasteurella genus were identified with phylogenetic clustering to *P. multocida* and other pathogenic species. Infection is associated with rhinitis, conjunctivitis, ear infections, pneumonia, genital infection, and/or septicemia (Suckow *et al.* 2002). Abscesses and septicemia may occur with little clinical signs and can lead to rapid decline and death. Additionally, if chronic infection occurs in genitalia, it can lead to reduced fertility in males and females, or spontaneous abortion (Suckow *et al.* 2002). A second rabbit disease of substantial concern is treponematosis, caused by bacterium of the *Treponema* genus and typically *Treponema paraluiscuniculi*. *T. paraluiscuniculi* is closely related to *T. pallidum*, the causal agent of human syphilis, and can manifest in urogenital lesions and visible infection of mucus membranes, such as the nares (Suckow *et al.* 2002). Other potential pathogens identified more commonly appear as rare or opportunistic infections; however, this data is based primarily on domesticated animals, farm livestock, or humans. How this translates to *Sylvilagus* species is unknown.

Interestingly, two zoonoses were also present, *Ehrlichia* sp. and *Rickettsia* sp. Both *Ehrlichia* and *Rickettsia* belong to the Proteobacteria phylum and are tick-transmitted. It is unknown if these bacteria induce disease in cottontails, but both are of substantial disease concern for humans. The *Ehrlichia* ASV identified was highly similar in sequence to *E. ewingii*, which is recognized as a tickborne human pathogen and the agent of human granulocytic

ehrlichiosis (Buller *et al.* 1999). The recovered *Rickettsia* ASV was identified to a clade encompassing six species of known pathogenicity and that cause various spotted fevers: *R. africae*, *R. amblyommii*, *R. conori*, *R. felis*, *R. prowazekii*, and *R. typhi*.

Several pathogenic bacteria of *a priori* interest were not detected (**Table S1**), including *Francisella tularensis*, *Anaplasma sp.*, and *Borrelia burgdorferi*. Anaplasma species have been detected in *S. floridanus* blood and spleen (Goethert & Telford 2003; Yabsley *et al.* 2006), and *F. tularensis* was cultured from *S. floridanus* spleen (Shoemaker *et al.* 1997). Therefore, it is possible that these bacteria are either not present in the small sample sizes used in this study, or that ear biopsy is not an appropriate tissue for detection.

It is imperative to acknowledge several limitations to the observations of this study. Modest sample sizes were used, and greater numbers are needed to ascertain accurate estimates of infection prevalence, particularly for rare microbial taxa. Increased sample sizes will also afford greater power to GWAS- and MHC-microbiota association studies. Relatedly, future work should include technical replicates to validate pathogen detection (Galan *et al.* 2016). While the use of non-invasive methods is appealing for imperiled species, it may select away from pathogenic reservoirs, such as spleen or other lymphoid organs. Future inclusion of more populations will also bolster findings and permit assessment of biogeographic signatures in both microbiome and genetic-microbiome associations.

This study has key implications for ongoing conservation efforts of NEC, as well as other natural systems. ddRAD and other reduced-representation genomic approaches are commonly employed in non-model systems, and there is great potential to wed these data with knowledge of microbial symbionts associated with these tissues. This information can simultaneously inform species' management and human public health risks, and greatly increase

our knowledge of genetic influences on the microbiome. Our data also provides valuable baseline data for numerous potential pathogens (**Table S1**) as well as public health knowledge concerning the presence of *Rickettsia* and *Ehrlichia*, both of which are of increasing concern to human infection suburban and urban settings (Comer *et al.* 2001).

Acknowledgements

We would like to thank collaborators who helped with EC and NEC specimen procurement: Massachusetts Department of Fish and Game and Roger Williams Park Zoo. We also thank Brown University undergraduate researchers Emily Longman and Paul Rizzo for their assistance with molecular wet lab work. Lastly, we thank Janet Atoyan of the URI Genomics Core for her assistance with Illumina MiSeq library preparation and sequencing. This research was supported by computational resources at Brown University's Center for Computation and Visualization. K.B.N. was supported by the NSF Graduate Research Fellowship Program (GRFP) and NSG IGERT Grant. A Bushnell Research and Education Award, Bushnell Dissertation Improvement Grant, and Brown Institute for Environment and Society (IBES) Research and Training Award also contributed to funding this work.

References

Alho JS, Välimäki K, Merilä J (2010) Rhh: An R extension for estimating multilocus heterozygosity and heterozygosity-heterozygosity correlation. *Molecular Ecology Resources*, **10**, 720–722.

Amato KR (2013) Co-evolution in context: The importance of studying gut microbiomes in wild animals. *Microbiome Science and Medicine*, **1**, 9–29.

Ayres JC, Feemster RF (1948) Epidemiology of tularemia in Massachusetts with a review of the literature. *New England Journal of Medicine Journal of Medicine*, **238**, 187–194.

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

Biedrzycka A, Sebastian A, Migalska M, Westerdahl H, Radwan J (2016) Testing genotyping strategies for ultra-deep sequencing of a co-amplifying gene family: MHC class I in a passerine bird. *Molecular Ecology Resources*.

Bolnick DI, Snowberg LK, Caporaso JG *et al.* (2014) Major Histocompatibility Complex class IIb polymorphism influences gut microbiota composition and diversity. *Molecular Ecology*, **23**, 4831–4845.

Bolyen E, Rideout JR, Dillon MR *et al.* (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, **37**, 852–857.

Boyce WM, Weisenberger ME, Penedo MCT, Johnson CK (2011) Wildlife translocation: the conservation implications of pathogen exposure and genetic heterozygosity. *BMC Ecology*, **11**.

Buffie CG, Pamer EG (2013) Microbiota-mediated colonization resistance against intestinal pathogens. *Nature Reviews Immunology*, **13**, 790–801.

Buller RS, Arens M, Hmiel SP *et al.* (1999) Ehrlichia ewingii, a Newly Recognized Agent of Human Ehrlichiosis. *The New England Journal of Medicine*, **341**, 148–155.

Byrd AL, Belkaid Y, Segre JA (2018) The human skin microbiome. *Nature Reviews Microbiology*, **16**, 143–155.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: An analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Chapman TR, Morgan RP (1973) Allen Press. *Wildlife Monographs*, **36**, 3–54.

Coltman DW, Pilkington JG, Smith JA, Pemberton JM (1999) Parasite-mediated selection

against inbred Soay sheep in a free-living, island population. *Evolution*, **53**, 1259–1267.

Comer JA, Paddock CD, Childs JE (2001) Urban zoonoses caused by Bartonella, Coxiella, Ehrlichia, and Rickettsia species. *Vector borne and zoonotic diseases (Larchmont, N.Y.)*, **1**, 91–118.

Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA (2012) The application of ecological theory toward an understanding of the human microbiome. *Science*, **336**, 1255–1262.

Cunningham AA (1996) Disease Risks of Wildlife Translocations. *Conservation Biology*, **10**, 349–353.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Dixon P (2003) VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, **14**, 927–930.

Dunn AM (2009) Parasites and Biological Invasions. *Advances in Parasitology*, **68**, 161–184.

Dunn AM, Hatcher MJ (2015) Parasites and biological invasions: parallels, interactions, and control. *Trends in Parasitology*, **31**, 189–199.

Ellison S, Rovito S, Parra-Olea G *et al.* (2019) The Influence of Habitat and Phylogeny on the Skin Microbiome of Amphibians in Guatemala and Mexico. *Microbial Ecology*, **78**, 257–267.

Fain M, Zhao T, Kindt T (2001) Improved typing procedure for the polymorphic single-copy RLA-DQA gene of the rabbit reveals a new allele. *Tissue Antigens*, **57**, 332–338.

Fallani M, Young D, Scott J *et al.* (2010) Intestinal microbiota of 6-week-old infants across Europe: Geographic influence beyond delivery mode, breast-feeding, and antibiotics.

Journal of Pediatric Gastroenterology and Nutrition, **51**, 77–84.

Feldman KA, Enscore RE, Lathrop SL *et al.* (2001) An Outbreak of Primary Pneumonic Tularemia on Martha's Vineyard. *The New England Journal of Medicine*, **345**, 1601–1606.

Fenderson LE, Kovach AI, Litvaitis J a., Litvaitis MK (2011) Population genetic structure and history of fragmented remnant populations of the New England cottontail (Sylvilagus transitionalis). *Conservation Genetics*, **12**, 943–958.

Galan M, Razzauti M, Bard E *et al.* (2016) 16S rRNA Amplicon Sequencing for Epidemiological Surveys of Bacteria in Wildlife. *mSystems*, **1**, 1–22.

Goethert HK, Telford SRI (2003) Enzootic Transmission of Anaplasma bovis in Nantucket Cottontail Rabbits. *Journal of Clinical Microbiology*, **41**, 3744–3747.

Goodrich JK, Davenport ER, Beaumont M *et al.* (2016) Genetic Determinants of the Gut Microbiome in UK Twins. *Cell Host and Microbe*, **19**, 731–743.

Grogan KE, McGinnis GJ, Sauther ML, Cuozzo FP, Drea CM (2016) Next-generation genotyping of hypervariable loci in many individuals of a non-model species: Technical and theoretical implications. *BMC Genomics*, **17**, 1–16.

Hall T (2011) BioEdit: An important software for molecular biology. *GERF Bulletin of Biosciences*, **2**, 60–61.

Hughes AL, Yeager M (1998) Natural selection and major histocompatibility complex loci of vertebrates. *Annual Review of Genetics*, **32**, 415–435.

Johnson RC, Nelson GW, Troyer JL *et al.* (2010) Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics*, **11**, 2–7.

Johnston JE (1972) Identification and distribution of cottontial rabbits in southern New England. University of Connecticut.

Judge J, Kyriazakis I, Greig A, Davidson RS, Hutchings MR (2006) Routes of intraspecies transmission of Mycobacterium avum subsp. paratuberculosis in Rabbits (Oryctolagus cuniculus): a Field Study. , **72**, 398–403.

Khan MAW, Zac Stephens W, Mohammed AD, Round JL, Kubinak JL (2019) Does MHC heterozygosity influence microbiota form and function? *PLoS ONE*, **14**, 1–23.

Kinross JM, Darzi AW, Nicholson JK (2011) Gut microbiome-host interactions in health and disease. *Genome medicine*, **3**, 14.

Koutsogiannouli EA, Moutou KA, Stamatis C, Walter L, Mamuris Z (2014) Genetic variation in the major histocompatibility complex of the European brown hare (Lepus europaeus) across distinct phylogeographic areas. *Immunogenetics*, **66**, 379–392.

Kubinak JL, Stephens WZ, Soto R *et al.* (2015) MHC variation sculpts individualized microbial communities that control susceptibility to enteric infection. *Nature Communications*, **6**.

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution*, **33**, 1870–1874.

Kurilshikov A, Wijmenga C, Fu J, Zhernakova A (2017) Host Genetics and Gut Microbiome: Challenges and Perspectives. *Trends in Immunology*, **38**, 633–647.

Ley RE, Hamady M, Lozupone C *et al.* (2008) Evolution of mammals and their gut microbes. *Science (New York, N.Y.)*, **320**, 1647–1651.

Li J, Zhan S, Liu X *et al.* (2018) Divergence of fecal microbiota and their associations with host phylogeny in Cervinae. *Frontiers in Microbiology*, **9**.

Lighten J, van Oosterhout C, Bentzen P (2014) Critical review of NGS analyses for de novo genotyping multigene families. *Molecular Ecology*, **23**, 3957–3972.

Litvaitis JA (2003) Are pre-Columbian conditions relevant baselines for managed forests in the

northeastern United States? *Forest Ecology and Management*, **185**, 113–126.

Litvaitis JA, Barbour MS, Brown AL *et al.* (2007) Testing Multiple Hypotheses to Identify Causes of the Decline of a Lagomorph Species : The New England Cottontail as a Case Study Natural History of New England Cottontails. In: *Lagomorph Biology: Evolution, Ecology, and Conservation* (eds Alves P, Ferrand N, Hacklander K), pp. 167–185. Berling Heidelberg.

Litvaitis JA, Johnson B, Jakubas W, Morris K (2003) Distribution and habitat features associated with remnant populations of New England cottontails in Maine. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **81**, 877–887.

Litvaitis JA, Tash JP, Litvaitis MK *et al.* (2004) A range-wide survey to determine the current distribution of New England cottontails. *Wildlife Society Bulletin*, **34**, 1190–1197.

Ludwig E, Reischl U, Janik D, Hermanns W (2009) Granulomatous pneumonia caused by mycobacterium genavense in a dwarf rabbit (Oryctolagus cuniculus). *Veterinary Pathology*, **46**, 1000–1002.

Mathews F, Moro D, Strachan R, Gelling M, Buller N (2006) Health surveillance in wildlife reintroductions. *Biological Conservation*, **131**, 338–347.

McCoy RH, Steenbergen F (1969) STAPHYLOCOCCUS EPIZOOTIC IN WESTERN OREGON COTTONTAILS. *Bulletin of the Wildlife Disease Association*, **5**, 11.

McMurdie PJ, Holmes S (2013) Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, **8**.

Mueller S, Saunier K, Hanisch C *et al.* (2006) Differences in Fecal Microbiota in Different European Study Populations in Relation to Age, Gender, and Country: a Cross-Sectional Study. *Applied and Environmental Microbiology*, **72**, 1027–1033.

Munson L, Cook RAR (1993) Monitoring, Investigation, and Surveillance of Diseases in Captive Wildlife. *Journal of Zoo and Wildlife Medicine*, **24**, 281–290.

Naik S, Bouladoux N, Wilhelm C *et al.* (2012) Compartmentalized Control of Skin Immunity by Resident Commensals. *Science*, **337**, 1115–1120.

O'Leary SJ, Puritz JB, Willis SC, Hollenbeck CM, Portnoy DS (2018) These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, **27**, 3193–3206.

Org E, Parks BW, Joo JWJ *et al.* (2015) Genetic and environmental control of host-gut microbiota interactions. *Genome Research*, **25**, 1558–1569.

Osebold JW, Gray DM (1960) Disseminated staphylococcal infections in wild jack rabbits (lepus californicus). *Journal of Infectious Diseases*, **106**, 91–94.

Paris JR, Stevens JR, Catchen JM (2017) Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*, **8**, 1360–1373.

Pedersen A, Fenton A, Pedersen AB, Fenton A (2007) Emphasizing the Ecology in Parasite Emphasizing the ecology in parasite community ecology. *TRENDS in Ecology and Evolution*, **22**, 133–139.

Pereira DG, Batista E, de Cristo TG *et al.* (2020) Aspiration Bronchopneumonia By Acinetobacter Baumannii in a Wildlife European Hare (Lepus Europaeus) in Brazil. *Journal of Zoo and Wildlife Medicine*, **51**, 253.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, **7**.

Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility

complex. *Heredity*, **96**, 7–21.

Poulin R, Mouillot D, Journal S, Jun N (2003) Host Introductions and the Geography of Parasite Taxonomic Diversity. *Journal of Biogeography*, **30**, 837–845.

Razzauti M, Galan M, Bernard M *et al.* (2015) A Comparison between Transcriptome Sequencing and 16S Metagenomics for Detection of Bacterial Pathogens in Wildlife. *PLOS Neglected Tropical Diseases*, **9**, e0003929.

Ren SY, Geng Y, Wang KY *et al.* (2014) Streptococcus agalactiae Infection in Domestic Rabbits, Oryctolagus cuniculus. *Transboundary and Emerging Diseases*, **61**, e92–e95.

Rochette NC, Catchen JM (2017) Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, **12**, 2640–2659.

Rosindell J, Hubbell SP, He F, Harmon LJ, Etienne RS (2012) The case for ecological neutral theory. *Trends in Ecology and Evolution*, **27**, 203–208.

Ross AA, Rodrigues Hoffmann A, Neufeld JD (2019) The skin microbiome of vertebrates. *Microbiome*, **7**, 1–14.

Schad J, Ganzhorn JU, Sommer S (2005) Parasite burden and constitution of major histocompatibility complex in the Malagasy mouse lemur, Microcebus murinus. *Evolution; international journal of organic evolution*, **59**, 439–450.

Schad J, Sommer S, Ganzhorn JU (2004) MHC variability of a small lemur in the littoral forest fragments of southeastern Madagascar. *Conservation Genetics*, **5**, 299–309.

Schoeb TR, Fox JG (1990) Enterocecocolitis associated with intraepithelial Campylobacter-like bacteria in rabbits (Oryctolagus cuniculus). *Veterinary pathology*, **27**, 73–80.

Schwender H, Li Q, Neumann C *et al.* (2014) Detecting Disease Variants in Case-Parent Trio Studies Using the Bioconductor Software Package trio. *Genetic Epidemiology*, **38**, 516–522.

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**, 498–510.

Sellers AJ, Ruiz GM, Leung B, Torchin ME (2015) Regional variation in parasite species richness and abundance in the introduced range of the invasive lionfish, Pterois volitans. *PLoS ONE*, **10**, 1–15.

Shoemaker D, Alan W, Kirkpatrick R, Cooper M (1997) Humoral immune response of cottontail rabbits naturally infected with Francisella tularensis in southern Illinois. *Journal of Wildlife Diseases*, **33**, 733–737.

Smulders MJM, Snoek LB, Booy G, Vosman B (2003) Complete loss of MHC genetic diversity in the Common Hamster (Cricetus cricetus) population in The Netherlands. Consequences for conservation strategies. *Conservation Genetics*, **4**, 441–451.

Sommer S, Courtiol A, Mazzoni CJ (2013) MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. *BMC Genomics*, **14**, 1.

Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Strauss A, White A, Boots M (2012) Invading with biological weapons: the importance of disease-mediated invasions. *Functional Ecology*, **26**, 1249–1261.

Suckow MA, Brammer DW, Rush HG, Chrisp CE (2002) *Biology and Diseases of Rabbits*.

Torchin ME, Lafferty KD, Dobson AP, McKenzie VJ, Kuris AM (2003) Introduced species and their missing parasites. *Nature*, **421**, 628–630.

Viggers K, Lindenmayer D, Spratt D (1993) The importance of disease in reintroduction

programs. *Wildl. Res.*, 687–98.

Woolf A, Shoemaker DR, Cooper M (1993) Evidence of tularemia regulating a semi-isolated cottontail rabbit population. *Journal of Wildlife Management*, **57**, 144–157.

Yabsley MJ, Romines J, Nettles VF (2006) Detection of Babesia and Anaplasma Species in Rabbits from Texas and Georgia, USA. *Vector-Borne and Zoonotic Diseases*, **6**, 7–13.

Yu G, Fadrosh D, Goedert JJ, Ravel J, Goldstein AM (2015) Nested PCR biases in interpreting microbial community structure in 16S rRNA gene sequence datasets. *PLoS ONE*, **10**, 1–12.

Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, **44**, 821–824.

Zoetendal EG, Akkermans ADL, Akkermans-van Vliet WM, De Visser JAGM, De Vos WM (2001) The host genotype affects the bacterial community in the human gastrointestinal tract. *Microbial Ecology in Health and Disease*, **13**, 129–134.

Supplementary Table

Table S1. Potential pathogens by ASV. ‡ documented pathogenicity in rabbits, hares, or cottontails. † described in rabbits, hares, or cottontails but pathogenicity unclear. *species is of known pathogenicity to one or more mammalian species.

