**Title**
Genomic approaches to confront disease-caused amphibian declines



**Author**
Poorten, Thomas

**Publication Date**
2015

Peer reviewed|Thesis/dissertation

# Genomic approaches to confront disease-caused amphibian declines

by

Thomas Poorten

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Environmental Science, Policy, and Management

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Erica Bree Rosenblum, Chair
Professor Steven R. Beissinger
Professor Rasmus Nielsen

Fall 2015

Genomic approaches to confront disease-caused amphibian declines

Abstract

Genomic approaches to confront disease-caused amphibian declines

by

Thomas Poorten

Doctor of Philosophy in Environmental Science, Policy, and Management

University of California, Berkeley

Professor Erica Bree Rosenblum, Chair

Recently emerged diseases in natural populations present novel problems for biodiversity conservation. Integrated approaches are needed to better understand disease-related threats, to mitigate these threats, and to assist population recovery. My dissertation research confronts the global amphibian biodiversity crisis. The recently emerged infectious disease chytridiomycosis, caused by the chytrid fungal pathogen *Batrachochytrium dendrobatidis* (Bd), infects hundreds of species around the world and is a major contributor to amphibian population declines. I use a multi-faceted approach to address critical issues of disease-caused amphibian declines. In this dissertation, I implement a novel methodology to gain insights on variability of host response to Bd (Chapter 2). Next, I analyze the spatial genetic structure of post-decline populations to aid conservation (Chapter 3). Finally, I examine the genetic underpinnings of pathogen attenuation – loss of virulence – using genomic sequencing (Chapter 4).

First, I examine differential disease progression and host response in two related species (Chapter 2). Determining how different hosts respond to infection by a widespread pathogen is essential for understanding - and ultimately limiting - the devastating effects of emerging infectious diseases. Previous work demonstrated that susceptibility to chytridiomycosis is variable among species, but the mechanism(s) that underlie the difference between winners and losers remains a mystery. I used an integrative approach to analyze host response to infection in two related toad species that are thought to differ in susceptibility: the invasive Cane Toad (*Bufo marinus*) and the threatened Boreal Toad (*Bufo boreas*). With my results, I characterize the nature of differential susceptibility and compare host response using genome-wide gene expression analysis. The susceptible *B. boreas* exhibited high pathogen loads, loss in body weight, severe changes in the epidermis, and dramatic transcriptomic changes without a robust immune response. Conversely, the resistant *B. marinus* exhibited low pathogen loads, stable body weight, only mild disruption of the epidermis and relatively few changes in transcriptomic profile. Together our results show intrinsic differences in host response between related species, which are likely to be an important factor in explaining variation in response to a deadly emerging pathogen in wild populations.

Second, I conducted a conservation genetics study of an endangered amphibian species in Yosemite National Park (Chapter 3). The most striking example of chytrid-associated population declines in North America is the mountain yellow-legged frog (*Rana muscosa* and *Rana sierrae*), including populations in Yosemite National Park. A clear picture of genetic structure and demography of remaining *R. sierrae* populations is critical to short-term management and

conservation. I conducted a study to describe phylogeographic patterns of *R. sierrae* in Yosemite NP in collaboration with ecologists and park biologists. I utilized a recently developed method for multilocus amplicon sequencing that allows sequence data collection from a vast collection of swabs that contain low quantities of input DNA. My analysis of population genetic structure suggests that three genetic clusters occur in Yosemite NP with a significant signature of isolation by distance. This analysis of population genetic structure adds a critical component to the population recovery plan and will assist management strategies such as translocations, reintroductions, and monitoring.

Third, I investigated the genomic changes associated with virulence attenuation in a lab-evolved Bd strain (Chapter 4). Despite recent efforts to characterize the diversity of Bd lineages, there are many questions that remain about the genetic underpinnings of pathogenicity. In a collaborative study, I take advantage of an accidental case of virulence reduction in a Bd strain that was lab passaged over many generations. I analyzed the genomic changes in strain samples cryo-archived before and after virulence attenuation. I found multiple patterns that may be linked to attenuation including decreases in chromosome copy number and mutations in putative virulence genes. These results contribute to the growing body of knowledge of how changes in pathogen genomes occur within a relatively short period of time, which has major implications for host-pathogen dynamics in natural systems.

In conclusion, my dissertation provides important new contributions to the study of host-pathogen interactions with specific relevance to the fields of disease biology, conservation genetics, and pathogen evolution. Integrating genomic tools into a variety of experimental methods enabled not only valuable novel insights, but also opened up many new opportunities for further exploration of disease-caused amphibian declines using the generated genomic and computational resources.

To Mom and Dad for unwavering support,
Chris for enthusiastic encouragement,
and Vanessa for boundless generosity

## Table of Contents

**Acknowledgements**

I thank Erica Bree Rosenblum for being a wonderful advisor. I truly appreciate all of the opportunities for professional and personal growth that Bree has afforded me. First as her lab manager at the University of Idaho, Bree provided an ideal work environment that combined enthusiastic scientific productivity, ample intellectual freedom, and healthy work-life balance. I was lucky to join a nascent lab group with a great deal of momentum in research output all while I was given chance after chance to build my skillset. I am grateful for the opportunity to receive further training as a graduate student at Idaho and at UC Berkeley. Bree has consistently exemplified ideal qualities of an academic advisor. She has been keen with her advice, empathetic with her guidance, and generous with her time. I hope to carry over these traits in my career.

I thank Ray Kuhn, my Master's advisor at Wake Forest University, for the encouragement to pursue a career in science. Ray's mentorship played a huge role for me and gave me the confidence to dive in to new lines of research. I thank past and present graduate students of the Rosenblum lab for all the support and friendship – Simone Des Roches, Kayla Hardwick, Tyler Hether, Lydia Gentry, Alex Krohn, Karina Klonoski, Maddie Girard, Allie Byrne, and Clay Noss. Rosenblum lab postdocs provided endless encouragement – Jamie Voyles, Jeanne Robertson, Suzanne Joneson, Christine Parent, Chris Martin, Eveline Diepeveen, Jeanine Refsnider-Streby, and Gideon Bradburd. I was fortunate to work with wonderful and energetic undergraduates including Karen Pohl, Isaiah Hoyer, Knut Hoversten, Shannon Torstrom, Brian Lohman, and Kendall Calhoun. I thank visiting Master's student Jesse Erens and lab manager Andrew Rothstein for lab support. I also thank the combined "RoHa" group of Rosenblum and Harmon labs at the Univ. of Idaho, which was a special place for intellectual growth in and critical appraisal of the fields of evolution biology and genomics; and for knowing how to have a good time together. At Idaho, several communities fostered my academic growth including the Institute for Bioinformatics and Evolutionary Studies; and the Bioinformatics & Computational Biology graduate program. I thank Matt Settles in particular for playing a huge role in my professional development with his steady mentorship and collaboration. I thank my qualifying exam and dissertation committees at UC Berkeley – Steve Beissinger and Rasmus Nielsen in particular – for facilitating my scholarly development as a scientist. I have felt privileged to be a part of several academic communities at Berkeley, in particular the Department of Environment Science, Policy, & Management; and the Museum of Vertebrate Zoology.

Finally, I thank my family for the huge amount of support they have given. My fiancé, Vanessa, has helped to make this dissertation possible in so many ways. I am deeply thankful for our partnership, especially in making the plunge to move to Berkeley together. Vanessa also made direct contributions to this dissertation as a wonderful partner in field work catching frogs in Yosemite. Whatever is next – I'm glad that we'll experience it together. I thank my parents for the constant support and for making so many cross country trips to spend vacation time together. I thank my brother Chris for his steady encouragement and friendship. I thank Grandma Jessie and Grandma Nancy for their love and expressing their pride in me in countless holiday cards. My extended family has been wonderfully supportive despite being geographically separated, in particular Aunt Patty who was an early inspiration for my ambition to pursue a graduate school degree.

Chapter 2 was co-authored with Erica Bree Rosenblum. Chapter 3 was co-authored with Roland Knapp and Erica Bree Rosenblum. Chapter 4 was co-authored with Jeanine Refsnider-Streby, Penny Langhammer, Patricia Burrowes, and Erica Bree Rosenblum and was previously published in *G3: Genes, Genomes, Genetics;* and re-printed here in accordance with the Creative Commons Attribution License applied by *G3*. I thank the National Science Foundation and the Yosemite Conservancy for funding.

## Chapter 1: General Introduction

With increased interconnectedness and globalization in the last century, infectious diseases can now rapidly spread around the world. The pressure of emerging infectious diseases (EIDs) is having major impacts on wildlife populations. Noteworthy examples of recent enigmatic fungal pathogens include white nose syndrome in bats and chytridiomycosis in amphibians (Fisher *et al.* 2012). White nose syndrome is a disease that has decimated bat populations in Eastern North America, with genetic evidence suggesting that the disease is caused by an exotic invasive pathogen. A parallel crisis in amphibians has emerged with the chytrid fungus, which has been the focal system for my research. As EIDs will likely continue to pose a major threat to biodiversity, research is urgently needed to preserve biodiversity in the face of this emerging problem.

The recently emerged infectious disease chytridiomycosis, caused by *Batrachochytrium dendrobatidis* (Bd), infects hundreds of species around the world and is a major contributor to amphibian population declines (Berger *et al.* 1998; James *et al.* 2009). Chytridiomycosis manifests as an epidermal infection and typically becomes lethal when pathogen burden reaches a high intensity. Numerous studies have documented population collapses as a direct result of Bd invasion (Skerratt *et al.* 2007). Skerratt et al. described the problem as "the most spectacular loss of biodiversity due to disease in recorded history" (2007). While researchers have learned a great deal about this system, there are many unanswered questions that need to be addressed in order to effectively triage outbreaks. For my dissertation, I use a multi-faceted approach to address different problems related to disease-associated declines in amphibian populations. In the field of disease-caused amphibian declines, persistent problems that I address are: (1) understanding differences in susceptibility among host species, (2) examining the genetic variability of post-decline populations to aid conservation, and (3) identifying underlying genomic correlates of virulence in the pathogen. With the recent advent of next-generation sequencing technology, it has become possible to study the genomics of host-pathogen interactions in non-model organisms. In my dissertation, I utilize genomic methods to address a variety of biological questions related to EIDs and species endangerment.

In Chapter 2, I examine differential disease progression and host response in two related species. In the wakes of Bd epidemic waves, researchers have observed that some host species suffer massive die-offs and others persist (Kilpatrick *et al.* 2010). The mechanism(s) that underlie differences in susceptible remain largely unknown thus far. Although environmental and ecological factors can affect susceptibility, previous work suggests that intrinsic differences in host response exist (Searle *et al.* 2011). Host response and pathology has been examined in few species thus far. I used an integrative approach to analyze host response to infection in two related toad species that are thought to differ in susceptibility: the invasive Cane Toad (*Bufo marinus*) and the threatened Boreal Toad (*Bufo boreas*). With my results, I characterize the nature of differential susceptibility and compare host response using genome-wide gene expression analysis. Gaining a better understanding of susceptibility and resistance to chytridiomycosis is critical to generating conservation strategies to protect amphibians. Differences in host susceptibility have implications not only for the pathophysiological effects of Bd, but also for modeling predictions in multiple host-Bd epidemics. For example, the presence of a competent reservoir host species can amplify the disease risk for other resident host species

by increasing the community-wide transmission process. Thus, infection experiments in a comparative framework will inform disease modeling and ultimately conservation strategies.

In Chapter 3, I conducted a conservation genetics study of an endangered amphibian species in Yosemite National Park. The most striking example of chytrid-associated population declines in North America is the mountain yellow-legged frog (*Rana muscosa* and *Rana sierrae*), including populations in Yosemite National Park (Vredenburg *et al.* 2010). Disease-caused declines compounded earlier population declines due to the introduction of non-native predatory trout in lakes and streams for sport-fishing. Building a solid foundation for an effective *R. sierrae* recovery program (that involves fish removal and frog reintroduction) depends critically on detailed knowledge about the relatedness of frog populations. The population genetic structure of *R. sierrae* populations has implications for making informed decisions for identifying suitable donor populations from which frogs are collected as part of future efforts to re-establish frogs in areas from which they were previously extirpated. I conducted a study to describe population structure of *R. sierrae* in Yosemite NP in collaboration with ecologists and park biologists. I utilized a recently developed method for multilocus amplicon sequencing that allows sequence data collection from a vast collection of swabs that contain low quantities of input DNA. For this assay, I designed 50 new nuclear genomic markers using a recently available transcriptome dataset. The results will be used in the restoration effort led by the National Park Service ecologists to guide park-wide translocation efforts in Yosemite NP, which has been identified as a critical conservation tool for *R. sierrae*.

In Chapter 4, I investigated the genomic changes associated with virulence attenuation in a lab-evolved Bd strain. Most infectious pathogens exhibit variation in virulence across space, time, or host species. This variation in infection outcome is affected by three interacting factors: environmental conditions such as temperature (Voyles *et al.* 2012), intrinsic differences in host susceptibility (e.g. Gervasi *et al.* 2013), and differences in virulence among *Bd* isolates (e.g. Berger *et al.* 2005). In this study, I focus on variation in virulence caused by intrinsic differences among *Bd* isolates. Specifically, I compared two laboratory-passaged *Bd* samples that originated from the same source but differed in passage history. Using genome resequencing, I identified genomic correlates of virulence attenuation and shed light on the genomic processes that may be important for the evolution of virulence in this host-pathogen system.

My dissertation utilizes a variety of methods to address critical issues involved with disease-caused amphibian declines. Clearly the challenges of recently emerged infectious diseases require an integrated approach to better understand the complexities of host-pathogen interactions in novel hosts and environments. Genomic methods will undoubtedly play an important role in future such studies. As technological advancements accelerate the pace and ease of applying genomic tools, the line between model and non-model organisms will continue to blur creating more opportunities to expand the scope of genomic studies of recently emerged diseases and biodiversity conservation.

**References**

Berger, L., G. Marantelli, L. F. Skerratt, and R. Speare, 2005 Virulence of the amphibian chytrid fungus Batrachochytrium dendrobatidis varies with the strain. Dis. Aquat. Organ. 68: 47–50.

Berger, L., R. Speare, P. Daszak, D. E. Green, A. A. Cunningham *et al.*, 1998 Chytridiomycosis causes amphibian mortality associated with population declines in the rain forests of

Australia and Central America. Proc. Natl. Acad. Sci. 95: 9031–9036.

Fisher, M. C., D. A. Henk, C. J. Briggs, J. S. Brownstein, L. C. Madoff *et al.*, 2012 Emerging fungal threats to animal, plant and ecosystem health. Nature 484: 186–94.

Gervasi, S., C. Gondhalekar, D. H. Olson, and A. R. Blaustein, 2013 Host Identity Matters in the Amphibian-Batrachochytrium dendrobatidis System: Fine-Scale Patterns of Variation in Responses to a Multi-Host Pathogen. PLoS One 8.:

James, T. Y., A. P. Litvintseva, R. Vilgalys, J. A. T. Morgan, J. W. Taylor *et al.*, 2009 Rapid global expansion of the fungal disease chytridiomycosis into declining and healthy amphibian populations. PLoS Pathog. 5: e1000458.

Kilpatrick, a M., C. J. Briggs, and P. Daszak, 2010 The ecology and impact of chytridiomycosis: an emerging disease of amphibians. Trends Ecol. Evol. 25: 109–18.

Searle, C. L., S. S. Gervasi, J. Hua, J. I. Hammond, R. A. Relyea *et al.*, 2011 Differential host susceptibility to Batrachochytrium dendrobatidis, an emerging amphibian pathogen. Conserv. Biol. 25: 965–74.

Skerratt, L. F., L. Berger, R. Speare, S. Cashins, K. R. McDonald *et al.*, 2007 Spread of Chytridiomycosis Has Caused the Rapid Global Decline and Extinction of Frogs. Ecohealth 4: 125–134.

Voyles, J., L. R. Johnson, C. J. Briggs, S. D. Cashins, R. a. Alford *et al.*, 2012 Temperature alters reproductive life history patterns in Batrachochytrium dendrobatidis, a lethal pathogen associated with the global loss of amphibians. Ecol. Evol. 2: 2241–2249.

Vredenburg, V. T., R. A. Knapp, T. S. Tunstall, and C. J. Briggs, 2010 Dynamics of an emerging disease drive large-scale amphibian population extinctions. Proc. Natl. Acad. Sci. U. S. A. 107: 9689–94.

## Chapter 2: Comparative study of host response to chytridiomycosis in susceptible and resistant toad species


### Abstract

In the past century, recently emerged infectious diseases have become major drivers of species decline and extinction. Amphibian declines have occurred due to the fungal disease chytridiomycosis, which has exacerbated the conservation crisis of this taxonomic group. Biologists are beginning to understand what traits are important for susceptibility to this disease, but more work is needed to determine why some species succumb to disease while others do not. We conducted a laboratory experiment to examine how two toad species respond to infection in controlled environment. We selected two related species thought to differ in susceptibility – *Bufo marinus* (an invasive and putatively resistant species) and *B. boreas* (an endangered and putatively susceptible species). We measured infection intensity, body weight, histological changes at the site of infection, and genome-wide gene expression changes using a custom assay developed from transcriptome sequencing. Our results confirmed that the two species differ in susceptibility. The more susceptible species, *B. boreas,* experienced higher infection intensities, loss in body weight, more dramatic histological changes, and larger perturbations in gene expression. We found key differences in skin expression responses in multiple pathways including up-regulation of skin integrity-related genes in the resistant *B. marinus*. Together our results show intrinsic differences in host response between related species, which are likely to be an important factor in explaining variation in response to a deadly emerging pathogen in wild populations.


### Introduction

The outcome of host-pathogen interactions results from a complex interplay between host, pathogen, and environmental effects. Because a pathogen is unlikely to have a universal effect on all hosts, characterizing differences in host response is critical for understanding population outcomes. The recently emerged disease of amphibians, chytridiomycosis, affects hundreds of host species. Chytridiomycosis, caused by the chytrid fungus *Batrachochytrium dendrobatidis* (Bd), has led to devastating population crashes in numerous amphibian species (Berger *et al.* 1998; Skerratt *et al.* 2007; Wake and Vredenburg 2008). Chytridiomycosis is a transmissible skin infection and can spread rapidly in host populations, causing high mortality rates. However, previous work has shown that disease outcome following Bd infection is variable among species. For example, differences in susceptibility have been found in North American species (Blaustein *et al.* 2005; Searle *et al.* 2011; Gervasi *et al.* 2013) and Australian species (Woodhams *et al.* 2007). Determining how different hosts respond to infection is essential for understanding - and ultimately limiting - the devastating effects of Bd.

Many different host traits could affect differences in susceptibility among amphibian species. For example, ecological traits including water dependence and population density can have major effects on pathogen exposure and transmission risks (Briggs *et al.* 2010; Murray *et al.* 2011). In addition, traits at the level of an individual host are integral to the host-pathogen interaction. At this level, hosts rely on physiological and immunological traits to prevent and fight off infection. Researchers are beginning to gain insights into the important components of host response (James *et al.* 2015). Variation in intrinsic host traits such as skin structure and

immune response have been proposed to play major roles at the interface of host and pathogen processes (van Rooij *et al.* 2012; Gervasi *et al.* 2013; James *et al.* 2015).

We performed an integrative study of host response in related species that differ in susceptibility. Using related species enables a comparative analysis of host response in orthologous genetic and physiological pathways. For our experiment, we selected two bufonid species thought to vary in susceptibility: *Bufo marinus* (Cane Toad) and *Bufo boreas* (Boreal Toad). *Bufo marinus* is reported to be resistant to chytridiomycosis as infected individuals exhibit only minor symptoms and suffer a low mortality rate (Fisher and Garner 2007). Additionally, further study on *B. marinus* is warranted given its status as a notorious invasive species in parts of Australia and the United States, and therefore may contribute to spreading Bd. In contrast, *B. boreas* typically displays very high susceptibility with a high mortality rate. Previous studies on wild *B. boreas,* a native species of grave conservation concern in the Western North America*,* have documented population crashes that resulted from Bd epidemics (Pilliod *et al.* 2010). Divergence time estimates place the split between these species at ~24.6 million years ago (Garcia-Porta *et al.* 2012).

We conducted a controlled laboratory infection experiment in common garden conditions and sequenced transcriptomes to generate resources for a genome-scale study of response to infection. During the course of the experiment, we tracked pathogen infection intensity and body weight and used histology and custom gene expression microarrays to characterize host response. Overall we found strong differences in pathogen infection intensity and body condition dynamics between these two species. Additionally, we found large differences in response to infection based on transcriptional and histological results, which are likely to be critical in determining host susceptibility.

## Materials and Methods

*Infection Experiment*

We performed an experimental infection of Bd with *B. marinus* and *B. boreas* in a controlled laboratory setting. We obtained 20 adult *B. marinus* from Carolina Biological Supply and 20 captive bred *B. boreas* from Colorado Division of Parks and Wildlife. We received a pathogen-free testing certification for the *B. boreas*, but we performed a swab-qPCR assay on all toads to ensure that they were all Bd-negative prior to the experiment. The toads were maintained in separate cages in a dedicated room that maintained the temperature at 20°C, humidity at 50%, and 12:12 hour light:dark cycle. We fed the toads vitamin-dusted crickets *ad libitum* twice per week. We cleaned the cages and changed the water after each feeding. We weighed each toad prior to starting the experiment to determine baseline body mass. The toads in each species were randomly and evenly divided into control and treatment groups. To inoculate the toads, we pipetted one million Bd zoospores per 40 grams of body mass on to the flanks of each toad. We adjusted the inoculation dose to account the size difference between the two species. For 24 hours we kept the toads in plastic containers that limited movement for the toads, which ensured that the toads were in close contact with the inoculum zoospores. Each container contained 10 mL of Holtfretter's solutions in order to maintain the osmolarity and pH during the incubation. We used Bd isolate "JEL275", which was collected from a *B. boreas* individual in Clear Creek, Colorado. We weighed and swabbed the toads on days 7, 14, and 18 following the inoculation. We used the established qPCR method to estimate pathogen infection intensity (Boyle *et al.* 2004). We let the experimental infection run until the first highly infected (>10,000 zoospore genomic equivalents [GE]) B. boreas host died, at which point all toads were sacrificed

by decapitation and immediately dissected for tissue preparation of histological and RNA samples. We used a linear mixed model to test for differences in infection intensity between species. The full model included two fixed effect parameters, species and time (number of days since inoculation), and one random effect parameter, subject identifier. The response variable was infection intensity in units of $\log_{10}$ zoospore equivalents. Control groups were excluded from this model as all individuals tested negative for Bd infection throughout the experiment. We tested for statistical significance using a likelihood ratio test to compare the full model to a null model without the parameter of interest – species. To test for differences in weight gain or loss between species, we again used a linear mixed model. The full model included three fixed effect parameters, species, experimental group (Bd exposure or control), and the interaction term. We treated subject identifier as a random effect in the model. The response variable was change in body weight over the course of the experiment in units expressed as $\log_2(\text{weight}_{\text{day18}} / \text{weight}_{\text{day0}})$, which provided a comparable metric for between-species comparisons. We used Tukey's post hoc test to compare weight change between pairs of experimental groups. We used the R package nlme to implement linear mixed models (Pinheiro *et al.* 2014). Both response variables in the mixed models (infection intensity and weight change) were approximately normally distributed after log scaling.

*Histology*

We conducted histological examination of frog ventral integument in control and experimentally infected individuals. To prepare the samples, we fixed freshly dissected skin samples in 4% paraformaldehyde for 24 hours and then sunk the samples in 30% sucrose overnight. Sectioning and staining with hematoxylin and eosin was performed by University of California, Davis Comparative Pathology Laboratory. A pathologist examined histological sections with 20X magnification light microscopy to identify secondary changes including hyperkeratosis (thickening of the top layer of the epidermis), hyperplasia (increased number of cells), dermatitis (inflammation), and spongiosis (fluid accumulation).

*Transcriptome sequencing and microarray design*

We performed transcriptome sequencing of the study species, *B. marinus* and *B. boreas*, because no genomic resources were available and then developed a custom Nimblegen (Madison, WI) gene expression microarray platform. We extracted total RNA from several tissues – liver, skin, and spleen – of several individuals in order to maximize transcript discovery. Tissues were preserved by flash freezing in liquid Nitrogen immediately after dissection. We used the Trizol protocol for RNA extraction as previously described, and then generated cDNA from each RNA sample. For each species, we pooled cDNA samples and added MID barcodes to each species. We used the Roche 454 sequencing platform to generate sequence data from the barcoded cDNA pool. We used custom scripts to preprocess the raw sequence data. We used the barcodes to separate the sequence data for each species into separate datasets. Then we cleaned the reads by removing adapters and filtering low quality reads with lucy (Chou and Holmes 2001). We performed de novo transcriptome assembly for each species using Newbler (Margulies *et al.* 2006). We obtained 8520 contigs for *B. marinus* and 6095 contigs for *B. boreas*. For the microarray design, we designed probes for contigs in three categories: *B. marinus*-specific contigs, *B. boreas*-specific contigs, and *X. tropicalis* transcripts (19,312 transcripts). We used this design strategy to maximize the diversity of contigs on the microarray and leverage the entire available probe area. The X. tropicalis transcripts were downloaded from Ensembl database (Xtrop v4.2), which also curate gene annotation, and processed to reduce paralogous transcripts by clustering with CD-HIT (parameters: cd-hit-est -T

8 -c .95 -s .95 -n 8 -d 200 -r 1 -g 1)(Li and Godzik 2006). The custom Nimblegen microarray design included 135,200 60-basepair probes (excluding control probes) targeting 33,822 transcript contigs. Most probe-sets included 4 probes (33,780/33,822). We used the 12-plex microarray platform (12 arrays per glass slide) to cost-effectively increase sample size in the experimental design.

Annotation was performed using the blat program to identify Bufo contigs with high sequence identity to *X. tropicalis* transcripts (Ensembl assembly v4.2). We ran blat with translated Bufo contigs and translated *X. tropicalis* transcripts (parameters: -q=dnax –t=dnax) in order to find orthologous sequences at the protein level, which is more conserved between divergent species. We parsed blat hits to remove short hits (<100 bp aligned to query) and low scoring hits (Expectation value >1.0e-12). We used the Ensembl Gene Identifier (e.g. ENSXETG00000XXXXXX) for the blat hit with longest alignment to the query as the final annotation call. We then imported annotation metadata with the biomart.org pipeline: Ensembl gene name and description, Gene Ontology (GO) terms, Ensembl protein families, Interpro protein domains. For both species combined, 65.4% (9565/14615; *B. marinus*: 64.4% - 5485/8520; *B. boreas* 66.9% - 4080/6095) contigs were annotated with a *X. tropicalis* Gene Identifier. Of the annotated contigs, 86.7% (8287/9565) contained at least one GO term. Next, we used the blast2go pipeline to annotate Bufo contigs without GO term metadata (6328 contigs). We used the blastx method with the NCBI nonredundant protein database to identify blast hits, and proceeded with blast2go pipeline with default settings (Conesa *et al.* 2005).

*Gene expression microarray pre-processing*

We processed 72 tissue samples in total: six biological replicates, three tissue types, two treatment groups, and two host species. Briefly, we extracted total RNA from flash frozen tissues samples with a standard Trizol protocol and assessed RNA quality with a Bioanalyzer 2100 Total RNA Pico assay (minimum RIN score = 8.0)(Agilent, Santa Clara, CA). Downstream processing was performed by University of Idaho Genomic Resources Core and followed Nimblegen standard protocols: double-stranded cDNA synthesis, Cy3 fluorescent dye labeling, sample-array hybridization, and data collection via fluorescence imaging. We visually assessed array images for fluorescence artifacts. Raw fluorescence intensities were extracted with NimbleScan v2.5. Downstream analyses were performed using Bioconductor microarray analysis packages in R. We used the R package "pdInfoBuilder" to construct a microarray design package for our custom microarray. We assessed raw data quality for each array by graphing intensity distributions, boxplots, and hierarchical clustering. In addition, we used the quality control pipeline in the R package "arrayQualityMetrics" (Kauffmann *et al.* 2009). Based on these analyses, we excluded one control group *B. marinus* individual, which we later determined to have an intradermal nematode infection based on histological analysis. For analyses in each species, we filtered out off-species probe-sets with redundant annotations as determined by Xenopus Ensembl Gene Identifier. For example, *B. marinus* analyses excluded redundant *B. boreas* and *X. tropicalis* probe-sets, but included uniquely annotated probe-sets from all three source species in order to maximize functional diversity of probe-sets in the analyses. We used RMA preprocessing algorithm in the "oligo" R package to perform three steps: background correction and probe level normalization for each array, and probe-set summarization via median polish (Carvalho and Irizarry 2010). To limit the number of probe-sets in the statistical analyses, we nonspecifically filtered out probe-sets where interquartile range < 0.5, which contain relatively low variability across all samples.

*Gene expression microarray analysis*

We analyzed gene expression responses to infection in three tissue types (ventral skin, liver, and spleen) at the final sampling point using a custom microarray. We conducted separate statistical analyses for each species and for each tissue type. In each statistical analysis, we assessed differential expression between the Bd-exposed and control groups, thus examining the response to infection in the Bd-exposed group. We conducted analyses at the two levels, probeset and gene set, which are analogous to gene and pathway levels, respectively. First, we statistically tested for differential expression with the R package "limma", which implements a linear model with an empirical Bayes adjustment to the variances (Ritchie *et al.* 2015). We used the microarray slide ID as a blocking factor as samples for each tissue were randomly placed in arrays across two glass slides. We controlled for the expected false discovery rate using the Benjamini and Hochberg (BH) method for multiple tests. We considered probe-sets to be significantly differentially expressed with BH-adjusted p-value <0.05 for *B. boreas* and <0.1 for *B. marinus*. We used the less stringent cutoff for *B. marinus* to detect more subtle changes in the transcription profile after we found in exploratory analyses that relatively few probesets contained strong expression changes. The validity of the cutoff was qualitatively confirmed based on comparison to the gene set level analysis results, as described below. We functionally profiled the differentially expressed probesets using GOstats R package to identify over-represented GO terms in probeset annotations (Falcon and Gentleman 2007); and reduced redundancy and overlap in gene sets using Revigo (Supek *et al.* 2011).

Our second level of differential expression – gene set – examined pathway level expression changes. We used GO term annotations to define gene sets in the Biological Process category of the GO framework. We estimated the response to infection at the gene set level using the Gene Set Variation Analysis (GSVA) methodology (Hänzelmann *et al.* 2013). GSVA produces gene set expressions scores, which summarize the expression level for each gene set in each sample. Expression scores are then used in the same differential expression framework as above using Bd-exposure as the main factor with microarray slide as the blocking factor. We defined differential expression as gene sets with BH-adjusted p-value <0.05. We reduced redundancy and overlap in differentially expressed gene sets using Revigo (Supek *et al.* 2011). We found concordant results when analyzing the dataset with the GSVA method. With the GSVA approach, we found differential expression – with the criterion of BH-corrected $p < 0.05$ – of the same or related gene sets as those described above in the GOstats enrichment test (supplementary Table 1). This observation qualitatively validates our use of a less stringent p-value cutoff (BH-corrected p-value=0.1) in the probeset-level analysis for *B. marinus*.

*Cross-species comparison*

We compared response to infection ($\log_2$ fold change ratios) between the two species using linear regression. We used gene set level $\log_2$ fold change ratios from the GSVA method described above. We used the Revigo method to reduce redundancy and overlap in gene sets prior to the regression analysis to limit pseudo-replication (Supek *et al.* 2011). We then regressed gene set $\log_2$ fold change ratios using a linear model. We repeated this analysis for 1000 iterations using permuted sample labels and recorded the regression slope at each iteration. We calculated the p-value as the number of permutation regression slope values greater than the test regression slope value (slope=-0.369) divided by 1000. We performed this test for the three tissue types.

## Results

*Infection intensity and body mass dynamics*

*Bufo boreas* experienced higher infection intensities than *B. marinus* through the 18-day course of the experiment, which confirmed the differential susceptibility between these species in terms of disease progression. All *B. boreas* tested positive for Bd infection at all three sampling points after inoculation, whereas eight out of ten *B. marinus* tested positive by the final sampling point (Fig 1). From the mixed effect model, we found that species identity affected Bd infection intensity ($\mathbf{X}^2(1) = 39.510$, $p < 0.0001$), increasing it by 2.379 +/- 0.212 $\log_{10}$ units in *B. boreas* relative to *B. marinus*. Adding the interaction term (species and time) did not significantly improve the model fit ($\mathbf{X}^2(1) = 2.144$, $p < 0.143$). The different body weight trajectories experienced between species also indicates differences in susceptibility. From the mixed effect model, we found that the interaction of Bd exposure and species affected change in body weight ($\mathbf{X}^2(1)= 17.69$, $p=0.000026$), lowering it by 0.273 +/- 0.057 $\log_2$ units. Post hoc Tukey tests confirmed that weight change for Bd-exposed *B. boreas* was significantly different than control *B. boreas* and both *B. marinus* experimental groups ($p < 0.001$). The weight loss effect for Bd-exposed *B. boreas* is illustrated in Figure 2. Below we further examine body condition changes and host response by examining histological changes at the site of infection – skin – and by analyzing gene expression changes in skin, liver, spleen tissues.

*Histological findings*

Histological examination of ventral skin sections revealed more dramatic effects of infection on *B. boreas* than *B. marinus* (Fig 3). Bd-exposed *B. boreas* samples contained a higher incidence of epidermal hyperplasia (increased number of cells) (3/5, Fig 3e) compared to exposed *B. marinus* (0/5). While both species experienced parakeratotic hyperkeratosis (thickening of the epidermis), the extent of the change tended to be higher in *B. boreas* (Table in S1 Table). Most *B. marinus* (4/5, Fig 3c) experienced minimal hyperkeratosis, while most *B. boreas* (4/5, Fig 3f) experienced mild hyperkeratosis with one individual achieving moderate hyperkeratosis (Fig 3e). In addition, spongiosis (intercellular edema) only occurred in exposed *B. boreas*, at mild to moderate levels (Fig 3e-f). Interestingly, both species showed minimal to mild evidence of chronic lymphoplasmacytic dermatitis (inflammation) in a subset of exposed and control individuals (Fig 3a,d). This observation suggests that the weak signal of inflammation in histological sections was not *necessarily* associated with response to infection, although the highest extent of dermatitis (mild to moderate) was observed in a highly infected individual (>10,000 zoospore equivalents, Table in S1 Table). Thus overall we observed sharp differences between species with more dramatic histological indicators of infection in B. boreas.

*Broad-scale gene expression patterns*

We measured gene expression changes in Bd-exposed individuals using a custom microarray developed from newly sequenced transcriptomes. Our host response analysis revealed a relatively weak transcriptional response in *B. marinus* and strong response in *B. boreas* (Fig 4). We analyzed gene expression responses to infection in three tissue types (ventral skin, liver, and spleen) at the end of the experiment. Bd-exposed *B. boreas* experienced major gene expression perturbations in the skin and liver, while Bd-exposed *B. marinus* experienced fewer and weaker gene expression changes only in the skin.

In the skin dataset, we observed contrasting expression responses to infection at the gene set level. We used a GSVA analysis of the expression profiles to compare gene set level (i.e. pathway) expression responses in the two species. We regressed the Bd-exposed vs. control

comparison log ratios ($log_2$ fold change values) between the two species and found a marginally significant negative association in gene set log ratios using a permutation test (regression slope = -0.369, p = 0.061, Fig 5). This pattern suggests that gene sets tended to be regulated in discordant directions between the two species in response to infection in the skin. We did not observe any trends for the liver and spleen datasets, which suggests that a general host response was not shared in these tissues either. Below we detail the gene expression results in the three focal tissues.

*Gene expression patterns in the skin*

At the site of infection, we found weak expression changes for *B. marinus* and massive perturbation in gene expression for *B. boreas*. In the resistant *B. marinus*, we identified differential expression in pathways that are likely linked to response to low-level infection. Our GOstats tests for gene set enrichment included 487 up-regulated and 70 down-regulated probesets, respectively. In the set of up-regulated probesets, we identified enriched GO terms related to skin structure maintenance and re-modeling: "wound healing", "epidermis development", and "biological adhesion" (Table 1). Relevant up-regulated probesets include integrin beta 3, tropomyosin 3, alpha-2-macroglobulin, coagulation factor 9, and collagen. We also identified enriched GO term "immune system development" (Table 1), which contained immunomodulatory genes (e.g. interferon regulatory factor 4 and zinc finger protein 36) that tend to have suppressive or regulatory effects on immune response. However, in individual probeset tests we found that three complement pathway probesets (complement c3, c8, c9) were up-regulated indicating some evidence of a weak innate inflammatory response, although the entire gene set was not found to be enriched using GOstats. We also found up-regulation of stress-related pathways based on enrichment of GO terms "response to stress" and "apoptotic signaling pathway". Finally, the small number of down-regulated probesets were enriched for metabolism-related gene sets (Table 1).

In the susceptible *B. boreas,* infected individuals experienced a relatively large degree of gene expression perturbations in the skin. At the probeset level, we found 1,108 up-regulated and 1,055 down-regulated probesets in the Bd-exposed group. From the GOstats analysis, a large number of gene sets were significantly enriched in these tissues. Notably, up-regulated probesets in the skin were enriched for immune and stress related gene sets. We found evidence of an innate immune response supported by enrichment of gene sets: "regulation of complement activation", "innate immune response", "response to external stimulus", and "response to yeast". Relevant up-regulated probesets included several classical complement pathway genes (components b, 2, 3, 4, and 7), toll like receptor 5, and tumor necrosis factor-a pathway genes. Other enriched gene sets were associated with stress: "response to stress" and "cell redox homeostasis". This analysis found some parallels with the *B. marinus* response to infection: wound healing, response to stress. With respect to down-regulated probesets, we found enrichment in several gene sets related to skin structure maintenance: "extracellular structure organization", "blood vessel development", "cell-matrix adhesion", and "epithelium development". This pattern involved down regulation of important structural proteins including 29 collagen probesets and 4 keratin probesets.

In our cross-species gene expression analysis, described above, we found gene sets with discordant and concordant expression patterns between species (Table 2). Discordant expression patterns reveal potentially important differences in host response between the two species. Gene sets that were up-regulated in *B. marinus* and down-regulated in *B. boreas* included skin integrity related pathways such as "cell-matrix adhesion" and "cortical actin cytoskeleton organization".

The between species comparison also uncovered pathways with concordant expression patterns, which point toward shared responses despite the general negative association. The four gene sets that were up-regulated in both species included “cellular component organization”, “histone modification”, “stem cell maintenance”, and “regulation of response to stimulus”.

*Gene expression patterns in the liver*

We found that infection status had little detectable impact on gene expression in the liver tissues of Bd-exposed *B. marinus*; and massive impact on gene expression for *B. boreas*. For the resistant *B. marinus*, we observed zero up-regulated probesets and only two down-regulated probesets (Table 1), which were too few to conduct a GOstats enrichment test; and similarly zero gene sets were differentially expressed based on the GSVA analysis at BH-corrected $p<0.05$. The two down-regulated probesets were annotated as MHC class 2 probesets, which suggests a weak or subtle reduction in the antigen processing and presentation process or a reduction in the number of antigen presenting cells in the tissue.

We detected a large number of differentially expressed probesets in the susceptible *B. boreas* liver samples: 1,947 up-regulated and 951 down-regulated probesets in the Bd-exposed group. We found substantially different expression patterns in liver tissue compared to skin. In the liver analysis, the list of down-regulated probesets was enriched for immune and defense related gene sets; whereas immune-related gene sets were up-regulated in the skin (Table 1). Down-regulated gene sets included: “positive regulation of T cell mediated cytotoxicity”; “antigen processing and presentation”; “defense response”; “immune response”. Relevant down-regulated probesets from these gene sets included several complement pathway genes, MHC class 1a and class 2, antibody receptor epsilon, and tapasin. Multiple stress-related gene sets were up-regulated including “response to stress” and “cell redox homeostasis”. We also observed up-regulation in several gene sets that typically have coordinated expression profiles with the stress response including “regulation of cell cycle”, “cellular metabolic process”, and “cellular respiration” (Table 1).

*Gene expression patterns in the spleen*

Finally, we found no differential expression activity for *B. marinus* and a very low level of differential expression in the *B. boreas* spleen dataset (only 3 up-regulated probesets). Notably, one of the up-regulated probesets is annotated as an Fc receptor-like 2, which has potential immunomodulatory effects on B-cell development. Given previous reports on the negative effects of Bd on splenocyte proliferation (Fites *et al.* 2013), we further examined the GSVA results for *B. boreas* at the uncorrected p-value level (i.e. without correction for multiple tests). Interestingly, we found that two out of four down-regulated gene sets were related to immune cell processes: “positive regulation of T cell mediated cytotoxicity” (uncorrected $p=0.01$); “B cell receptor signaling” (uncorrected $p=0.02$).

## Discussion

*Differential susceptibility between species*

Studying the variability in host response is critical for gaining a better understanding of the epidemiological patterns of emerging infectious diseases. With recent appreciation for the range of outcomes that species experience following Bd infection, there is large interest in uncovering mechanisms of resistance (Berger *et al.* 2010; Voyles *et al.* 2011; McMahon *et al.* 2014). To address this problem, our study utilized an integrative approach to examine disease progression and host response in a related species pair. The focal species are of major interest

due to concerns regarding conservation status (*B. boreas*) and invasiveness (*B. marinus*), thus warranting detailed study on host response to Bd. Although previous studies have documented susceptibility of *B. boreas*, the comparative framework used here allowed us to polarize the relatively severe susceptibility of *B. boreas* to the resistance of *B. marinus*. The multiple assays in this study led to a further characterization of the nature of differential susceptibility – at the host level – between these species. Below we discuss the implications and potential mechanisms of strong resistance in *B. marinus* with host response focused locally at the site of infection; and strong susceptibility in *B. boreas* with evidence of local and systemic host response.

Previous work on the more-studied *B. boreas* generally supports our findings regarding infection intensity and disease progression. An earlier study found high mortality rate of juvenile *B. boreas* between 15 and 25 days after inoculation in multiple experiments (Carey *et al.* 2006). Furthermore, in our study the first *B. boreas* mortality occurred in the individual that first crossed the 10,000 zoospore genomic equivalents threshold, thus consistent with the so-called "Vredenburg's 10,000 rule" (Vredenburg *et al.* 2010). The reduction in body weight for the infected *B. boreas* toads demonstrates that body condition deteriorated with Bd infection. This observation has been documented in other species (Retallick and Miera 2007; Harris *et al.* 2009), but not in all susceptible species (Voyles *et al.* 2009) indicating infection can be deadly even without decline in body condition in some species. Our histological results were consistent with our qPCR-estimated infection intensities. The highest degree of epidermal hyperkeratosis was observed in the *B. boreas* individual with the highest infection intensity. Furthermore, Bd-exposed *B. boreas* toads experienced spongiosis, while *B. marinus* did not. This histological change may be variable among species or associated with higher infection intensities. While we cannot distinguish these possibilities as all infected *B. boreas* achieved higher infection intensities than *B. marinus*, we expect that the increased infection intensities may be associated with increased inflammatory changes including spongiosis. Previous studies have found spongiosis, as well as hyperkeratosis, to be a common feature of Bd infection (Berger *et al.* 1998, 2005; Carey *et al.* 2006; Voyles *et al.* 2011). Together the differences in infection intensities, body condition, and histological effects characterize the nature of differential susceptibility between *B. marinus* and *B. boreas*. These differences are reflected our analyses of host response pathways at the gene expression level as discussed below.

*Comparing gene expression responses in the skin*

Our skin expression analysis revealed a trend in contrasting expression responses between the susceptive and resistant species at the gene set level. The host response in the skin has undoubtedly critical importance not only because it is the site of infection but also because of documented pathological effects of pathogen-mediated skin disruption including catastrophic fluid and electrolyte imbalance (Voyles *et al.* 2009). As such, maintaining skin integrity during infection is critical to maintain physiological homeostasis. In the present study, we found that the contrasting host response trend included a discordant pattern in skin integrity pathway genes (e.g. "cell-matrix adhesion" gene set). This key finding provides support for the previously hypothesized importance of this pathway from earlier work. For example, a gene expression study involving multiple species found that the gene coexpression module enriched for "cell adhesion" was up-regulated in the most resistant species in the experiment and down-regulated in the three more susceptible species (Ellison *et al.* 2015). In addition, the down-regulation pattern of skin integrity genes was also observed in a gene expression study involving only susceptible species (Rosenblum *et al.* 2012). In that study, several keratin and collagen genes were down-regulated in skin samples of Bd-exposed frogs. Likewise, we observed down-

regulated of 29 collagen probesets and 4 keratin probesets in *B. boreas* skin. The regulation of skin integrity pathways appears to be an important marker of infection outcome and more work is needed to reveal whether the (dys)regulation causes or coincides with susceptibility.

The between species comparison also uncovered pathways with concordant expression patterns, which may play a shared or general role in host response. An up-regulated gene set in both species – "regulation of response to stimulus" – was also a top enriched gene set for three species in a recent study (Ellison *et al.* 2015). These results indicate that genes involved in regulating response to stimuli (e.g. alpha-2-macroglobin, TNF-R12) represent a common response to Bd at the site of infection. We also observed some notable shared expression changes between species from the results of the GOstats analysis. For example, we found that both species up-regulated probesets enriched for gene sets related to skin repair: wound healing and apoptosis-related gene sets. Wound healing and apoptosis pathways, in particular, may have a more general role in host response to Bd as these pathways were also up-regulated in three out of four species in a previous study (Ellison *et al.* 2015). The coordinated up-regulation of these gene sets are likely to be linked to coping with the infection given their importance in skin repair. In addition, both species responded to infection with up-regulation of stress-related pathways. Some gene expression studies have found differential expression of stress pathways in skin and liver of susceptible species, while other studies have found increases in physiological markers of stress (i.e. corticosterone) during infection (Gabor *et al.* 2013; Peterson *et al.* 2013). More research is needed to further examine how stress pathways may locally interact with skin maintenance and immune-related pathways. As studies on resistant and susceptible species continue to mount, the generalities and subtleties of host response among species will emerge.

*Implications of immune-related gene expression responses*

We detected signals of immune-related responses in both species, but with different specific pathways in the sampled tissues. In the resistant *B. marinus* skin dataset, while we observed some expression changes related to immune function, we did not detect a robust immune or inflammatory response as expected given the resistant nature of this species. In the GOstats analysis, we observed up-regulation of immune regulatory pathways that typically have modulatory effects on immune response. In contrast, the susceptible *B. boreas* experienced stronger up-regulation of innate immune related pathways in the skin dataset. Recent studies have also documented different patterns of regulation of these pathways in different species (e.g., the complement pathway has been shown to be activated in response to Bd infection in some species (Ellison *et al.* 2015), but down-regulated in other species (Rosenblum *et al.* 2009, 2012). Although a general pattern of complement pathway activation has not emerged, the mounting evidence of differential expression supports a role – but not necessarily effective role – in host response.

From the *B. boreas* liver dataset analysis, we observed a striking pattern of down-regulation of immune response related pathways in contrast to the skin results. This pattern may be due to either host-mediated immune regulatory pathways or pathogen-mediated immunosuppression. In some cases, strong immune regulation can serve to modulate an ongoing immune response to limit collateral immunopathology. While this process is conceivable given the high Bd infection intensities, the lack of detectable expression changes in the spleen does not support this hypothesis as a coordinated splenic response would be predicted. Pathogen-mediated immunosuppression is also a possible mechanism, as this process has been documented in recent studies albeit in the context of spleen-derived cells. Previous work suggests that Bd factors inhibit *in vitro* growth of frog spleen-derived T-cells and B-cells (Fites *et al.* 2013) and Bd

infection can lead to down-regulation of immune response and T-cell activation pathways in spleens of infected frogs (Ellison *et al.* 2014, 2015). Interestingly, we observed expression patterns suggestive of immune suppression pathways in the *B. boreas* spleen dataset. Out of only 3 up-regulated probesets, one of the up-regulated probesets is annotated as an Fc receptor-like 2, which has immunomodulatory effects on B-cell development. In the gene set level analysis, down-regulation of immune cell signaling pathways was also suggestive of immunosuppression, albeit with weak statistical support. In addition, for the *B. marinus* liver dataset the only two significantly down-regulated probesets were annotated for MHC class 2, which is an antigen presentation protein necessary for T-cell activation. This pattern suggests either a reduction in the antigen processing and presentation process or a reduction in the number of antigen presenting cells in the tissue. These expression results for both species motivate more work to examine the effects of Bd factors on immune activation and suppression of liver- and spleen-derived cells.

The weak expression response for *B. marinus* was puzzling given the large difference in infection intensity dynamics between species. Two non-mutually exclusive mechanisms may explain the basis of resistance given the weak response we observed. First, *B. marinus* may mount an immediate innate immune response at the site of infection, which limits the colonization and proliferation of the pathogen. This process may involve cellular-mediated response (e.g. macrophage, neutrophils) or secreted antimicrobial peptides. The infection can then persist at low levels as Bd may continue to grow at a slow rate on the host or on skin sloughs off of the host. As the low level infection is contained by an innate immune response, the infection may not trigger a strong systemic adaptive immune response. Local responses maintain skin structural integrity while pathological inflammation is avoided due to weak or modulated pro-inflammatory signals. Second, certain structural or molecular features of the epidermis may reduce the ability of Bd to colonize and proliferate. Although Bd has been documented to infect a wide range of species, it is conceivable that infectivity varies among host species. Several factors may be involved including: lower binding capacity of Bd receptors for species-specific epidermal features; effects of mucus composition; and presence of bacterial communities that limit Bd growth. It is worth noting that the expression profiling results for Bd-resistant *A. callidryas* in a previous study provide the best comparison to date (Ellison *et al.* 2015). *Agalychnis callidryas* generated a generally weak expression response across tissues (although livers were not assayed) – as seen in *B. marinus*. However *A. callidryas* generated substantial splenic expression responses – in contrast to *B. marinus* – with evidence of down-regulated immune response pathways except for up-regulation of T-cell activation. The weak expression response observed in *A. callidryas* and in *B. marinus*, in the present study, motivates further exploration of the mechanistic basis of resistance at the host level, especially in light of recent lab-based work that shows evidence of adaptive immune response in some species (Ramsey *et al.* 2010; McMahon *et al.* 2014). Clearly more work in needed to examine the early stages of infection in resistant species in order to better understand how some species curb the infection process. In addition, we note that further studies involving more sampling time points (e.g. longer interval between inoculation and sampling) and other tissue types (e.g. thymus, kidney) may indeed reveal some form of systemic host response to Bd infection in *B. marinus*.

*Integrating histology and gene expression assays*

Our integrated use of histology and gene expression increased our ability to interpret the results of our analyses. For example, an unexpected histological finding was the observation of minimal to mild chronic lymphoplasmacytic dermatitis in all experimental groups – including the

control groups. Lymphoplasmacytic dermatitis refers to skin inflammation involving the elevated numbers of lymphocytes and plasma cells in tissue (ie. T cells and B cells). This evidence suggests that some experimental condition(s) unrelated to the Bd inoculation led to a weak inflammation reaction in the ventral skin. This pattern may have weakened our ability to uncover transcriptional changes associated with inflammation. However, some immune- and inflammation-related GO gene sets were enriched in both exposed treatment groups for skin samples suggesting that we still detected acute changes due to the Bd infection and not only due to a baseline condition. Another unexpected histological finding was the presence of an intradermal nematode in one control *B. marinus* host. Interestingly, microarray results from this individual revealed an outlier transcriptional profile, which may be explained by the helminthic infection. This individual was removed from the gene expression analyses described above. These observations highlight the importance of pairing transcriptomic assays with histologic examination to aid interpretation of gene expression results.

*Conservation implications*

Given the recent worldwide expansion of Bd, the difference in Bd infection intensity dynamics found in this study is likely to have ecological consequences for amphibian communities in the wild. The striking differences in disease progression between the two congeneric species, combined with differential susceptibility in multiple assays, which suggests that these species are at near opposite ends of the susceptibility-resistance spectrum – at the host level. While the *B. marinus* hosts reached lower Bd infection intensities, we note that the persistent infection in most hosts over 18 days may have transmission implications for wild populations. As asymptomatic carriers of Bd infection, *B. marinus* could function as spreaders of the disease to the environment and to other species. For *B. boreas*, the high infection intensities support the scenario for high transmission leading to rapid population collapse, as has been observed in the wild (Pilliod *et al.* 2010).

Further work in linking host response variability and epidemic outcome is likely to be fruitful to better understand mechanisms of resistance and develop potential management interventions. While accelerating such studies is needed to further characterize individual level host response, it will be critical to extend genomics methods into field-based studies in order to "ground truth" lab-based results and to investigate host-pathogen interactions in natural conditions. Research is this field will undoubtedly have a large role in studying and mitigating emerging infectious diseases of wildlife.

**Acknowledgements**

We thank Colorado Parks & Wildlife for providing surplus *Bufo boreas* for research; Karen Pohl, Isaiah Hoyer, and Jamie Voyles for assistance with lab work and animal husbandry; Tyler Hether, Denim Jochimsen, Stephen Spear, and Vanessa Wimmer for assistance with tissue collection; Cindy Carey for providing tissue samples for transcriptome sequencing; Matt Settles for assistance with microarray development; Rosenblum lab members and Steve Beissinger for constructive comments on the manuscript. All live animal work was conducted at the University of Idaho with the approval of the University of Idaho Animal Care and Use Committee. This study was supported by National Science Foundation grant IOS-13542421.

## Data Availability

Transcriptome sequencing data have been deposited to NCBI SRA (accession SRP065907, SRP065908). Microarray data have been deposited to NCBI GEO (accession GSE74788).