Genus [Family] (Phylum)	ASV ID	Biology	No. of positive samples		Closest species matches
			EC	NEC	
Acinetobacter‡† [Moraxellaceae] (Proteobacteria)	0342ea32ea76f6b0e88fbb5c0e64624b	Common soil bacteria; Opportunistic pathogens that can cause pneumonia, bacteremia, meningitis, and endocarditis	0	1	
	05ab4587c01c91120329efce9f18a62b		0	1	
	236d2b669d0b22f1f682ed37153a53af		1	0	
	2f2eb9ff996ea176335eaab55ef8572d		0	1	
	368d74e30b8c5ef97d2105d3fa90ffb4		0	1	
	37c03af0c93baa76926221206678ba9e		4	8	*berenziae*
					courvalinii
					*junii**
					modestus
					proteolyticus
					viviani
	3a70840924e06d2055bed46452f38bd7		3	2	
	4f945ead995891d7c82efb1d5e007eba		0	1	*johnsonii*
	6371a55dbbb498d996f6e1c7ee868dae		4	5	
	66a959f7f685bdd092f897408f80e016		3	0	*bohemicus*
	693305b1e19f919a356c92aa57b08bac		3	2	*baumannii**‡
	90c1ffa8d3678882bd15a6e6b01d2817		0	2	*dispersus*
					*haemolyticus**
					parvus
	93c484472e6e5c4e1afecbc0c2a50375		0	1	*radiosistens*
	9d1dbc7f3a8a963b294b70d41987125b		0	1	
	a355375ffbf75c266bd550841be84eb4		1	1	*pragensis*
					celticus
	b994ec18b15f2ab584f37b591352037d		1	0	*bouvetii*
					chinensis
					gandensis

	bb985247faa82f8edb2ccd436507732b		3	5	
Aeromonas ‡ [Aeromonadaceae] (Proteobacteria)	54739e4791c30dc94be4ead9b3dc6df5	Common in soil and water; opportunistic pathogens that can cause gastroenteritis, bacteremia, and necrotizing fasciitis	0	1	
Anaplasma† (Proteobacteria)		Tickborne; zoonotic; cause disease anaplasmosis	0	0	
Bacillus ‡† [Bacillaceae] (Firmicutes)	11f5cc1590e46280017f5f2afc5bf3ed	Saprophyte; common environmental bacteria; some species pathogenic	0	1	
	36f031680d2ce8bab939e6bffcad6e47		0	1	
	4b84dc79e02d95c6411ecd2c59d6c11c		2	0	*mojavensis*
	5a3ebaf94159f3c1b4e7bf9fe173c298		0	1	
	6c49f7a52ecfbe93ada5566c9f9ad1ce		0	1	*horikoshii*
	8daee363bb729a897eba3699df2ae386		4	8	
	8ffc9da4907ca52b11b2926b445a3f13		1	0	*kokeshiiformis* *thermolactis*
	c5b9a58c78ba2039a63663bac8011998		1	1	*butanolivorans* *horikoshi* *muralis* *simplex*
	dfc3fa01e65c1a1c5de95aec5ca97767		0	1	*haynesii* *oceanisediminis* *paralicheniformis* *piscis* *swezeyi*
Bartonella† [Bartonellaceae] (Proteobacteria)		Tickborne; zoonotic; some species pathogenic or opportunistic pathogens	0	0	
Bordetella† [Alcaligenaceae] (Proteobacteria)		Multiple pathogenic species; causes whooping cough and respiratory disease	0	0	
Borrelia†‡ [Spitochaetaceae] (Spirochaetes)		Multiple pathogenic species; causes Lyme disease, borreliosis, bacteremia, and relapsing fevers	0	0	
Brucella† [Brucellaceae] (Proteobacteria)		Multiple pathogenic species; zoonotic; causes brucellosis,	0	0	

		gastrointestinal illness, endocarditis, meningitis, and spontaneous abortion	0	0	
Burkholderia [Burkholderiaceae] (Proteobacteria)		Saprophyte; some species pathogenic; causes melioidosis and pulmonary infection	0	0	
Campylobacter‡ [Campylobacteraceae] (Proteobacteria)	5547dd220f8a2aa775a46aa299c4ec00	Often commensal; opportunistic pathogen; zoonotic	3	7	
Clostridium‡ [Clostridiaceae] (Firmicutes)		Naturally found in gut, reproductive tract, and environment; Multiple pathogenic species; causes botulism and tetanus	0	0	
Corynebacterium‡† [Corynebacteriaceae] (Actinobacteria)	4452fa5c03513bc99fe79de73d8f1831	Saprophyte and common commensal; some species pathogenic or opportunistic pathogens	0	1	*pelargi* *pseudotuberculosis** *ulcerans** *kutscheri* *diphtheriae**
	ae740a2e22255bc76415f84be5d060c0		0	1	*lactis*
	ddd34dc7f565e5697c80d1aca9c40f93		0	1	*bovis**‡ *jeikeium** *renale**
Coxiella‡ [Coxiellaceae] (Proteobacteria)		Some species pathogenic; zoonotic; can cause Q fever, endocarditis	0	0	
Ehrlichia [Anaplasmataceae] (Proteobacteria)	c9a62860d3ef6b6186826fc76b45ccfe	Tickborne; zoonotic; multiple pathogenic species; causes ehrlichiosis	0	1	*ewingii**
Enterococcus‡† [Enterococcaceae] (Firmicutes)	028e2c1b650e2e2b5847154a153e0166	Commensal; opportunistic pathogens; can cause endocarditis, bacteremia, meningitis, and peritonitis.	3	4	*casseliflavus* *durans* *faecalis**‡† *gallinarum** *hirae** *mundtii* *saccharolyticus*
	8babd348a3b7ba726e8f1a9a132fe47a		6	8	
	df7d3bc94e4f0f1cbef7257b0da5a4c6		1	0	

Eubacterium [Eubacteriaceae] (Firmicutes)	8261e8a258f04149bd57e115608d58cc	Typically commensal, can be opportunistic pathogen	1	0	*limosum*
Escherichia-Shigella [Enterobacteriaceae] (Proteobacteria)	4c8288bfbd76958c0c094d87b97650f8	Normal flora of gut and feces; opportunistic pathogen; can cause gastroenteritis, urinary tract infections, colitis, and meningitis	6	3	
	f43f69cc15abe96324e20b38c29563d2		0	1	
Francisella‡ [Francisellaceae] (Proteobacteria)		Causal agent of tularemia	0	0	
Haemophilus‡ [Pasteurellaceae] (Proteobacteria)	ce10bc5c943070e0610ad4310958321e	Commensal to mucosal membranes of respiratory tract, intestines, and vagina; opportunistic pathogen	0	1	*haemoglobinophilus*
Helicobacter [Helicobacteraceae] (Proteobacteria)	04a2ccf6aed90985912e53d443edcd7f	Commensal to upper gastrointestinal tract and liver; some strains are pathogenic; can cause cancer and gastritis	1	1	
	6e2d41c236d5b2f31b55483602432762		0	1	
	e5987670630befc2e0315bd1d753125b		0	1	
Klebsiella [Enterobacteriaceae] (Proteobacteria)		Common environmental bacteria; saprophyte; opportunistic pathogen	0	0	
Legionella [Legionellaceae] (Proteobacteria)	0bf3f25c648260e9cb935bd2709a6776	Common saprophyte; some strains can cause disease; may cause Legionnaires' disease, Pontiac fever, and pneumonia-like illness	1	0	
	33866eb004cf9363f7589ab3d41fd38c		0	1	
	45eab0db972753133f8537d6b4bac27a		0	1	*nagasakiensis*
	718d5b633afb992e98ff5d7347af6002		0	1	
	827bbca5224fa997b1401db1df6846f9		1	0	
	829d685640e34c758cb755a9652cbf6d		0	1	
	85f9cf6ac5b04b877bcbc0c166d249e3		0	1	
	867b6dd038c3f5eeb2c3d838204d9394		1	0	*donaldsonii feeleii**
	99890460449ea196363cf4828fdb92f8		0	1	
	a74bc5a6a7455633cf5e41fe82b58b27		0	1	
	b5e04f1fe4ee73a5c42ed8246f621690		1	0	
	bf539f9cb3059d065dacccabcdf29050		1	0	
	c615404bfb60616b71c32a0d51a76d44		0	1	
	c95abfb921445e16274fcc23c2efe18e		1	0	
	e5669445220036e091ebbee01f845977		0	1	

	ea184b1df098db56af28401e57941f66		0	1	
	f46603fbc859cbaf6b3627caa7e75431		1	0	
	fc28eaad8022bb595025b5feaed58928		1	0	
Leptospira‡ [Leptospiraceae] (Spirochaetes)		Saprophyte; can cause disease, including leptospirosis	0	0	
Leptotrichia [Leptrotichiaceae] (Fusobacteria)	96a0c44812b55dd95c4fc27a29ff1d39	Often commensal; can cause disease and endocarditis	0	2	*goodfellowii**
Listeria‡ [Listeriaceae] (Firmicutes)	122c4248578d1ed17ae0e48a1afc0efe	Saprophyte; some species pathogenic or opportunistic pathogens; can cause listeriosis and meningitis	1	0	*innocua* *ivanovii* *monocytogenes*‡* *seeligeri*
	5adc2daa2981bc07e21b14d53496ff67		2	0	*innocua* *monocytogenes**
Mannheimia [Pasteurellaceae] (Proteobacteria)		Saprophyte; opportunistic pathogen	0	0	
Micrococcus† [Micrococcaceae] (Actinobacteria)		Saprophyte; skin commensal; opportunistic pathogen	0	0	
Moraxella† [Moraxellaceae] (Proteobacteria)	0110fc6ca0347cd659c8b81dfef9e868	Commensal of mucosa; some species pathogenic or opportunistic pathogens; can cause pneumonia, bronchitis, and conjunctivitis	1	2	
	0b13f3ebd9394e30862fc8986c9e0adf		1	0	
	51bf1579df9e129ab8105218f8881687		1	1	*cuniculi*
	7cdb3cc64015f4a13bbc0d6f5c6ee013		7	3	
	7d1ce6efa9e77ef56f5f993830cba674		1	0	
	9257dfb02c102d3bad3bc41f8563ee11		0	2	
	ba1e824872c61acd930ce9482970e534		0	1	
	cb754768bbaa852a9f22e6bf52bc3d52		0	1	
	f6a0fa281b3458132a9078216104628a		7	6	
	f83c087eb4a4c68bcfa8c7f58a4fb824		0	1	
Mycobacterium‡ [Mycobacteriaceae] (Actinobacteria)	e04dd8f94154c2083702fbf3ecb0e3af	Saprophyte; common in water and food; some species pathogenic and opportunistic pathogens; can cause tuberculosis, leprosy, and enteritis	0	1	*madagascariense*
Mycoplasma‡	1d18cd6f38fa28c3ad79536f957714b2		1	0	*bovis**

[Mycoplasmataceae] (Tenericutes)		Saprophyte; commensal; pathogenic and opportunistic species; can cause pneumonia, pelvic inflammatory disease, arthritis, mastitis, and spontaneous abortion	1	0	*capricolum** *felis* *fementans* *leopharyngis* *mycoides** *hyopneaumoniae** *hyorhinis** *pullorum* *pulminis*
	52dbb59a624e6735ff13ec510d706d5c		1	0	*cavipharyngis* *insons*
	6a630f63b77f6e84c827960869740dee		1	0	
	dfeac46dfc4f0216f031332bd1b291b7		1	1	*cavipharyngis*
Neisseria† [Neisseriaceae] (Proteobacteria)	4c2834609383bf1b78ac226a22fe76c6	Commensal of mucosal surfaces; pathogenic species; can cause meningococcal disease, sepsis, and gonorrhea	0	1	
	6e8234d29b248dd1bfa786ebc81df25f		5	10	
Nocardia [Nocardiaceae] (Actinobacteria)		Some species pathogenic or opportunistic; can cause nocardiosis, pneumonia, and encephalitis	0	0	
Orientia [Rickettsiaceae] (Proteobacteria)		Some species pathogenic; can cause scrub typhus	0	0	
Pasteurella‡† [Pasteurellaceae] (Proteobacteria)	928fb89274f12f3691fe593c89f8bbc5	Species can be commensal, pathogenic, or opportunistic; zoonotic; can cause pasteurellosis, respiratory disease, cholera, and septicemia	9	9	*canis** *dagmatis** *multocida**‡
	aa1f21924032b1bb713c2b7c8c763518		0	1	*canis** *dagmatis** *multocida**‡
	ccbd6937a4048a9ce6ba1fcbcd8eaa2a		0	1	*canis** *dagmatis** *multocida**‡
	cd9306bde777da5a8860f9ec662ba45e		3	1	*canis** *dagmatis** *multocida**‡
	d1374a18875c514cb9c396ba8cdbb2c7		0	1	*canis**

					*dagmatis** *multocida*‡*
Pseudomonas‡† [Pseudomonadaceae] (Proteobacteria)	009792680a20775a699473e2ce682761	Diverse genus with saprophytes, soil bacteria, and opportunistic pathogens	25	14	
	0cff17dc8f0cbb1eccd6a0f808cbe4cc		0	1	
	1601b63ea158342776c4cd442b1fe96c		3	1	
	17db5fb5a38a6f1770ffc299bcba710e		0	1	*orientalis*
	1dd0de0176ddf4ce0f0b8ab5fb4d5bde		2	1	*aeruginosa*‡*
	1eb82b44b4ff5d651d11898c23c6a585		0	1	
	1f0fb4082ffc1c983824335b07bc4fa0		1	0	*pyschrotolerans*
	232a51f5276645771dbfbc959d1e8a32		0	1	
	28c63a62d58f95683bda68e1af09d2a0		0	2	
	30f4ec7c7a4b1a68e38b3f26020595e4		0	3	*batumici* *chlororaphis* *moraviensis*
	313708917b41d6751e9448f546f5549a		0	1	
	3209883c6ba8f9003adae3d0489fcdb4		1	0	
	3695b1dda7433526fc3d7e95ee43620a		0	1	
	3f2738e7eb8bfc23b09e89dfcc6ec1a3		8	0	
	49df6cdec06cb4ed2df781073e2b19bc		1	0	*gingeri*
	4bb029f06bdd2f369c72e15f4789113d		0	1	
	4e92f89b5b551cf8c85f5d1327f0e242		0	1	*lini*
	52389d158f00bae3f23b84402d7b03a7		1	5	
	5adf7069178cea5d32662c2b50acacf2		0	1	
	5d4a167c6b7b7322c3066b9d357e4497		0	1	*extremaustralis* *proteolytica* *veronii*
	5ded73994d6c6db89d54f8a869ee38fb		5	11	
	6827ed4b2e96c1ea82167b54cb2d0527		0	1	
	68f64871fade18c35a089770e5015a37		3	3	*extremaustralis* *proteolytica* *veronii*
	69686976c3a2a7876234986e4714fcd3		1	2	
	69d9bd0fff4fa0bfc373eca44023973d		7	17	
	7cd996bb047a00ee3a1a96fa23427a0c		0	1	
	7da94d47fee93a7f6447c0c0dde2b5f0		1	0	

	858faf8ba1de6d0d613095e525f24424		1	0	
	8954ce86e6c903c0f10705585bde9da4		0	1	
	8b209128f5aaaf96c6811f8ca9190267		10	21	
	9332a45a3ddcc0005bd507c9d095585e		1	1	*reidholzensis*
	953eda5ca99a9a63c7a738a8575b8f65		2	5	
	a06fd28618ae12842004968bab06a766		1	1	
	a170e22bd06530aecbb1a71d737b7543		1	1	*alkylphenolica* *cichorii* *laurentiana* *vranovensis* *wadenswilerensis*
	a8a489f9c77e295872f5787dfc6535c6		11	7	*guguanensis*
	aa335e63c66f918058e69257cce84d3d		0	1	
	b57fd93fd9d1e7837a11cfce4383fb0e		2	1	
	ba2f5bc3e6b9698a0bd86847bd5d9c40		4	12	
	c06f6a48ce9a25814b592df7829c3a39		1	0	
	c1c47613e6618cf652b86ef1ab53720f		1	0	
	c3edd09182a86cdd78963a3307011fac		0	1	
	cdf69b9f1e56acb68cdfd3542a83ae60		4	12	
	cfbd5d428b352e900a39dc0da67ebd3a		1	0	
	d46365bc2fc1c3d3a6dc1c6b161f4333		0	1	
	df9a5388e9c0322b1a619648fb479cdf		0	3	
	e1661704d8f7a3a702f70d4b8353a256		1	0	*aeruginosa**‡ *furukawaii* *guguanensis* *peli* *resinovorans*
	f2ba83acecd6beee4c18ef3cba385a18		0	2	
	f8ef89b0ef91e703f5a73299b597887b		3	8	*graminis* *viridiflava*
Rhodococcus [Nocardiaceae] (Actinobacteria)	6109ecb534a70ca103d4aee9875d087a	Common in soil and water; saprophyte; some strains pathogenic	0	1	
	aa609e4d569307c490efe0d53bf877f9		0	2	*erythropolis* *qingshengii* *degredans*
	b35556a43f36f85349cb9d779b109b54		0	2	*corynebacterioides* *trifolii*

	c292cc07cbb7c242682dd6064e5447a1		2	0	
Rickettsia† [Rickettsiaceae] (Proteobacteria)	ecc7c776997c56087616034e99df2668	Pathogenic; zoonotic; causes typhus, pox, spotted fevers	0	1	*africae** *amblyommii** *bellii* *conori** *felis** *prowazekii** *typhi**
Salmonella [Enterobacteriaceae] (Proteobacteria)		Saprophyte and opportunistic pathogen	0	0	
Sphingomonas [Sphingomonadaceae] (Proteobacteria)	0fe59612948140b14867954b8762c27a	Common and widespread environmental bacteria; opportunistic pathogen	0	1	*oligoaromativorans* *polyaromaticivorans*
	1cc70068e55b83c44c4cc2233343229b		0	1	*abaci* *cynarae* *roseiflava*
	1e0f325e8b24e5b9396fd23990d53938		0	3	*oligoaromativorans* *polyaromaticivorans*
	37b90a559870427419c4b10ebec07a2c		0	3	
	40115e05d0dc8f2e0a0fc8521d307626		0	1	*abaci* *cynarae* *roseiflava*
	4711420ea5416c1b6935904216a58a50		1	1	*aerolata* *olei*
	4813199881e90f7e1ade29bec3134311		0	1	
	49bea5eaf6cb56ca4012858ad4ef2758		0	2	*oligoaromativorans* *polyaromaticivorans*
	4b956734dadcbd24d264d397add0b283		0	1	*endophytica* *phyllosphaerae*
	4a80eb89bd903a2abec371a2ff6bec0f		0	1	
	4baca23af25cb9721a340d79bd2a4b9d		0	3	*oligoaromativorans* *polyaromaticivorans*
	4c1ca629d90819d7c7dc96fa6693aae2		0	1	
	51c27abcd11c292bc0eb9ac7db0bee8d		0	1	*arantia* *montana* *oligoaromativorans*

			0	1	*polyaromaticivorans* *prati*
	56e169624d7104a34526225dd4e4f241		0	2	*abaci* *cynarae* *roseiflava*
	5d7a9e56ac840a5ee787cc93f7fbac98		0	1	
	60bb6a32d56cf380228d4a71fd163df2		0	1	*aerolata* *olei*
	647a6e7563ec946cc06f35747679e526		0	1	
	792cf373040017cd83912797b8d0ce00		0	1	*arantia* *montana* *prati*
	827a568edc2cd70b311cdc07dab7acda		1	1	
	8c715c15875c554a34a8fd7cd30a10d9		0	1	*oerolata* *Olei*
	8dd6ae90f41b0427aaa5938511171fbf		0	1	*abaci* *cynarae* *roseiflava*
	96e23b574722b4dba7a918263fc1ecc6		0	1	*endophytica* *phyllosphaerae*
	9842d9cf9d14a75deabe902f13d10cdb		0	1	*arantia* *montana* *oligoaromativorans* *polyaromaticivorans* *prati*
	a640291c122b34247d73ce6252154b5e		0	2	*oligoaromativorans* *polyaromaticivorans* *wittichii*
	ab4a2f8da36d4df7f0208e28f29603b1		0	2	
	b4f9138d30af941e268274aafaba4803		0	2	*oligoaromativorans* *polyaromaticivorans*
	b7ace396aa0b7393072bc0597a099038		0	1	
	c89efecc79747ad47c17747b823372d5		1	2	*glacialis* *psychrolutea*
	ccaadd6b6a19c706f9420e7e9a72a2e0		0	1	
	ce0d7d93fc323ca5c7b1fc18071e55d4		0	1	*arantia*