## References


Berger, L., A. D. Hyatt, R. Speare, and J. E. Longcore, 2005 Life cycle stages of the amphibian chytrid Batrachochytrium dendrobatidis. Dis. Aquat. Organ. 68: 51–63.

Berger, L., R. Speare, P. Daszak, D. E. Green, A. A. Cunningham *et al.*, 1998 Chytridiomycosis causes amphibian mortality associated with population declines in the rain forests of Australia and Central America. Proc. Natl. Acad. Sci. 95: 9031–9036.

Berger, L., R. Speare, A. Pessier, J. Voyles, and L. F. Skerratt, 2010 Treatment of chytridiomycosis requires urgent clinical trials. Dis. Aquat. Organ. 92: 165–174.

Blaustein, A. R., J. M. Romansic, E. A. Scheessele, B. A. Han, A. P. Pessier *et al.*, 2005 Interspecific variation in susceptibility of frog tadpoles to the pathogenic fungus Batrachochytrium dendrobatidis. Conserv. Biol. 19: 1460–1468.

Boyle, D. G., D. B. Boyle, V. Olsen, J. A. T. Morgan, and A. D. Hyatt, 2004 Rapid quantitative detection of chytridiomycosis (Batrachochytrium dendrobatidis) in amphibian samples using real-time Taqman PCR assay. Dis. Aquat. Organ. 60: 141–148.

Briggs, C. J., R. A. Knapp, and V. T. Vredenburg, 2010 Enzootic and epizootic dynamics of the chytrid fungal pathogen of amphibians. Proc. Natl. Acad. Sci. U. S. A. 107: 9695–9700.

Carey, C., J. E. Bruzgul, L. J. Livo, M. L. Walling, K. A. Kuehl *et al.*, 2006 Experimental Exposures of Boreal Toads (Bufo boreas) to a Pathogenic Chytrid Fungus (Batrachochytrium dendrobatidis). Ecohealth 3: 5–21.

Carvalho, B. S., and R. a. Irizarry, 2010 A framework for oligonucleotide microarray preprocessing. Bioinformatics 26: 2363–2367.

Chou, H. H., and M. H. Holmes, 2001 DNA sequence quality trimming and vector removal. Bioinformatics 17: 1093–104.

Conesa, A., S. Götz, J. M. García-Gómez, J. Terol, M. Talón *et al.*, 2005 Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674–3676.

Ellison, A. R., A. E. Savage, G. V. DiRenzo, P. Langhammer, K. R. Lips *et al.*, 2014 Fighting a losing battle: Vigorous immune response countered by pathogen suppression of host defenses in the chytridiomycosis-susceptible frog Atelopus zeteki. G3 Genes Genomes Genet. 4: 1275–1289.

Ellison, A. R., T. Tunstall, G. V DiRenzo, M. C. Hughey, E. A. Rebollar *et al.*, 2015 More than skin deep: functional genomic basis for resistance to amphibian chytridiomycosis. Genome Biol. Evol. 7: 286–98.

Falcon, S., and R. Gentleman, 2007 Using GOstats to test gene lists for GO term association. Bioinformatics 23: 257–258.

Fisher, M. C., and T. W. J. Garner, 2007 The relationship between the emergence of Batrachochytrium dendrobatidis, the international trade in amphibians and introduced amphibian species. Fungal Biol. Rev. 21: 2–9.

Fites, J. S., J. P. Ramsey, W. M. Holden, S. P. Collier, D. M. Sutherland *et al.*, 2013 The invasive chytrid fungus of amphibians paralyzes lymphocyte responses. Science 342: 366–9.

Gabor, C. R., M. C. Fisher, and J. Bosch, 2013 A Non-Invasive Stress Assay Shows That Tadpole Populations Infected with Batrachochytrium dendrobatidis Have Elevated Corticosterone Levels. PLoS One 8: 1–5.

Garcia-Porta, J., S. N. Litvinchuk, P. a. Crochet, a. Romano, P. H. Geniez *et al.*, 2012 Molecular phylogenetics and historical biogeography of the west-palearctic common toads (Bufo bufo species complex). Mol. Phylogenet. Evol. 63: 113–130.

Gervasi, S., C. Gondhalekar, D. H. Olson, and A. R. Blaustein, 2013 Host Identity Matters in the Amphibian-Batrachochytrium dendrobatidis System: Fine-Scale Patterns of Variation in Responses to a Multi-Host Pathogen. PLoS One 8.:

Hänzelmann, S., R. Castelo, and J. Guinney, 2013 GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 14: 7.

Harris, R. N., R. M. Brucker, J. B. Walke, M. H. Becker, C. R. Schwantes *et al.*, 2009 Skin microbes on frogs prevent morbidity and mortality caused by a lethal skin fungus. ISME J 3: 818–824.

James, T. Y., L. F. Toledo, D. Rödder, D. da Silva Leite, A. M. Belasen *et al.*, 2015 Disentangling host, pathogen, and environmental determinants of a recently emerged wildlife disease: lessons from the first 15 years of amphibian chytridiomycosis research. Ecol. Evol. 5: 4079–4097.

Kauffmann, A., R. Gentleman, and W. Huber, 2009 arrayQualityMetrics - A bioconductor package for quality assessment of microarray data. Bioinformatics 25: 415–416.

Li, W., and A. Godzik, 2006 Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22: 1658–1659.

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader *et al.*, 2006 Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. Nat. Biotechnol. 437: 376–380.

McMahon, T. A., B. F. Sears, M. D. Venesky, S. M. Bessler, J. M. Brown *et al.*, 2014 Amphibians acquire resistance to live and dead fungus overcoming fungal immunosuppression. Nature 511: 224–227.

Murray, K. A., R. W. R. Retallick, R. Puschendorf, L. F. Skerratt, D. Rosauer *et al.*, 2011 Assessing spatial patterns of disease risk to biodiversity: Implications for the management of the amphibian pathogen, Batrachochytrium dendrobatidis. J. Appl. Ecol. 48: 163–173.

Peterson, J. D., J. E. Steffen, L. K. Reinert, P. a. Cobine, A. Appel *et al.*, 2013 Host Stress Response Is Important for the Pathogenesis of the Deadly Amphibian Disease, Chytridiomycosis, in Litoria caerulea. PLoS One 8: 1–7.

Pilliod, D. S., E. Muths, R. D. Scherer, P. E. Bartelt, P. S. Corn *et al.*, 2010 Effects of amphibian chytrid fungus on individual survival probability in wild boreal toads. Conserv. Biol. 24: 1259–67.

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and R. C. Team, 2014 nlme: Linear and Nonlinear Mixed Effects Models. R Packag. version 3.1-118.

Ramsey, J. P., L. K. Reinert, L. K. Harper, D. C. Woodhams, and L. A. Rollins-Smith, 2010 Immune Defenses against Batrachochytrium dendrobatidis, a Fungus Linked to Global Amphibian Declines, in the South African Clawed Frog, Xenopus laevis. Infect. Immun. 78: 3981–3992.

Retallick, R. W. R., and V. Miera, 2007 Strain differences in the amphibian chytrid Batrachochytrium dendrobatidis and non-permanent, sub-lethal effects of infection. Dis. Aquat. Organ. 75: 201–207.

Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law *et al.*, 2015 limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43: e47.

van Rooij, P., A. Martel, K. D'Herde, M. Brutyn, S. Croubels *et al.*, 2012 Germ tube mediated invasion of Batrachochytrium dendrobatidis in Amphibian skin is host dependent. PLoS One 7: 1–8.

Rosenblum, E. B., T. J. Poorten, M. Settles, and G. K. Murdoch, 2012 Only skin deep: shared genetic response to the deadly chytrid fungus in susceptible frog species. Mol. Ecol. 21: 3110–20.

Rosenblum, E. B., T. J. Poorten, M. Settles, G. K. Murdoch, J. Robert *et al.*, 2009 Genome-Wide Transcriptional Response of Silurana (Xenopus) tropicalis to Infection with the Deadly Chytrid Fungus (G. Butler, Ed.). PLoS One 4: 10.

Searle, C. L., S. S. Gervasi, J. Hua, J. I. Hammond, R. A. Relyea *et al.*, 2011 Differential host susceptibility to Batrachochytrium dendrobatidis, an emerging amphibian pathogen. Conserv. Biol. 25: 965–74.

Skerratt, L. F., L. Berger, R. Speare, S. Cashins, K. R. McDonald *et al.*, 2007 Spread of Chytridiomycosis Has Caused the Rapid Global Decline and Extinction of Frogs. Ecohealth 4: 125–134.

Supek, F., M. Bošnjak, N. Škunca, and T. Šmuc, 2011 REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One 6: e21800.

Voyles, J., E. B. Rosenblum, and L. Berger, 2011 Interactions between Batrachochytrium dendrobatidis and its amphibian hosts: a review of pathogenesis and immunity. Microbes Infect. Inst. Pasteur 13: 25–32.

Voyles, J., S. Young, L. Berger, C. Campbell, W. F. Voyles *et al.*, 2009 Pathogenesis of chytridiomycosis, a cause of catastrophic amphibian declines. Science 326: 582–585.

Vredenburg, V. T., R. A. Knapp, T. S. Tunstall, and C. J. Briggs, 2010 Dynamics of an emerging disease drive large-scale amphibian population extinctions. Proc. Natl. Acad. Sci. U. S. A. 107: 9689–94.

Wake, D. B., and V. T. Vredenburg, 2008 Are we in the midst of the sixth mass extinction ? A view from the world of amphibians. Proc. Natl. Acad. Sci. U. S. A. 105: 11466–11473.

Woodhams, D. C., K. Ardipradja, R. A. Alford, G. Marantelli, L. K. Reinert *et al.*, 2007 Resistance to chytridiomycosis varies among amphibian species and is correlated with skin peptide defenses. Anim. Conserv. 10: 409–417.

## Tables

**Table 1: Summary of gene expression results from GO stats enrichment analysis**

| Group | No. of DE Probesets[a] | No. of Enriched GO Terms[b] | Selected Enriched GO Terms |
|---|---|---|---|
| **Skin: Up-regulated** | | | |
| *B. marinus* | 487 | 66 | epidermis development; wound healing; cell proliferation; apoptotic signaling pathway; response to stress; metabolic process; biological adhesion; immune system development |
| *B. boreas* | 1108 | 68 | regulation of complement activation; response to stress; wound healing; cell redox homeostasis; response to external stimulus; response to yeast; hematopoietic or lymphoid organ development; leukocyte migration; apoptotic process; cellular metabolic process; innate immune response; coagulation |
| **Skin: Down-regulated** | | | |
| *B. marinus* | 70 | 2 | cellular localization; metabolic process |
| *B. boreas* | 1055 | 58 | collagen catabolic process; extracellular structure organization; blood vessel development; response to wounding; hemostasis; cell-matrix adhesion; tissue development; epithelium development; actin cytoskeleton organization |
| **Liver: Up-regulated** | | | |
| *B. marinus* | 0 | NA | - |
| *B. boreas* | 1947 | 70 | vesicle-mediated transport; protein folding; regulation of cell cycle; cellular metabolic process; RNA processing; cellular respiration; cell proliferation; response to stress |
| **Liver: Down-regulated** | | | |
| *B. marinus* | 2 | NA | - |
| *B. boreas* | 951 | 74 | actin filament organization; wound healing; response to stress; immune system process; complement activation; antigen processing and presentation of peptide antigen; innate immune response; blood coagulation; blood vessel morphogenesis |
| **Spleen: Up-regulated** | | | |
| *B. marinus* | 0 | NA | - |
| *B. boreas* | 0 | NA | - |
| **Spleen: Down-regulated** | | | |
| *B. marinus* | 0 | NA | - |
| *B. boreas* | 3 | NA | - |

a: Differential expression (DE) threshold criteria for B. marinus was BH-corrected p value < 0.1; and for B. boreas was BH-corrected p value < 0.05

b: list of "Biological Process" category GO terms was reduced using Revigo to remove semantic redundancies

**Table 2. Between species comparative analysis of gene set expression patterns**

| **Differential expression directions** | **GO term** | **GO description** |
|---|---|---|
| *B. marinus* up; *B. boreas* down | GO:0008104 | protein localization |
| | GO:0006139 | nucleobase-containing compound metabolic process |
| | GO:0001889 | liver development |
| | GO:0031290 | retinal ganglion cell axon guidance |
| | GO:0001501 | skeletal system development |
| | GO:0060070 | canonical Wnt signaling pathway |
| | GO:0007160 | cell-matrix adhesion |
| | GO:0009887 | organ morphogenesis |
| | GO:0030866 | cortical actin cytoskeleton organization |
| | GO:0007015 | actin filament organization |
| | GO:0048699 | generation of neurons |
| | GO:0001843 | neural tube closure |
| | GO:0061314 | Notch signaling involved in heart development |
| *B. marinus* up; *B. boreas* up | GO:0016043 | cellular component organization |
| | GO:0016570 | histone modification |
| | GO:0019827 | stem cell maintenance |
| | GO:0048583 | regulation of response to stimulus |
| *B. marinus* down; *B. boreas* up | GO:0022904 | respiratory electron transport chain |

## Figures

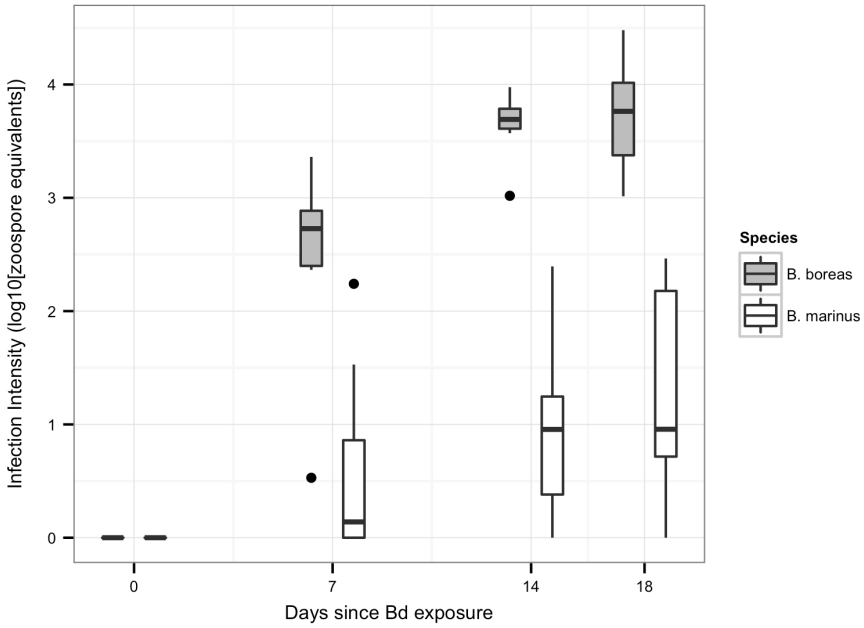



**Figure 1. Time course of Bd infection intensities in *B. boreas* and *B. marinus*.**
Boxplots of infection intensities are in units of $\log_{10}$ zoospore equivalents measured via qPCR on days 7, 14, and 18 after Bd exposure. Results are shown for only the Bd-exposed groups.

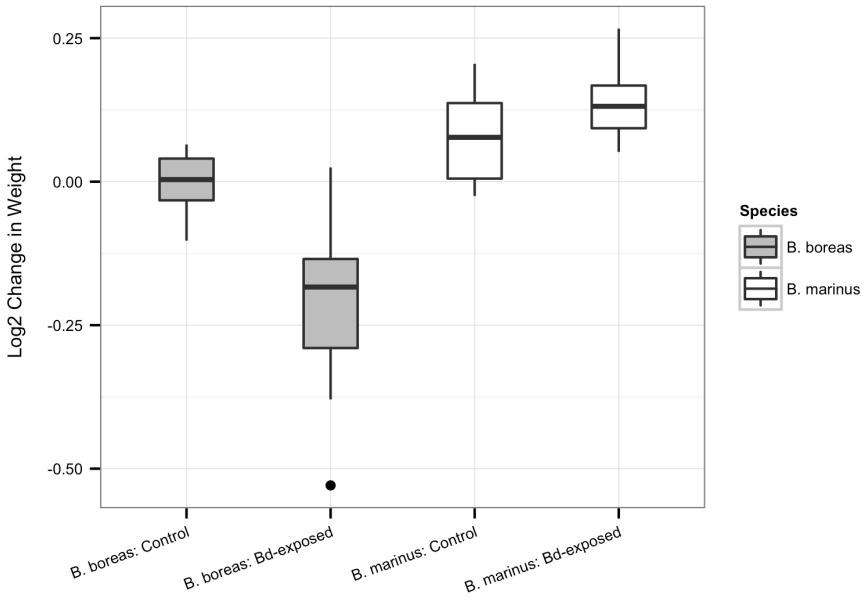



**Figure 2. Bd exposure led to loss in body weight in *B. boreas* only.**
Boxplot of change in body weight in units of $\log_2(\text{weight}_{\text{day}18} / \text{weight}_{\text{day}0})$ for Bd-exposed and control groups for both species.

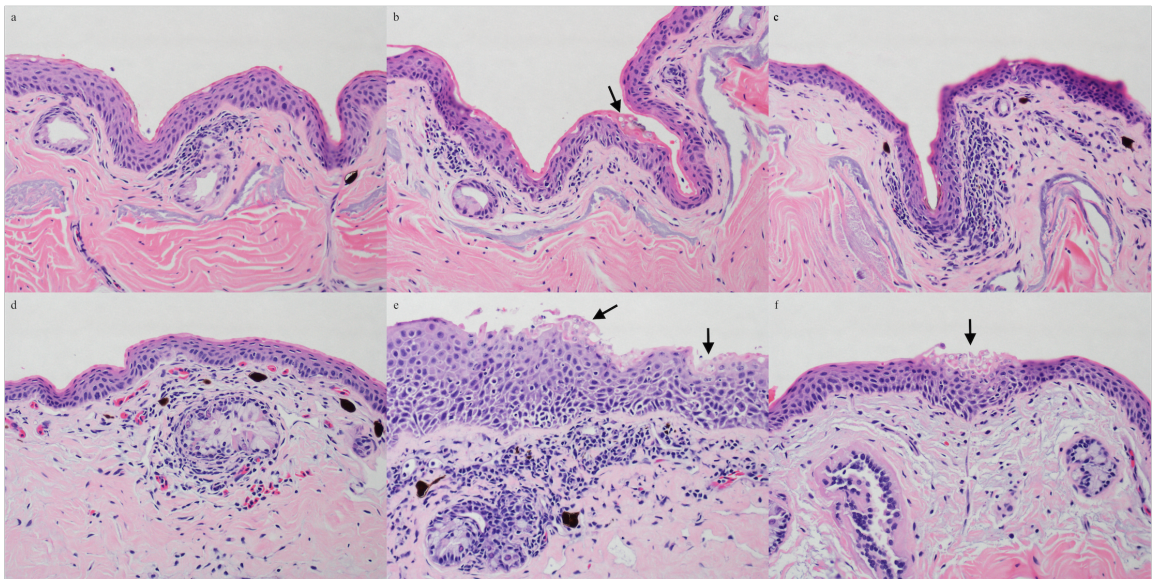



**Figure 3. Histology micrographs at 20X magnification showing stained skin sections.** (a) *B. marinus* control group, (b,c) *B. marinus* Bd-exposed group, (d) *B. boreas* control group, and (e,f) *B. boreas* Bd-exposed group. The epidermis is oriented upward in each panel. Arrows indicate sites of Bd infection.

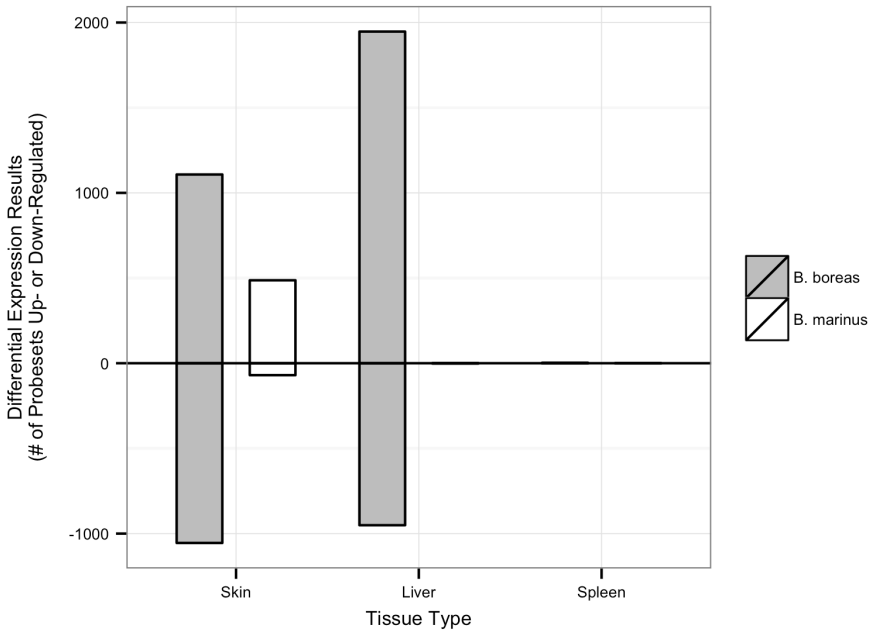



**Figure 4. Summary of gene expression microarray results for skin, liver, and spleen in both species.**

Bars show number of differentially expressed probesets after correcting of multiple tests with Benjamini-Hochberg p-value adjustment. Positive values represent up-regulated probesets in the Bd-exposed group relative to the control group; while negative values represent down-regulated probesets. Note that different alpha values used for each species: alpha = 0.05 for *B. boreas*; alpha = 0.1 for *B. marinus*.

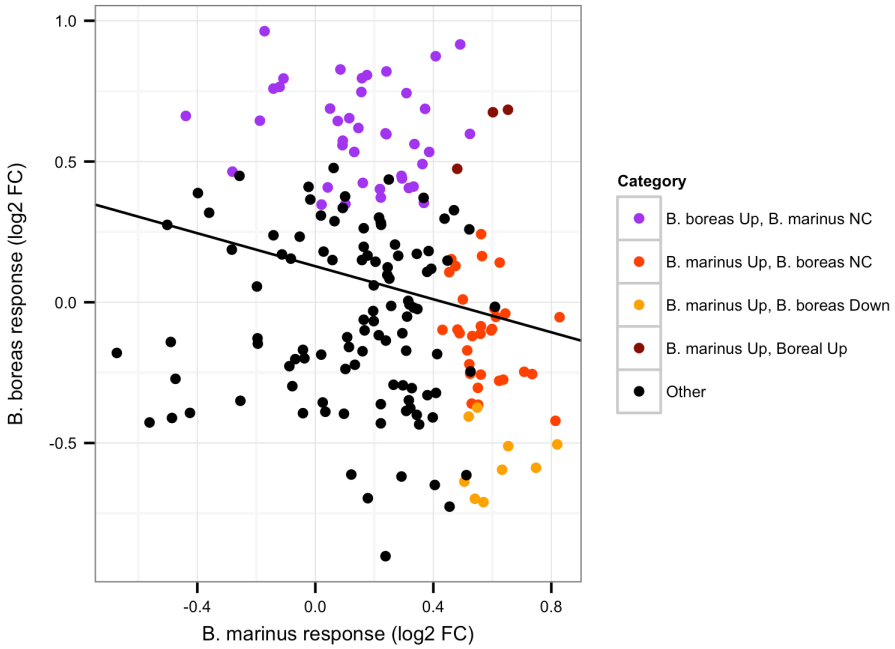



**Figure 5. Comparison of gene set level differential expression between species.**
Points represent $\log_2$ fold change ratios for gene sets and are colored by differential expression category as defined in the legend. NC: no change in expression; $\log_2$ FC: $\log_2$ fold change ratios.

## Chapter 3: Population genetic structure of the endangered Sierra Nevada yellow-legged frog (*Rana sierrae*) in Yosemite National Park based on multi-locus nuclear markers


### Abstract

The mountain yellow-legged species complex (*Rana sierrae* and *Rana muscosa*) has declined precipitously in distribution and abundance during the last century. The two primary threats are chytrid epidemic-associated population collapses and predation from the introduction of non-native trout. Widespread declines have occurred throughout the range of these species, including populations of *R. sierrae* in Yosemite National Park. A clear picture of genetic structure of remaining Yosemite *R. sierrae* populations is critical to short-term management and conservation. We conducted a population genetics study that included samples from 23 geographic sites distributed throughout the range of *R. sierrae* in Yosemite NP. We used non-invasive swab samples to collect genetic data from mitochondrial and nuclear DNA via sequencing and analyzed the distribution of genetic variation in a geographic context. From our mtDNA analysis, our results confirm a previous study's result that two haplotype groups occur in Yosemite. However, we discovered that the two haplotypes are not completely geographically partitioned into the two main drainages (Merced drainage and Tuolumne drainage), as previously postulated. From our nuclear DNA analysis, we found a general pattern a genetic isolation-by-distance in the dataset, where genetic differentiation was correlated with geographic distance between sites. In addition, our analyses suggest that three clusters of genetically cohesive sites occur in the study area. Understanding population genetic patterns of variability will inform management strategies such as translocations, reintroductions, and monitoring for this endangered frog.


### Introduction

Historically, mountain yellow-legged frogs (*Rana muscosa* and *Rana sierrae*) were one of the most abundant vertebrate species in the high elevation Sierra Nevada (Grinnell and Storer 1924). *Rana muscosa/sierrae* currently are found at only 7% of their historical localities, despite the remote and undeveloped characteristics of the associated landscape (e.g. national parks, wilderness areas) (Vredenburg *et al.* 2007). As a consequence of this precipitous decline, both species are now listed as "endangered" under the federal Endangered Species Act, while the California Endangered Species Act lists *R. sierrae* as "threatened" and *R. muscosa* in the Sierra Nevada is listed as "endangered". The two primary drivers of population declines are the introduction of non-native predatory trout in lakes and streams, and the arrival of the infectious disease chytridiomycosis caused by the amphibian chytrid fungus (*Batrachochytrium dendrobatidis*; "Bd"). Starting in the mid-1800s trout were introduced into thousands of naturally fishless lakes in the Sierra Nevada, many of which contained native *R. muscosa/sierrae* populations. Co-occurrence of introduced trout and frog populations typically resulted in subsequent frog population decline and extirpation due to predation by the non-native trout (Knapp and Matthews 2000; Knapp 2005). Although fish stocking ceased in National Parks by the early 1990s, many introduced fish populations persist in aquatic habitats that would otherwise be suitable habitat for *R. muscosa/sierrae* (e.g., lakes, low-gradient streams). The effect of Bd on mountain yellow-legged frogs provides one of the most striking examples of chytrid-associated amphibian population declines in North America. During the past several decades Bd has spread across the Sierra Nevada and caused the decline or extirpation of

hundreds of mountain yellow-legged frog populations (Vredenburg *et al.* 2010). In response to these population declines, academic scientists and natural resource managers from state and federal agencies have led several management efforts including population monitoring, non-native fish removal, Bd surveys, captive breeding, and bacterial augmentation experiments to reduce Bd infection intensities.

A thorough understanding of the genetic structure of remaining *R. muscosa/sierrae* populations is critical to management and conservation of these endangered species. Information on population structure will be important to determine if subpopulations are demographically independent and therefore warrant consideration as separate management units. This information has implications for making informed decisions for identifying suitable donor populations from which frogs are collected as part of future efforts to (1) re-establish frogs in areas from which they were previously extirpated and (2) provide supplemental frogs to areas with low population size. Previous work by Vredenburg et al. (2007) analyzed the phylogeography of the *R. muscosa*/sierrae species complex using a mitochondrial DNA (mtDNA) marker. This analysis found three mtDNA clades of each species (five clades in the Sierra Nevada, one clade in southern California) in a pattern concordant with geographic distribution of the populations. Although this study provided important insights into the genetic structure of this species complex, for several reasons it nonetheless likely provides an incomplete picture of *R. muscosa/sierrae* population structure in the Sierra Nevada – and Yosemite National Park in particular. First, this study relied on a single mtDNA marker and may not be representative of results obtained from a broader sampling of the nuclear genome. Second, the different nature of inheritance of mtDNA versus nuclear DNA (mtDNA is maternally inherited) can sometimes lead to discordant inferences when mtDNA is used in population genetic analyses. Third, the number of samples available for the Sierra Nevada was relatively low and may therefore have missed important genetic structure. For example, *R. sierrae* in Yosemite National Park was represented by only seven samples, leaving the majority of the park's geography unrepresented.

The present study seeks to better resolve the genetic structure of *R. sierrae* populations in Yosemite National Park using multi-locus nuclear and mitochondrial genetic data. We sampled frogs from 23 sites spread across Yosemite, with samples collected from 3-15 adult frogs per site. Vredenburg et al. (2007) found that Yosemite contains two mtDNA clades, which are separated into the two drainages in Yosemite: Tuolumne River drainage in the northern region and Merced River drainage in the southern region. We test this *a priori* two-population hypothesis with our dataset that contains a larger sample size and multi-locus genetic data. We also use genetic clustering methods that are agnostic to sampling location in order to let the genetic data reveal the structure of populations in Yosemite. We then analyze the spatial distribution of genetic variation using spatially explicit methods. Our results show *R. sierrae* populations follow trend of isolation-by-distance and contain substantial within-population admixture. These results suggest that using frogs from several nearby, within-cluster populations in translocations could be a reasonable strategy to re-establish extirpated *R. sierrae* populations that retain as much of the regional genetic diversity as possible.

## Methods

*Sampling Scheme*

For genetic samples of *R. sierrae* individuals, we utilized archived genetic samples obtained from skin swabs that were previously collected for monitoring Bd infection in *R. sierrae* populations in Yosemite National Park (results reported in Briggs et al. 2010,

Vredenburg et al. 2010). In addition, we supplemented the swab archive by collecting additional swab samples from under-sampled geographic areas of Yosemite. For nuclear DNA sequencing, our sampling scheme included 192 individuals from 23 *R. sierrae* populations with 3-15 samples per location (Fig. 1, Table 1). Figure 1 shows the sample grouping scheme, where the sampling sites are labeled with numbers that are roughly ordered in a clockwise pattern on the landscape. Sites represent either single sample locations (e.g. a single lake) or contain groupings of sample locations within relatively close proximity. We slightly revised the site numbering scheme after performing the genetic data analyses to reflect the spatial pattern of variation. We note that genetic clustering data analyses (described below) were agnostic to sampling site assignment, and therefore avoided biased grouping of samples. This swab archive represented an invaluable resource for collecting genetic data given the extensive sampling in the cited previous studies, which includes some sampling locations that suffered subsequent population collapses. However, collecting genetic data from archived swab extracts is challenging given the generally low quality and purity of nucleic acids in such samples. We therefore chose to use a targeted sequencing approach where nuclear genomic markers, developed from a transcriptome dataset, were PCR amplified in a high throughput platform and subsequently sequenced with next-generation sequencing.

*Nucleic acid preparation and purification*

Swabs were stroked 30 times on the dorsal and ventral surface of the sampled frogs, and nucleic acids were extracted from swabs with the Prepman Ultra reagent (Hyatt *et al.* 2007). To further purify DNA from the swab extracts, we used an isopropanol precipitation protocol as described in the Prepman Ultra manual appendix. Briefly, we centrifuged the extract at 16,000 x *g* for 2 minutes to pellet debris, and moved the supernatant to a fresh low-adhesion tube. We added low TE buffer to a final volume of 447.5 uL (10 mM Tris-HCl, 0.1 mM EDTA; note that EDTA concentration is lower than typical in order to avoid PCR inhibition). Then we added 2.5 uL of 20 ug/uL glycogen and 50 uL of 3M NaOAc, pH 5.2, gently mixed, then added 500 uL of isopropanol, and gently mixed again. After 20 minutes of room temperature incubation, we centrifuged the tubes at 13,000 x *g* for 10 minutes at room temperature. We discarded the supernatant, then washed the pellet twice with 500 uL 70% ethanol. Each wash included a 5 minute incubation and a 30 second centrifugation at 13,000 x *g*. After letting the residual ethanol evaporate for approximately 5 minutes, we added 25 uL low TE buffer and incubated the samples overnight to allow thorough resuspension of the pellet.

*Mitochondrial DNA sequencing*

Our initial investigation of genetic variation included mtDNA sequencing using the locus previously developed by Vredenburg et al.: ND2 fragment (2007). This was done to allow direct comparison between our results and those of the Vredenburg et al. study. We followed the PCR protocol in Vredenburg et al. (2007) and used the conventional Sanger sequencing platform at the UC Berkeley DNA Sequencing Facility. Using the sequence analysis software Geneious, we quality trimmed the raw read data and aligned the sequences with MUSCLE algorithm (Edgar 2004; Kearse *et al.* 2012). After removing low quality sequences, we analyzed this 941 bp marker in 106 individuals spread across the 23 sampling sites. After removing duplicate haplotypes present within sampling sites, we then inferred the phylogeny of 41 haplotypes using a maximum likelihood approach in the program RAXML (Stamatakis 2006). We used the "GTRCAT" model of sequence evolution, identified the optimal tree from 20 separate ML searches, and conducted 500 bootstrapping replicates to evaluate node support (parameters: '-f d -b 500'). We used three reference sequences from Genbank for the outgroup (AF314027,

AF314029, AF314030), which includes one frog from Sixty Lakes Basin in Kings Canyon National Park (*R. muscosa*) and two frogs from southern California Transverse Range localities (*R. muscosa*).

*Nuclear genetic marker development*

We designed primers for 50 nuclear targets in a de novo transcriptome assembly of *Rana muscosa/sierrae*. The targets primarily included 3' terminal exons and flanking 3' untranslated regions (UTR). We targeted UTR regions due to (1) the relatively long length among spliced transcript regions and (2) the relatively high polymorphism rate among transcript regions. We began by generating de novo contigs (putative transcripts) using the mira assembler (Chevreux *et al.* 2004) from RNAseq 454 next-generation sequencing reads previously generated by Rosenblum et al. (2012). To identify 3' UTR sequences in *R. sierrae* contigs, we followed a previously developed approach where exon alignments are generated from reference genomes (Zieliński *et al.* 2014). We identified exon-intron boundaries by aligning *R. sierrae* contigs to two reference transcriptomes: *Xenopus tropicalis* (Ensembl assembly v. 4.2) and *Anolis carolensis* (Ensembl assembly v. 2.0). We downloaded reference transcripts for *X. tropicalis* and *A. carolensis* from Biomart using the "cDNA" setting with 500bp of 3' downstream flanking sequence (Smedley *et al.* 2009). We used the blastclust program to generate alignments using the following parameters: 70% identity threshold, 50% overlap length, require coverage on both neighbors = FALSE. The resulting sequence alignment clusters were parsed in R and prepared for importation into Geneious (Kearse et al. 2012). We prioritized alignment clusters with single hits in each reference genome to limit selection of genes prone to paralogy. To prepare *Xenopus* and *Anolis* transcript annotation information for these clusters, the gene annotation files (in "gtf" format) were downloaded from Ensembl and parsed in R for importation into Geneious. Working in Geneious, we selected *R. sierrae* contigs with either a long exon or a long 3' UTR (> 300 bp) on a cluster-by-cluster basis. The exon-intron boundaries were mapped onto each cluster alignment using annotation information from the reference sequences. The final set of targets included 50 genomic targets. We then designed primers according to the manual for the target amplification platform (Fluidigm Access Array).

*Sequencing approach*

Our sequencing approach utilized microfluidic PCR amplification of genomic targets (Fluidigm Access Array 48.48) and next-generation sequencing (Illumina Miseq). The main advantages of the Access Array platform are the high throughput amplification of targets in 48 samples (in isolated microfluidic chambers) and the production of a sequencer-ready DNA library during the target amplification, which includes simultaneous sequencing adapter incorporation and barcode tagging of each amplicon. Access Array amplification and sequencing were performed at the University of Idaho IBEST Genomic Resources Core. To enrich the target regions prior to microfluidic PCR in the Access Array, we performed preamplification reactions for each sample according to manufacturer's protocol with slight modification. Briefly, three modifications included using untagged primers (lacking "common sequence" tags used for incorporating adapter), maintaining the annealing temperature at 60°C for all cycles, and adding one cycle to the PCR preamplification thermocycling protocol. Preamplification products were then cleaned with Exosap-it and diluted 1:5 in accordance with the manufacturer's protocol. We then submitted the "preamplified" samples for Access Array amplification and sequencing. Given that the Access Array platform carries out PCR amplifications in a microfluidic grid of 48 samples x 48 primer pair pools, our primer pooling scheme included 2 2-plex primer pools and 46 1-plex primer pools. For the Illumina Miseq sequencing run, we used 2 x 300 bp paired-end

reads from ¼ of the sequencing plate (~4.5 million reads), which was sufficient to generate ~470X coverage for each unique amplicon (includes every combination of sample and target). Although this sequencing coverage might be excessive in many applications, given the variable quality of our samples we expected that the high number of reads used would minimize per-sample missing data.

*Raw sequence processing and variant calling*

The bioinformatics pipeline starting from raw sequencing reads included adapter and primer trimming, variant calling, and variant filtering and phasing. We used the dbcAmplicons software (https://github.com/msettles/dbcAmplicons) for trimming adapter and primer sequences from the raw data. Downstream bioinformatics steps through variant calling were performed for each sample. The paired-end reads were merged to yield extended reads that spanned the length of the target using the flash2 software (ref). We used bwa software ("mem" mode) to align the reads to the reference targets (Li and Durbin 2009). We then followed the GATK ver. 3.4 best practices pipeline for calling variants (SNPs and INDELs) (Van der Auwera *et al.* 2013). The HaplotypeCaller tool in GATK calls SNPs and INDELs using local re-assembly of haplotypes for each sample. We then merged variant calls across all samples with the GenotypeGVCFs tool in GATK, and performed several levels of filtering to the raw variant calls. Our downstream filtering and analyses included the SNP calls only. We filtered SNP sites using standard quality control parameters (BaseQRanksSum < -5, MQRankSum < -3, ReadPosRankSum < -4, AlleleNumber < 200). Next we removed six targets that contained two types of evidence for paralogy: (1) an excess in heterozygous calls across the majority samples and (2) an excess of mapped reads across the majority of samples. Additionally one target was removed due to low PCR success rate (>10X coverage in only 46% of samples). To determine phase of the variants, we used the ReadBackedPhasing tool in GATK for each individual. We removed samples from downstream analyses that contained a high proportion of missing data (>50%), which left 164 samples in the dataset for downstream analyses.

*Genetic clustering*

We first investigated genetic variation among samples using principal components analysis (PCA) and the Bayesian clustering method STRUCTURE. Both of these methods can identify genetic clusters of samples in the dataset. We used the R function *prcomp* to compute principle component (PC) scores, and then plotted the first two PCs with sampling locations used as labels. Biallelic genotypes were encoded as: 0 for homozygous reference allele; 1 for heterozygous; and 2 for homozygous alternative allele. Missing genotypes were changed to the mean genotype per SNP. Next we conducted an analysis using STRUCTURE (ver. 2.3.4), a multi-locus clustering program, to infer the presence of distinct populations in the dataset (Pritchard *et al.* 2000). This program uses a model to cluster individuals into populations that are in Hardy-Weinberg equilibrium and allows for admixture within individuals. We used Evanno's delta K method to select the best value of number of populations (K) with the program StructureHarvester (Evanno *et al.* 2005; Earl and vonHoldt 2012). Using the phased haplotypes encoded as alleles, we ran the program under the admixture model ten times for each value of K = {2,3,4,5,6}. Each run included 20,000 MCMC steps with 10,000 steps as burn-in. Using the best value of K, we ran five long runs to assess population assignments and admixture estimates with 100,000 MCMC steps with 10,000 steps as burn-in. The results were summarized and plotted using the Clumpak program (Kopelman *et al.* 2015).

*Spatial analysis of genetic variation*

We used spatial PCA (sPCA) to analyze between-individual genetic variability in a spatially explicit context (Jombart *et al.* 2008). We utilized the sPCA pipeline in the adegenet R package in order to identify spatial autocorrelation in allele frequencies in a georeferenced dataset. We used the Euclidean distance method to create a network map used to define distances between sampling locations. Following Jombart et al. (2008), we first tested for spatial autocorrelation in conventional PCA scores using Moran's I test. Next we used the sPCA framework to conduct global (e.g. clines or patches) and local (e.g. differentiation between neighbors) tests and examined significant results by mapping spatial eigenvectors on to the geographic map. We then analyzed the hierarchical partitioning of genetic variation using spatial AMOVA, as implemented for multi-locus datasets in the spads program (Dellicour and Mardulyn 2013). We defined hierarchical levels by sampling sites and the clusters inferred from STRUCTURE using K = 3.

*Isolation by Distance test*

We tested for Isolation by Distance (IBD) in the dataset, which is a positive relationship between geographic distance and genetic differentiation for pairwise comparisons of populations. We used linearized Fst (ie. 1/(1-Fst)) as the genetic differentiation measure for sampling location. Fst was estimated using the multilocus weighted Fst method implemented in the snpStats R package (Clayton 2014). We used a Mantel test to assess the relationship between genetic and geographic pairwise distance matrices, using 10,000 permutations to estimate the significance value. Furthermore using a Partial Mantel test, we tested for the effect of a barrier between the two drainages in Yosemite: Tuolumne River drainage and Merced River drainage. The potential barrier is formed largely by the Cathedral Range ridgeline. We also used a Partial Mantel test to separate the effect of Euclidean distance from the population assignments, based on STRUCTURE results. In addition, we used multiple matrix regression (MMR) with permutation tests as another method to assess the effects of geography, drainage, and population assignment on genetic distance (Wang 2013). We repeated the regression analysis with within-drainage and between-drainage population comparisons in separate models in order to test for differences in IBD trends that may reflect the presence of a barrier between drainages. Partial Mantel and MMR computations were performed using functions in the PopGenReport R package (Adamack and Gruber 2014). Finally, we constructed a neighbor-joining tree based on pairwise Fst to graphically illustrate genetic differentiation patterns among populations with the phangorn package in R (Schliep 2011).

## Results

*Mitochondrial DNA analysis*

The general pattern in the mtDNA phylogenetic tree is a bipartition between Tuolumne R. and Merced R. drainages, with no strong pattern of substructure within drainages (Fig. 2). However two sites violate the general pattern in the tree with a split between two drainage-specific haplotype groups. Sampled individuals from Site 10 (Unicorn Lake) carried Merced-specific haplotypes despite being situated in Tuolumne drainage. Our sampling included haplotypes from 6 individuals with 1 haplotype represented in the tree (due to identical haplotypes among individuals). The other exception involved Site 14 (near Gallison Lk.) where the reverse discordant pattern occurred: some individuals located in the Merced drainage carried Tuolumne-specific haplotypes. In this case, there was a mixture of haplotypes, in which 15 out of 17 individuals carried Tuolumne haplotypes. We note that this set of samples combines two

surveys: in 2006 4 out of 6 carried Tuolumne haplotypes; in 2014 11 out of 11 carried Tuolumne haplotypes. Considering these two exceptions suggests that between-drainage migration occurs (or occurred) at least occasionally near the drainage boundary. In our sampling scheme for the nuclear genetic sequencing, we include slightly higher sample sizes for sites near the boundary when possible to increase the chance of detecting between drainage connectivity.

*Nuclear Genetic Clustering*

The nuclear SNP dataset included 1,251 SNPs across 43 targets after variant filtering with an average of 29 +/- 16 (mean +/- sd) SNPs per contig; and after removing singletons 20 +/- 10 SNPs per contig. The target lengths spanned 287-465 bp with an average of 375 +/- 49 bp (mean +/- sd). The results of the PCA and STRUCTURE analyses contain similar patterns. Primarily, both analyses suggest that three clusters occur in the nuclear genetic dataset. The PCA plot shows a general pattern of clustering: sampling sites 1-7.1 (Tuolumne), 7.2-14.2 (Central), 15-20 (Merced) (Fig. 3). In addition, the most notable pattern in the PCA plot is that the two-dimensional summary of the genetic dataset generally mirrors the geography of the sampling locations, which is a pattern observed in other genetic studies (Novembre et al. 2008). In Figure 3, we plot PC1 on the y-axis and PC2 on the x-axis to facilitate visual comparison to the sample map in Figure 1. We also note that PC1, which contains the highest proportion of variance in the dataset (by definition), generally separates individuals by drainage (Tuolumne: PC1>1; Merced: PC1<2) and strongly aligns with latitude. Given the orthogonal nature of PC axes, PC2 therefore aligns with longitude. While PC scores can reflect demographic histories (e.g. less migration along PC1), we note that previous studies have shown that PCA may artificially separate subpopulations when sampling is not spatially uniform (Novembre and Stephens 2008). This phenomenon is especially relevant for interpreting the PCA results of the Central cluster. Although PC1 shows separation of sampling sites by drainage within this cluster, we refrain from inferring demographic processes from the results (e.g. reduced migration due to a barrier) given the discontinuous nature of the habitat and therefore sampling scheme. As such, we interpret the PCA plot to indicate three clusters of sampling sites in the dataset.

The results of the STRUCTURE analysis also indicate that three clusters occur in the genetic dataset. Using Evanno's delta K method, the value of K=3 is best supported by the data. We show the results for different values of K={2,3,4}, but describe the biogeographic patterns for K=3 (Fig. 4). Individuals from sites 1 – 7.1 (in numerical order) cluster together and share an ancestral population under the model used by STRUCTURE. These sites cluster geographically in the northwestern and northern regions of Yosemite in different tributaries of the Tuolumne drainage. Sites 7.2 – 14.2 cluster together and are located in the northeastern and eastern regions of Yosemite NP. Also situated in the eastern region, sites 15 – 17 share ancestry with sites 7.2 – 14.2, but also contain ancestry from the southern cluster: sites 16-20. We note that sites 1 – 12 are located in the Tuolumne drainage and sites 14.1-20 are located in the Merced drainage. Pairwise Fst for the three clusters are: sites 1-7.1 (Tuolumne) vs. sites 16-20 (Merced): Fst=0.248; sites 1-7.1 (Tuolumne) vs. sites 7.2-14.2 (Central): Fst=0.0901; sites 7.2-14.2 (Central) vs. sites 16-20 (Merced): Fst=0.0986. Although STRUCTURE indicates that a model with three genetic clusters provides the best fit to the data, we note that we cautiously interpret the results of this analysis pending our subsequent analyses described below. Work by the authors of STRUCTURE and other researchers suggests that genetic patterns generated by isolation-by-distance processes can "fool" to the model into artificially binning individuals into clusters when the actual population history is the result of more continuous processes (e.g. stepping stone model). Indeed, the K=2 plot is suggestive of a gradient in ancestry proportions

through the sampling sites. Our downstream analyses that follow seek to determine if these partitions in genetic variation are in fact artifacts.

*Spatial analysis of genetic variation*

The results of the sPCA indicated significant "global" spatial structure in dataset. Using the Moran's I method, the conventional PCA score identified significant spatial autocorrelations on PC1 and PC2 ($p<0.001$; $p=0.007$, respectively). Given this, we retained the first two positive eigenvectors from sPCA output. The global test of the sPCA results revealed a significant positive spatial autocorrelation ($p<0.001$), which is suggestive of clines or patches in the dataset. Figure 5a shows an interpolated map of sPC1 scores where the first eigenvector separates the Tuolumne and Merced sites, supporting the conventional PCA above. Likewise the sPC2 scores separate the Central cluster of sites from the Southern and Northern clusters (Fig. 5b), also similar to the conventional PCA. These results suggest that the genetic dataset is spatially partitioned along two axes, as identified in the conventional PCA, with stronger differentiation in allele frequencies between the northern Tuolumne and southern Merced sampling locations than along the border of the two drainages. Notably, the sPCA results did not identify a significant pattern strong differentiation between neighboring sites, which is termed "local" structure in this framework ($p=0.3$). Barriers would tend to create local structure due stronger than expected differentiation between neighboring sites bisected by the barrier. As such, our *a priori* hypothesis regarding the Cathedral Range as a strong local barrier to migration is not supported by these results. Although sPC1 roughly divides the sampling locations into two drainages, the location of the y-axis intercept could be the combined result of the discontinuous sampling scheme and a general isolation-by-distance pattern. In addition, we repeated the analysis with only the Central cluster of sites. We did this to 1) test if the spatial partitioning occurs at a finer scale and 2) remove the strong signal in the dataset from the differentiation of individuals at the extremes of the sampling scheme (northern and southern sampling locations). Moran's I test indicates significant positive spatial autocorrelation ($p=0.005$), where sPC1 scores contain some but not complete partitioning by drainage (Fig. 5c). The global and local tests for spatial autocorrelation yielded non-significant results. While this analysis does not provide support for the strong barrier hypothesis, it remains a possibility that there are undetected, location-specific barrier(s) in this region.

The results of the spatial AMOVA test largely agreed with the STRUCTURE cluster assignments when K is set at 3, where the AMOVA cluster assignments were: cluster 1 contains sites 1-7.1; cluster 2 contains sites 7.2-17; and cluster 3 contains 18.1-20. The highest proportion of variance occurred among populations in different clusters ($\mathbf{\Phi}st = 0.41$). The other partitions also contain substantial amount of the variance: among populations within the same cluster: $\mathbf{\Phi}sc = 0.20$; and among clusters: $\mathbf{\Phi}ct = 0.26$. This suggests that a substantial amount of genetic variation occurs within clusters as well as among clusters.

*Isolation-by-Distance*

Isolation-by-distance is a prominent pattern in the dataset at the population level as well as individual level, as indicated in the sPCA results. Fst, which is a measure of population differentiation, is significantly associated with Euclidean geographic distance (Mantel test: $r = 0.632$; $p < 0.001$; Fig. 6a). When geographic distance is controlled for using a Partial Mantel test, the three cluster assignment is a significant predictor variable of Fst ($r=0.23$, $p = 0.0062$), which supports the three-cluster scenario predicted by STRUCTURE. Drainage (Tuolumne vs. Merced) was also a statistically significant predictor of Fst when controlling for Euclidean distance ($r=0.417$, $p<0.001$). In addition, the regression of Fst on distance is stronger for between-

drainage pairs than for within-drainage pairs (Fig. 6b). These measures suggest that the divide between the Tuolumne and Merced drainages, including the Cathedral Range, has reduced gene flow between drainages, although the patterns do not appear indicative of a strong impermeable barrier. We also constructed a neighbor-joining tree with the pairwise Fst values to assess the branching patterns of population units at the level of sampling sites (Fig. 7). The dominant pattern in the unrooted NJ tree is akin to a stepping stone model of branching.

## Discussion

*Implications of genetic variation in mtDNA*

Our mtDNA results bring a clearer picture of haplotype variation among *R. sierrae* populations in Yosemite. First we confirmed that two major groups of haplotypes occur among the populations in Yosemite, as was also suggested by Vredenburg et al. (2007). The two groups are generally divided into two drainages (Tuolumne and Merced), but with two exceptions in our dataset. The two exceptions occurred in populations located near the boundary between the two drainages and included both directions of discordance. These results suggest that migration of at least female individuals occurs (or occurred) in this region near the drainage boundary. Prior to population declines over the last century, connectivity between drainages may have been more plausible than today, given the large *R. sierrae* population sizes in Yosemite (Grinnell and Storer 1924). This observation provided strong incentive to further investigate patterns of connectivity and population clustering in a set of nuclear genome markers.

*Implications of genetic variation in nuclear DNA*

Our multi-locus nuclear DNA analysis contained three key findings. First, we found that genetic variation among sampling locations reflects geography. Both types of PCA (conventional and spatially-explicit) showed a pattern that resembled that the actual geographic locations of the samples. The PCA plots suggest that there is a semi-continuous pattern of genetic differentiation between sampling locations, rather than the strong discontinuous pattern between the two drainages seen in the mtDNA phylogenetic tree. The primary axis of differentiation is nearly aligned with the North-South axis. This aligns with the general pattern throughout the range of *R. sierrae* and *R. muscosa*, as seen in the mtDNA phylogenetic tree from Vredenburg et al. (2007), where clade splitting occurs along the axis of the Sierra Nevada mountain range and south into the Transverse Range of southern California. We note that this north-south axis of differentiation is correlated with the between-drainage differentiation, which is discussed in more detail below. The secondary axis, with an east-west orientation, also shows a semi-continuous pattern of differentiation among sampling locations. The patterns depicted in the PCA plots were also observed at the population level in two analyses: in the (1) isolation-by-distance analysis and (2) the NJ tree of pairwise Fst. The isolation-by-distance analysis showed a significant relationship between genetic differentiation and geographic distance between pairs of sampling locations. Likewise, the NJ tree of pairwise Fst values showed the sampling locations in an approximate stepping-stone pattern, and therefore supports the semi-continuous nature of differentiation.

The second key finding is that the analyses indicate that three clusters occur in the dataset. The STRUCTURE clustering analysis showed that a model with K=3 provided the best fit for the dataset. As expected given the PCA results, the clusters generally reflect the geographic structure of sampling locations. Interestingly, the Central Cluster included sampling locations from both drainages, which supports the hypothesis of ongoing or relatively recent connectivity among populations located along the drainage boundary, as suggested by the mtDNA results. The three-cluster scenario is supported by the spatial PCA analysis where the

two significant axes of spatial autocorrelation (north-south and east-west) suggested that the Central cluster contained significant differentiation. In addition, the three-cluster scenario is supported by the Partial Mantel test, which showed a significant correlation between genetic distance and cluster assignment when geographic distance was controlled for. The STRUCTURE cluster assignments also included substantial amounts of admixture in several sites. This suggests that genetic connectivity exists between these sampling locations and the other clusters.