			0	1	*montana* *prati*
	d410da853445281efc6db49a978a10d6		1	0	*aerolata* *olei*
	d5f46e8aa089d5c4a7e2f5dd6dd4382b		1	0	*arantia* *montana* *prati*
	d92235d8c3d9cae7713114c414e86291		0	1	
	e1178e9ca11ed2e9b7512a6451d9b042		0	1	
	e867b8a6006646ac1441a89272805839		0	1	
	ee4d8c239717384199dcbf4552be9e7e		0	2	
	f0a367b7722ffdfbe55a3061d853b701		0	1	*hankyongensis*
	f0bcf604e080d2bc8bf781b10dbf5fd8		0	1	*arantia* *montana* *prati*
	f19b3a33eaad3e784641de933f3ebd7b		0	2	*glacialis* *psychrolutea*
	f751509f2488e053ae842a1acae9702e		0	4	*oligoaromativorans* *polyaromaticivorans*
Spiroplasma [Spiroplasmataceae] (Tenericutes)		Common gut symbiont; can cause disease in rodents	0	0	
Staphylococcus‡† [Staphylococcaceae] (Firmicutes)	1355705d5f5537fdd6063e8aef08650d	Common skin and mucosal commensal; pathogenic and opportunistic pathogen	4	5	*aureus**‡† *caprae** *epidermis** *hominis** *lugdunensis** *pasteuri*
	19a139f6eef38d65c37ac7297e222b17		2	2	*vitulinis*
	2ad39a6148a52137531084b50b74ab9b		1	0	*arlettae* *kloosii*†
	2b998f812c268ebcf105cd49acd67990		1	0	
	4606815b23ed5535782faba4eea2a513		0	1	*stepanovicii*
	4f69c663fc25039048f1ffc4a369ec79		0	1	*vitulinis*
	54192b86acd2062c6d0f3be602cf37e3		1	1	

	59c19ba971bf8e1cfaaa0f6e73a97825		1	0	
	748cab1a44a80e38adbba16931df032e		4	5	
	9900aee560e191b16e56a45d42158a49		0	1	*equorum*†
	c0c415f73589a908a8c48d6df1e2c91d		0	1	
	c415efb8f8d0cbc3c5c2b9d9450e0afb		1	0	*arlettae* *kloosii*†
	deebad5ffd1f3a2afc4151471d054636		1	0	
	f745d265f36ea7eac26d97280e690796		12	11	
Stenotrophomonas† [Xanthomonadaceae] (Proteobacteria)	46aaf3735d747a329482df089438b2b8	Soil and plant associated; saprophyte; opportunistic pathogen	0	1	*pavanii*
	4c5acc235c1777d411ede7cb41f88676		0	1	
	628eb483be4174d48a09070f5640d121		1	0	
	70ba427bc7ef9110c2a1c49eb3177910		0	1	
	78e0f808e9b9a0ce5682a5f825001061		0	1	
	8fb925e40b88d6314bbce04bece83804		25	18	
	95a8fdad7e371d1061cdafdb448f61b4		1	0	
	9b0cdb331ee6a6af7bf19db1a3d840fa		0	1	
	b9879d325f89aea8dd84cd331deb2f37		1	0	
	d4bc50f3bc51661b9111bbdcfc977916		1	3	*rhizophila*
	f1bc452ea849acfd2b2689800e235c07		1	0	
	f9d5150646e55bd83664fd85164c2275		1	0	
Streptococcus‡† [Streptococcaceae] (Firmicutes)	000e38f477632fa6299901badf2e0f06	Soil and water bacteria; saprophyte; common skin and oral commensal; some pathogenic or opportunistic pathogens; can cause septicemia, endocarditis, meningitis, pneumonia, and dental caries	1	1	*gallolyticus**
	349b10a604cd2aba88db3ecbb4eb65a0		0	2	
	4ee45ebbdb2ef72675cdb466a5fd4d9a		1	0	*gallinaceus*
	4f81a404de7546a59679497a1e8d21cf		0	1	
	84b629c4b3d34cce9cb318640466be05		1	13	
	88d45f1da91c832b51439f2e46284345		1	0	
	a754b7955dbbfb64b01277011ffca768		1	0	
	ab60931f99194481c9230dcbd89dbeb		0	1	
	ad2c511bdcd3f0a816d1e7141a9bdab3		0	3	
	bd2ebc70501f7d867c204f94c4e483da		1	2	*salivarius* *thermophilus*
	df33c34c10e8d19cc5505005674fec19		1	0	
	e3e480e457abded49e322b2c4d9b4b8c		0	1	
	e467d524b08b74dca23e6f415d60110c		1	1	
	f562590cbd1d81981183744986aaa011		6	6	

Treponema‡ [Spirochaetaceae] (Spirochaetes)	b4012526cfaf071f40aac9717833e489	Commensal; can cause syphilis, skin disease, periodontitis, and lesions	1	0	*maltophilum** *parvum*
	da7323be6c28864891b90a50bdfcf4c6		0	1	*rectale*
	ddb7505c161248db3928e4fce284a94c		0	2	*rectale*
	f00e8abdef54e30313e915acb0c33ee3		1	0	*rectale*
Ureaplasma [Mycoplasmataceae] (Tenericutes)		Urogenital commensal; opportunistic pathogen	0	0	
Vibrio† [Vibrionaceae] (Proteobacteria)		Saprophyte; opportunistic pathogen; can cause foodborne illness, cholera	0	0	
Yersinia [Yersiniaceae] (Proteobacteria)		Saprophyte; some species pathogenic; can cause the plague	0	0	

Supplementary References

Allison K, Clarridge JE (2005) Long-term respiratory tract infection with canine-associated Pasteurella dagmatis and Neisseria canis in a patient with chronic bronchiectasis. *Journal of Clinical Microbiology*, **43**, 4272–4274.

Angelakis E, Billeter SA, Breitschwerdt EB, Chomel BB, Raoult D (2010) Potential for Tick-borne Bartonelloses. *Emerging Infectious Diseases*, **16**, 385–391.

Antunes LCS, Visca P, Towner KJ (2014) Acinetobacter baumannii: Evolution of a global pathogen. *Pathogens and Disease*, **71**, 292–301.

Arias CA, Contreras GA, Murray BE (2010) Management of multidrug-resistant enterococcal infections. *Clinical Microbiology and Infectious Diseases*, **16**.

Avsever ML, Seza E, Tunaligil S (2015) Aeromonas hydrophila isolation from Holland lop (Oryctolagus cuniculus) rabbits. *Etlik Vet Mikrobiyol Derg*, **26**, 21–22.

Balikoglu-Yilmaz M, Yilmaz T, Esen AB, Engin KN, Taskapili M (2012) Pasteurella canis and Granulicatella adiacens conjunctivitis outbreak resistant to empirical treatment in a child welfare agency. *Journal of Pediatric Ophthalmology and Strabismus*, **49**, 314–319.

Bourguet A, Desprez I, Volait L *et al.* (2019) Conjunctival bacterial and fungal flora and cutaneous fungal flora in healthy domestic rabbits (Oryctolagus cuniculus). *Journal of Small Animal Practice*, **60**, 417–422.

Ciceroni L, Pinto A, Rossi C *et al.* (1988) Rickettsiae of the spotted fever group associated with the host-parasite system Oryctolagus cuniculi/Rhipicephalus pusillus. *Zentralblatt fur Bakteriologie Mikrobiologie und Hygiene - Abt. 1 Orig. A*, **269**, 211–217.

Comer JA, Paddock CD, Childs JE (2001) Urban zoonoses caused by Bartonella, Coxiella, Ehrlichia, and Rickettsia species. *Vector borne and zoonotic diseases*, **1**, 91–118.

Cooke MM, Jackson R, Coleman JD (1993) Tuberculosis in a free-living brown hare (Lepus europaeus occidentalis). *New Zealand Veterinary Journal*, **41**, 144–146.

Cooper SC, McLellan GJ, Rycroft AN (2001) Conjunctival flora observed in 70 healthy domestic rabbits (Oryctolagus cuniculus). *The Veterinary record*, **149**, 232–235.

Cutler SJ, Ruzic-Sabljic E, Potkonjak A (2017) Emerging borreliae – Expanding beyond Lyme borreliosis. *Molecular and Cellular Probes*, **31**, 22–27.

Doughari HJ, Ndakidemi PA, Human IS, Benade S (2011) The Ecology, Biology, and Pathogenesis of Acinetobacter spp.: An Overview. *Microbes and Environments*, **26**, 101–112.

Fisher K, Phillips C (2009) The ecology, epidemiology and virulence of Enterococcus. *Microbiology*, **155**, 1749–1757.

Galan M, Razzauti M, Bard E *et al.* (2016) 16S rRNA Amplicon Sequencing for Epidemiological Surveys of Bacteria in Wildlife. *mSystems*, **1**, 1–22.

Greig A, Stevenson K, Perez V *et al.* (1997) Paratuberculosis in wild rabbits (Oryctolagus cuniculus). *Veterinary Record*, **140**, 141–143.

Hoskisson PA (2018) Microbe Profile: Corynebacterium diphtheriae - an old foe always ready to seize opportunity. *Microbiology*, **164**, 865–867.

Judge J, Kyriazakis I, Greig A, Davidson RS, Hutchings MR (2006) Routes of intraspecies transmission of Mycobacterium avum subsp. paratuberculosis in Rabbits (Oryctolagus cuniculus): a Field Study. , **72**, 398–403.

Kernif T, Parola P, Ricci J-C, Raoult D, Rolain J-M (2010) Molecular detection of Bartonella alsatica in rabbit fleas, France. *Emerging Infectious Diseases*, **16**.

Lartigue M-F, Monnet X, Le Flèche A *et al.* (2005) Corynebacterium ulcerans in an

immunocompromised patient with diphtheria and her dog. , **43**, 999–1001.

Lim YK, Kweon OJ, Kim HR, Lee M-K (2016) Leptrotrichia goodfellowii infection: case report and literature review. *Annals of Clinical & Laboratory Science*, **46**, 83–86.

Linde H-J, Joachim H, Holler E, Reischl U, Lehn N (2002) Septicemia due to Acinetobacter junii. *Journal of Clinical Microbiology*, **40**, 2696–2697.

Lopetuso LR, Scaldaferri F, Petito V, Gasbarrini A (2013) Commensal Clostridia: leading players in the maintenance of gut homeostasis. *Gut Pathogens*, **5**.

Ludwig E, Reischl U, Janik D, Hermanns W (2009) Granulomatous pneumonia caused by mycobacterium genavense in a dwarf rabbit (Oryctolagus cuniculus). *Veterinary Pathology*, **46**, 1000–1002.

McCoy RH, Steenbergen F (1969) STAPHYLOCOCCUS EPIZOOTIC IN WESTERN OREGON COTTONTAILS. *Bulletin of the Wildlife Disease Association*, **5**, 11.

Moore JE, Corcoran D, Dooley JSG *et al.* (2005) Campylobacter. *Veterinary Research*, **36**, 351–382.

Osebold JW, Gray DM (1960) Disseminated staphylococcal infections in wild jack rabbits (lepus californicus). *Journal of Infectious Diseases*, **106**, 91–94.

Parker JL, Shaw JG (2011) Aeromonas spp. clinical microbiology and disease. *Journal of Infection*, **62**, 109–118.

Paul GC, Friend DG (2019) Clostridial enterotoxemia and coccidiosis in weanling cottontail rabbits (Sylvilagus audubonii, Sylvilagus floridanus, Sylvilagus nuttallii) from Colorado, USA. *Journal of Wildlife Diseases*, **55**, 189–195.

Pereira DG, Batista E, de Cristo TG *et al.* (2020) Aspiration Bronchopneumonia By Acinetobacter Baumannii in a Wildlife European Hare (Lepus Europaeus) in Brazil. *Journal*

of Zoo and Wildlife Medicine, **51**, 253.

Perlman SJ, Hunter MS, Zchori-Fein E (2006) The emerging diversity of Rickettsia. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 2097–2106.

Poyart C, Lambert T, Morand P *et al.* (2002) Native valve endocarditis due to Enterococcus hirae. *Journal of Clinical Microbiology*, **40**, 2689–2690.

Quinton JF, Lennox A, Guillon L *et al.* (2014) Results of bacterial culture and sensitivity testing from nasolacrimal duct flushes in one hundred and three both healthy and clinically ill pet rabbits (oryctolagus cuniculus). *International Journal of Applied Research in Veterinary Medicine*, **12**, 107–120.

Razzauti M, Galan M, Bernard M *et al.* (2015) A Comparison between Transcriptome Sequencing and 16S Metagenomics for Detection of Bacterial Pathogens in Wildlife. *PLOS Neglected Tropical Diseases*, **9**, e0003929.

Ren SY, Geng Y, Wang KY *et al.* (2014) Streptococcus agalactiae Infection in Domestic Rabbits, Oryctolagus cuniculus. *Transboundary and Emerging Diseases*, **61**, e92–e95.

Sato S, Brinkerhoff RJ, Hollis E *et al.* (2020) Detection of zoonotic Bartonella pathogens in rabbit fleas, Colorado, USA. *Emerging Infectious Diseases*, **26**, 778–781.

Schlievert PM, Gahr PJ, Assimacopoulos AP *et al.* (1998) Aggregation and binding substances enhance pathogenicity in rabbit models of Enterococcus faecalis endocarditis. *Infection and Immunity*, **66**, 218–223.

Schoeb TR, Fox JG (1990) Enterocecocolitis associated with intraepithelial Campylobacter-like bacteria in rabbits (Oryctolagus cuniculus). *Veterinary pathology*, **27**, 73–80.

Suckow MA, Brammer DW, Rush HG, Chrisp CE (2002) Biology and Diseases of Rabbits. In: *Laboratory Animal Medicine* (eds Fox JG, Anderson LC, Loew FM, Quimby FW), pp. 329–

364.

Washburn LR, Cole BC, Gelman MI, Ward JR (1980) Chronic arthritis of rabbits induced by mycoplasmas I. clinical, microbiologic, and histologic features. *Arthritis & Rheumatism*, **23**, 825–836.

Wising C, Mölne L, Jonsson IM, Ahlman K, Lagergård T (2005) The cytolethal distending toxin of Haemophilus ducreyi aggravates dermal lesions in a rabbit model of chancroid. *Microbes and Infection*, **7**, 867–874.

Yabsley MJ, Romines J, Nettles VF (2006) Detection of Babesia and Anaplasma Species in Rabbits from Texas and Georgia, USA. *Vector-Borne and Zoonotic Diseases*, **6**, 7–13.

Yáñez A, Martínez-Ramos A, Calixto T *et al.* Animal model of Mycoplasma fermentans respiratory infection.

CHAPTER 3

Species identity dominates in structuring fecal microbiome of sympatric invasive and imperiled cottontails

Kimberly B. Neil, Thomas J. McGreevy, Jr., and David M. Rand

Introduction

The association between gut microbes and their hosts profoundly impacts species' health, physiology, development, and behavior (Kinross *et al.* 2011; Costello *et al.* 2012; Amato 2013). A healthy gut microbiome can exclude pathogens, and many mammals depend on gut symbionts for nutrient acquisition. Numerous factors shape the gut microbial community and its stability over time, including both host-associated elements (e.g., nutritional needs, immune system, genetics) and environmental constraints (e.g., diet, local exposures). For example, host genotype and evolutionary relationships influence microbial community composition both within species and across higher-order taxonomic relationships (Zoetendal *et al.* 2001; Ley *et al.* 2008) and geography can also explain different microbiome communities across populations and yield insight into local microbial exposure sources (Mueller *et al.* 2006; Fallani *et al.* 2010).

Due to the pervasive influence of both host and environment in shaping the gut microbiome, it can be difficult to translate laboratory studies to natural systems. Knowledge of microbiota associated with wild, non-model species is rapidly advancing, and its application to systems of conservation relevance holds particular importance for both advancing ecological theory and protecting imperiled species (Redford *et al.* 2012; Amato 2013; Bahrndorff *et al.* 2016; Trevelline *et al.* 2019).

Species of conservation concern may face multiple challenges in their natural habitats that can impact microbiome composition. Habitat and anthropogenic pressures may vary across a species' range, resulting in biogeographic signals in microbiota that are reflective of local conditions. For example, habitat fragmentation can alter microbial community compositions through altered foraging opportunities, environmental exposures, and species' interactions (Amato *et al.* 2013; Barelli *et al.* 2015; Becker *et al.* 2017; Teyssier *et al.* 2018). Allopatric

populations may harbor distinct microbiota due to neutral processes acting independently across populations (e.g., ecological drift) (Lankau *et al.* 2012; Baldo *et al.* 2018), adaptation to divergent local conditions, such as varying foraging material (Duncan *et al.* 2003), or more closely related individuals harboring more similar microbiomes due to maternal transmission (Funkhouser & Bordenstein 2013) or heritable host genotype effects (Bolnick *et al.* 2014, 2015). However, many species possess an identifying "core" gut microbial community that supersedes environmental variation (Ley *et al.* 2008; Cariveau *et al.* 2014). Indeed, species-level signatures can be viewed in a larger evolutionary framework (Ley *et al.* 2008), and relationships among recovered microbial communities may form patterns congruent with host phylogeny (Kohl 2020). Due to the broad functional role of the microbiome as well as potential coevolution between hosts and their microbes, knowledge of species' microbiomes can also yield valuable insight into their evolutionary histories and contemporary physiological demands (Colston & Jackson 2016).

Here, I explore microbiome composition across two congeners of conservation concern: the imperiled New England cottontail (*Sylvilagus transitionalis*; NEC) and the non-native eastern cottontail (*S. floridanus*; EC). NEC is the present target of a reintroduction program back into its native New England range. Since the 1960s, NEC have declined in abundance and have contracted to just five geographically distinct remnant groups (Litvaitis *et al.* 2004; Fenderson *et al.* 2011). NEC population decay is largely attributed to habitat loss as a consequence of land use change (Litvaitis *et al.* 2003) but is also coincident with the intentional introduction of ECs. Beginning in the early 1900s, ECs were released into the northeastern United States to supplement small game hunting (Chapman & Morgan 1973). Records suggest that multiple subspecies of ECs were introduced in numbers in excess of 200,000 (Litvaitis *et al.* 2007), until

concerns of disease halted their import in the 1940s (Ayres & Feemster 1948; Johnston 1972). Currently, ECs are widespread throughout much of New England, creating a patchy distribution where some populations occur sympatrically whereas others exist in allopatry.

I investigate the role of species' identity, biogeography, and reintroduction in structuring microbiome communities across this imperiled-invasive cottontail system. I expect to detect a "core" species microbiome, as has been documented for other species (e.g., Cariveau *et al.* 2014; Easson *et al.* 2020). However, whether shared regional environments also result in increased similarity of microbial phylotypes between sympatric species as compared to other populations of conspecifics is unknown. Microbiomes of sympatric EC and NEC may display increased convergence due to regional habitat and/or diet overlap. Alternatively, selective pressure to maintain a "core" microbial community that meets key species-specific physiological and health demands may override shared local exposures, resulting in greater microbial community similarity within species than between sympatric populations of different species. Little is known about EC and NEC foraging or nutritional requirements (Dalke & Sime 1941), but ECs are more ecologically flexible relative to habitat specialized NEC (Smith & Litvaitis 2000; Litvaitis *et al.* 2007; Cheeseman *et al.* 2018). This could translate into greater range-wide gut microbial variability relative to NEC. Therefore, it is possible that geographic spread may not reciprocally impact NEC and EC microbiomes across populations.

Whether reintroduced NEC display a distinct microbiome relative to their wild counterparts is also unknown. NEC have been translocated to Patience Island, Rhode Island since 2012, and the populations continued to be annually supplemented by captive breeding up to 2018. Captive breeding environments are typically highly artificial, and species' diets, ambient microbial exposures, and interactions are significantly altered compared to natural contexts

(Nelson *et al.* 2013; Kohl & Dearing 2014; Kohl *et al.* 2014; Loudon *et al.* 2014; Xie *et al.* 2016). Given that gut microflora reflect coevolved relationships that aid in host nutrient acquisition and pathogen defense, perturbations to these microbial communities could be maladaptive. Reintroduced cottontails possess a shared ancestry through captive breeding stock and experience the same contemporary ecological selective pressures, yet captive-born offspring did not experience natural conditions or foraging opportunities during early life. I predict that reintroduced NEC will possess a similar fecal microbiome to wild-born individuals due to the critical role of microflora in species' health, and particularly in nutrient acquisition by herbivores (Stumpf *et al.* 2016). However, I also predict distinct microbial taxa and/or microbial diversity levels will be present relative to their wild conspecifics.