The third key finding relates to the potential role of the divide between the Tuolumne and the Merced drainages, including the Cathedral Range, as a barrier to migration, which would maintain separation among locations in these two drainages. Our results suggest that this divide has a generally weak effect on migration. The fact that the Central cluster contains sampling locations on both the Tuolumne and Merced sides of the Cathedral Range suggests that this potential barrier is weak enough to allow connectivity. However, our isolation-by-distance analysis indicated that the among-drainage pairs of locations tended to have higher differentiation after controlling for geographic distance. Considering these two analyses suggests that there are effective barriers in some contexts (i.e., between certain population pairs).

Previous work on the Yosemite toad (*Bufo canorus*), another anuran in Yosemite, provides additional context for our results. An analysis of population structure in the Yosemite toad found some similar patterns as in our study (Shaffer *et al.* 2000). There was a general pattern of genetic variation partitioned between Tuolumne and Merced River drainages with some between-drainage haplotype sharing in populations near the Cathedral range. A significant pattern of isolation-by-distance was also observed among the populations sampled across Yosemite. These patterns suggest that commonalities occur in biogeographic histories between *B. canorus* and *R. sierrae*. A likely driver for biogeography patterns is the Pleistocene glaciation events, which has been proposed for several other clades found in the Sierra Nevada (e.g. Rovito 2010, Schoville et al. 2012).

*Reconciling mitochondrial discordance across drainages*

The unexpected distribution of mitochondrial haplotypes across the Cathedral Range at two sites could have occurred with different demographic scenarios. The two sites with alternate mtDNA haplotypes in our dataset were Gallison lake area and Unicorn Lake. One scenario is the very recent migration occurred with possible human-assisted activity. It is conceivable that frogs or tadpoles hitch-hiked with trout during fish transplant efforts in small containers (e.g. coffee can). However, the nuclear DNA results are not consistent with the "coffee can" scenario, where we would expect to find that recently introduced frogs would have nuclear DNA haplotypes closely related to a particular population on the opposite side of Cathedral Range, along with a high degree of admixture. Instead we find that frogs from Gallison (site 14.1) and Unicorn (site 10) group more closely with frogs collected at geographically proximate sites in their respective drainages (Fig. 3). The Unicorn individuals contained a low degree of admixture, while Gallison individuals contained moderate levels of admixture with contributions from the "Merced" cluster (Fig. 4). The alternative scenario is that migration occurred sometime in the more distant past. The observed nuclear DNA results are more consistent with this scenario. Several processes could explain the mitochondrial-nuclear discordance including: asymmetric dispersal, mating, or offspring production; and adaptive introgression of mtDNA.

*Noninvasive sampling approach*

Noninvasive sampling techniques are especially critical for conservation studies involving threatened or endangered species. The present study represents an advance in the use of swabs from previously conducted epidemiological surveys (chytridiomycosis) for studying

host population genetics. This advance, which required thorough protocol optimization, will likely be useful for future studies on host genetics of other amphibian species given the vast number of swab-based chytridiomycosis surveys over the last decade. This method allows researchers to study host population genetics without needing to re-survey populations if swabs are available and avoid inflicting harm on study subjects. In addition, researchers can use previously collected swabs to study the genetics of extirpated populations that collapsed due to an epidemic. We also note that this study advanced previous work on *R. sierrae* by providing a new set of nuclear sequence markers, which enabled a fine-scale analysis of population structure.

*Implications for management*

Gaining a better understanding of the spatial genetic structure is important for conservation management of the remaining populations. For example, in the case of translocating individuals the preferred source population for a particular recipient population can be selected based on relatively high historical gene flow and/or short divergence time. This will minimize the primary genetic risks of translocations: outbreeding depression and loss of locally adapted alleles. Our first key finding – geographic structure with isolation-by-distance – suggests that translocations between neighboring geographic locations would be preferred over long-distance translocations. Our second key finding – the three cluster scenario – suggests that translocation within-clusters would be preferred over between-cluster translocations. For the "Central" cluster, a conservative approach would be to translocate individuals within basin given (1) the broad pattern of mtDNA partitioning by basin and (2) evidence of weak substructure in nuclear DNA results among sampling locations. In general, the patterns of isolation-by-distance and substantial within-population admixture suggest that using frogs from several nearby, within-cluster populations in translocations could be a reasonable strategy to re-establish extirpated *R. sierrae* populations that retain as much of the regional genetic diversity as possible.

**Acknowledgements**

This work would not have been possible without the swab archive generated in previous studies led by Roland Knapp. We thank Mary Toothman for assistance in providing archived samples, Jeanine Refsnider for assistance with experimental design, Rob Grasso for helpful comments on the manuscript. All sample collection (noninvasive skin swab) was led by Roland Knapp and authorized by research permits provided by Yosemite National Park and the Institutional Animal Care and Use Committee at University of California, Santa Barbara (UCSB).

**References**

Adamack, A. T., and B. Gruber, 2014 PopGenReport: simplifying basic population genetic analyses in R (S. Dray, Ed.). Methods Ecol. Evol. 5: 384–387.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel *et al.*, 2013 From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Curr. Protoc. Bioinforma.

Briggs, C. J., R. A. Knapp, and V. T. Vredenburg, 2010 Enzootic and epizootic dynamics of the chytrid fungal pathogen of amphibians. Proc. Natl. Acad. Sci. U. S. A. 107: 9695–9700.

Chevreux, B., T. Pfisterer, B. Drescher, A. J. Driesel, W. E. G. Müller *et al.*, 2004 Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP

detection in sequenced ESTs. Genome Res. 14: 1147–1159.

Clayton, D., 2014 snpStats: SnpMatrix and XSnpMatrix classes and methods.

Dellicour, S., and P. Mardulyn, 2013 spads 1.0: A toolbox to perform spatial analyses on DNA sequence data sets. Mol. Ecol. Resour.

Earl, D. A., and B. M. vonHoldt, 2012 STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4: 359–361.

Edgar, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792–7.

Evanno, G., S. Regnaut, and J. Goudet, 2005 Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14: 2611–2620.

Grinnell, J., and T. Storer, 1924 *Animal life in the yosemite.* University of California Press, Berkeley, CA.

Hyatt, A. D., D. G. Boyle, V. Olsen, D. B. Boyle, L. Berger *et al.*, 2007 Diagnostic assays and sampling protocols for the detection of Batrachochytrium dendrobatidis. Dis. Aquat. Organ. 73: 175–192.

Jombart, T., S. Devillard, A.-B. Dufour, and D. Pontier, 2008 Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity (Edinb). 101: 92–103.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung *et al.*, 2012 Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647–1649.

Knapp, R. A., 2005 Effects of nonnative fish and habitat characteristics on lentic herpetofauna in Yosemite National Park, USA. Biol. Conserv. 121: 265–279.

Knapp, R. A., and K. R. Matthews, 2000 Non-native fish introductions and the decline of the mountain yellow-legged frog from within protected areas. Conserv. Biol. 14: 428–438.

Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose, 2015 Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol. Ecol. Resour. n/a–n/a.

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–1760.

Pritchard, J. K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. Genetics 155: 945–59.

Rosenblum, E. B., T. J. Poorten, M. Settles, and G. K. Murdoch, 2012 Only skin deep: shared genetic response to the deadly chytrid fungus in susceptible frog species. Mol. Ecol. 21: 3110–20.

Rovito, S. M., 2010 Lineage divergence and speciation in the Web-toed Salamanders (Plethodontidae: Hydromantes) of the Sierra Nevada, California. Mol. Ecol. 19: 4554–4571.

Schliep, K. P., 2011 phangorn: Phylogenetic analysis in R. Bioinformatics 27: 592–593.

Schoville, S. D., G. K. Roderick, and D. H. Kavanaugh, 2012 Testing the "Pleistocene species pump"in alpine habitats: lineage diversification of flightless ground beetles (Coleoptera: Carabidae: Nebria) in relation to altitudinal zonation. Biol. J. Linn. Soc. 107: 95–111.

Shaffer, H. B., G. M. Fellers, A. Magee, and S. R. Voss, 2000 The genetics of amphibian declines: population substructure and molecular differentiation in the Yosemite Toad, Bufo canorus (Anura, Bufonidae) based on single-strand conformation polymorphism analysis (SSCP) and mitochondrial DNA sequence data. Mol. Ecol. 9: 245–257.

Smedley, D., S. Haider, B. Ballester, R. Holland, D. London *et al.*, 2009 BioMart--biological queries made easy. BMC Genomics 10: 22.

Stamatakis, A., 2006 RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688–2690.

Vredenburg, V. T., R. Bingham, R. Knapp, J. A. T. Morgan, C. Moritz *et al.*, 2007 Concordant molecular and phenotypic data delineate new taxonomy and conservation priorities for the endangered mountain yellow-legged frog. J. Zool. 271: 361–374.

Vredenburg, V. T., R. A. Knapp, T. S. Tunstall, and C. J. Briggs, 2010 Dynamics of an emerging disease drive large-scale amphibian population extinctions. Proc. Natl. Acad. Sci. U. S. A. 107: 9689–94.

Wang, I. J., 2013 Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation. Evolution (N. Y). 67: 3403–3411.

Zieliński, P., M. T. Stuglik, K. Dudek, M. Konczal, and W. Babik, 2014 Development, validation and high-throughput analysis of sequence markers in nonmodel species. Mol. Ecol. Resour. 14: 352–360.

## Figures

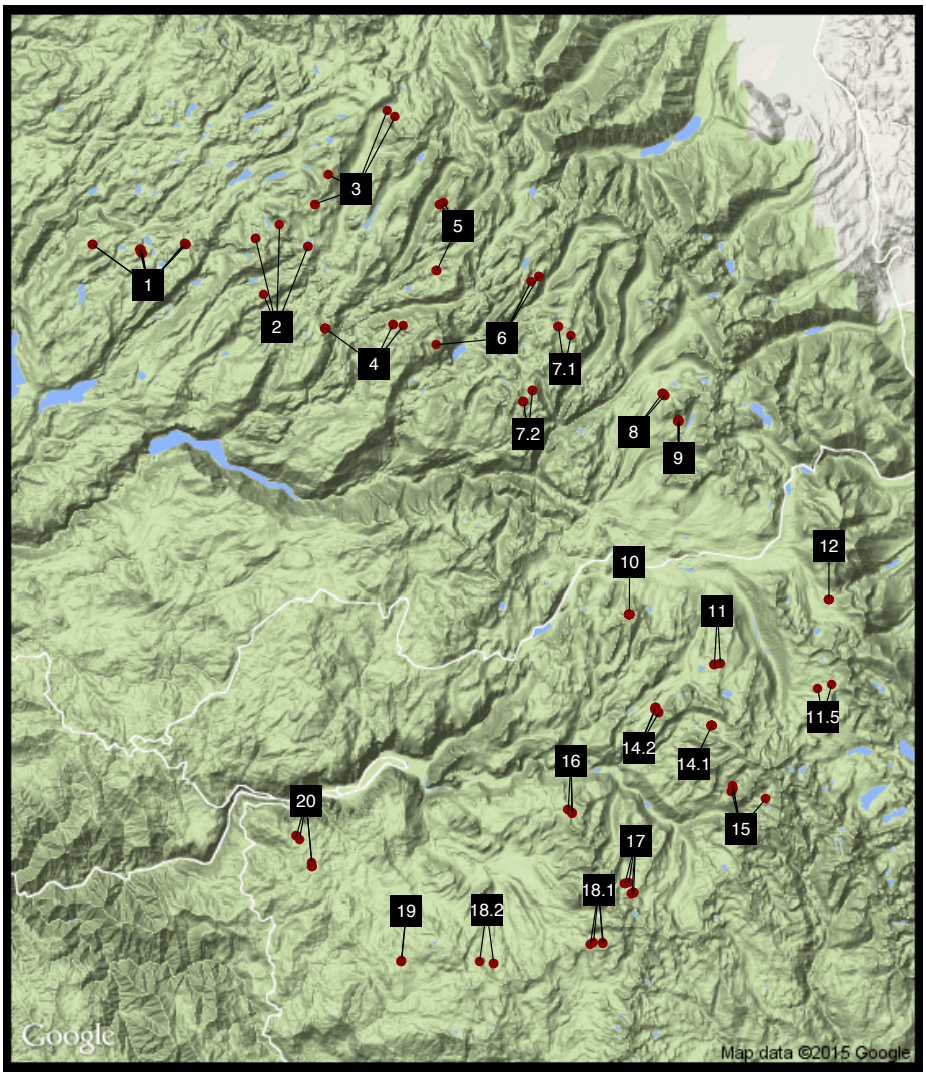



**Figure 1**. Yosemite National Park study area, showing the sample locations (red circles). Sites labels (black squares) represent either single sample locations or contain groupings of nearby sample locations as indicated by connecting lines to sampling locations.

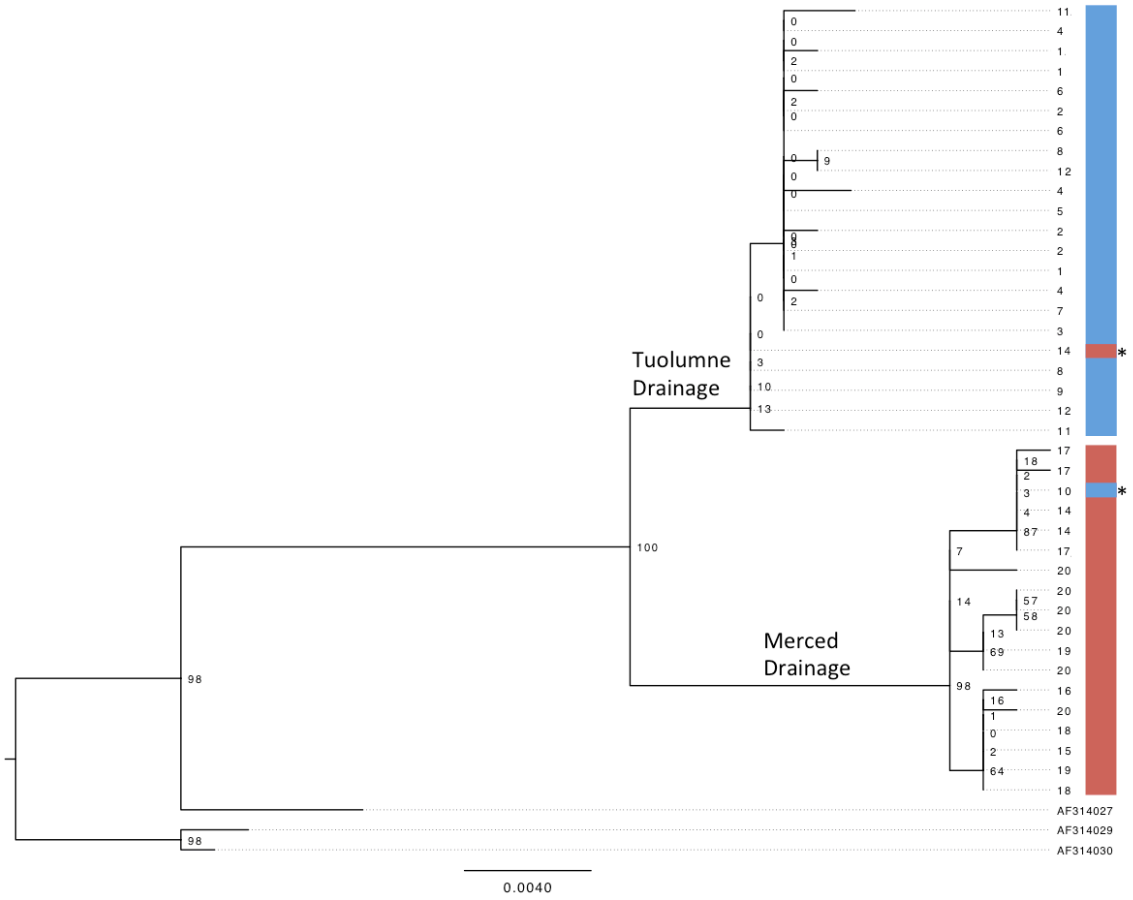



**Figure 2**. Consensus phylogenetic tree inferred from mitochondrial DNA locus (ND2). Colored bars indicate drainage location of individuals: red=Merced; blue=Tuolumne. Terminal nodes show respective site label of individuals. Asterisks indicate drainage-discordant haplotypes.

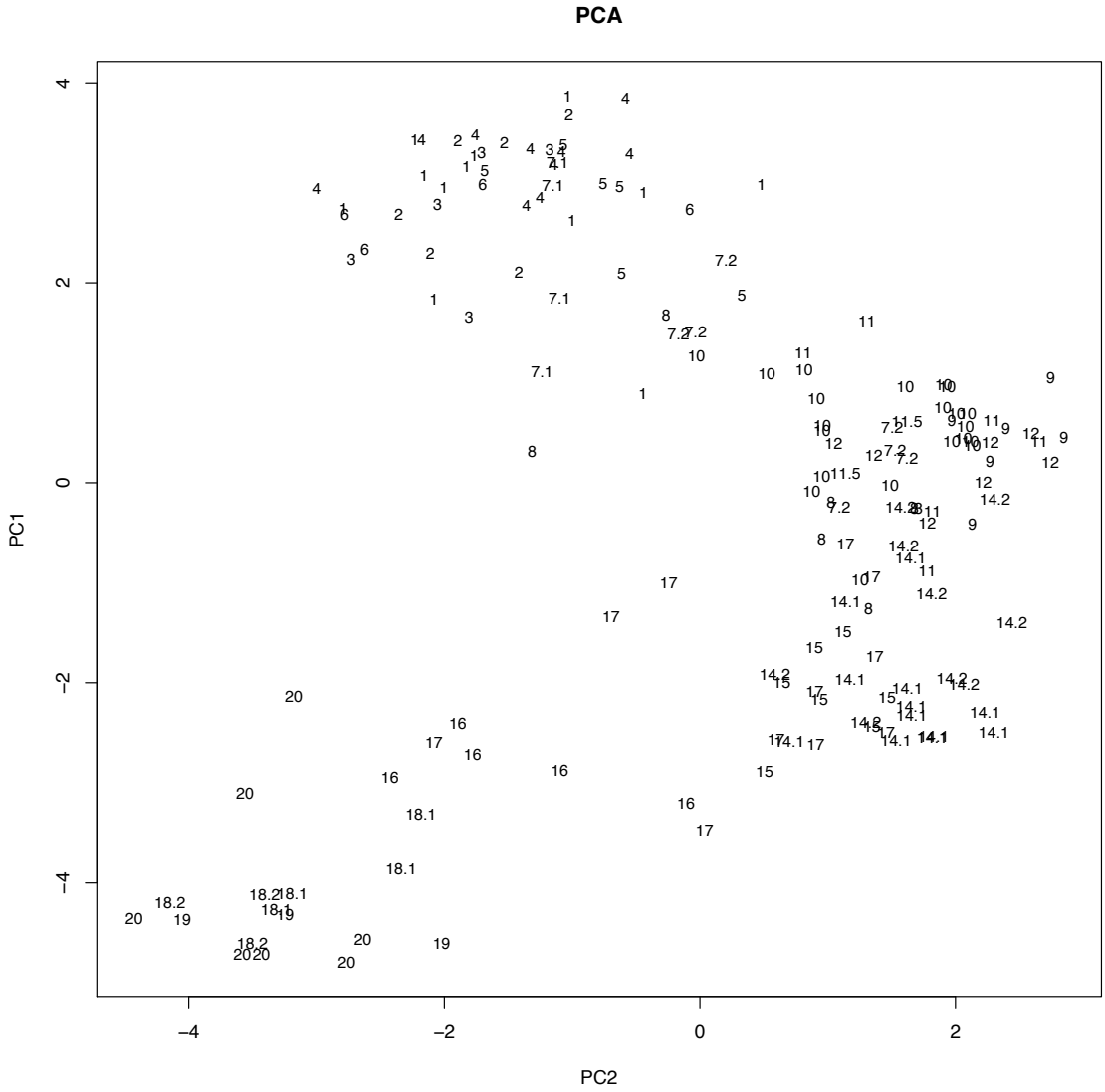



**Figure 3**. Principal components analysis (PCA) of the nuclear genetic dataset, showing that individual frogs tend to cluster together according to sampling location. Note that the plot was rotated for easier comparison to the study area: PC1 is on the y-axis and PC2 is on the x-axis.

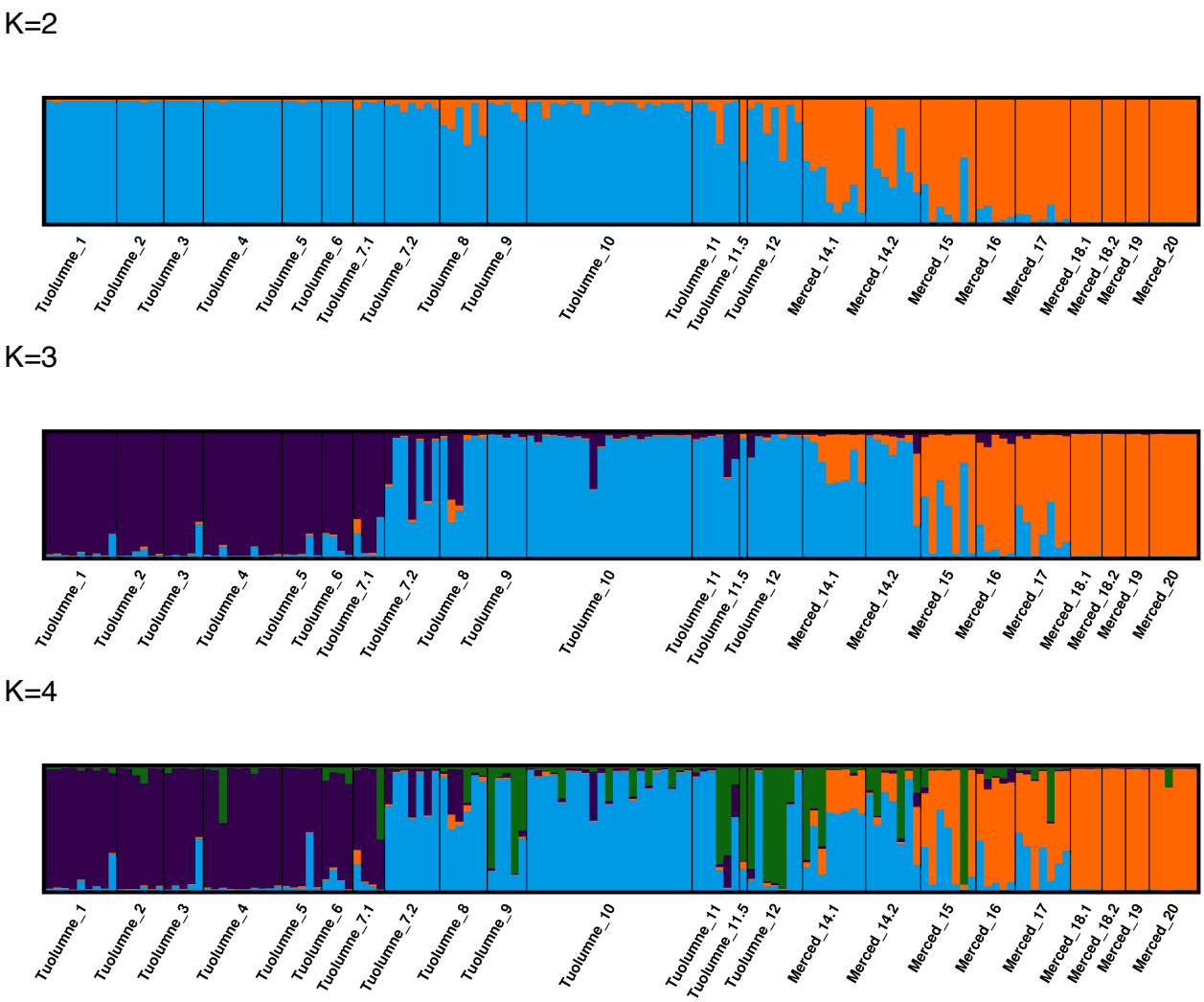
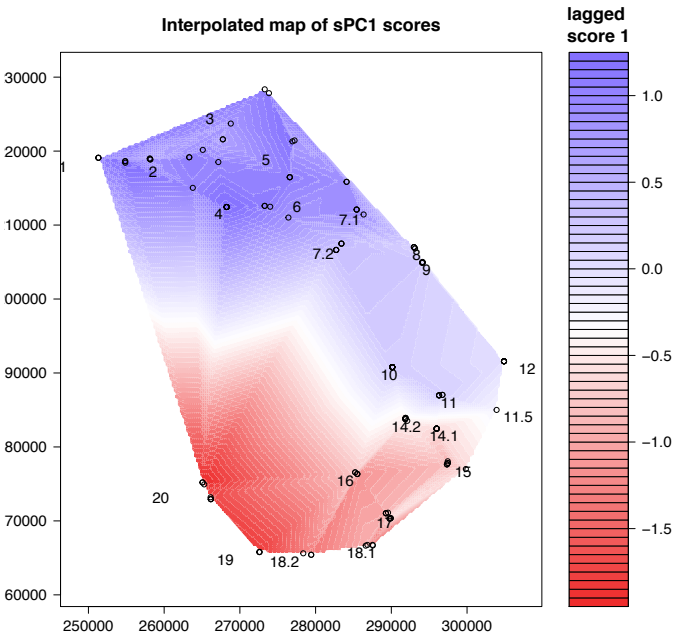
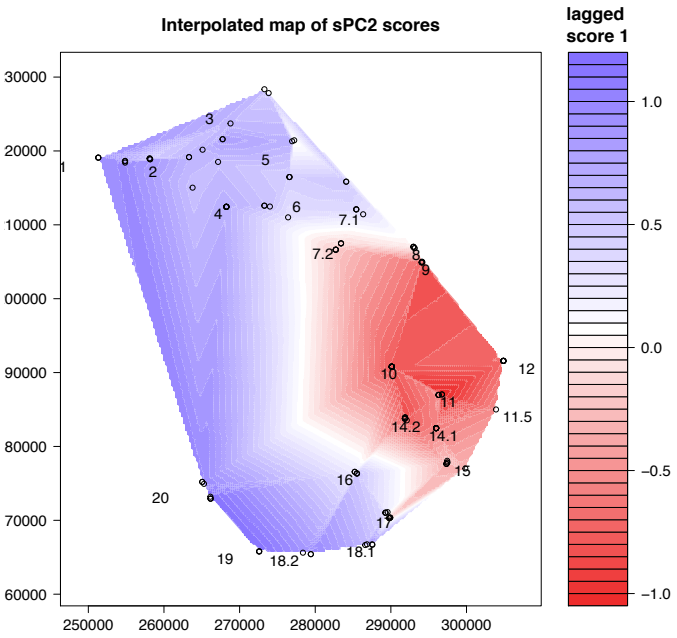



**Figure 4**. STRUCTURE cluster assignment results for two, three, and four clusters (K). Each bar represents an individual frog, and colors represent inferred population ancestry. K=3 was chosen as the best value of K by Evanno's method (Earl and vonHoldt 2012).