Methods

Power analysis

A power analysis was conducted in RStudio v.1.1.463 using preliminary data consisting of seven EC fecal pellets and eight NEC fecal pellets, as in Xia *et al.* 2018. The minimum number of samples per population required to detect a difference in alpha diversity between species was four (alpha = 0.05, power = 0.80, t-test) **(Figure S1).** The package "HMP" was used to compare taxa frequencies of NEC and EC based on the Dirichlet-multinomial parameters obtained from the preliminary sample dataset. Alpha was set to 0.05, and sample sizes per species and read depths were varied. This analysis indicated that as few as six individuals were needed to detect the effect size observed in the preliminary dataset **(Figure S1)**.

Sample collection

Fecal pellets were obtained through archives in the Wildlife Genetics and Ecology Laboratory (WGEL) at the University of Rhode Island (URI). Fecal pellets were limited to the

winter of 2018-2019, spanning December 2018 and January-March of 2019. Samples were collected by state wildlife agencies and/or WGEL research staff, placed in sterile tubes, transported to WGEL at 4°C, and stored at -80°C until processing. To reduce the likelihood of resampling the same individual, samples were chosen that displayed different mitochondrial genotypes, as determined by WGEL (data not shown), or were collected more than two miles apart. A total of 34 EC and 39 NEC samples were used (**Figure 1**), spanning three sympatric, wild populations: western Connecticut (WCT; $N_{EC} = 14$, $N_{NEC} = 13$), eastern Connecticut (ECT; $N_{EC} = 8$, $N_{NEC} = 6$), and Cape Cod, Massachusetts (MACC; $N_{EC} = 12$, $N_{NEC} = 9$), and one reintroduced NEC population located on Patience Island, Rhode Island (PI; $N_{NEC} = 11$).

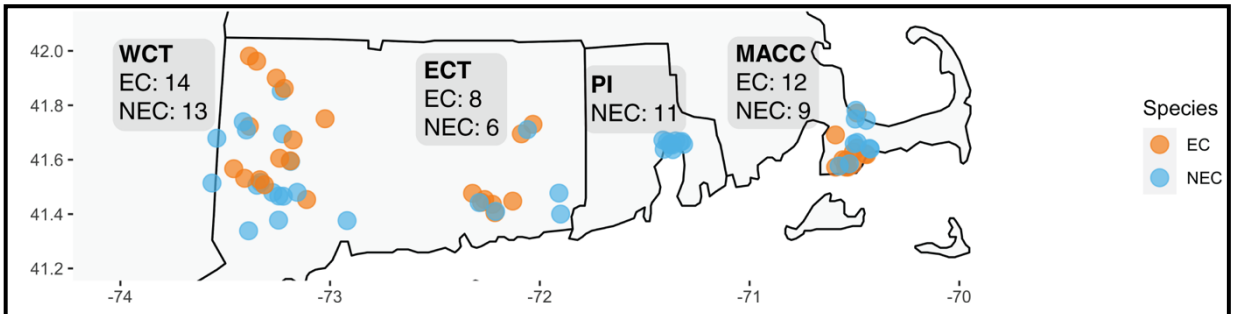

Figure 1. Sampling sites and sampling sizes for EC and NEC. Samples encompass three wild sympatric populations of EC and NC: 1) western Connecticut (WCT), 2) eastern Connecticut (ECT), 3) Cape Cod, Massachusetts (MACC), and one reintroduced NEC population: Patience Island, Rhode Island (PI).

DNA extraction and sequencing

DNA was extracted from each fecal sample using the ZymoBIOMICS DNA Mini Kit (Zymo Research, Cat. No. D4300) according to manufacturers' instructions. Negative controls were included in each batch, consisting of only kit reagents and no sample. Quality was assessed by gel electrophoresis and nanodrop. Amplification of the V4 region of the 16s-rRNA gene was conducted using Phusion High-Fidelity DNA Polymerase (New England BioLabs) and primers 518F and 806R (Caporaso *et al.* 2012). Negative PCR controls were included that consisted of molecular-grade water in place of DNA. Amplified products were purified using AMPure beads

(Beckman Coulter) according to manufacturer instruction, and shipped to the URI Genomics Core for further library preparation then sequencing on the Illumina MiSeq platform (2 x 250 base pair (bp) paired-end). Briefly, the Nextera XT Index Kit was used to attach full indices and adapters, and products were again purified by AMPure bead cleanup according to manufacturer instruction. Samples were analyzed by the Agilent BioAnalyzer DNA 1000 chip and quantified using Qubit (Invitrogen) fluorometry. Samples were normalized to equimolar ratio before pooling, and the final pooled library was quantified using qPCR prior to sequencing.

Quality filtering and read processing

Forward and reverse reads were imported into Qiime2 (version 2020.2) for standard quality control, read assembly, denoising, and sequence classification. The Qiime2 Dada2 plugin was used to denoise, remove chimeras and identify amplicon sequence variants (ASVs). Taxonomy was assigned using the Silva 132 reference database, and sequences identified as chloroplasts or mitochondria were pruned from the dataset. A phylogenetic tree was generated for beta diversity estimates. Upon visualization of raw data and negative control abundance, a Patescibacteria ASV was identified as a contaminant and removed from the dataset (**Figure S2**).

Statistical analyses

All statistical analyses were conducted in RStudio v.1.1.463. Discrete population groups were assigned for each species based on known population genetic groupings that cluster with geography. Compositional differences among species were visualized by non-metric multidimensional scaling (NMDS) and among populations by principal coordinates analysis (PCoA) using the R package 'phyloseq' (McMurdie & Holmes 2013). To test whether explanatory variables impacted ordination patterns, samples were sub-sampled to a depth of 5,000 ASVs; based on rarefaction curves, this depth was appropriate to capture the majority of

microbial diversity while also retaining the maximum number of samples possible (**Figure S3**). We performed a nonparametric multivariate analysis of variance (permanova) using the package 'vegan'. Multiple independent models were run to compare the effects of species identity, population, sample month, latitude, longitude, and whether the population was wild or reintroduced. When appropriate, pairwise differences among groups were explored using the package 'pairwiseAdonis'. Bray-Curtis, weighted Unifrac, and unweighted Unifrac dissimilarity matrices were calculated using 'phyloseq' (McMurdie & Holmes 2013). To determine which microbes were driving any differences, a similarity percentage analysis (SIMPER) was conducted at the ASV level.

Alpha diversity estimates included Shannon index, observed ASVs, and Simpson diversity. For each sample, we calculated the mean of 20 iterations at a sub-sampling depth of 5,000 ASVs (between species) or 10,000 ASVs (within species). Differences were assessed with the Kruskal-Wallis test and a Bonferroni correction for multiple testing using base R functions.

Geographic distances were calculated between pairs as the Euclidean distance in kilometers between latitude and longitudes of collection coordinates. Comparisons of individual Bray-Curtis dissimilarities and geographic distances were performed per species, using the Mantel function in the package 'vegan'. At the population level, a Kruskal-Wallis test was used to investigate whether Bray-Curtis distances differed between sympatric and allopatric populations.

To assess the conservation of bacterial taxa across populations, we assessed the presence and abundance of ASVs within and across populations. The core microbiome of a species was calculated using the package 'microbiome' and a detection threshold of 0.001% and a prevalence threshold of 0.80. This detection threshold is frequently employed when identifying core

microbiota (e.g., Antwis *et al.* 2018; Griffiths *et al.* 2019), whereas the prevalence threshold may vary depending on system characteristics. As no taxa were shared within species at 100% prevalence and few at 90%, 80% was selected as appropriate for both EC and NEC.

I assessed differentially abundant bacterial taxa using Linear Discriminant Analysis (LDA) Effect Size (LEfSe). LEfSe first compares the relative abundance of ASVs using a Kruskal-Wallis test, and ASVs identified as significantly different are then piped into a linear discriminant analysis (LDA) to determine their effect sizes. I applied a minimum logarithmic LDA score (e.g., effect size) of 2.5 and an alpha of 1E-5 for species-level comparisons and 0.001 for population-level comparisons.

To explore functional profiles of host-associated fecal microbiome, I used the bioinformatics tool PICRUSt, which uses 16s-rRNA amplicon data to predict gene family abundances. We used an ASV table with associated Silva taxonomy and performed separate analyses to compare samples at the species-level and the population-level. LEfSe was used to identify enriched functions using the thresholds specified previously.

Results

Three samples were excluded due to low numbers of reads or failure to sequence. Of the remaining samples, between 1,515 and 85,227 reads were assembled (mean = 35,811 ± 1,898). In total, 4,097 ASVs were found across all samples, encompassing 17 bacterial phyla plus reads identified as bacterial but not assigned to a specific phylum (**Figure 2**). The most dominant phylum for both NEC and EC was Firmicutes, which was primarily composed of the family Ruminococcaceae (**Figure 2**). The second most abundant phyla was Bacteroidetes, which was

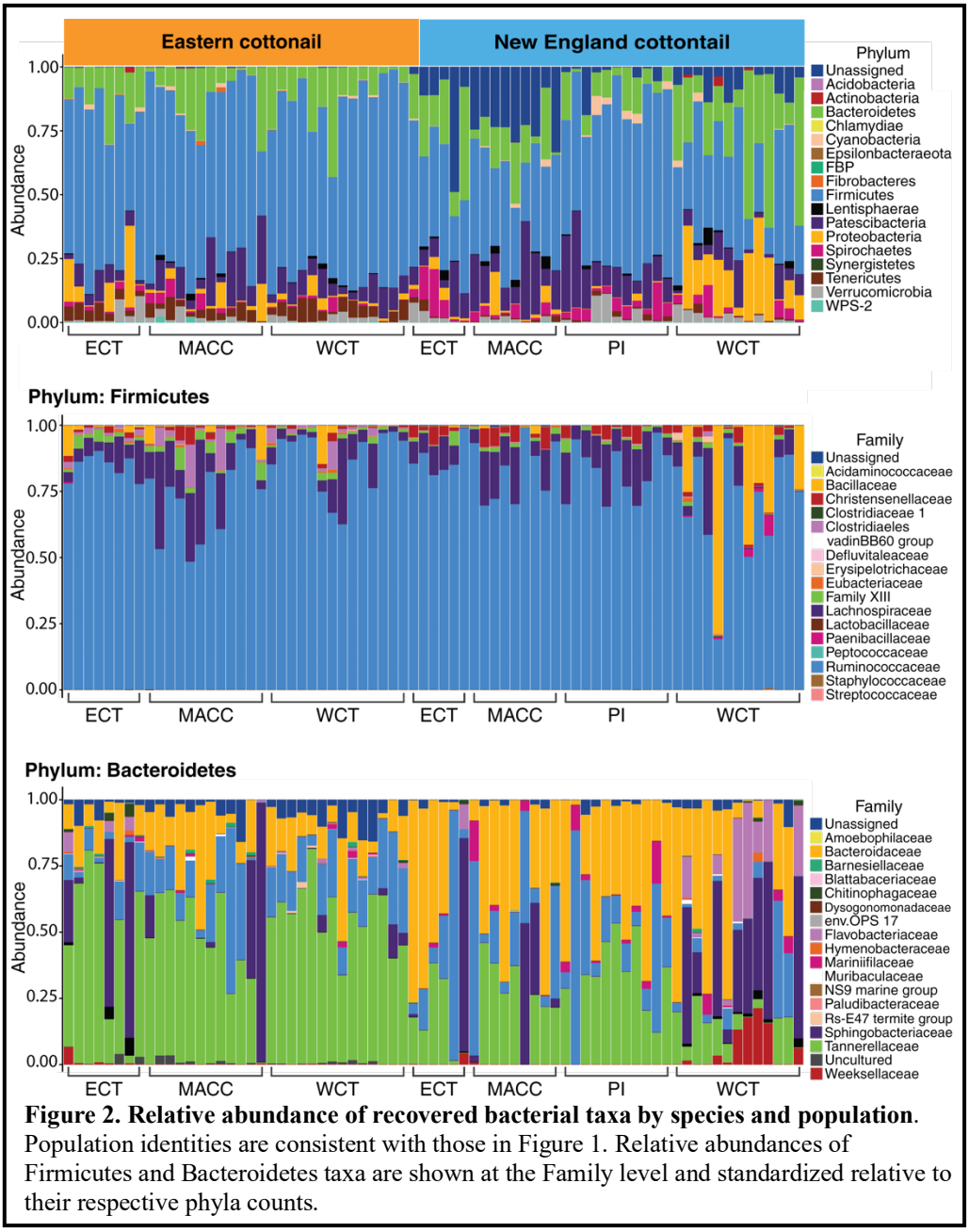

Figure 2. Relative abundance of recovered bacterial taxa by species and population. Population identities are consistent with those in Figure 1. Relative abundances of Firmicutes and Bacteroidetes taxa are shown at the Family level and standardized relative to their respective phyla counts.

composed of primarily Tannerellaceae and Bacteroidaceae, followed by Patescibacteria, which consisted entirely of Saccharimonadaceae. Hereafter each species diverged in the most abundant phyla. For NEC, unassigned phyla, Proteobacteria, and Spirochaetes were the next most abundant, whereas EC were composed of Tenericutes, Proteobacteria, and Verrucomicrobia. For NEC, the phyla Verrucomicrobia, Lentisphaerae, Cyanobacteria, Tenericutes, Actinobacteria, WPS-2, Synergistetes, Acidobacteria, and Chlamydiae were also detected at low levels. EC

additionally displayed these phyla, but also were also composed of low levels of FBP, Epsilonbacteraeota, and Fibrobacteres.

Cottontail microbiomes were found to differ by host species. Species clustered strongly in NMDS ordination using Bray-Curtis distances **(Figure 3)**, and species identity explained the most variation in microbial community differences according to permanova analysis (**Table 1**).

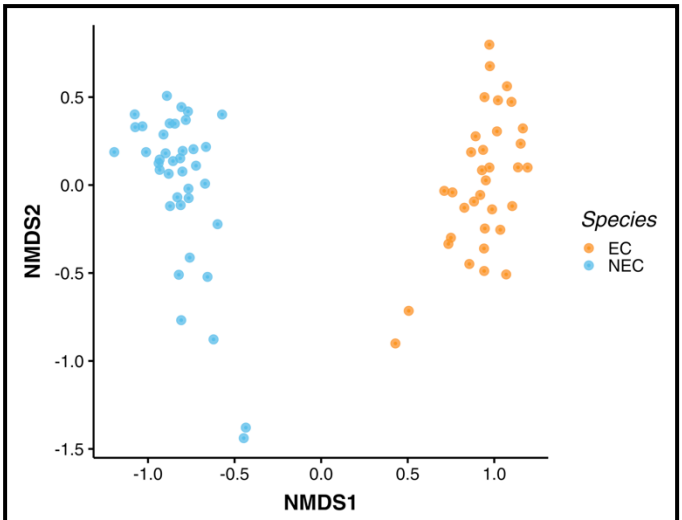

Figure 3. Non-metric multidimensional scaling (NMDS) plot by species. Circles indicate the location of individuals within microbiome space and are colored by species.

EC and NEC were each found to possess a core microbiome **(Figure 4).** For NEC, this consisted of an unassigned bacterial ASV at a mean species abundance of 5.1%, followed by *Ruminococcus* sp., *Bacteroides* sp., *Treponema* sp., *Parabacteroides* sp., *Ruminococcaceae UCG-014* sp., and multiple *Candidatus Saccharimonas* spp. For EC, core microbes included *Bacteroides* sp. at a mean abundance of 5.6%, followed by multiple *Candidatus Saccharimonas* spp., multiple *Ruminococcaceae UCG-014* spp., *Family XII AD3011 group* sp., and *Marvinbryantia* sp. Two *Candidatus Saccharimonas* spp. were shared between the two species, and all other core species were distinct ASVs. Simper analysis revealed several of the core species-specific ASVs as the drivers of compositional differences among species (**Figure 5)**, and LEfSe uncovered that core species members were also among the top 30 most differentially abundant taxa between species, as determined by LDA score (**Figure 6**).

Table 1. PERMANOVA results for Bray-Curtis dissimilarity and weighted Unifrac distance. Significant terms and *p*-values bolded.

Row	Species	Explanatory variables	Bray-Curtis			Weighted Unifrac		
			F	R^2	P	F	R^2	P
1	EC + NEC	**Species**	21.7	0.24	**0.001**	12.0	0.15	**0.001**
2	EC + NEC	**Species**	23.0	0.24	**0.001**	13.0	0.15	**0.001**
		Population	2.39	0.07	**0.001**	2.98	0.10	**0.002**
3	EC + NEC	**Species**	23.9	0.24	**0.001**	13.4	0.15	**0.001**
		Population	2.49	0.07	**0.001**	3.07	0.10	**0.001**
		Species*Population	2.34	0.05	**0.001**	2.05	0.04	**0.026**
4	EC + NEC	**Species**	22.3	0.24	**0.001**	12.6	0.15	**0.001**
		Latitude	1.18	0.01	0.202	0.69	0.01	0.608
		Longitude	2.70	0.03	**0.003**	5.21	0.06	**0.003**
5	EC	**Population**	1.91	0.11	**0.001**	2.12	0.12	**0.011**
		ECT vs. MACC	1.97	0.10	**0.021**	2.13	0.11	0.114
		ECT vs. WCT	1.95	0.09	**0.006**	2.02	0.09	**0.015**
		MACC vs. WCT	1.84	0.07	**0.006**	2.17	0.08	0.108
6	EC	**Population**	1.93	0.11	**0.001**	2.14	0.12	**0.004**
		Sampling Month	1.11	0.09	0.160	1.10	0.09	0.316
7	EC	Latitude	1.16	0.03	0.174	1.07	0.03	0.330
		Longitude	1.88	0.06	**0.004**	2.31	0.07	**0.017**
8	NEC	**Population**	3.01	0.21	**0.001**	2.34	0.17	**0.001**
		ECT vs. MACC	1.44	0.10	0.882	1.40	0.08	1.000
		ECT vs. PI	2.26	0.14	**0.006**	1.93	0.16	**0.030**
		ECT vs. WCT	1.71	0.09	0.462	2.86	0.08	0.696
		MACC vs. PI	3.95	0.19	**0.006**	1.26	0.18	**0.006**
		MACC vs. WCT	4.03	0.17	**0.006**	2.61	0.09	0.102
		PI vs. WCT	3.92	0.19	**0.006**	3.70	0.12	**0.006**
9	NEC	**Population**	3.06	0.21	**0.001**	2.37	0.17	**0.001**
		Sampling Month	1.20	0.08	0.156	1.16	0.08	0.245
10	NEC	Latitude	1.16	0.03	0.276	1.64	0.04	0.061
		Longitude	4.49	0.11	**0.001**	2.00	0.05	**0.010**
11	NEC	**Wild vs. Reintroduced**	3.42	0.09	**0.001**	3.85	0.10	**0.001**

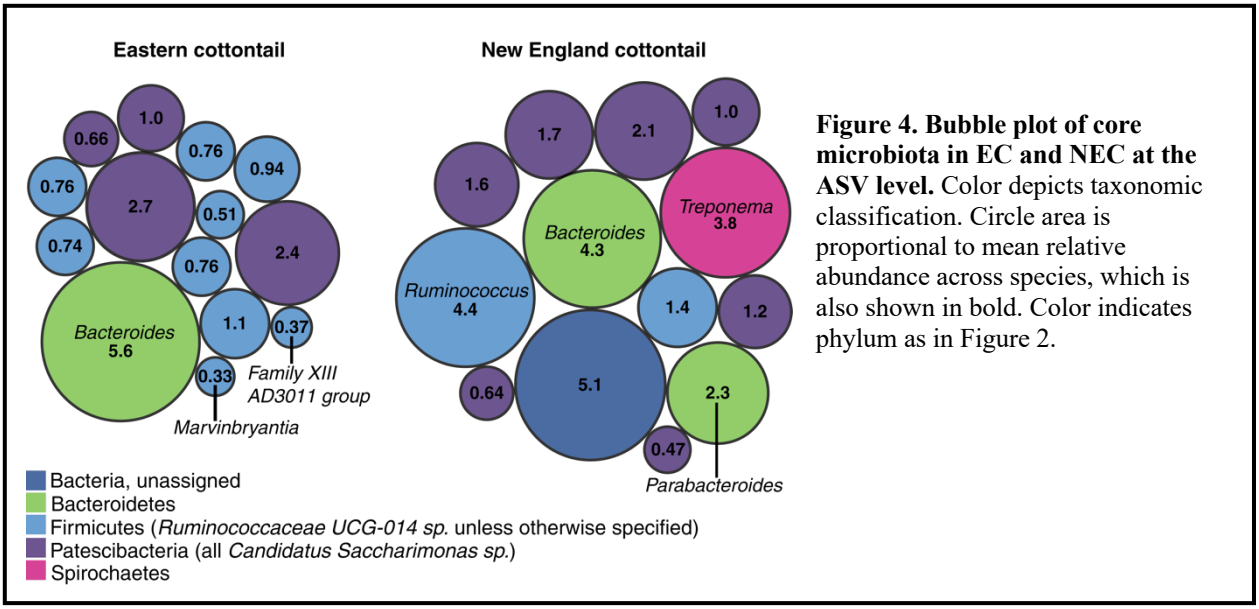

Figure 4. Bubble plot of core microbiota in EC and NEC at the ASV level. Color depicts taxonomic classification. Circle area is proportional to mean relative abundance across species, which is also shown in bold. Color indicates phylum as in Figure 2.