**(a)**


Interpolated map of sPC1 scores
lagged score 1
1.0
0.5
0.0
−0.5
−1.0
−1.5
30000
20000
10000
00000
90000
80000
70000
60000
250000
260000
270000
280000
290000
300000
1
2
3
4
5
6
7.1
7.2
9
10
11
11.5
12
14.1
14.2
15
16
17
18.1
18.2
19
20

**(b)**


Interpolated map of sPC2 scores
lagged score 1
1.0
0.5
0.0
−0.5
−1.0
30000
20000
10000
00000
90000
80000
70000
60000
250000
260000
270000
280000
290000
300000
1
2
3
4
5
6
7.1
7.2
9
10
11
11.5
12
14.1
14.2
15
16
17
18.1
18.2
19
20

**(c)**

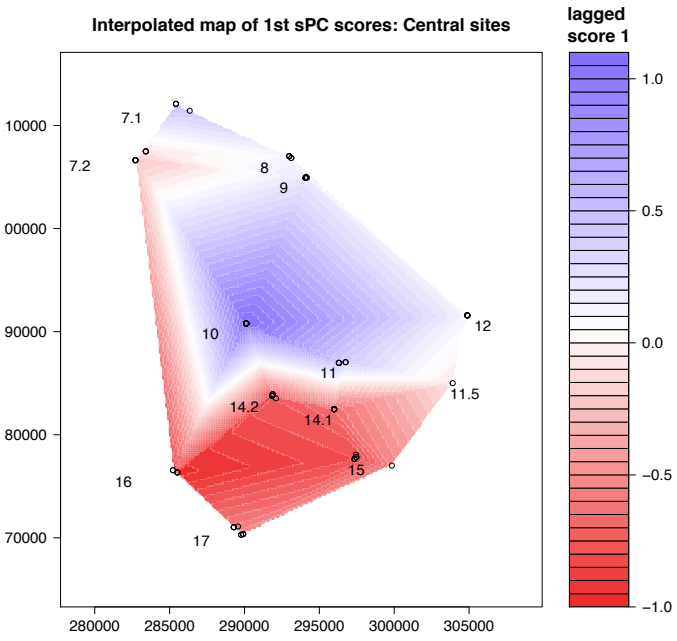
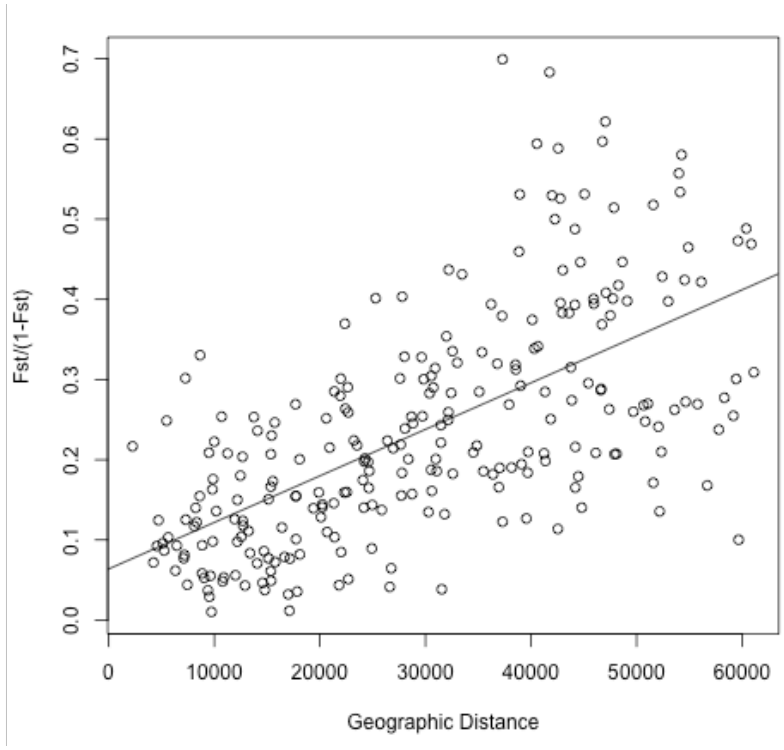



**Figure 5**. Interpolated maps of spatial PCA scores for the first (a) and second (b) eigenvalues projected on the study area. Sampling locations are shown with open circles and numbers indicate locations of sample groupings. (c) Interpolated map of sPC1 for the subset of sites in Central region (7.1 – 17).

**(a)**




**(b)**

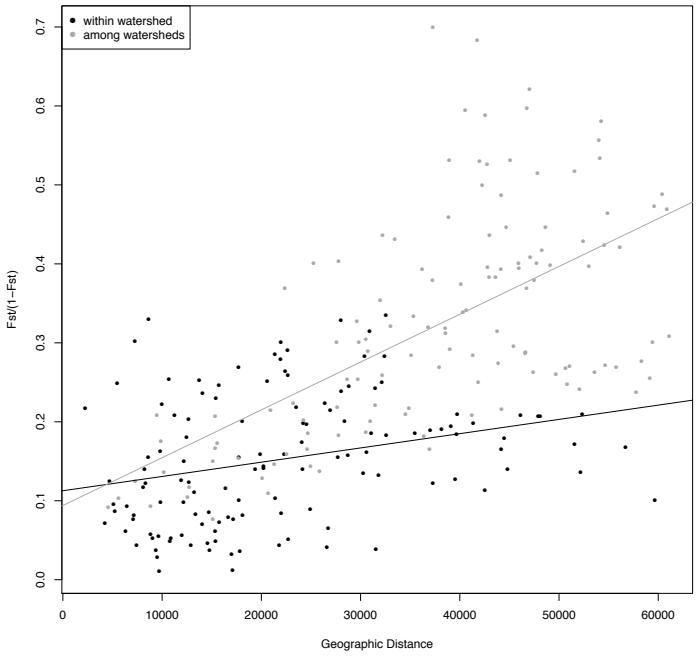



**Figure 6**. (a) Regression of linearized Fst versus geographic distance between sample groupings, showing a pattern of isolation-by-distance. (b) Regression as in (a) split by drainage showing a stronger IBD relationship for between-drainage pairs of sample groupings.

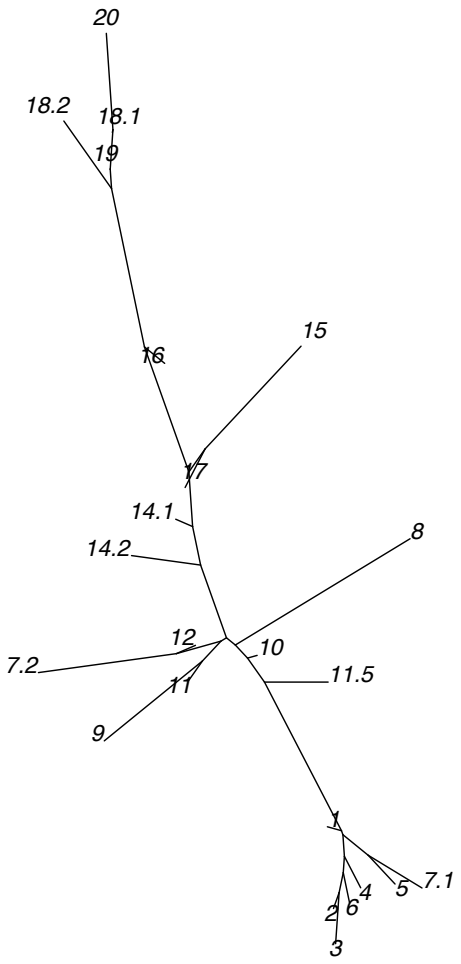



**Figure 7**. Neighbor-joining tree based on pairwise Fst (linearized) of sample groupings. The pattern resembles a stepping stone model of population relationships (rather than strong bipartitioning).

## Chapter 4: Genomic correlates of virulence attenuation in the deadly amphibian chytrid fungus, *Batrachochytrium dendrobatidis*


### Abstract

Emerging infectious diseases pose a significant threat to global health, but predicting disease outcomes for particular species can be complicated when pathogen virulence varies across space, time, or hosts. The pathogenic chytrid fungus *Batrachochytrium dendrobatidis* (*Bd*) has caused worldwide declines in frog populations. Not only do *Bd* isolates from wild populations vary in virulence, but virulence shifts can occur over short timescales when *Bd* is maintained in the laboratory. We leveraged changes in *Bd* virulence over multiple generations of passage to better understand mechanisms of pathogen virulence. We conducted whole-genome resequencing of two samples of the same *Bd* isolate, differing only in passage history, to identify genomic processes associated with virulence attenuation. The isolate with shorter passage history (and higher virulence) had higher chromosome copy numbers than the isolate maintained in culture for longer, suggesting that virulence attenuation may be associated with loss of chromosome copies. Our results suggest that genomic processes proposed as mechanisms for rapid evolution in *Bd* are correlated with virulence attenuation in laboratory culture within a single lineage of *Bd*. Moreover, these genomic processes can occur over extremely short timescales. On a practical level, our results underscore the importance of immediately cryo-archiving new *Bd* isolates and using fresh isolates, rather than samples cultured in the lab for long periods, for laboratory infection experiments. Finally, when attempting to predict disease outcomes for this ecologically important pathogen, it is critical to consider existing variation in virulence among isolates and the potential for shifts in virulence over short timescales.


### Introduction

Most infectious pathogens exhibit variation in virulence across space, time, or host species. For example, pathogens may affect hosts differentially (e.g., Johnson *et al*. 2012), cause different outcomes under different environmental conditions (e.g., Mitchell *et al*. 2005), and evolve increased or decreased virulence over time (e.g., Boots *et al*. 2004). Understanding variation in pathogen virulence is essential to create accurate epidemiological models, predict disease outcomes, and understand mechanisms of pathogenicity (Jones *et al*. 2008).

*Batrachochytrium dendrobatidis* (*Bd*) is an ecologically important pathogen that provides an opportunity to understand mechanisms of variation in virulence. *Bd* is a chytrid fungus that attacks amphibian hosts and has led to population declines and extirpations around the world (e.g., Berger *et al*. 1998; Bosch *et al*. 2001; Muths *et al*. 2003; Lips *et al*. 2006). However, there is dramatic variation in the outcome of *Bd* infection: some populations are decimated by *Bd* (e.g., Schloegel *et al*. 2006; Crawford *et al*. 2010; Vredenburg *et al*. 2010), while others persist (Retallick *et al*. 2004; Woodhams *et al*. 2007; Lam *et al*. 2010). This variation in infection outcome is affected by three interacting factors: environmental conditions such as temperature (Voyles *et al*. 2012; Spitzen-van der Sluijs *et al*. 2014), intrinsic differences in host susceptibility (e.g., Bishop *et al*. 2009; Searle *et al*. 2011; Voyles *et al*. 2011; Ghal *et al*. 2012; Gervasi *et al*. 2013), and differences in virulence among *Bd* isolates (e.g., Berger *et al*. 2005; Retallick & Miera 2005; Fisher *et al*. 2009; Farrer *et al*. 2011; Voyles 2011).

Here we focus on variation in virulence caused by intrinsic differences among *Bd* isolates. There is increasing evidence that *Bd* isolates differ in virulence even in controlled

common garden conditions (e.g., when hosts and environmental conditions are kept constant; Berger *et al*. 2005; Retallick & Miera 2005; Fisher *et al*. 2009; Farrer *et al*. 2011; Voyles 2011). Moreover, recent studies have revealed substantial genetic variation within *Bd* and dynamic genomic processes that could provide a mechanism for rapid shifts in virulence (Farrer *et al*. 2011, 2013; Rosenblum *et al*. 2013; Piovia-Scott *et al*. 2014).

An ideal framework for studying mechanisms of pathogen virulence is when shifts in virulence occur within a single pathogen lineage. More specifically, when virulence shifts can be induced in the laboratory, mechanisms of virulence can be examined directly without the confounding effects of spatial genetic structure, environmental variation, host-specific processes, or uncertain ancestry. Several recent studies have reported an attenuation of virulence in *Bd* over very short timescales in cultures maintained in the laboratory (Brem *et al*. 2013; Langhammer *et al*. 2013). Virulence attenuation has also been observed in other pathogenic fungi maintained in laboratory conditions (Lord & Roberts 1986; Butt *et al*. 2006; Safavi 2012; Janbon *et al*. 2014). Here we make use of changes in *Bd* virulence that occurred in a laboratory setting to interrogate the mechanisms of variation in pathogen virulence. Specifically, we compared two *Bd* samples that originated from the same source but differed in passage history. One sample, JEL427-P9, was cryo-archived shortly after isolation (having undergone nine laboratory passages), and a second sample, JEL427-P39, was maintained in culture for six years (39 laboratory passages). The two isolates differ in phenotypes relevant for virulence (e.g., higher zoospore production rate in the JEL427-P9 isolate, Langhammer *et al*. 2013) as well as in their effect on frogs, with the JEL427-P9 isolate causing increased mortality (Langhammer *et al*. 2013). We undertook a genome resequencing study to identify genomic correlates of virulence attenuation and shed light on the genomic processes that may be important for the evolution of virulence in this host-pathogen system.

**Materials and Methods**

We extracted genomic DNA from two *Bd* isolates of strain JEL427 isolates that had undergone nine and 39 laboratory passages (JEL427-P9, the ancestral isolate, and JEL427-P39, the derived isolate, respectively). Prior to DNA extraction, we cultured each isolate for five days on two replicate 1% tryptone and 1% agar plates. We flooded the plates for zoospores and concentrated zoospores using a tabletop centrifuge. We extracted genomic DNA using a phenol-chloroform protocol (Zolan & Pukkila 1986) modified for use in *Bd* (Joneson *et al*. 2011). Genomes from the two isolates were sequenced using the Illumina MiSeq platform (2 x 250 bp paired end reads) at the University of Idaho's Genomics Resources Core Facility.

Sequence alignment and SNP-calling were performed as in Rosenblum *et al*. (2013). Briefly, we used Seqyclean v. 1.8.10 (Zhbannikov *et al*. unpublished) to clean reads; remove PCR duplicates, contaminants, and adaptors; and trim sequences by quality scores. Reads were aligned to the reference genome of JEL423 (Broad Institute v. 17-Jan-2007) using Bowtie 2 v. 2.1.0 (Langmead *et al*. 2009), and we used best practice protocol for variant calling in GATK v. 1.4 (McKenna *et al*. 2010). We marked duplicate reads with Picard, and we realigned reads containing insertions/deletions (indels) with GATK walkers RealignerTargetCreator and IndelRealigner. We made final variant calls, and filtered false positives, using GATK UnifiedGenotyper and VariantFiltration walkers with the same filter parameter value as in Rosenblum *et al*. (2013). We then used snpEff software (version 2.0.5, Cingolani *et al*. 2012) and custom R scripts to extract summary information from the SNP dataset (e.g., genomic position, synonymous vs nonsynonymous mutations) to compare JEL427-P9 and JEL427-P39.

We also used bedtools software (Quinlan and Hall 2010) to identify indels that occurred in coding exonic regions using the gene feature annotation file downloaded from the Broad Institute's online database (https:www.broadinstitute.org).

We inferred phylogenetic relationships among isolates using the parsimony method described previously (Rosenblum *et al*., 2013). Briefly, the SNPs were encoded to distinguish three character states (0-2): homozygous with respect to reference allele, heterozygous, and homozygous for the alternate allele. We performed 200 bootstrap replicates to generate node support values under the parsimony optimality criterion.

We used a previously developed method to predict genomic regions affected by past "loss of heterozygosity" (LOH) events (Rosenblum *et al*. 2013), which result in large chromosomal regions that are homozygous. We implemented a hidden Markov model method to identify genomic regions with long stretches of homozygosity in the 15 largest supercontigs. Variation in chromosome copy number is a change in the number of copies of a particular chromosome without a change in ploidy of the entire chromosome set. We estimated copy number for each supercontig, which are analogous to chromosomal segments in *Bd*, using SNP allele frequencies (as in Rosenblum *et al*. 2013). From the variant call format file, we extracted the mean allele frequency across SNPs for each supercontig. For each supercontig, we plotted the distribution of all SNP allele frequencies for that supercontig, and we used Gaussian kernel density in the R package KernSmooth to smooth the distribution. Each supercontig was then assigned a copy number based on the expected distribution of all SNP allele frequencies for a supercontig that was monosomic, disomic, trisomic, tetrasomic, or pentasomic. That is, allele frequencies would have a unimodal distribution centered at 0.5 for disomic chromosomes, a bimodal distribution with peaks at 0.33 and 0.67 for trisomic chromosomes, a trimodal distribution with peaks at 0.25, 0.5, and 0.75 for tetrasomic chromosomes, and a tetramodal distribution with peaks at 0.2, 0.4, 0.6, and 0.8 for pentasomic chromosomes.

Finally, we evaluated the potential functional relevance of SNPs and indels using Gene Ontology (GO) categories and gene expression results from previous microarray experiments (Rosenblum *et al*. 2012). With the gene expression data, we asked whether sequence changes were more likely to occur in genes that were found previously to be up-regulated on frog skin. We hypothesized that, over the generations grown in culture (rather than on host tissue), selection on virulence-associated genes may be relaxed, and could be evidenced by an elevated substitution rate relative to more highly constrained genes. To test this hypothesis for the SNP dataset, we conducted regression by implementing a generalized linear model assuming a negative binomial distribution of the number of nucleotide changes per gene in five differential expression categories: strong down-regulation, moderate down-regulation, no change, moderate up-regulation, and strong up-regulation. To test this hypothesis for the indel dataset, we used only three differential expression categories due to the relatively low number of indels in the dataset: down-regulation, no change, up-regulation. We repeated the enrichment test with all indel-containing genes given the small number of indels. For all analyses, we included gene length as a covariate and generated models and estimated parameter 95% confidence intervals for both mutation types. For the SNPs, we partitioned the dataset into three types of nucleotide changes: nonsynonymous, synonymous, and combined. We also looked at the functional relevance of genes in our dataset that contained nonsynonymous mutations or indels and were up-regulated on frog skin in Rosenblum *et al*. (2012) by conducting an over-representation test of GO terms. We used the hypergeometric test in the GOstats R package (Falcon & Gentleman

2007) with the GO annotation assignments from Rosenblum *et al*. (2013) and filtered the enrichment results to exclude enriched GO terms that contained only one gene in the gene list.

**Results**

We obtained a depth of coverage for aligned reads of 33X and 20X for the JEL427-P9 and JEL427-P39 isolates, respectively. Using a phylogeny that included 29 isolates from a previous study (Rosenblum *et al*. 2013), we found that the two JEL427 isolates clustered together with high bootstrap support and nested within the most recently derived and widely distributed clade, the Global Panzootic Lineage (Fig. 1). We note that the terminal branch-lengths of the JEL427 isolates had a prima facie counter-intuitive pattern. This pattern is likely explained by the trend of genotype changes in identified LOH regions. For LOH regions across the genome, the ancestral isolate (i.e., JEL427-P9) tended to contain the highest density of heterozygous sites, and the derived isolate (i.e., JEL427-P39) showed a loss of allelic diversity. This trend would lead to the branch-length pattern seen in Fig. 1, where JEL427-P9 contains the longer terminal branch.

The most compelling genomic patterns we observed were chromosomal copy number variation in the ancestral isolate. Chromosomal copy number for the JEL427-P39 isolate varied from disomic to tetrasomic, while the JEL427-P9 isolate varied from disomic to pentasomic. Most notably, the JEL427-P39 isolate showed a decrease in copy number for 9 of 17 supercontigs compared to the JEL427-P9 isolate (Fig. 2). Specifically, supercontigs 6, 7, 8, 9, 10, 13, 14, 15, and 16 showed a higher chromosome copy number in the JEL427-P9 isolate when compared to the JEL427-P39 isolate, and when compared to the average of other isolates studied to date (Rosenblum *et al*. 2013; Fig. 2).

After SNP filtering, we observed 2,231 changes between JEL427-P9 and JEL427-P39 (Suppl. Table 1). The snpEff results were qualitatively similar to the pattern observed in the Rosenblum *et al*. 2013 dataset of 29 global isolates. Roughly half of the SNPs occurred in regions either up- or down-stream of genes (within 1 kb). In coding regions, ~9% of all SNPs caused nonsynonymous and ~20% caused synonymous changes, and ~18% of SNPs occurred in introns. Overall, the SNPs were distributed across supercontigs in a similar pattern for both comparisons. However, a few supercontigs (i.e., supercontigs 7, 14, and 17) contained an elevated number of SNPs. Our indel analysis uncovered 470 indels between JEL427-P9 and JEL427-P39, 149 of which occurred in coding exonic regions.

To determine whether loss-of-heterozygosity (LOH) patterns affected the rate of nucleotide changes, we calculated the synonymous change rate in LOH and non-LOH regions. LOH patterns did not appear to have a major effect on genotype change, as the synonymous change rates were 0.00102 and 0.00105 in LOH and non-LOH regions, respectively.

The analysis including previous gene expression results (from Rosenblum *et al*. 2012) suggested that the differential expression (on frog skin) coefficient of a gene was significantly correlated with the number of nucleotide changes observed in this study. Genes that were moderately up-regulated when grown on frog skin had a higher number of nonsynonymous changes compared to genes without differential expression in different growth conditions (Fig. 3, $p= 0.00648$ in negative binomial regression analysis). In all, 84 genes met the two criteria of being up-regulated and containing $\geq 1$ nonsynonymous change. The hypergeometric test for over-representation of GO terms did not reveal any notable patterns (Suppl. Table 2), which was likely due to the relatively low number of annotated genes in the gene list. However, the gene

set contained one M36 metallopeptidase gene, a gene family that has previously been hypothesized to play a role in *Bd* virulence (Joneson *et al*. 2011).

In the gene expression analysis with indels, there was a weak significant effect of increased gene expression level on indel occurrence (Fig 4, p=0.048). We found a significant enrichment of GO terms "proteolysis" and "protein metabolic process" in the gene list of genes with increased expression and indels, however we note that there were only 2 genes in the proteolysis gene set and 3 genes in the protein metabolic process gene set (p=0.029 for both GO terms). For the analysis including all indel-containing genes (regardless of expression), we found the list of genes containing exonic indels was over-enriched for the Gene Ontology terms related to proteolysis and metabolism (Table 1).

**Discussion**

Determining how quickly pathogens can exhibit shifts in virulence, and the mechanisms underlying virulence shifts, is critical for understanding and mitigating emerging infectious disease threats. Our study focused on identifying the genomic correlates of rapid virulence shifts within a single lineage of the pathogenic fungus, *Bd*.

Studying the evolution of genomic mechanisms within individual *Bd* isolates is key to understanding variation in pathogen virulence. Our genomic data placed the two JEL427 isolates together and solidly within the Global Panzootic Lineage (GPL, Fig. 1), the *Bd* clade that is most recent evolutionarily and the most widespread geographically. Previous studies have demonstrated substantial genetic variation and the potential for rapid genomic evolution within the GPL (Rosenblum *et al*. 2013; Farrer *et al*. 2013). The genetic variation identified within the GPL may reflect functional variation, and indeed the *Bd* isolates used in our study show differences in virulence-related phenotypes that accumulated rapidly, over only 30 generations (~6 years, Langhammer *et al*. 2013). The isolate with a longer passage history (JEL427-P39) was less virulent compared to the ancestral isolate that was cryo-archived immediately upon isolation (JEL427-P9; Langhammer *et al*. 2013). Specifically, JEL427-P9 produced zoospores at a higher rate and caused increased mortality of frogs compared to JEL427-P39 (Langhammer *et al*. 2013). Studying a rapid virulence shift within a single lineage controls for many confounding factors common in studies of pathogens in natural systems (e.g., environmental conditions, host susceptibility, and pathogen ancestry) and provides an opportunity to assess the genomic changes associated with attenuation of virulence. We therefore used our whole genome resequencing data to assess whether specific genomic mechanisms were associated with observed virulence attenuation.

Lability in chromosomal copy number and heterozygosity have been identified as genomic process that could lead to rapid evolution in *Bd* (Rosenblum *et al*. 2013; Farrer *et al*. 2013). We therefore tested for differences in chromosome copy number and loss of heterozygosity between *Bd* isolates differing in laboratory passage history. We did not find evidence that loss of heterozyosity was associated with attenuation of virulence, but we did find striking changes in chromosomal copy number over the short timescale studied here (Fig. 2). Specifically, we found that the derived isolate JEL427-P39, which was maintained in culture for ~6 years, had, in general, fewer copies of chromosomes than the ancestral isolate JEL427-P9. Decreased chromosomal copy number in the less virulent isolate with a longer passage history was a highly consistent pattern, observed in more than half of the analyzed chromosomal segments (i.e., 9 of the 17 largest super-contigs). The complementary pattern of higher chromosome copy number in highly virulent *Bd* isolates in nature has also been observed

(Piovia-Scott *et al*. 2014), suggesting that copy number variation is potentially an important mechanism for rapid changes in virulence more generally in this pathogen. While an increase in chromosome copy number can lead to increased virulence, possibly because it results in more copies of genes related to pathogenicity, such a process may entail costs to the pathogen. Therefore in laboratory culture media (i.e., without natural selection imposed by a host), loss of chromosomal copies may occur in a predictable manner. Attenuation of virulence has been associated with chromosomal changes in other fungal pathogens, for example losses of mini-chromosomes containing genes related to pathogenicity in fungal pathogens of insects and plants (Wang *et al*. 2002, 2003; Akamatsu *et al*. 1999), and karyotype changes and loss of chromosome segments in *Cryptococcus*, a fungal pathogen of humans (Franzot *et al*. 1998; Janbon *et al*. 2014). Thus our observation that virulence attenuation was associated with loss of chromosome copies may reflect a fairly general pattern.

Results regarding copy number changes from another recent study of *Bd* genomics bear specific mention here. Farrer *et al.* (2013) also compared replicate lines of a *Bd* isolate, which were passaged in the lab for 40 generations with and without exposure to amphibian antimicrobial peptides. In contrast to our results, Farrer *et al.* found fewer changes in chromosome copy number in their isolates after 40 generations (i.e., their control isolate lost a copy of supercontig 4 and gained a copy of supercontig 5 and their isolate passaged with antimicrobial peptides gained a copy of supercontig 5). However, we note two differences between the studies that may explain why we observed more changes in chromosomal copy number. First the isolates were derived from two divergent lineages of *Bd* (*Bd*GPL in our study and *Bd*CH in Farrer et al. 2013), which may exhibit different dynamics of genome evolution. Second, the more ancestral isolate in our study (JEL427-P9) contain higher chromosome copy numbers than the ancestral isolates in Farrer et al. 2013 (ACON): mean copy number 3.5 vs. 2.8, respectively, for supercontigs 1-15. The higher starting chromosome copy number in our study may have made the isolate more prone to losing chromosome copies. These hypotheses can be tested with formal meta-analyses as additional laboratory passage experiments are conducted with *Bd*. Specifically laboratory passage experiments with different isolates under different selection regimes are needed, particularly those that characterize virulence phenotypes and genomic changes over time.