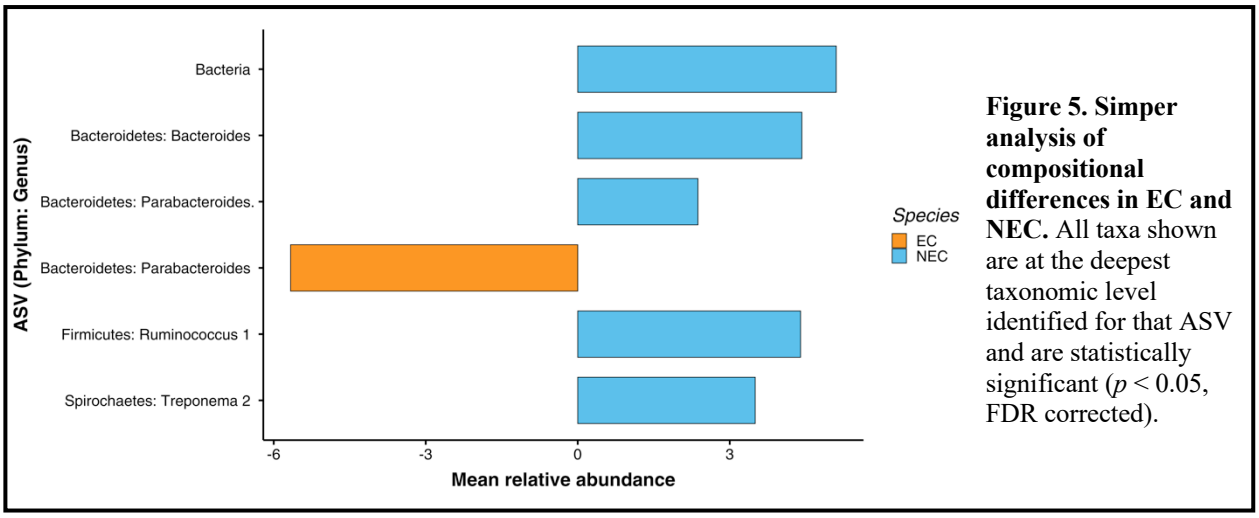

Figure 5. Simper analysis of compositional differences in EC and NEC. All taxa shown are at the deepest taxonomic level identified for that ASV and are statistically significant ($p < 0.05$, FDR corrected).

Alpha diversity estimates also significantly differed between species, with EC demonstrating greater alpha diversity across the number of observed ASVs (mean ASVs ± SE: EC, 227.1 ± 10.6; NEC, 152.3 ± 6.5; Kruskal-Wallis rank sum test $p < 0.0001$), Simpson diversity (mean Simpson ± SE: EC, 0.970 ± 0.003; NEC, 0.955 ± 0.004; Kruskal-Wallis rank sum test $p < 0.0001$), and Shannon diversity (mean Shannon ± SE: EC, 4.43 ± 0.065; NEC, 3.97 ± 0.060; Kruskal-Wallis rank sum test $p < 0.0001$) (**Figure 7**).

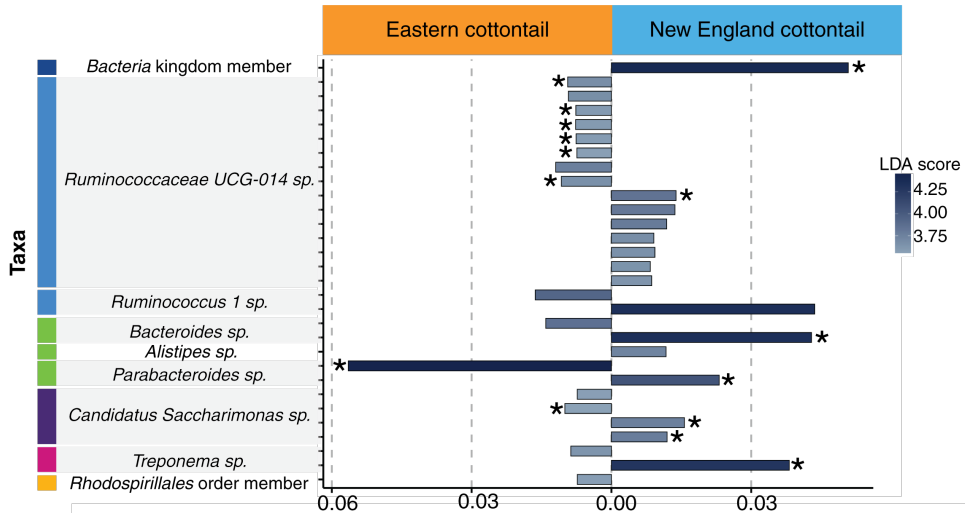

Figure 6. LEfSe analysis between EC and NEC. Bars depict the 30 most differentially abundant taxa between EC and NEC, depicted at the deepest taxonomic level identified. Phylum is indicated by color as in Figure 2. Bars are filled by LDA score, and height reflects the mean relative abundance of a given ASV across a given species. Asterisk indicates ASV is a member of the species core microbiome.

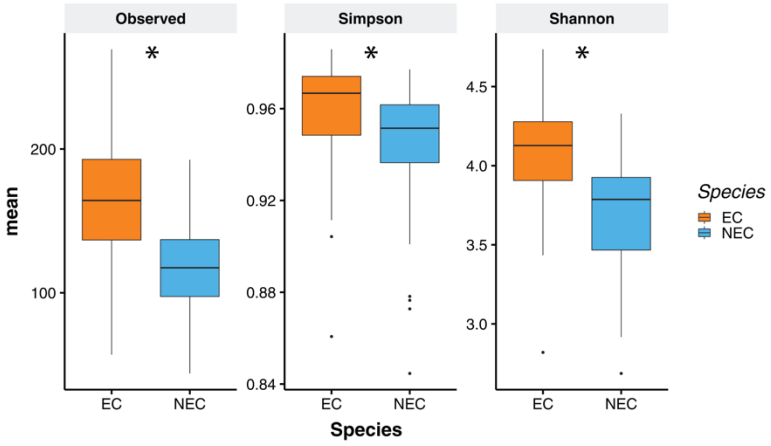

Figure 7. Alpha diversity estimates between species. Data were rarified to 5,000 ASVs per sample. Boxplot depicts the median and interquartile range. All species comparisons were significant ($p < 0.0001$).

Within species, fecal alpha diversity did not differ between EC or NEC populations, sampling months, or between reintroduced and wild NEC populations (**Figures S4 & S5**). Fecal microbiome composition structured loosely by geography in both species. PCoA plots of Bray-Curtis distance matrices demonstrate loose clustering by population with significant overlap between populations (**Figure S6**). Population identity explained the majority of intraspecific compositional variation for both EC and NEC relative to other factors (EC: $R^2 = 0.110$, $p = 0.001$; NEC: $R^2 = 0.210$, $p = 0.001$; **Table 1**). All three EC populations significantly differed

both globally and in pairwise analyses corrected for multiple comparisons (**Table 1**). According to Simper analysis, compositional differences between populations were primarily driven by a *Ruminococcus 1 sp.* (Firmicutes) and *Massilia sp.* (Proteobacteria), although following FDR correction these were no longer significant (p = 0.081 and p = 0.31, respectively). Differences in enriched taxa between EC populations were noted for ASVs primarily of the phylum Firmicutes, followed by single observations of an unassigned bacterial ASV, Spirochaetes, Actinobacteria, and Bacteroidetes (**Figure 8A**).

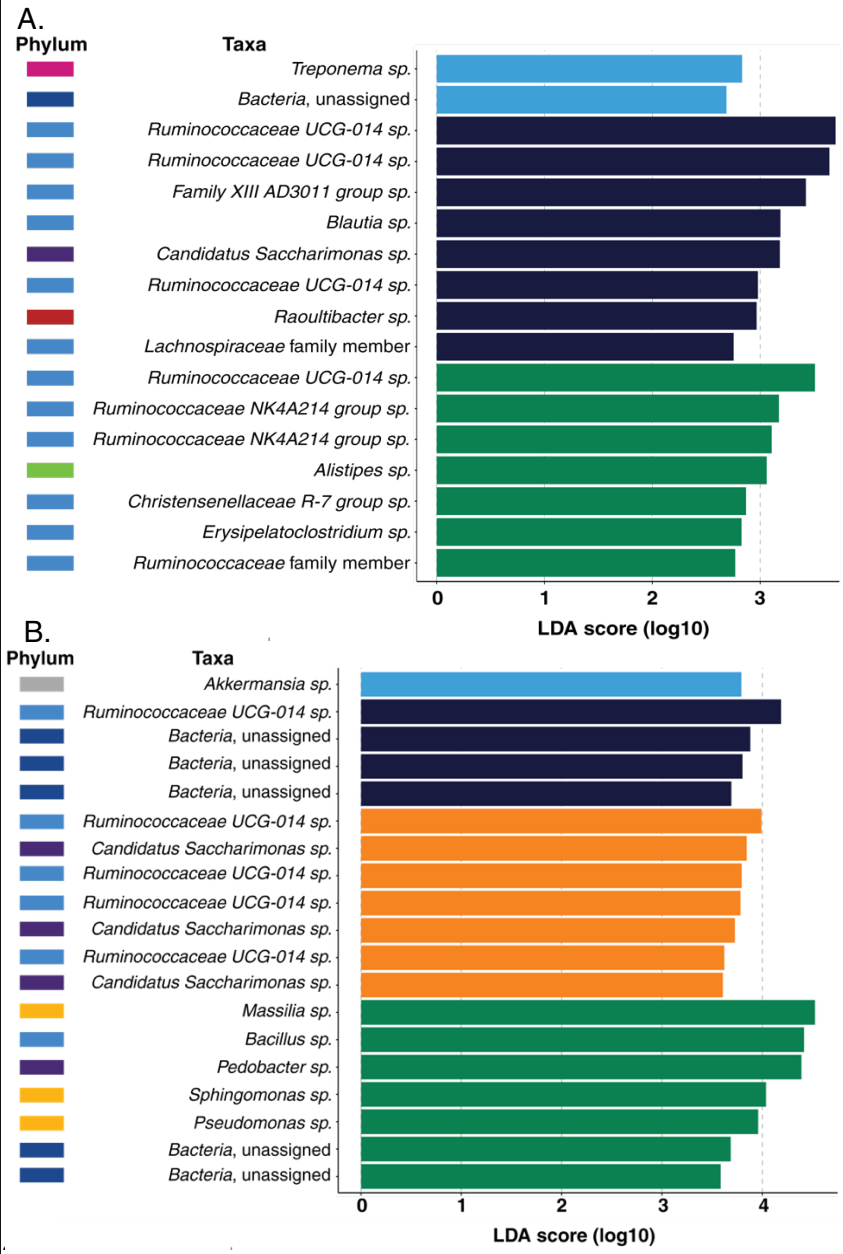

Figure 8. LEfSe analysis between EC (A) and NEC (B) populations. Bars depict top differentially abundant taxa between populations, depicted at the deepest taxonomic level identified. Light blue = ECT, teal = WCT, navy = MACC, and orange = PI. Phylum is indicated by color as in Figure 2.

NEC populations also differed significantly according to permanova testing, and pairwise analyses revealed that differences between the

reintroduced PI population and each wild population were primarily driving this signal, along with a significant difference between the composition of the WCT and MACC populations (**Table 1**). Simper analysis revealed 11 bacterial taxa as the drivers of compositional differences between populations (**Figure 9**). Many of these taxa overlapped with those identified by LEfSe as differentially enriched within populations (**Figure 8B**). Encompassed taxa primarily were from the phylum Firmicutes, followed by unassigned bacterial ASVs, Patescibacteria, and Proteobacteria, and a single Verrucomicrobia species.

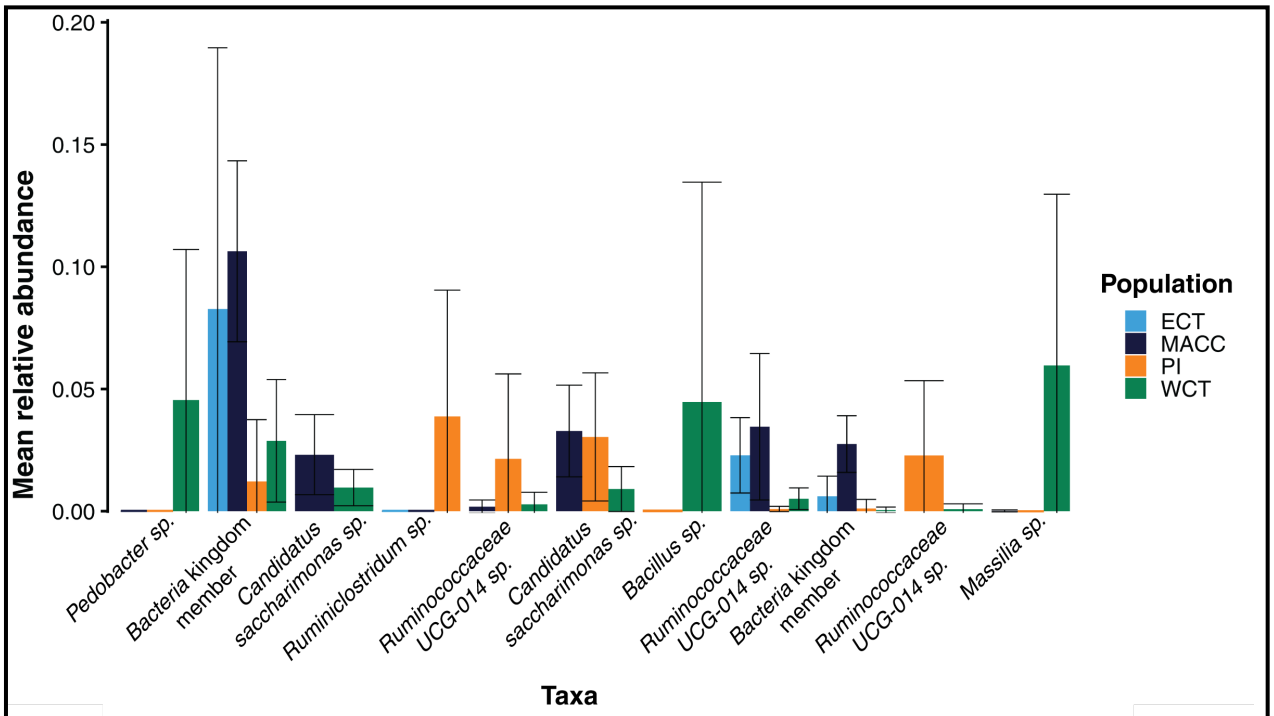

Figure 9. Simper analysis for NEC population differences. Taxa are shown if identified as significantly different between populations shown following FDR correction. Populations not displayed for a given taxa were not statistically different from any population at that ASV. Bars indicate standard deviation in mean relative abundance.

For both EC and NEC, geographic distance between individuals was significantly but weakly correlated with Bray-Curtis dissimilarity (EC: *Mantel* $r = 0.147$, $p < 0.001$; NEC: *Mantel* $r = 0.1954$, $p < 0.001$). At the population level, EC microbiome distances were equally similar between sympatric and allopatric individuals ($p = 0.816$), whereas NEC were significantly more

similar within populations than between allopatric conspecific groups (p = 0.006; **Figure 10**). Within groups, EC also displayed significantly greater dissimilarity compared to NEC within group values (p = 0.015). For both species, intraspecific differences were significantly lower when compared to differences between either sympatric (EC: p = 0.00; NEC: p = 0.00) or allopatric congeners (EC: p = 0.00; NEC: p = 0.00)

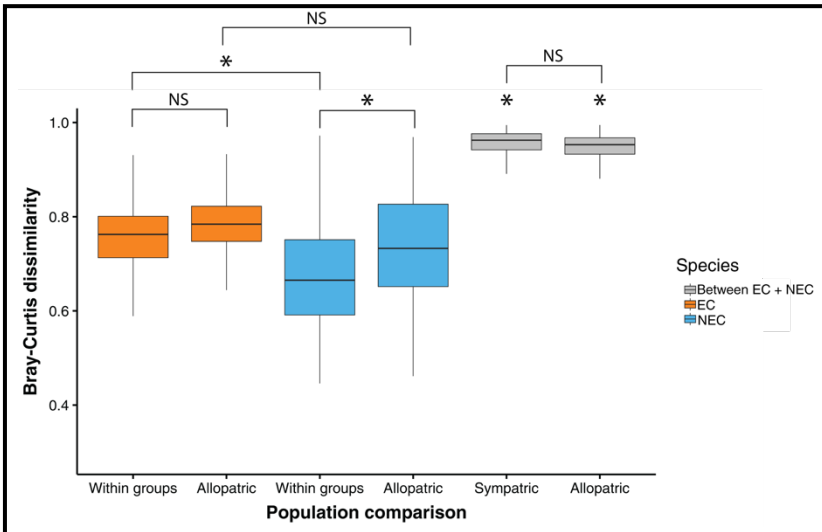

Figure 10. Microbiome distance as a function of species and geography. Bray-Curtis dissimilarity stratified according to 1) within-population conspecifics ("Within groups", orange/blue), 2) allopatric conspecifics ("Allopatric", orange/blue) 3) Sympatric congeners ("Sympatric", gray), and 4) Allopatric congeners ("Allopatric", gray). Asterisks indicate significant comparisons at p<0.01, whereas NS indicates non-significant comparisons (p > 0.05).

PICRUSt analysis uncovered 26 functional pathways that were differentially abundant between EC and NEC (**Figure 11**). These pathways were primarily composed of a broad biosynthesis function, followed by generation of precursor metabolites and energy, then degradation/utilization/assimilation, and a single pathway involved in macromolecule modification.

Discussion

Microbiome composition is consistent with other Leporidae

This study focuses on the fecal microbiota of an imperiled mammal and its invasive congener across native and reintroduced populations, and is additionally the first study to investigate microbiota in any member of the genus *Sylvilagus*. EC and NEC displayed different

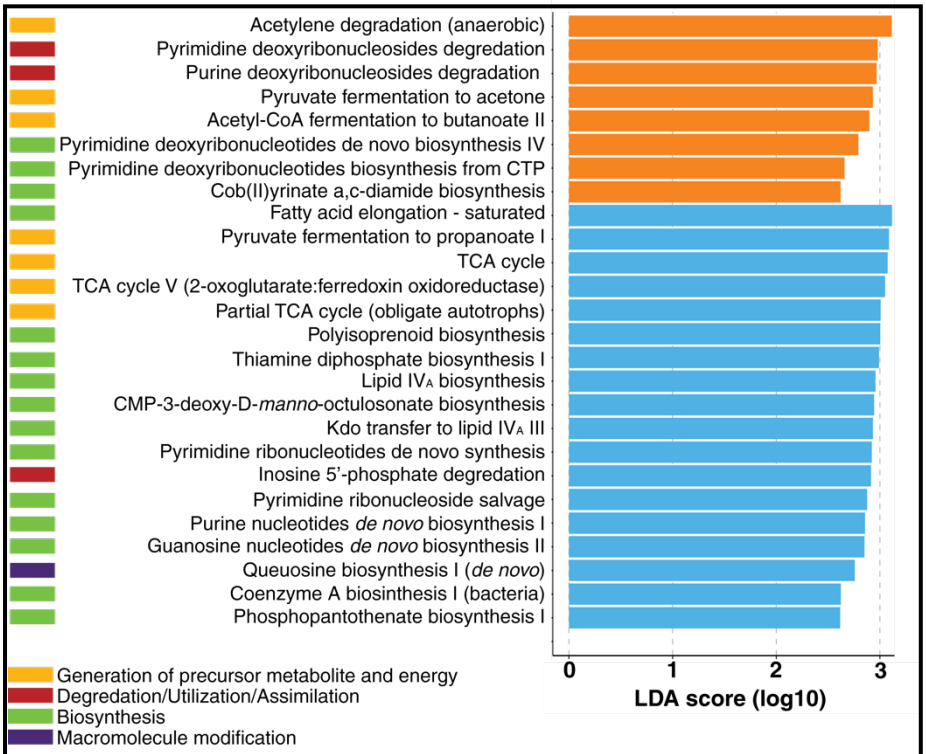

Figure 11. LEfSe analysis of PICRUSt functional data in EC and NEC. Bars depict differentially abundant functions between species based on PICRUSt annotation. Broader functional category indicated by color.

microbiomes, but both species' microbiota were consistent with what has been reported for other Leporidae. The domination of Firmicutes and Bacteroidetes is consistent with other studies in European rabbits (*Oryctolagus cuniculus*) and European hares (*Lepus europaeus*) (Bäuerl *et al.* 2014; Combes *et al.* 2014; Crowley *et al.* 2017; Kylie *et al.* 2018; Read *et al.* 2019; Shanmuganandam *et al.* 2019; Stalder *et al.* 2019). The dominance of the Firmicutes families Ruminococcaceae and Lachnospiraceae is consistent with folivory and herbivory, as both families contain numerous pathways designed to degrade complex plant material, including otherwise indigestible cellulose and hemicellulose plant products (Brulc *et al.* 2009; Biddle *et al.* 2013). Additionally, during fermentation of fibers and sugars, Ruminococcaceae produce the product butyrate (Louis *et al.* 2010), which is a source of energy for gut barrier epithelium (Rivière *et al.* 2016) and may have an anti-inflammatory effect (Le Poul *et al.* 2003; Kasubuchi *et al.* 2015).

Bacteroidetes was the next most abundant phylum for both EC and NEC and was comprised primarily of the families Tannerellaceae and Bacteroidaceae, although some samples were dominated by Sphingobacteriaceae (**Figure 2**). This is consistent with the findings of Shanmuganandam *et al.* 2019 who reported that both European rabbits and hares shared Tannerellaceae as high abundance Bacteroidetes families.

Patescibacteria was found in high prevalence and abundance across EC and NEC samples (**Figure 2**) and has been detected in Leporidae fecal microbiota previously (Combes *et al.* 2014; Shanmuganandam *et al.* 2019). Patescibacteria has also been reported at high abundance in the prairie vole microbiome (Curtis *et al.* 2018), but its functional role is unclear. Patescibacteria can ferment sugars (Albertsen *et al.* 2013), and *Candidatus Saccharimonas* spp. may be involved in pathways that support energy metabolism (Ogunade *et al.* 2019). However, the phylum has also been associated with various diseases, including Crohn's disease, colitis, and periodontitis (Brinig *et al.* 2003; Kuehbacher *et al.* 2008; Elinav *et al.* 2011).

Proteobacteria and Spirochaetes have also been reported in Leporidae before, both achieving relative abundances as high as seven percent (Combes *et al.* 2014; Crowley *et al.* 2017; Stalder *et al.* 2019). Other phyla detected at low abundance were also previously detected in Leporidae fecal microbiomes, including Verrucomicrobia, Tenericutes, Actinobacteria, Fibrobacteres, and Synergistes (Combes *et al.* 2014; Crowley *et al.* 2017).

EC and NEC microbiomes differ by host species

EC microbiome is characterized by a greater number of observed ASVs and greater diversity metrics relative to NEC (**Figure 7 & Figure 10**). NEC are habitat specialists restricted to primarily early successional habitat, whereas EC are more ecologically flexible and are found in a variety of habitats and land types (Probert & Litvaitis 1996; Smith & Litvaitis 2000; Litvaitis

et al. 2007; Cheeseman *et al.* 2018). Therefore, differential resource selection between the two species may contribute to observed microbiome divergence. While knowledge of EC and NEC foraging is limited (Dalke & Sime 1941), it is possible that EC enjoy a more varied diet or more varied nutritional needs, and that this is reflected in their fecal microbiota. New England's invasive EC populations span multiple subspecies that were introduced and subsequently admixed (Litvaitis *et al.* 2007), whereas NEC populations are genetically isolated and depauperate (Fenderson *et al.* 2011). A more genetically diverse host species may also produce greater variety in host-associated microbial outcomes or maternally-transmitted microbial lineages.

EC and NEC each possessed a distinct core microbiome (**Figure 4**). While two Patescibacteria ASVs were shared between the two species, other shared taxa represented distinct ASVs. For example, while *Bacteroides* and *Ruminococcaceae UCG-014* were detected in both cottontails, they represented different ASVs that were not shared between the two species. *Ruminococcaceae UCG-014* has been reported in healthy European rabbits (Jin *et al.* 2018), and multiple *Bacteroides* spp. contribute to the core microbiome in the European hare (Stalder *et al.* 2019). However, the functional role of *Bacteroides* is less clear. Zeng *et al.* 2015 reported that *Bacteroides* were enriched in healthy, high weight rabbits, whereas Bäuerl *et al.* 2014 and Jin *et al.* 2018 found that *Bacteroides* abundance was associated with the severe rabbit pathology Epizootic Rabbit Enteropathy (ERE).

Within the EC core microbiome, Family XIII AD3011 group was detected, which has also been identified in other herbivores' microbiomes (Haworth *et al.* 2019; Suzuki *et al.* 2019). Although the functional role is currently unknown, taxonomic members of the Family XIII have been detected in domesticated rabbits and correlate with both diet changes and somatotropin

growth hormone levels (North *et al.* 2019). Lastly, *Marvinbryantia* has previously been reported in rabbit fecal microbiomes (Kylie *et al.* 2018; Wang *et al.* 2019), and Kylie *et al.* 2018 found that the genus was significantly enriched in the winter relative to summer months.

Within the NEC core microbiome, the most abundant ASV consisted of an unassigned bacterial taxon, followed by *Ruminococcus. Ruminococcus* spp. can occupy key roles in degrading cellulose (Julliand *et al.* 1999; Flint *et al.* 2008) and have been detected previously among both rabbit and hare fecal microbiota (Bäuerl *et al.* 2014; Fang *et al.* 2019; Stalder *et al.* 2019). *Parabacteroides* was also detected in the NEC core microbiome and was previously reported as one of the top 50 most abundant operational taxonomic units (OTUs) in the European hare (Stalder *et al.* 2019). Interestingly, *Treponema* was detected as a core microbiota member of NEC. In termites, *Treponema* spp. may assist with lignin and xylan degradation (Warnecke *et al.* 2007), and it has been suggested that *Treponema* spp. may enhance polysaccharide extraction from high-fiber plants or vegetables (De Filippo *et al.* 2010; Schnorr *et al.* 2014). *Treponema* is also responsible for the sexually-transmitted disease syphilis, including two virulent forms specific to Leporidae, *T. paraluiscuniculi* and *T. paraluisleporis* (Lumeij *et al.* 2013). When *Treponema* ASVs recovered in this study are mapped into a phylogenetic tree (**Figure S7**), they cluster primarily in their own clade and are more closely related to non-pathogenic strains. Therefore, it is likely that the *Treponema* spp. recovered in this study are non-pathogenic.

Notably, the species-specific core microbiota and species-specific ASVs of shared genera were driving observed differences between EC and NEC microbiome according to both Simper analysis (**Figure 5**) and LEfSe (**Figure 6**). Perpetuation of species-specific ASVs could indicate that these particular strains provide an adaptive advantage to their respective hosts and consequently have been maintained by species-specific selection, or it is possible that ecological

drift or chains of maternal inheritance propagate these observed differences (Lankau *et al.* 2012; Funkhouser & Bordenstein 2013).

Compositional and abundance differences between the two species also translate into differential functions, as identified by PICRUSt and LEfSe (**Figure 11**). EC were enriched for fewer PICRUSt categories relative to NEC, and differences primarily fell under the category of biosynthesis, followed by generation of precursor metabolite and energy.

Within species, fecal microbiomes structure by geography

Intraspecific microbiome variation was primarily driven by biogeographic population variation for both EC and NEC (**Table 1**). EC populations were enriched for different taxa that primarily hailed from the Firmicutes phylum and/or were from genera naturally found at low abundances (**Figure 8A**). *Treponema*, *Ruminococcaceae UCG-014*, the *Family XIII AD3011 group*, and *Candidatus Saccharimonas* were found to differ in abundance among populations and were previously described as common members of Leporidae fecal microbiota with key functional roles. In addition to these taxa, *Blautia* was also detected as enriched in MACC EC and has been detected in rabbit and hare fecal microbiomes (Fang *et al.* 2019; Stalder *et al.* 2019). Its absence associates with ERE (Jin *et al.* 2018), and *Blautia* has been inversely correlated with other disease, including *Clostridium difficile* infection (Shankar *et al.* 2014) and acute hemorrhagic disease (Suchodolski *et al.* 2012). *Alistipes sp.*, *Ruminococcaceae KN4A214 group spp.*, and *Erysipelatoclostridium sp*. have also been detected previously in the fecal microbiome of rabbits or hares (Bäuerl *et al.* 2014; Zhu *et al.* 2015; Jin *et al.* 2018; Fang *et al.* 2019; Read *et al.* 2019; Shanmuganandam *et al.* 2019), but any functional implications are unclear. Jin *et al.* 2018 reported that lower *Ruminococcaceae KN4A214 group* abundances were found in rabbits with ERE, whereas Read *et al.* 2019 found it to be one of the dominant

Bacteroidetes genera present. *Erysipelatoclostridium* may also associate with low weaning weights in commercial rabbits (Fang *et al.* 2019).

NEC populations also structured loosely by geography but not every population significantly differed (**Table 1**). It is possible that this is due to low sample sizes in the ECT population, as the power analysis conducted was executed to discern species-level differences whereas population variation is expected to be more subtle. Permanova uncovered significant differences between the reintroduced PI population and all wild populations, as well as between the two most geographically distant populations, WCT and MACC (**Table 1**). NEC-specific Simper analysis (**Figure 9**) and LEfSe (**Figure 8B**) identified multiple taxa in the PI and WCT populations as drivers of compositional differences or differential enrichment. *Pedobacter*, *Sphingomonas*, *Massilia*, and *Pseudomonas* were drivers of WCT differences relative to other populations, and all three species are commonly reported as microbiome contaminants associated with water, soil, or the laboratory environment (Gallego *et al.* 2006; Kwon *et al.* 2011; Salter *et al.* 2014; Weyrich *et al.* 2019). However, *Pedobacter spp*. have been reported as true members of other mammalian fecal microbiomes (Angelakis *et al.* 2016; Finlayson-Trick *et al.* 2017) and is found at increased abundances in wild mammals (Alessandri *et al.* 2019). *Sphingomonas*, *Massilia*, and *Pseudomonas* have also been reported in rabbit and/or hare fecal microbiota (Jin *et al.* 2018; Stalder *et al.* 2019; Yuan *et al.* 2019) and in other mammalian microbiomes (Angelakis *et al.* 2016; Finlayson-Trick *et al.* 2017; Zhu *et al.* 2018). *Bacillus* was also identified as a driver of WCT population divergence and is associated with increased growth, digestive capabilities, and immune function in commercial rabbits (Wang *et al.* 2020). Indeed, *Bacillus spp*. are frequently proposed as a healthy probiotic option (Urdaci *et al.* 2004; Riddell *et al.* 2010).

Geographic distance correlated with individual microbial distance for both species, as

reported by the significant Mantel tests. Mantel's *r* was greater for NEC ($r = 0.1954$, EC: $r = 0.147$), but both species' values were relatively low. Therefore, while microbial community structure varies across populations, dependence on geographic distance was weak. At the population level, different patterns emerged between EC and NEC. EC microbiome distances were equally similar within and between conspecific populations, whereas NEC were significantly more similar in composition if from the same population than if from an allopatric site (**Figure 10**). Within-population distance was also significantly greater for EC than NEC. These patterns could arise due to differing population structure between the species; whereby EC are continuous across much of New England's landscape, NEC are geographically isolated to distinct remnant populations. If EC enjoy a more diverse diet or greater breadth of habitats, it is possible that high microbiome distances are driven by diets or environmental exposures that are highly varied both within and across populations. For both species, microbiome composition was more similar between conspecifics than congeners, regardless of whether comparisons were made sympatrically or allopatrically.

Conservation Implications

The world is experiencing rapid anthropogenic-driven change, and incorporating microbiome research into the conservation toolkit is widely urged (Redford *et al.* 2012; Amato 2013; Bahrndorff *et al.* 2016; Trevelline *et al.* 2019). This study is one of the first to investigate the microbiome across an imperiled native species and its closely related, successful invader. Since NEC are part of an active and ongoing reintroduction program, it is also important to consider the role of microbiota on population-scales. Here, we uncover population-level differences that could have functional consequences if mismatches between individuals' microbiome and the environment are maladaptive. Alberdi *et al.* 2016 suggests that host-

associated microbiome may source opportunities for rapid local adaptation, and therefore conservation studies should also consider preserving host-associated microbes in addition to genetic diversity. Relatedly, the highly diverse microbiome found in EC may have aided in its successful establishment and expansion throughout New England. Taxa commonly associated with rabbit disease include *Escherichia coli*, *Clostridium spiriforme*, *C. piliforme*, and *Lawsonia intracellularis* (Peeters *et al.* 1984, 1986), and were not detected in either EC or NEC. This study can provide valuable baseline information in the event of any future diarrheal or pathogen outbreak. Indeed, routine pathogen monitoring is regularly called for by conservation experts yet rarely implemented in practice (source). Additionally, the *Treponema* species recovered are likely non-pathogenic due to their high prevalence and clustering with other non-pathogenic species (**Figure S7**), yet this warrants further study.

Acknowledgements

We would like to thank collaborators who helped with EC and NEC specimen procurement: New Hampshire Fish and Game, Massachusetts Department of Fish and Game, Connecticut Department of Energy and Environmental Protection, Rhode Island Department of Environmental Management, Roger Williams Park Zoo, and the Queens Zoo. Lastly, we thank Janet Atoyan of the URI Genomics Core for her assistance with Illumina MiSeq library preparation and sequencing. This research was supported by computational resources at Brown University's Center for Computation and Visualization. K.B.N. was supported by the NSF Graduate Research Fellowship Program (GRFP) and NSF IGERT Grant.

References

Alberdi A, Aizpurua O, Bohmann K, Zepeda-Mendoza ML, Gilbert MTP (2016) Do Vertebrate Gut Metagenomes Confer Rapid Ecological Adaptation? *Trends in Ecology and Evolution*,

31, 689–699.

Albertsen M, Hugenholtz P, Skarshewski A *et al.* (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology*, **31**, 533–538.

Alessandri G, Milani C, Mancabelli L *et al.* (2019) The impact of human-facilitated selection on the gut microbiota of domesticated mammals. *FEMS Microbiology Ecology*, **95**, 1–13.

Amato KR (2013) Co-evolution in context: The importance of studying gut microbiomes in wild animals. *Microbiome Science and Medicine*, **1**, 9–29.

Amato KR, Yeoman CJ, Kent A *et al.* (2013) Habitat degradation impacts black howler monkey (Alouatta pigra) gastrointestinal microbiomes. *The ISME Journal*, **7**, 1344–1353.

Angelakis E, Yasir M, Bachar D *et al.* (2016) Gut microbiome and dietary patterns in different Saudi populations and monkeys. *Scientific Reports*, **6**, 1–9.

Antwis RE, Lea JMD, Unwin B, Shultz S (2018) Gut microbiome composition is associated with spatial structuring and social interactions in semi-feral Welsh Mountain ponies. *Microbiome*, **6**, 207.

Ayres JC, Feemster RF (1948) Epidemiology of tularemia in Massachusetts with a review of the literature. *New England Journal of Medicine Journal of Medicine*, **238**, 187–194.

Bahrndorff S, Alemu T, Alemneh T, Lund Nielsen J (2016) The Microbiome of Animals: Implications for Conservation Biology. *International Journal of Genomics*, **2016**.

Baldo L, Riera JL, Mitsi K, Pretus JL (2018) Processes shaping gut microbiota diversity in allopatric populations of the endemic lizard Podarcis lilfordi from Menorcan islets (Balearic Islands). *FEMS microbiology ecology*, **94**, 1–14.

Barelli C, Albanese D, Donati C *et al.* (2015) Habitat fragmentation is associated to gut

microbiota diversity of an endangered primate: implications for conservation. *Scientific Reports*, **5**, 14862.

Bäuerl C, Collado MC, Zúñiga M, Blas E, Martínez GP (2014) Changes in cecal microbiota and mucosal gene expression revealed new aspects of epizootic rabbit enteropathy. *PLoS ONE*, **9**.

Becker CG, Longo A V., Haddad CFB, Zamudio KR (2017) Land cover and forest connectivity alter the interactions among host, pathogen and skin microbiome. *Proceedings of the Royal Society B: Biological Sciences*, **284**.

Biddle A, Stewart L, Blanchard J, Leschine S (2013) Untangling the genetic basis of fibrolytic specialization by lachnospiraceae and ruminococcaceae in diverse gut communities. *Diversity*, **5**, 627–640.

Bolnick DI, Snowberg LK, Caporaso JG, Lauber C (2015) Major Histocompatibility Complex class IIb polymorphism influences gut microbiota composition and diversity. *Molecular Ecology*, **23**, 4831–4845.

Bolnick DI, Snowberg LK, Hirsch PE *et al.* (2014) Individual diet has sex-dependent effects on vertebrate gut microbiota. *Nature Communications*, **5**.

Brinig MM, Lepp PW, Ouverney CC, Armitage GC, Relman DA (2003) Prevalence of bacteria of division TM7 in human subgingival plaque and their association with disease. *Applied and Environmental Microbiology*, **69**, 1687–1694.

Brulc JM, Antonopoulos DA, Berg Miller ME *et al.* (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 1948–1953.

Caporaso JG, Lauber CL, Walters W a *et al.* (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, **6**, 1621–1624.

Cariveau DP, Powell JE, Koch H, Winfree R, Moran NA (2014) Variation in gut microbial communities and its association with pathogen infection in wild bumble bees (Bombus). *The ISME Journal*, **8**, 2369–2379.

Chapman TR, Morgan RP (1973) Allen Press. *Wildlife Monographs*, **36**, 3–54.

Cheeseman AE, Ryan SJ, Whipps CM, Cohen JB (2018) Competition alters seasonal resource selection and promotes use of invasive shrubs by an imperiled native cottontail. *Ecology and Evolution*, **8**, 11122–11133.

Colston TJ, Jackson CR (2016) Microbiome evolution along divergent branches of the vertebrate tree of life: what is known and unknown. *Molecular Ecology*, **25**, 3776–3800.

Combes S, Gidenne T, Cauquil L, Bouchez O, Fortun-Lamothe L (2014) Coprophagous behavior of rabbit pups affects implantation of cecal microbiota and health status. *Journal of Animal Science*, **92**, 652–665.

Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA (2012) The application of ecological theory toward an understanding of the human microbiome. *Science*, **336**, 1255–1262.