We also tested for the enrichment of mutations (SNPs and indels) in genes with putative virulence effects. We found that the list of indel-bearing genes contained an over-enrichment for GO term annotations related to proteolysis and metabolism (Table 1). The occurrence of indels in protease genes was particularly notable given the hypothesized role of proteases in *Bd* virulence (Rosenblum *et al*. 2013; Joneson *et al*. 2011). Given that indels are likely to have deleterious effects on protein production and function, the occurrence of indels may be indicative of loss-of-function effects for such genes. This supposition requires further testing but is consistent with the hypothesis that the laboratory-culturing environment relaxes selection on virulence-related genes relative to the host environment. With regard to SNPs in our dataset, we found that *Bd* genes previously demonstrated to be up-regulated when grown in frog skin (Rosenblum *et al*. 2012) had a higher frequency of mutational changes in the two JEL427 isolates than genes not up-regulated in frog skin (Fig. 3). In particular, one of the genes that was both up-regulated in the presence of frog skin and contained at least one nonsynonymous change between the JEL427-P9 and JEL427-P39 samples was a M36 metallopeptidase, a gene family that has previously been hypothesized to play a role in *Bd* virulence (Rosenblum *et al*. 2013; Joneson *et al*. 2011).

Here we show that the genomic processes previously proposed as mechanisms for rapid evolution in *Bd* are correlated with attenuation of virulence in laboratory culture within a single lineage of *Bd*. The fact that virulence attenuates in vitro suggests that maintaining virulence genes and extra copies of chromosomes may be costly to *Bd* at the cellular level. Moreover, the speed with which shifts in virulence can apparently occur in *Bd* suggests that this pathogen has a considerable capacity to respond to novel conditions. Over six years of laboratory passage, we observed 0.096 changes per kb, which translates into a mutation rate of $1.6 \times 10^{-5}$ changes per site per year (Table S1). A mutation rate of this magnitude is high compared to other fungi (e.g., *Saccharomyces cerevisiae*; Zeyl & DeVisser 2001) but similar to other fungal pathogens such as *Cryptococcus neoformans* (Xu 2002) and *Mycosphaerella graminicola* (Stukenbrock *et al*. 2006). Future work should endeavor to identify proximate triggers of rapid genomic changes in laboratory settings and investigate dynamics of genomic evolution of *Bd* in nature. From a practical standpoint, our study suggests that controlled infection experiments designed to predict disease outcomes should use recently isolated samples, rather than samples cultured in the lab for long periods. In addition, researchers should cryo-archive *Bd* samples immediately after isolation. *Bd* has a demonstrated ability to respond quickly to novel hosts and environmental conditions through relatively rapid increases or decreases in virulence, which has likely contributed to its spread around the world. The results of our study emphasize the importance of considering virulence on an isolate-specific basis when creating epidemiological models and predicting disease outcomes for this ecologically important pathogen.

**Acknowledgements**

This research was funded by the National Science Foundation (award IOS-1354241 to E.B.R. and Postdoctoral Research Fellowship # 1202725 to J.M.R.) and Proyecto Coqui (to P.A.B.). P.A.B. would like to thank B. Bolaños at UPR-Medical Campus for space to maintain *Bd* cultures.

**References**

Akamatsu, H., M. Taga, M. Kodama, R. Johnson, H. Otani *et al*., 1999 Molecular karyotypes for *Alternaria* plant pathogens known to produce host-specific toxins. Curr. Genet. 35: 647-656.

Berger, L., G. Marantelli, L. L. Skerratt, and R. Speare, 2005 Virulence of the amphibian chytrid fungus *Batrachochytrium dendrobatidis* varies with the strain. Dis. Aquat. Org. 68: 47-50.

Berger, L., R. Speare, P. Daszak, D. E. Green, A. A. Cunningham *et al*., 1998 Chytridiomycosis causes amphibian mortality associated with population declines in the rain forests of Australia and Central America. Proc. Natl. Acad. Sci. 95: 9031-9036.

Bishop, P. J., R. Speare, R. Poulter, M. Butler, B. J. Speare *et al*., 2009 Elimination of the amphibian chytrid fungus *Batrachochytrium dendrobatidis* by Archey's frog *Leiopelma archeyi.* Dis. Aquat. Org. 84: 9-15.

Boots, M., P. J. Hudson, and A. Sasaki, 2004 Large shifts in pathogen virulence relate to host population structure. Science 303: 842-844.

Bosch, J., I. Martínez-Solano, and M. García-París, 2001 Evidence of a chytrid fungus infection involved in the decline of the common midwife toad (*Alytes obstetricans*) in protected areas of central Spain. Biol. Conserv. 97: 331-337.

Brem, F. M., M. J. Parris, and G. E. Padgett-Flohr, 2013 Re-isolating *Batrachochytrium dendrobatidis* from an amphibian host increases pathogenicity in a subsequent exposure. PLoS ONE 8: e61260.

Butt, T. M., C. Wang, F. A. Shah, and R. Hall, 2006 Degeneration of entomogenous fungi, pp. 213-226 in *An Ecological and Societal Approach to Biological Control*, edited by J. Eilenberg and H. M. T. Hokkanen. Springer Verlag, New York.

Cingolani, P., A. Platts, L. Wang, M. Coon, T. Nguyen *et al*., 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 6: 80-92.

Crawford, A. J., K. R. Lips, and E. Bermingham, 2010 Epidemic disease decimates amphibian abundance, species diversity, and evolutionary history in the highlands of central Panama. Proc. Natl. Acad. Sci. 107: 13777-13782.

Falcon, S. and R. Gentleman, 2007 Using GOstats to test gene lists for GO term association. Bioinformatics 23: 257-258.

Farrer, R. A., D. A. Henk, T. W. J. Garner, F. Balloux, D. C. Woodhams *et al*., 2013 Chromosomal copy number variation, selection and uneven rates of recombination reveal cryptic genome diversity linked to pathogenicity. PLoS Genet. 9: e1003703.

Farrer, R. A., L. A. Weinert, J. Bielby, T. W. J. Garner, F. Balloux *et al*., 2011 Multiple emergences of genetically diverse amphibian-infecting chytrids include a globalized hypervirulent recombinant lineage. Proc. Natl. Acad. Sci. 108: 18732-18736.

Fisher, M. C., J. Bosch, Z. Yin, D. A. Stead, J. Walker *et al*., 2009 Proteomic and phenotypic profiling of the amphibian pathogen *Batrachochytrium dendrobatidis* shows that genotype is linked to virulence. Molec. Ecol. 18: 415-429.

Franzot, S. P., J. Mukherjee, R. Cherniak, L.-C. Chen, J. S. Hamdan *et al*., 1998 Microevolution of a standard strain of *Cryptococcus neoformans* resulting in differences in virulence and other phenotypes. Infect. Immun. 66: 89-97.

Ghal, M. K., J. E. Longcore, and J. E. Houlahan, 2012 Varying responses of northeastern North American amphibians to the chytrid pathogen *Batrachochytrium dendrobatidis*. Conserv. Biol. 26: 135-141.

Gervasi, S. S., C. Gondhalekar, D. H. Olson, and A. R. Blaustein, 2013 Host identity matters in the amphibian-*Batrachochytrium dendrobatidis* system: fine-scale patterns of variation in responses to a multi-host pathogen. PLoS ONE 8: e54490.

Janbon, G., K. L. Ormerod, D. Paulet, E. J. Byrnes, III, V. Yadav *et al*., 2014 Analysis of the genome and transcriptome of *Cryptococcus neoformans* var. *grubii* reveals complex RNA expression and microevolution leading to virulence attenuation. PLoS Genet. 10: e1004261.

Johnson, P. T. J., J. R. Rohr, J. T. Hoverman, E. Kellermanns, J. Bowerman *et al*., 2012 Living fast and dying of infection: host life history drives interspecific variation in infection and disease risk. Ecol. Lett. 15: 235-242.

Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk *et al*., 2008 Global trends in emerging infectious diseases. Nature 451: 990-994.

Joneson, S., J. E. Stajich, S-H. Shiu, and E. B. Rosenblum, 2011 Genomic transition to pathogenicity in chytrid fungi. PLoS Pathogens 11: e1002338.

Lam, B. A., J. B. Walke, V. T. Vredenburg, and R. N. Harris, 2010 Proportion of individuals with anti-*Batrachochytrium dendrobatidis* skin bacteria is associated with population persistence in the frog *Rana muscosa*. Biol. Conserv. 143: 529-531.

Langhammer, P. F., K. R. Lips, P. A, Burrowes, T. Tunstall, C. M. Palmer *et al*., 2013 A fungal pathogen of amphibians, *Batrachochytrium dendrobatidis*, attenuates in pathogenicity with in vitro passages. PLoS ONE 8: e77630.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg, 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10: R25.

Lips, K. R., F. Brem, R. Brenes, J. D. Reeve, R. A. Alford *et al*., 2006 Emerging infectious disease and the loss of biodiversity in a Neotropical amphibian community. Proc. Natl. Acad. Sci. 103: 3165-3170.

Lord, J. C., and D. W. Roberts, 1986 The effects of culture-medium quality and host passage on zoosporogenesis, oosporogenetis, and infectivity of *Lagenidium giganteum* (Oomycetes, Lagenidiales). J. Invert. Pathol. 48: 355-361.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al*., 2010 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303.

Mitchell, S. E., E. S. Rogers, T. J. Little, and A. F. Read, 2005 Host-parasite and genotype-by-environment interactions: temperature modifies potential for selection by a sterilizing pathogen. Evolution 59: 70-80.

Muths, E., P. S. Corn, A. P. Pessier, and D. E. Green, 2003 Evidence for disease related amphibian decline in Colorado. Biol. Conserv. 110: 357-365.

Piovia-Scott, J., K. Pope, S. J. Worth, E. B. Rosenblum, T. Poorten *et al*., 2014 Phenotypic and genomic correlates of virulence in the frog-killing fungus *Batrachochytrium dendrobatidis*: evidence from a California amphibian die-off. ISME J. 2014: 1-9.

Quinlan, A. R. and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-842.

Retallick, R. W. R., and V. Miera, 2005 Strain differences in the amphibian chytrid *Batrachochytrium dendrobatidis* and non-permanent, sub-lethal effects of infection. Dis. Aquat. Org. 75: 201-207.

Retallick, R. W. R., H. McCallum, and R. Speare, 2004 Endemic infection of the amphibian chytrid fungus in a frog community post-decline. PLoS Biol. 2: e351.

Rosenblum, E. B., T. Y. James, K. R. Zamudio, T. J. Poorten, D. Ilut *et al*., 2013 Complex history of the amphibian-killing chytrid fungus revealed with genome resequencing data. Proc. Natl. Acad. Sci. 110: 9385-9390.

Rosenblum, E. B., T. J. Poorten, S. Joneson, and M. Settles, 2012 Substrate-specific gene expression in *Batrachochytrium dendrobatidis*, the chytrid pathogen of amphibians. PLoS ONE 7: e49924.

Safavi, S. A., 2012 Attenuation of the entomopathogenic fungus *Beauveria bassiana* following serial in vitro transfers. Biologica 67: 1062-1068.

Schloegel, L. M., J. M. Hero, L. Berger, R. Speare, K. McDonald *et al*., 2006 The decline of the sharp-snouted day frog (*Taudactylus acutirostris*): the first documented case of extinction by infection in a free-ranging wildlife species. EcoHealth 3: 35-40.

Searle, C. L., S. S. Gervasi, J. Hua, J. I. Hammond, R. A. Relyea *et al*., 2011 Differential host susceptibility to *Batrachochytrium dendrobatidis*, an emerging amphibian pathogen. Conserv. Biol. 25: 965-974.

Spitzen-Van der Sluijs, A., A. Martel, C. A. Hallmann, W. Bosman, T. W. J. Garner *et al*., 2014 Environmental determinants of recent endemism of *Batrachochytrium dendrobatidis* infections in amphibian assemblages in the absence of disease outbreaks. Conserv. Biol. 28: 1302-1311.

Stukenbrock, E. H., S. Banke, M. Javan-Nikkhah, and B. A. McDonald, 2006 Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. Mol. Biol. Evol. 24: 398-411.

Voyles, J. 2011 Phenotypic profiling of *Batrachochytrium dendrobatidis*, a lethal fungal pathogen of amphibians. Fungal Ecol. 4: 196-200.

Voyles, J., L. R. Johnson, C. J. Briggs, S. D. Cashins, R. A. Alford *et al*., 2012 Temperature alters reproductive life history patterns in *Batrachochytrium dendrobatidis*, a lethal pathogen associated with the global loss of amphibians. Ecol. Evol. 2: 2241-2249.

Voyles, J., E. B. Rosenblum, and L. Berger, 2011 Interactions between *Batrachochytrium dendrobatidis* and its amphibian hosts: a review of pathogenesis and immunity. Microbes Immun. 13: 25-32.

Vredenburg, V. T., R. A. Knapp, T. S. Tunstall, and C. J. Briggs, 2010 Dynamics of an emerging disease drive large-scale amphibian population extinctions. Proc. Natl. Acad. Sci. 107: 9689-9694.

Wang, C., A. Skrobek, and T. M. Butt, 2003 Concurrence of losing a chromosome and the ability to produce destruxins in a mutant of *Metarhizium anisopliae*. FEMS Microbiol. Lett. 226: 373-378.

Wang, C., M. A. Typas, and T. M. Butt, 2002 Detection and characterization of pr1 virulent gene deficiencies in the insect pathogenic fungus *Metarhizium anisopliae*. FEMS Microbiol. Lett. 213: 251-255.

Woodhams, D. C., V. T. Vredenburg, M-A. Simon, D. Billheimer, B. Shakhtour *et al*., 2007 Symbiotic bacteria contribute to innate immune defenses of the threatened mountain yellow-legged frog, *Rana muscosa*. Biol. Conserv. 138: 390-398.

Xu, J., 2002 Estimating the spontaneous mutation rate of loss of sex in the human pathogenic fungus *Cryptococcus neoformans*. Genetics 162: 1157-1167.

Zeyl, C. and J. A. G. M. DeVisser, 2001 Estimates of the rate and distribution of fitness effects of spontaneous mutation in *Saccharomyces cerevisiae*. Genetics 157: 53-61.

Zhbannikov, I .Y., S. S. Hunter, and M. L. Settles, SeqyClean: a Software Tool for Comprehensive Preprocessing of Sequence Data. https://bitbucket.org/izhbannikov/seqyclean.

Zolan, M. E., and P. J. Pukkila, 1986 Inheritance of DNA methylation in *Coprinus cinereus*. Mol. Cell Biol. 6: 195-200.

**Tables**

Table 1. Over-enriched Gene Ontology (GO) terms in genes containing exonic indels

| GO Category | GO Term | P value | Count | GO Term Size | GO Term Name |
|---|---|---|---|---|---|
| Biological Process | GO:0006508 | 0.0010 | 9 | 282 | proteolysis |
| | GO:0043170 | 0.0296 | 18 | 1290 | macromolecule metabolic process |
| | GO:0008152 | 0.0431 | 25 | 2147 | metabolic process |
| Molecular Function | GO:0008236 | 0.0035 | 4 | 64 | serine-type peptidase activity |
| | GO:0016787 | 0.0104 | 17 | 991 | hydrolase activity |
| | GO:0004190 | 0.0267 | 4 | 115 | aspartic-type endopeptidase activity |
| | GO:0070011 | 0.0435 | 5 | 214 | peptidase activity, acting on L-amino acid peptides |

## Figures

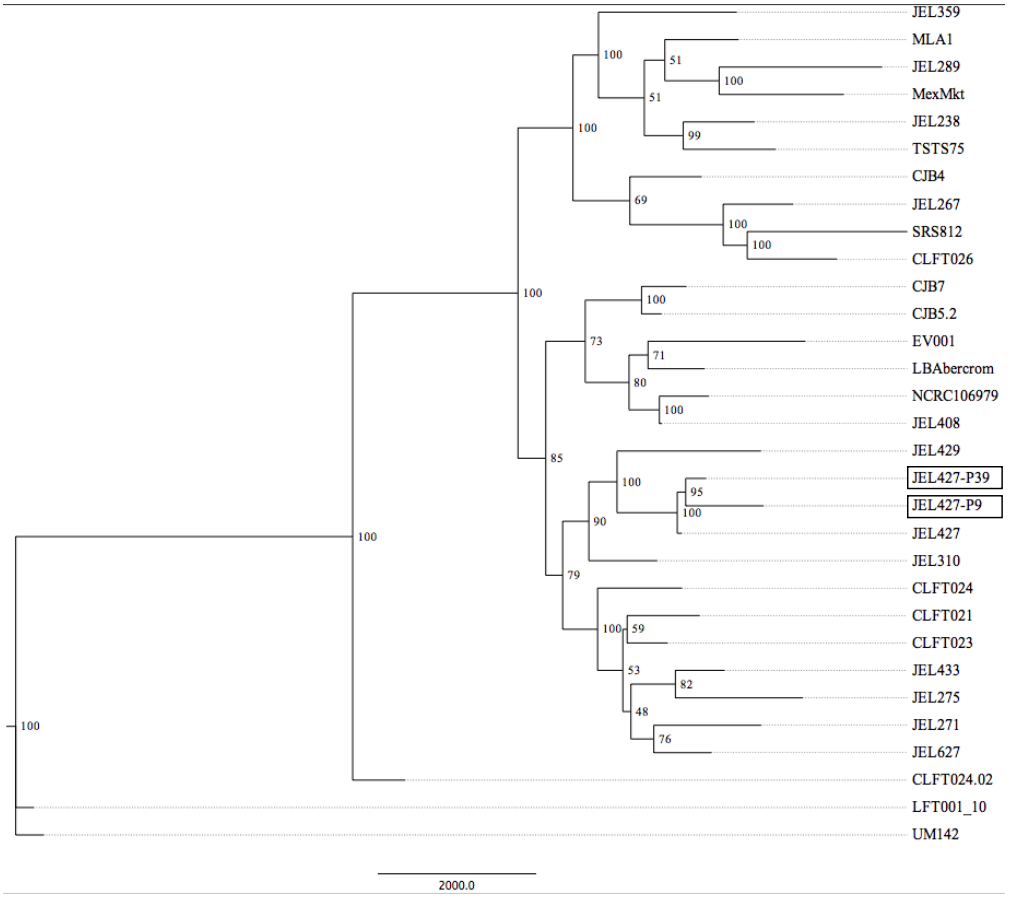



**Figure 1.** Rooted *Bd* phylogeny based on genomic data. Maximum parsimony tree is shown with nodal support values generated from 200 bootstrap replicates. The two JEL427 isolates (highlighted with boxes) cluster together and are solidly nested within the Global Panzootic Lineage (GPL). The isolate labeled "JEL427" was previously sequenced in Rosenblum *et al*. (2013).

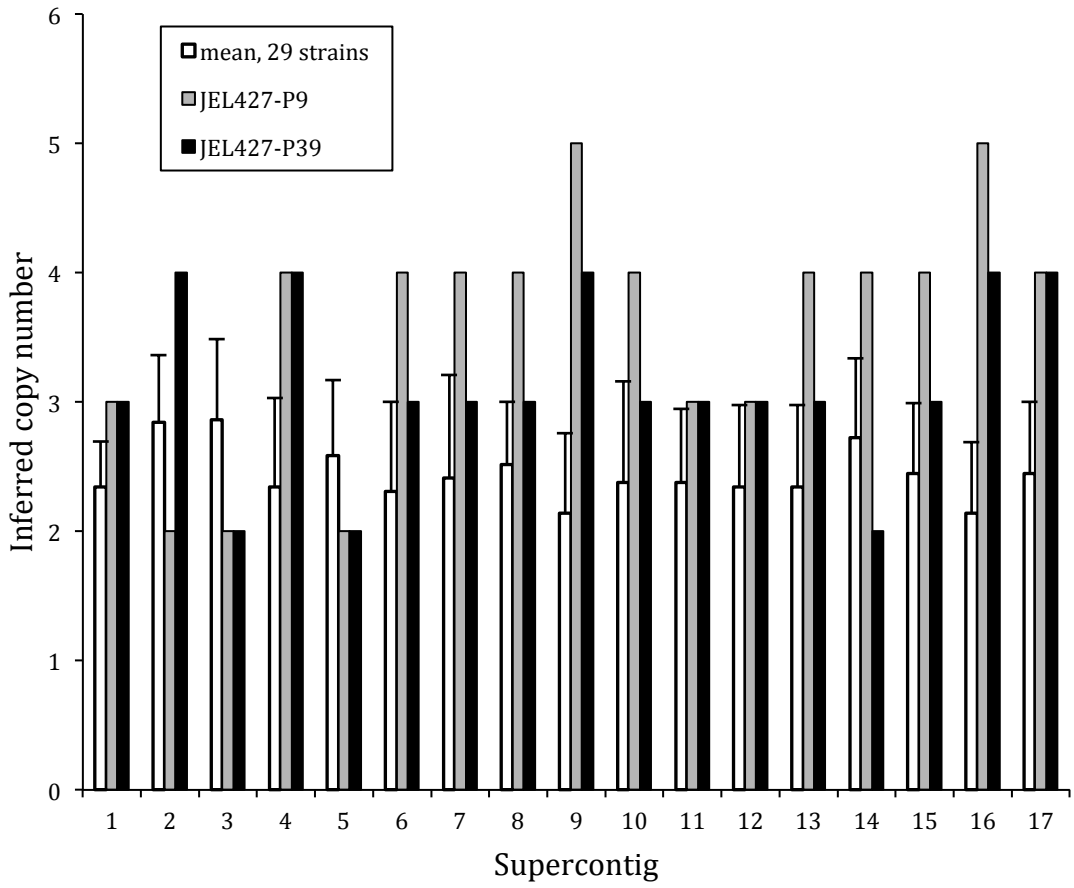



**Figure 2.** Evidence for decrease in chromosomal copy numbers between the ancestral *Bd* isolate (JEL427-P9) and the derived isolate (JEL427-P39), which differ only in laboratory passage history. Mean copy numbers from 29 *Bd* isolates resequenced in Rosenblum *et al*. (2013) are shown for comparison. Supercontigs are analogous to chromosomes in *Bd*.

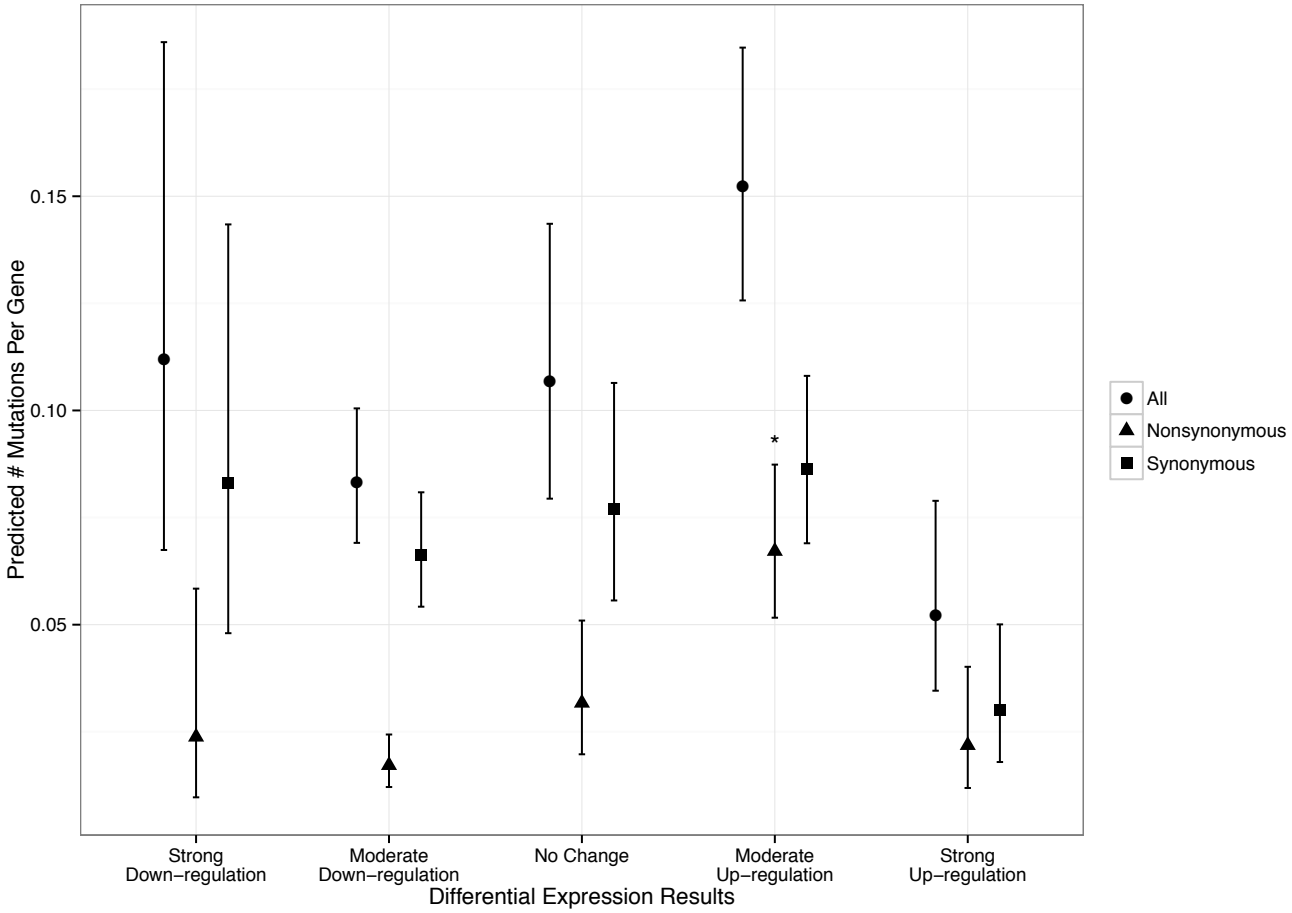



**Figure 3.** *Bd* genes that were moderately upregulated in the presence of frog skin showed greater than expected frequency of nonsynonymous mutation in the JEL427-P39 (compared to the JEL427-P9) isolate. The x-axis clusters genes based on their differential expression (DE) profiles in a previous study (Rosenblum *et al*. 2012), where "up-regulation" implies increased gene expression when grown on frog skin compared to standard tryptone growth medium. The y-axis shows number of mutations at all sites (circles), nonsynonymous sites (triangles) and synonymous sites (squares). The asterisk indicates the genes that were moderately up-regulated in frog skin had a significantly higher number of nonsynonymous mutational changes.