Crowley EJ, King JM, Wilkinson T *et al.* (2017) Comparison of the microbial population in rabbits and guinea pigs by next generation sequencing. *PLoS ONE*, **12**, 1–14.

Curtis JT, Assefa S, Francis A, Köhler GA (2018) Fecal microbiota in the female prairie vole (Microtus ochrogaster). *PLoS ONE*, **13**, 1–22.

Dalke PD, Sime PR (1941) Food Habits of the Eastern and New England Cottontails. *The Journal of Wildlife Management*, **5**, 216–228.

Duncan SH, Scott KP, Ramsay AG *et al.* (2003) Effects of alternative dietary substrates on competition between human colonic bacteria in an anaerobic fermentor system. *Applied and Environmental Microbiology*, **69**, 1136–1142.

Easson CG, Chaves-Fonnegra A, Thacker RW, Lopez J V. (2020) Host population genetics and biogeography structure the microbiome of the sponge Cliona delitrix. *Ecology and Evolution*, 2007–2020.

Elinav E, Strowig T, Kau AL *et al.* (2011) NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell*, **145**, 745–757.

Fallani M, Young D, Scott J *et al.* (2010) Intestinal microbiota of 6-week-old infants across Europe: Geographic influence beyond delivery mode, breast-feeding, and antibiotics. *Journal of Pediatric Gastroenterology and Nutrition*, **51**, 77–84.

Fang S, Chen X, Zhou L *et al.* (2019) Faecal microbiota and functional capacity associated with weaning weight in meat rabbits. *Microbial Biotechnology*, **12**, 1441–1452.

Fenderson LE, Kovach AI, Litvaitis J a., Litvaitis MK (2011) Population genetic structure and history of fragmented remnant populations of the New England cottontail (Sylvilagus transitionalis). *Conservation Genetics*, **12**, 943–958.

De Filippo C, Cavalieri D, Di Paola M *et al.* (2010) Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 14691–14696.

Finlayson-Trick ECL, Getz LJ, Slaine PD *et al.* (2017) Taxonomic differences of gut microbiomes drive cellulolytic enzymatic potential within hind-gut fermenting mammals. *PLoS ONE*, **12**, 1–22.

Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA (2008) Polysaccharide utilization by gut

bacteria: Potential for new insights from genomic analysis. *Nature Reviews Microbiology*, **6**, 121–131.

Funkhouser LJ, Bordenstein SR (2013) Mom Knows Best: The Universality of Maternal Microbial Transmission. *PLoS Biology*, **11**, 1–9.

Gallego V, Sánchez-Porro C, García MT, Ventosa A (2006) Massilia aurea sp. nov., isolated from drinking water. *International Journal of Systematic and Evolutionary Microbiology*, **56**, 2449–2453.

Griffiths SM, Antwis RE, Lenzi L *et al.* (2019) Host genetics and geography influence microbiome composition in the sponge Ircinia campana. *Journal of Animal Ecology*, **88**, 1684–1695.

Haworth SE, White KS, Côté SD, Shafer ABA (2019) Space, time and captivity: Quantifying the factors influencing the fecal microbiome of an alpine ungulate. *FEMS Microbiology Ecology*, **95**, 1–12.

Jin DX, Zou HW, Liu SQ *et al.* (2018) The underlying microbial mechanism of epizootic rabbit enteropathy triggered by a low fiber diet. *Scientific Reports*, **8**, 1–15.

Johnston JE (1972) Identification and distribution of cottontial rabbits in southern New England. University of Connecticut.

Julliand V, De Vaux A, Millet L, Fonty G (1999) Identification of Ruminococcus flavefaciens as the predominant cellulolytic bacterial species of the equine cecum. *Applied and Environmental Microbiology*, **65**, 3738–3741.

Kasubuchi M, Hasegawa S, Hiramatsu T, Ichimura A, Kimura I (2015) Dietary gut microbial metabolites, short-chain fatty acids, and host metabolic regulation. *Nutrients*, **7**, 2839–2849.

Kinross JM, Darzi AW, Nicholson JK (2011) Gut microbiome-host interactions in health and

disease. *Genome medicine*, **3**, 14.

Kohl KD (2020) Ecological and evolutionary mechanisms underlying patterns of phylosymbiosis in host-associated microbial communities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **375**.

Kohl KD, Dearing MD (2014) Wild-caught rodents retain a majority of their natural gut microbiota upon entrance into captivity. *Environmental Microbiology Reports*, **6**, 191–195.

Kohl KD, Skopec MM, Dearing MD (2014) Captivity results in disparate loss of gut microbial diversity in closely related hosts. *Conservation Physiology*, **2**, 1–11.

Kuehbacher T, Rehman A, Lepage P *et al.* (2008) Intestinal TM7 bacterial phylogenies in active inflammatory bowel disease. *Journal of Medical Microbiology*, **57**, 1569–1576.

Kwon SW, Son JA, Kim SJ *et al.* (2011) Pedobacter rhizosphaerae sp. nov. and Pedobacter soli sp. nov., isolated from rhizosphere soil of Chinese cabbage (Brassica campestris). *International Journal of Systematic and Evolutionary Microbiology*, **61**, 2874–2879.

Kylie J, Weese JS, Turner P V. (2018) Comparison of the fecal microbiota of domestic commercial meat, laboratory, companion, and shelter rabbits (Oryctolagus cuniculi). *BMC Veterinary Research*, **14**, 1–15.

Lankau EW, Hong PY, MacKie RI (2012) Ecological drift and local exposures drive enteric bacterial community differences within species of Galápagos iguanas. *Molecular Ecology*, **21**, 1779–1788.

Ley RE, Hamady M, Lozupone C *et al.* (2008) Evolution of mammals and their gut microbes. *Science (New York, N.Y.)*, **320**, 1647–1651.

Litvaitis JA, Barbour MS, Brown AL *et al.* (2007) Testing Multiple Hypotheses to Identify Causes of the Decline of a Lagomorph Species : The New England Cottontail as a Case

Study Natural History of New England Cottontails. In: *Lagomorph Biology: Evolution, Ecology, and Conservation* (eds Alves P, Ferrand N, Hacklander K), pp. 167–185. Berling Heidelberg.

Litvaitis JA, Johnson B, Jakubas W, Morris K (2003) Distribution and habitat features associated with remnant populations of New England cottontails in Maine. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **81**, 877–887.

Litvaitis JA, Tash JP, Litvaitis MK *et al.* (2004) A range-wide survey to determine the current distribution of New England cottontails. *Wildlife Society Bulletin*, **34**, 1190–1197.

Loudon AH, Woodhams DC, Parfrey LW *et al.* (2014) Microbial community dynamics and effect of environmental microbial reservoirs on red-backed salamanders (Plethodon cinereus). *The ISME Journal*, **8**, 830–840.

Louis P, Young P, Holtrop G, Flint HJ (2010) Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl-CoA:acetate CoA-transferase gene. *Environmental Microbiology*, **12**, 304–314.

Lumeij JT, Mikalová L, Šmajs D (2013) Is there a difference between hare syphilis and rabbit syphilis? Cross infection experiments between rabbits and hares. *Veterinary Microbiology*, **164**, 190–194.

McMurdie PJ, Holmes S (2013) Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, **8**.

Mueller S, Saunier K, Hanisch C *et al.* (2006) Differences in Fecal Microbiota in Different European Study Populations in Relation to Age, Gender, and Country: a Cross-Sectional Study. *Applied and Environmental Microbiology*, **72**, 1027–1033.

Nelson TM, Rogers TL, Carlini AR, Brown M V. (2013) Diet and phylogeny shape the gut

microbiota of Antarctic seals: a comparison of wild and captive animals. *Environmental Microbiology*, **15**, 1132–1145.

North MK, Dalle Zotte A, Hoffman LC (2019) Composition of rabbit caecal microbiota and the effects of dietary quercetin supplementation and sex thereupon. *World Rabbit Science*, **27**, 185–198.

Ogunade I, Schweickart H, McCoun M, Cannon K, McManus C (2019) Integrating 16s rRNA sequencing and LC–MS-based metabolomics to evaluate the effects of live yeast on rumen function in beef cattle. *Animals*, **9**.

Peeters JE, Geeroms R, Carman RJ, Wilkins TD (1986) SIGNIFICANCE OF CLOSTRIDIUM SPIROFORME IN THE ENTERITIS-COMPLEX OF COMMERCIAL RABBITS. *Veterinary Microbiology*, **12**, 25–31.

Peeters JE, Pohl P, Charlier G (1984) Infectious agents associated with diarrhoea in commercial rabbits: A field study. *Annales de Recherches Veterinaires*, **15**, 335–340.

Le Poul E, Loison C, Struyf S *et al.* (2003) Functional characterization of human receptors for short chain fatty acids and their role in polymorphonuclear cell activation. *Journal of Biological Chemistry*, **278**, 25481–25489.

Probert BL, Litvaitis JA (1996) Behavioral interactions between invading and endemic lagomorphs: Implications for conserving a declining species. , **76**, 289–295.

Read T, Fortun-Lamothe L, Pascal G *et al.* (2019) Diversity and Co-occurrence pattern analysis of cecal microbiota establishment at the onset of solid feeding in young rabbits. *Frontiers in Microbiology*, **10**, 1–13.

Redford KH, Segre JA, Salafsky N, Martinez del Rio C, McAloose D (2012) Conservation and the Microbiome. *Conservation Biology*, **26**, 195–197.

Riddell JB, Mcleod KR, Cv SA De (2010) Addition of a Bacillus based probiotic to the diet of preruminant calves: Influence on growth, health, and blood parameters. *International Journal of Applied Research in Veterinary Medicine*, **8**, 78–85.

Rivière A, Selak M, Lantin D, Leroy F, De Vuyst L (2016) Bifidobacteria and butyrate-producing colon bacteria: Importance and strategies for their stimulation in the human gut. *Frontiers in Microbiology*, **7**.

Salter SJ, Cox MJ, Turek EM *et al.* (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, **12**, 1–12.

Schnorr SL, Candela M, Rampelli S *et al.* (2014) Gut microbiome of the Hadza hunter-gatherers. *Nature Communications*, **5**.

Shankar V, Hamilton MJ, Khoruts A *et al.* (2014) Species and genus level resolution analysis of gut microbiota in Clostridium difficile patients following fecal microbiota transplantation. *Microbiome*, **2**, 1–10.

Shanmuganandam S, Hu Y, Strive T, Schwessinger B, Hall RN (2019) Uncovering the microbiome of invasive sympatric European brown hares and European rabbits in Australia. *BioRxiv*.

Smith DF, Litvaitis JA (2000) Foraging strategies of sympatric lagomorphs: implications for differential success in fragmented landscapes. *Canadian Journal of Zoology*, **78**, 2134–2141.

Stalder GL, Pinior B, Zwirzitz B *et al.* (2019) Gut microbiota of the European Brown Hare (Lepus europaeus). *Scientific Reports*, **9**, 1–15.

Stumpf RM, Gomez A, Amato KR *et al.* (2016) Microbiomes, metagenomics, and primate conservation: New strategies, tools, and applications. *Biological Conservation*, **199**, 56–66.

Suchodolski JS, Markel ME, Garcia-Mazcorro JF *et al.* (2012) The Fecal Microbiome in Dogs with Acute Diarrhea and Idiopathic Inflammatory Bowel Disease. *PLoS ONE*, **7**.

Suzuki A, Ueda K, Segawa T, Suzuki M (2019) Fecal microbiota of captive Antillean manatee Trichechus manatus manatus. *FEMS Microbiology Letters*, **366**, 1–5.

Teyssier A, Rouffaer LO, Saleh Hudin N *et al.* (2018) Inside the guts of the city: Urban-induced alterations of the gut microbiota in a wild passerine. *Science of the Total Environment*, **612**, 1276–1286.

Trevelline BK, Fontaine SS, Hartup BK, Kohl KD (2019) Conservation biology needs a microbial renaissance: A call for the consideration of host-associated microbiota in wildlife management practices. *Proceedings of the Royal Society B: Biological Sciences*, **286**.

Urdaci MC, Bressollier P, Pinchuk I (2004) Bacillus clausii probiotic strains: antimicrobial and immunomodulatory activities. *Journal of clinical gastroenterology*, **38**, 2–7.

Wang Q, Fu W, Guo Y *et al.* (2019) Drinking Warm Water Improves Growth Performance and Optimizes the gut Microbiota in Early Postweaning Rabbits during Winter. *Animals*, **9**.

Wang J, Ni X, Wen B *et al.* (2020) Bacillus strains improve growth performance via enhancing digestive function and anti-disease ability in young and weaning rex rabbits. *Applied Microbiology and Biotechnology*, **104**, 4493–4504.

Warnecke F, Luginbühl P, Ivanova N *et al.* (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature*, **450**, 560–565.

Weyrich LS, Farrer AG, Eisenhofer R *et al.* (2019) Laboratory contamination over time during low-biomass sample analysis. *Molecular Ecology Resources*, **19**, 982–996.

Xia Y, Sun J, Chen D-G (2018) Power and Sample Size Calculations for Microbiome Data. In: *Statistical Analysis of Microbiome Data with R*, pp. 129–165. Springer Nature Singapore

Pte Ltd.

Xie Y, Xia P, Wang H *et al.* (2016) Effects of captivity and artificial breeding on microbiota in feces of the red-crowned crane (Grus japonensis). *Scientific Reports*, **6**, 33350.

Yuan M, Liu S, Wang Z *et al.* (2019) Effects of particle size of ground alfalfa hay on caecal bacteria and archaea populations of rabbits. *PeerJ*, 1–21.

Zeng B, Han S, Wang P *et al.* (2015) The bacterial communities associated with fecal types and body weight of rex rabbits. *Scientific reports*, **5**, 9342.

Zhu Y, Wang C, Li F (2015) Impact of Dietary Fiber/Starch Ratio in Shaping Cecal Microbiota in Rabbits. *Canadian Journal of Microbiology*, **61**, 771–84.

Zhu L, Yang Z, Yao R *et al.* (2018) Potential Mechanism of Detoxification of Cyanide Compounds by Gut Microbiomes of Bamboo-Eating Pandas. *mSphere*, **3**, 1–11.

Zoetendal EG, Akkermans ADL, Akkermans-van Vliet WM, De Visser JAGM, De Vos WM (2001) The host genotype affects the bacterial community in the human gastrointestinal tract. *Microbial Ecology in Health and Disease*, **13**, 129–134.

Supplemental Figures

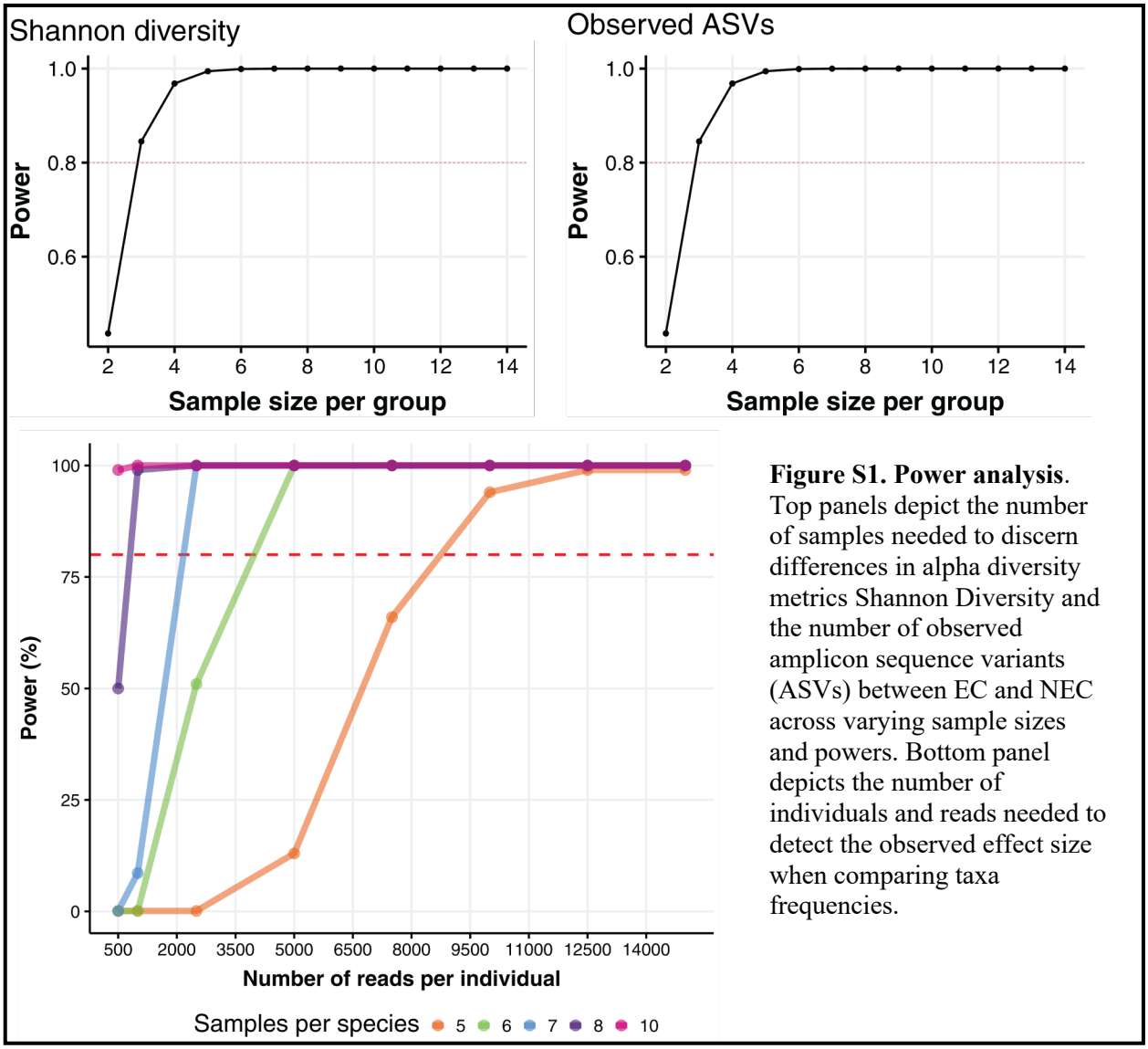

Figure S1. Power analysis. Top panels depict the number of samples needed to discern differences in alpha diversity metrics Shannon Diversity and the number of observed amplicon sequence variants (ASVs) between EC and NEC across varying sample sizes and powers. Bottom panel depicts the number of individuals and reads needed to detect the observed effect size when comparing taxa frequencies.

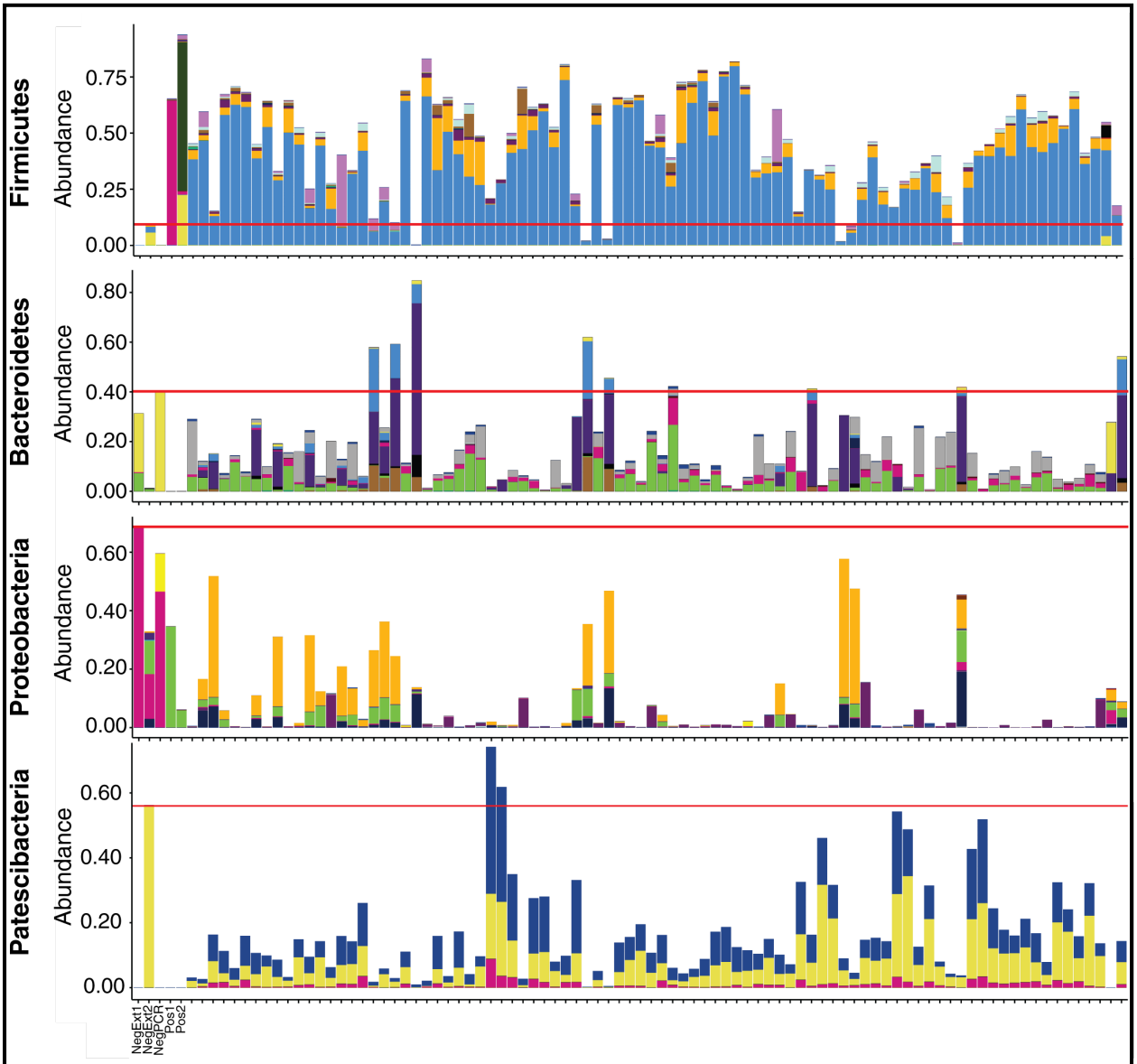

Figure S2. Exploratory relative abundance of taxa by phylum. Each panel depicts relative abundances of taxa across samples and controls (negative extraction controls, negative PCR controls, and positive controls of known compositions). Taxonomic level shown is the most specific taxonomic level in which discernment between negative control taxa and sample taxa is achieved. Different colored bars depict different bacteria at the specified taxonomic level. The red horizontal line depicts the maximum relative abundance detected in a negative control sample. Note that for Patescibacteria, an ASV (yellow) is 1) detected at high abundance, and 2) shared across samples, relative to contaminants that appear in other phyla at low abundance and/or as rare taxa in true samples.

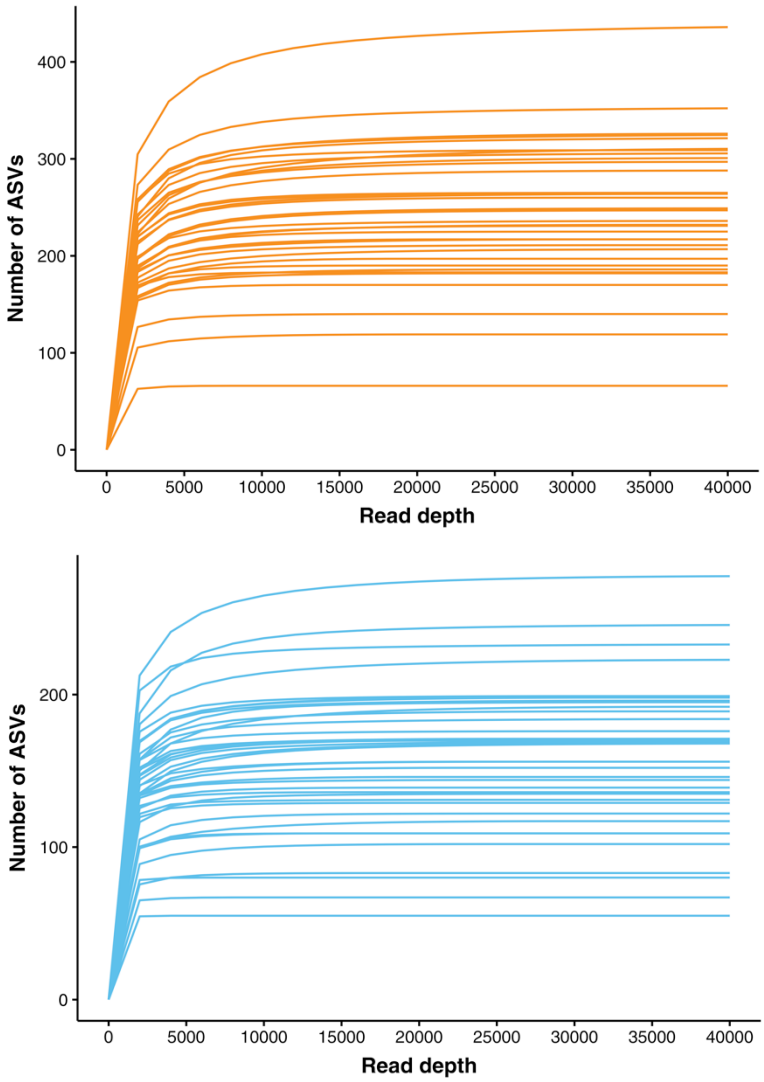

Figure S3. Rarefaction curves for EC and NEC. The number of ASVs recovered with increasing read depth is shown for EC (orange) and NEC (blue). Only samples with a minimum of 40,000 reads are included.

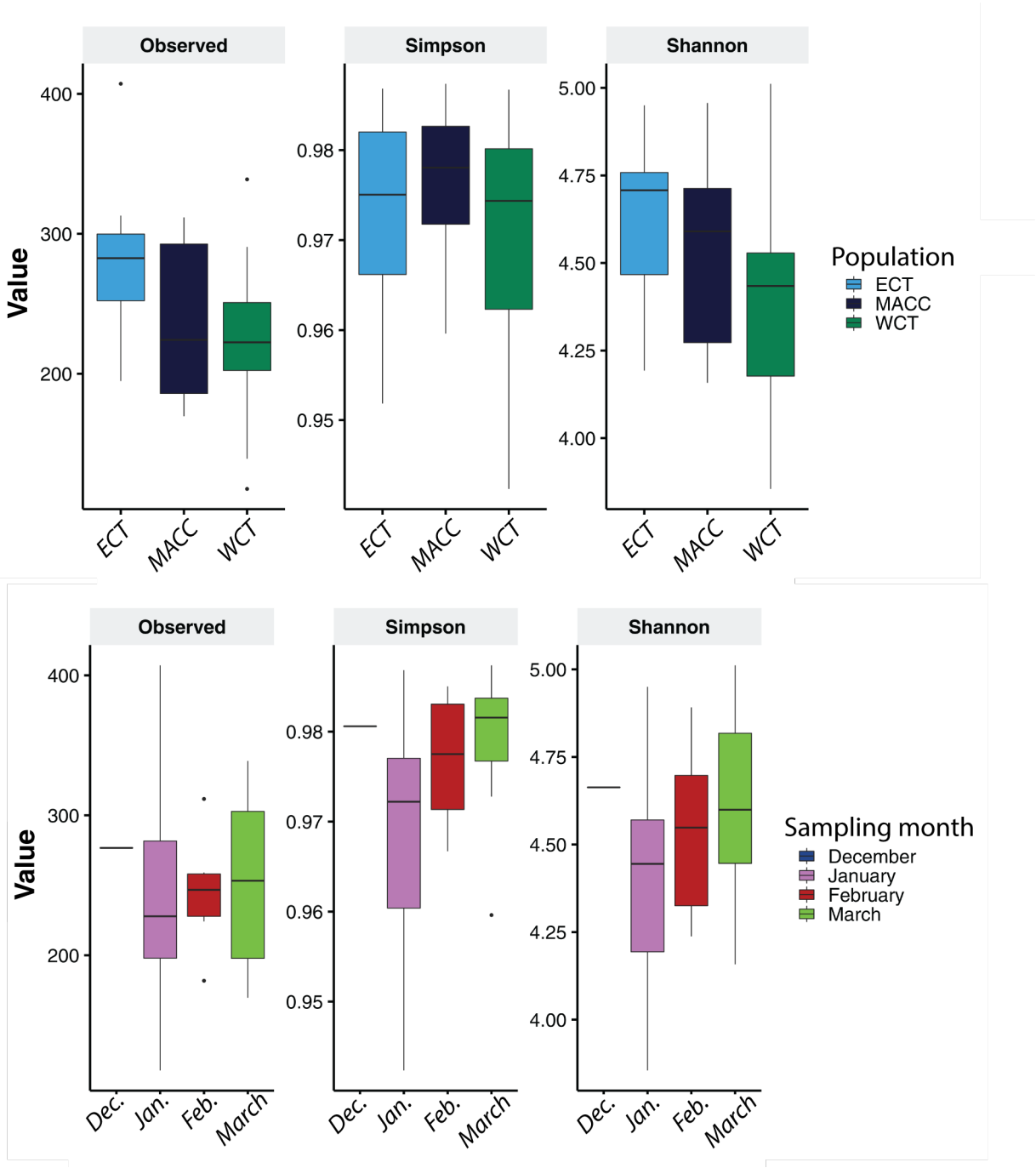

Figure S4. Alpha diversity estimates of EC fecal microbiome across a) population and b) sampling month. Observed ASVs, Simpson Index, and Shannon Diversity are shown, along with the median and interquartile range for each variable.

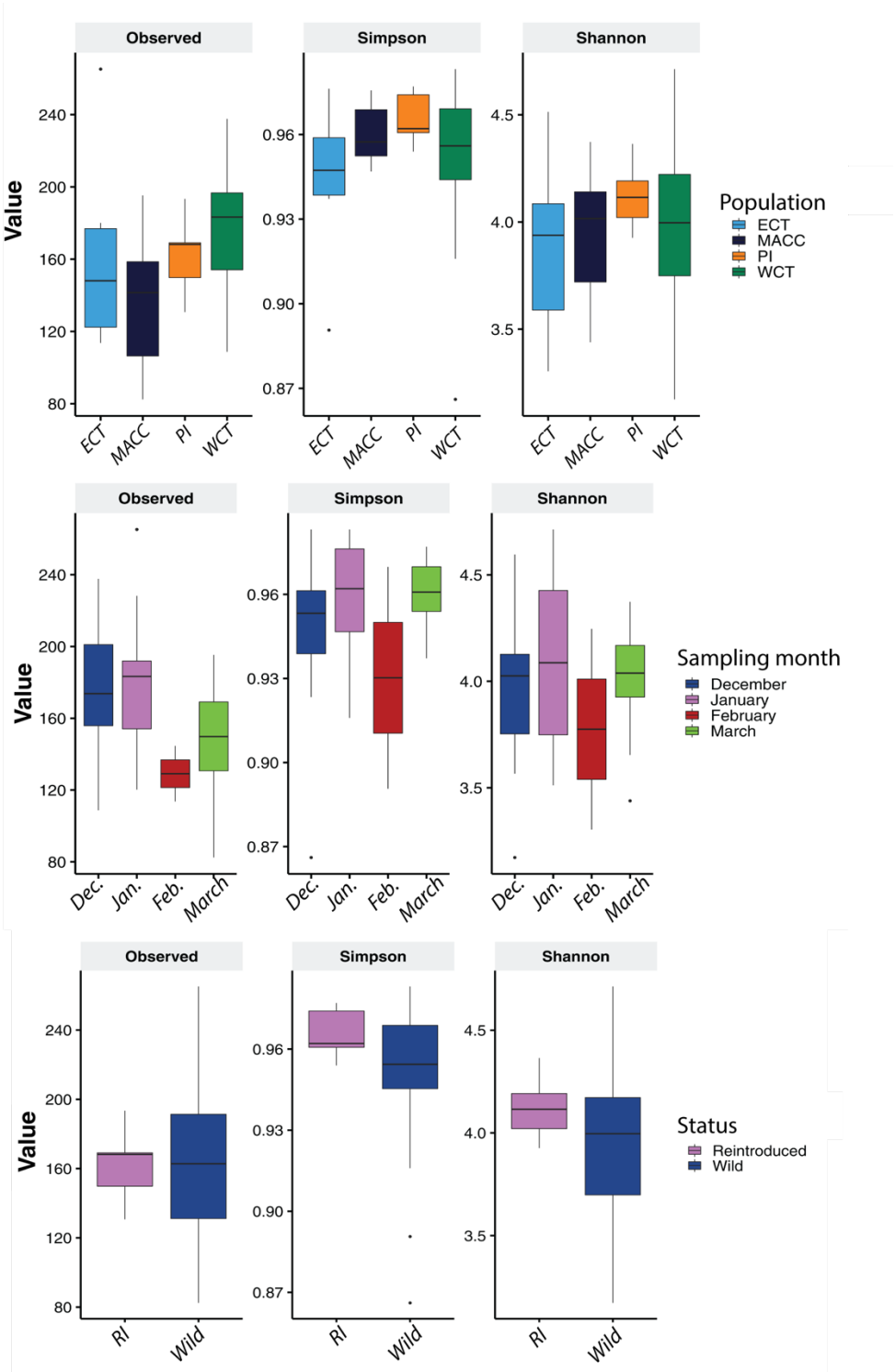

Figure S5. Alpha diversity estimates of NEC fecal microbiome across a) population, b) sampling month, and c) reintroduction status. Observed ASVs, Simpson Index, and Shannon Diversity are shown, along with the median and interquartile range for each variable.

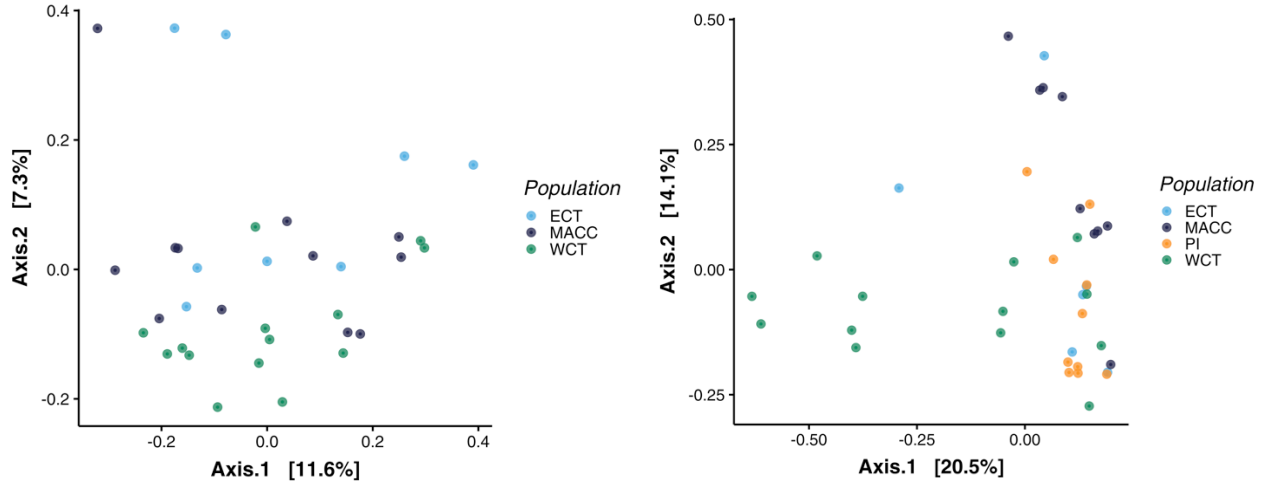

Figure S6. PCOA of EC (left) and NEC (right) by population. Bray-Curtis dissimilarity used. Color indicates population identity.

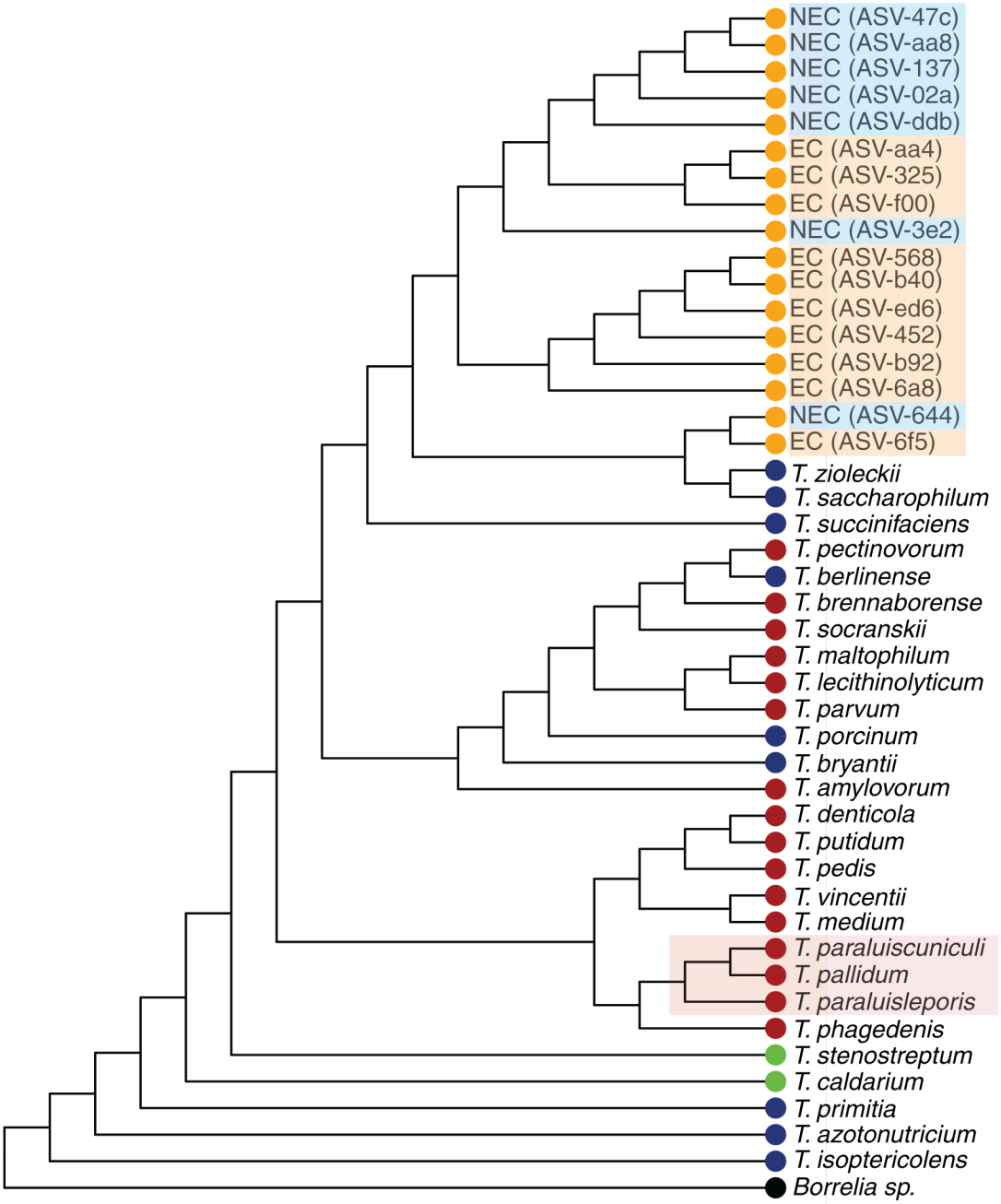

Figure S7. Treponema phylogeny using 16s rRNA gene segment. Sequences not from this study were obtained from Genbank. Sequences were trimmed, aligned, and placed into a phylogeny by RAXML. Circle color indicates pathogenicity: Red = known pathogen, blue = rumen bacteria, green = environmental/avirulent, black = outgroup, orange = unknown from this study. The red box indicates species implicated in syphilis, blue box indicates ASVs recovered only from NEC, and orange box indicates ASVs recovered only from EC.