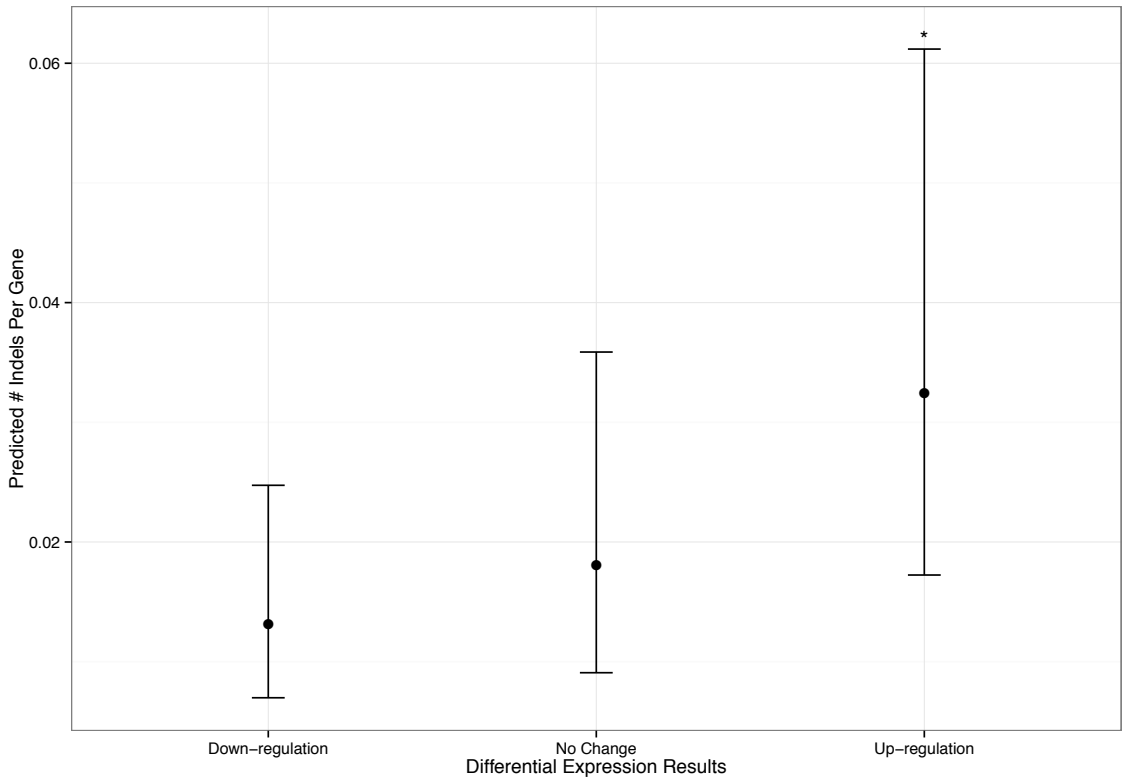



**Figure 4.** *Bd* genes that were upregulated in the presence of frog skin showed greater than expected frequency of indels in the JEL427-P39 (compared to the JEL427-P9) isolate. The x-axis clusters genes based on their differential expression (DE) profiles in a previous study (Rosenblum *et al*. 2012), where "up-regulation" implies increased gene expression when grown on frog skin compared to standard tryptone growth medium. The y-axis shows number of indels. The asterisk indicates the genes that were up-regulated in frog skin had a significantly higher number of indels.

## Appendices

Appendix for Chapter 2

S1 Table. Combined experimental data table including Bd infection intensities, body weights, and histological results from stained skin sections.

| Subject ID | Species | Treatment | Histology selection | Histology results: Chytrid thalli | Histology results: Approximate burden | Histology results: Secondary changes (hyperkeratosis) | Histology results: Secondary changes (dermatitis) | Histology results: Secondary changes (spongiosis) | Notes |
|---|---|---|---|---|---|---|---|---|---|
| B_boreas1 | Bufo boreas | Bd-exposed | **yes** | yes | **3+** | mild, multifocal parakeratotic hyperkeratosis | - | mild spongiosis | |
| B_boreas2 | Bufo boreas | Bd-exposed | **yes** | yes | **1+** | **minimal** multifocal parakeratic hyperkeratosis with **mild epidermal hyperplasia** | - | mild spongiosis | |
| B_boreas7 | Bufo boreas | Bd-exposed | **yes** | yes | **3+** | **moderate**, multifocal, parakeratotic hyperkeratosis **with epidermal hyperplasia** | **mild to moderate**, multifocal, chronic lymphoplasmacytic dermatitis | **moderate** spongiosis | |
| B_boreas8 | Bufo boreas | Bd-exposed | **yes** | yes | **2+** | mild, multifocal parakeratotic hyperkeratosis **with epidermal hyperplasia** | **mild, focal**, chronic lymphoplasmacytic dermatitis | **moderate** spongiosis | |
| B_boreas12 | Bufo boreas | Bd-exposed | **yes** | yes | **2+** | mild, multifocal, parakeratotic hyperkeratosis | - | **moderate** spongiosis | |
| B_boreas3 | Bufo boreas | Control | **yes** | no | NA | - | **minimal, focal**, chronic lymphoplasmacytic dermatitis | - | |
| B_boreas4 | Bufo boreas | Control | **yes** | no | NA | - | - | - | |
| B_boreas6 | Bufo boreas | Control | **yes** | no | NA | - | **minimal, focal**, chronic lymphoplasmacytic dermatitis | - | |
| B_boreas9 | Bufo boreas | Control | **yes** | no | NA | - | - | - | |
| B_boreas10 | Bufo boreas | Control | **yes** | no | NA | - | - | - | |
| B_marinus1 | Bufo marinus | Bd-exposed | **yes** | no | NA | minimal, multifocal parakeratotic hyperkeratosis | - | - | |

| Subject ID | Species | Treatment | Histology selection | Histology results: Chytrid thalli | Histology results: Approximate burden | Histology results: Secondary changes (hyperkeratosis) | Histology results: Secondary changes (dermatitis) | Histology results: Secondary changes (spongiosis) | Notes |
|---|---|---|---|---|---|---|---|---|---|
| B_marinus2 | Bufo marinus | Bd-exposed | **yes** | no | NA | minimal, multifocal parakeratotic hyperkeratosis | minimal, multifocal, chronic lymphoplasmacytic dermatitis | - | |
| B_marinus6 | Bufo marinus | Bd-exposed | **no** | NA | NA | - | - | - | |
| B_marinus8 | Bufo marinus | Bd-exposed | **yes** | yes | 1+ (mostly empty thalli) | **mild**, multifocal parakeratotic hyperkeratosis | minimal, multifocal chronic lymphoplasmacytic dermatitis | - | |
| B_marinus14 | Bufo marinus | Bd-exposed | **yes** | no | NA | minimal, multifocal parakeratotic hyperkeratosis | minimal, multifocal chronic lymphoplasmacytic dermatitis | - | |
| B_marinus20 | Bufo marinus | Bd-exposed | **yes** | yes | 1+ (single thallus) | minimal, multifocal parakeratotic hyperkeratosis | **mild focal** chronic **lymphohistiocytic** dermatitis | - | |
| B_marinus3 | Bufo marinus | Control | **yes** | no | NA | - | mild, multifocal, chronic lymphoplasmacytic dermatitis | **mild spongiosis** | Intradermal nematode (incidental) |
| B_marinus4 | Bufo marinus | Control | **yes** | no | NA | - | mild, multifocal, chronic lymphoplasmacytic dermatitis | - | |
| B_marinus5 | Bufo marinus | Control | **yes** | no | NA | - | - | - | |
| B_marinus7 | Bufo marinus | Control | **yes** | no | NA | - | mild, multifocal, chronic lymphoplasmacytic dermatitis | - | |
| B_marinus9 | Bufo marinus | Control | **yes** | no | NA | - | **minimal**, multifocal, chronic lymphoplasmacytic dermatitis | - | |

Appendix for Chapter 3

Table S1. Samples included in the nuclear DNA data analyses

| Sample ID | Lake ID | Site ID (as in Figure 1 map) | Cluster ID (inferred from analyses) | UTMe | UTMn |
|---|---|---|---|---|---|
| RKS1014 | 70335 | 1 | 1 | 251316.419457 | 4219097.382417 |
| RKS2862 | 70335 | 1 | 1 | 251316.419457 | 4219097.382417 |
| RKS7081 | 70508 | 1 | 1 | 258097 | 4218951 |
| RKS7091 | 70399 | 1 | 1 | 254829 | 4218600 |
| RKS7102 | 70508 | 1 | 1 | 258113 | 4219009 |
| RKS7170 | 70355 | 1 | 1 | 254969.885516 | 4218317.10368 |
| RKS7360 | 70399 | 1 | 1 | 254751 | 4218641 |
| RKS7608 | 70355 | 1 | 1 | 254864 | 4218421 |
| RKS7981 | 74062 | 1 | 1 | 254865.350026 | 4218657.367395 |
| RKS9599 | 70508 | 1 | 1 | 258172.731165 | 4218873.133056 |
| RKS9721 | 70508 | 1 | 1 | 258172.731165 | 4218873.133056 |
| TP34 | 70335 | 1 | 1 | 251316 | 4219097 |
| RKS1126 | 70529 | 2 | 1 | 263314.659345 | 4219165.45252 |
| RKS1134 | 70558 | 2 | 1 | 265104.651562 | 4220166.878809 |
| RKS1137 | 70529 | 2 | 1 | 263314.659345 | 4219165.45252 |
| RKS1200 | 70197 | 2 | 1 | 267156.540976 | 4218515.762895 |
| RKS1217 | 70197 | 2 | 1 | 267156.540976 | 4218515.762895 |
| RKS4107 | 70184 | 2 | 1 | 263784.090299 | 4215031.303103 |
| RKS1798 | 70660 | 3 | 1 | 273280.540694 | 4228354.323323 |
| RKS5933 | 72135 | 3 | 1 | 273839.491837 | 4227831.441091 |
| RKS5937 | 72313 | 3 | 1 | 268797.9096 | 4223722.513274 |
| RKS5939 | 70518 | 3 | 1 | 267753.288222 | 4221591.17745 |
| RKS5940 | 70518 | 3 | 1 | 267753.288222 | 4221591.17745 |
| RKS1441 | 70597 | 4 | 1 | 274007.526332 | 4212475.234129 |
| RKS1459 | 70149 | 4 | 1 | 273274.115735 | 4212594.319657 |
| RKS1464 | 70149 | 4 | 1 | 273274.115735 | 4212594.319657 |
| RKS9690 | 70439 | 4 | 1 | 268252.161265 | 4212438.848885 |
| RKS9692 | 70439 | 4 | 1 | 268252.161265 | 4212438.848885 |
| RKS9693 | 70439 | 4 | 1 | 268252.161265 | 4212438.848885 |
| RKS9694 | 70439 | 4 | 1 | 268252.161265 | 4212438.848885 |
| RKS9695 | 70439 | 4 | 1 | 268252.161265 | 4212438.848885 |
| RKS9697 | 70439 | 4 | 1 | 268252.161265 | 4212438.848885 |
| RKS9699 | 70439 | 4 | 1 | 268252.161265 | 4212438.848885 |
| RKS1293 | 70528 | 5 | 1 | 276591.480374 | 4216469.419183 |

| | | | | | |
|---|---|---|---|---|---|
| RKS1332 | 70528 | 5 | 1 | 276591.480374 | 4216469.419183 |
| RKS1364 | 70528 | 5 | 1 | 276591.480374 | 4216469.419183 |
| RKS5942 | 70629 | 5 | 1 | 276944.401135 | 4221315.211266 |
| RKS5945 | 70403 | 5 | 1 | 277223.924801 | 4221440.054708 |
| RKS5956 | 70629 | 5 | 1 | 276944.401135 | 4221315.211266 |
| RKS1039 | 72105 | 6 | 1 | 283512.384861 | 4215479.719931 |
| RKS1044 | 71967 | 6 | 1 | 284092.49214 | 4215842.114759 |
| RKS1222 | 71967 | 6 | 1 | 284092.49214 | 4215842.114759 |
| RKS1454 | 71465 | 6 | 1 | 276412.460087 | 4211013.154429 |
| RKS17527 | 70215 | 7.1 | 2 | 285424 | 4212093 |
| RKS17529 | 70215 | 7.1 | 2 | 285424 | 4212093 |
| RKS17530 | 70215 | 7.1 | 2 | 285424 | 4212093 |
| RKS17531 | 70448 | 7.1 | 2 | 286340 | 4211428 |
| RKS17508 | 70525 | 7.2 | 2 | 282716 | 4206635 |
| RKS17509 | 70525 | 7.2 | 2 | 282716 | 4206635 |
| RKS17510 | 70251 | 7.2 | 2 | 283409 | 4207490 |
| RKS17511 | 70251 | 7.2 | 2 | 283409 | 4207490 |
| RKS17516 | 70251 | 7.2 | 2 | 283409 | 4207490 |
| RKS17524 | 70525 | 7.2 | 2 | 282716 | 4206635 |
| RKS17525 | 70525 | 7.2 | 2 | 282716 | 4206635 |
| RKS5436 | 70255 | 8 | 2 | 292984.29802 | 4207019.668843 |
| RKS5949 | 70255 | 8 | 2 | 292984.29802 | 4207019.668843 |
| RKS5950 | 70255 | 8 | 2 | 292984.29802 | 4207019.668843 |
| RKS5952 | 70255 | 8 | 2 | 292984.29802 | 4207019.668843 |
| RKS5965 | 74335 | 8 | 2 | 293129 | 4206846 |
| RKS5968 | 74335 | 8 | 2 | 293129 | 4206846 |
| RKS5969 | 74335 | 8 | 2 | 293129 | 4206846 |
| RKS7434 | 72996 | 9 | 2 | 294105.879469 | 4204959.849999 |
| RKS7937 | 72996 | 9 | 2 | 294105.879469 | 4204959.849999 |
| RKS9344 | 72996 | 9 | 2 | 294167 | 4204929 |
| RKS9540 | 72996 | 9 | 2 | 294070 | 4204904 |
| RKS9543 | 72996 | 9 | 2 | 294112 | 4205012 |
| RKS9544 | 72996 | 9 | 2 | 294070 | 4204909 |
| MTS13067 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13068 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13069 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13070 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13072 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13075 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13082 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13084 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |

| | | | | | |
|---|---|---|---|---|---|
| MTS13090 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13091 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13093 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13094 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13095 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13097 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13098 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13099 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13100 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13101 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13102 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13103 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| MTS13104 | 72808 | 10 | 2 | 290122.2866 | 4190793.714 |
| RKS2820 | 70397 | 11 | 2 | 296751.115506 | 4187040.512809 |
| RKS2822 | 70397 | 11 | 2 | 296751.115506 | 4187040.512809 |
| TP106 | 70397 | 11 | 2 | 296306 | 4186973 |
| TP111 | 70397 | 11 | 2 | 296306 | 4186973 |
| TP112 | 70397 | 11 | 2 | 296306 | 4186973 |
| TP16 | 70397 | 11 | 2 | 296306 | 4186973 |
| RKS3062 | 72476 | 11.5 | 2 | 304944.60743 | 4185301.884014 |
| RKS3125 | 74269 | 11.5 | 2 | 303903.502238 | 4185005.129603 |
| MTS13035 | 72425 | 12 | 2 | 304880.202156 | 4191583.301797 |
| MTS13037 | 72425 | 12 | 2 | 304880.202156 | 4191583.301797 |
| MTS13042 | 72425 | 12 | 2 | 304880.202156 | 4191583.301797 |
| MTS13043 | 72425 | 12 | 2 | 304880.202156 | 4191583.301797 |
| MTS13044 | 72425 | 12 | 2 | 304880.202156 | 4191583.301797 |
| RKS2829 | 70543 | 12 | 2 | 304880.202156 | 4191583.301797 |
| TP1 | 72425 | 12 | 2 | 304880.202156 | 4191583.301797 |
| RKS2809 | 74281 | 14.1 | 2 | 295985.623643 | 4182475.450867 |
| RKS2811 | 74281 | 14.1 | 2 | 295985.623643 | 4182475.450867 |
| RKS2815 | 74281 | 14.1 | 2 | 295985.623643 | 4182475.450867 |
| RKS2816 | 74281 | 14.1 | 2 | 295985.623643 | 4182475.450867 |
| RKS2817 | 74281 | 14.1 | 2 | 295985.623643 | 4182475.450867 |
| TP100 | 74281 | 14.1 | 2 | 295985.623643 | 4182475.450867 |
| TP85 | 74281 | 14.1 | 2 | 295985.623643 | 4182475.450867 |
| TP87 | 74281 | 14.1 | 2 | 295985.623643 | 4182475.450867 |
| TP89 | 74281 | 14.1 | 2 | 295985.623643 | 4182475.450867 |
| TP92 | 74281 | 14.1 | 2 | 295985.623643 | 4182475.450867 |
| TP94 | 74281 | 14.1 | 2 | 295985.623643 | 4182475.450867 |
| TP96 | 74281 | 14.1 | 2 | 295985.623643 | 4182475.450867 |
| RKS2801 | 70113 | 14.2 | 2 | 291851.054486 | 4183763.144507 |

| | | | | | |
|---|---|---|---|---|---|
| RKS2803 | 70113 | 14.2 | 2 | 291851.054486 | 4183763.144507 |
| RKS2804 | 70113 | 14.2 | 2 | 291851.054486 | 4183763.144507 |
| RKS2805 | 70239 | 14.2 | 2 | 292104.564245 | 4183529.44028 |
| RKS2808 | 70239 | 14.2 | 2 | 292104.564245 | 4183529.44028 |
| TP2 | 70239 | 14.2 | 2 | 291881 | 4183923 |
| TP3 | 70239 | 14.2 | 2 | 291881 | 4183923 |
| TP4 | 70239 | 14.2 | 2 | 291881 | 4183923 |
| TP8 | 70239 | 14.2 | 2 | 291881 | 4183923 |
| RKS3839 | 72472 | 15 | 2 | 299842.850955 | 4177011.108601 |
| RKS3876 | 70062 | 15 | 3 | 297468.600526 | 4177822.469059 |
| RKS3894 | 70567 | 15 | 2 | 297345.989245 | 4177644.657047 |
| RKS4031 | 70567 | 15 | 2 | 297345.989245 | 4177644.657047 |
| RKS4040 | 70567 | 15 | 3 | 297345.989245 | 4177644.657047 |
| RKS4041 | 70284 | 15 | 2 | 297450.172993 | 4178061.264695 |
| RKS4046 | 70062 | 15 | 3 | 297468.600526 | 4177822.469059 |
| RKS5755 | 70122 | 16 | 3 | 285499.676354 | 4176352.57762 |
| RKS5758 | 70122 | 16 | 3 | 285499.676354 | 4176352.57762 |
| RKS5760 | 70074 | 16 | 3 | 285526.309622 | 4176321.822717 |
| RKS5762 | 72986 | 16 | 3 | 285222.944042 | 4176561.718995 |
| RKS5763 | 72986 | 16 | 3 | 285222.944042 | 4176561.718995 |
| RKS7194 | 70343 | 17 | 3 | 289285.81745 | 4171029.821552 |
| RKS7196 | 70343 | 17 | 3 | 289285.81745 | 4171029.821552 |
| RKS7199 | 70343 | 17 | 3 | 289285.81745 | 4171029.821552 |
| RKS7204 | 72759 | 17 | 3 | 289774.447214 | 4170300.699672 |
| RKS7205 | 72759 | 17 | 3 | 289774.447214 | 4170300.699672 |
| RKS7206 | 72759 | 17 | 3 | 289774.447214 | 4170300.699672 |
| RKS7249 | 70583 | 17 | 3 | 289564.442505 | 4171109.829995 |
| RKS7314 | 70278 | 17 | 3 | 289910.01271 | 4170363.289444 |
| RKS7317 | 70278 | 17 | 3 | 289910.01271 | 4170363.289444 |
| RKS7320 | 70278 | 17 | 3 | 289910.01271 | 4170363.289444 |
| RKS7321 | 70278 | 17 | 3 | 289910.01271 | 4170363.289444 |
| RKS7370 | 70563 | 18.1 | 3 | 286844 | 4166752 |
| RKS7458 | 70563 | 18.1 | 3 | 286615 | 4166653 |
| RKS7505 | 70736 | 18.1 | 3 | 287560 | 4166686 |
| RKS8779 | 70736 | 18.1 | 3 | 287522 | 4166755 |
| JR02 | 70488 | 18.2 | 3 | 279418 | 4165419 |
| JR03 | 70488 | 18.2 | 3 | 279418 | 4165419 |
| RKS2858 | 70326 | 18.2 | 3 | 278375.144156 | 4165617.783099 |
| TP64 | 70541 | 19 | 3 | 272604 | 4165806 |
| TP65 | 70541 | 19 | 3 | 272579 | 4165782 |
| TP66 | 70541 | 19 | 3 | 272579 | 4165782 |

| | | | | | |
|---|---|---|---|---|---|
| TP40 | 70314 | 20 | 3 | 266160 | 4172937 |
| TP41 | 70314 | 20 | 3 | 266160 | 4172937 |
| TP42 | 70314 | 20 | 3 | 266160 | 4172937 |
| TP48 | 70314 | 20 | 3 | 266142 | 4173197 |
| TP58 | 70727 | 20 | 3 | 265059 | 4175198 |
| TP59 | 70727 | 20 | 3 | 265059 | 4175198 |
| TP60 | 70727 | 20 | 3 | 265292 | 4174936 |

# Appendix for Chapter 4

**Supporting Information, Table 1.** Summary of 2,231 nucleotide changes between two Bd isolates, JEL427-P9 and JEL427-P39, differing only in laboratory passage history (A). Following SNP filtering, summary information from the SNP dataset (e.g., genomic position, synonymous vs. nonsynonymous mutations) were extracted using snpEff software (version 2.0.5, Cingolani et al. 2012) and custom R scripts to compare the two isolates. The snpEff results were qualitatively similar to the pattern observed in the Rosenblum et al. 2013 dataset of 29 global isolates (B).

**A: JEL427-P9 to JEL427-P39**

**Changes by type**

| | # of changes | Per cent of changes |
|---|---|---|
| **DOWNSTREAM** | 894 | 25.88% |
| **INTERGENIC** | 17 | 0.49% |
| **INTRON** | 641 | 18.56% |
| **NON_SYNONYMOUS_CODING** | 300 | 8.69% |
| **SPLICE_SITE_ACCEPTOR** | 1 | 0.03% |
| **SPLICE_SITE_DONOR** | 1 | 0.03% |
| **START_GAINED** | 1 | 0.03% |
| **STOP_GAINED** | 2 | 0.06% |
| **STOP_LOST** | 0 | 0.00% |
| **SYNONYMOUS_CODING** | 642 | 18.59% |
| **SYNONYMOUS_STOP** | 2 | 0.06% |
| **UPSTREAM** | 930 | 26.93% |
| **UTR_3_PRIME** | 17 | 0.49% |
| **UTR_5_PRIME** | 6 | 0.17% |

| Chromosome | Length | Changes | Change rate | Change per kb | Normalized |
|---|---|---|---|---|---|
| bden_JEL423_supercont1.1 | 4,440,149 | 503 | 8,827 | 0.113 | 0.051 |
| bden_JEL423_supercont1.2 | 2,313,122 | 31 | 74,616 | 0.013 | 0.006 |
| bden_JEL423_supercont1.3 | 1,829,408 | 62 | 29,506 | 0.034 | 0.015 |
| bden_JEL423_supercont1.4 | 1,803,316 | 160 | 11,270 | 0.089 | 0.040 |
| bden_JEL423_supercont1.5 | 1,707,251 | 42 | 40,648 | 0.025 | 0.011 |
| bden_JEL423_supercont1.6 | 1,545,501 | 194 | 7,966 | 0.126 | 0.056 |
| bden_JEL423_supercont1.7 | 1,398,854 | 263 | 5,318 | 0.188 | 0.084 |
| bden_JEL423_supercont1.8 | 1,069,847 | 49 | 21,833 | 0.046 | 0.021 |
| bden_JEL423_supercont1.9 | 1,057,463 | 123 | 8,597 | 0.116 | 0.052 |
| bden_JEL423_supercont1.10 | 1,012,305 | 137 | 7,389 | 0.135 | 0.061 |
| bden_JEL423_supercont1.11 | 979,369 | 58 | 16,885 | 0.059 | 0.027 |

| | | | | | |
|---|---|---|---|---|---|
| bden_JEL423_supercont1.12 | 937,107 | 150 | 6,247 | 0.160 | 0.072 |
| bden_JEL423_supercont1.13 | 898,261 | 158 | 5,685 | 0.176 | 0.079 |
| bden_JEL423_supercont1.14 | 857,155 | 89 | 9,630 | 0.104 | 0.047 |
| bden_JEL423_supercont1.15 | 557,602 | 100 | 5,576 | 0.179 | 0.080 |
| bden_JEL423_supercont1.16 | 498,254 | 57 | 8,741 | 0.114 | 0.051 |
| bden_JEL423_supercont1.17 | 243,426 | 50 | 4,868 | 0.205 | 0.092 |
| bden_JEL423_supercont1.21 | 28,371 | 4 | 7,092 | 0.141 | 0.063 |
| bden_JEL423_supercont1.46 | 7,014 | 1 | 7,014 | 0.143 | 0.064 |
| **Total** | **23,183,77** | **2,231** | **10,391** | **0.096** | |
| **Ts/Tv ratio** | **5.4109** | | | | |

**Genotype changes**

| | # | proportion |
|---|---|---|
| het -> hom | 646 | 0.299 |
| hom -> het | 1477 | 0.683 |
| hom1 -> hom2 | 38 | 0.018 |
| **Total** | **2161** | |

**B: Results from Rosenblum *et al*. 2013 (Proc. Natl. Acad. Sci.)**

**Changes by type**

| | # of changes | Per cent of changes |
|---|---|---|
| **DOWNSTREAM** | 49,685 | 26.59% |
| **INTERGENIC** | 1,970 | 1.05% |
| **INTRON** | 26,688 | 14.28% |
| **NON_SYNONYMOUS_CODING** | 19,756 | 10.57% |
| **SPLICE_SITE_ACCEPTOR** | 59 | 0.03% |
| **SPLICE_SITE_DONOR** | 75 | 0.04% |
| **START_GAINED** | 50 | 0.03% |
| **STOP_GAINED** | 114 | 0.06% |
| **STOP_LOST** | 26 | 0.01% |
| **SYNONYMOUS_CODING** | 38,397 | 20.55% |
| **SYNONYMOUS_STOP** | 51 | 0.03% |
| **UPSTREAM** | 48,665 | 26.05% |
| **UTR_3_PRIME** | 973 | 0.52% |
| **UTR_5_PRIME** | 306 | 0.16% |

| Chromosome | Length | Changes | Change rate | Change per kb | Normalized |
|---|---|---|---|---|---|
| bden_JEL423_supercont1.1 | 4,440,149 | 23,541 | 188 | 5.302 | 0.043 |
| bden_JEL423_supercont1.2 | 2,313,122 | 10,157 | 227 | 4.391 | 0.036 |

| | | | | | |
|---|---|---|---|---|---|
| bden_JEL423_supercont1.3 | 1,829,408 | 9,588 | 190 | 5.241 | 0.043 |
| bden_JEL423_supercont1.4 | 1,803,316 | 9,178 | 196 | 5.090 | 0.042 |
| bden_JEL423_supercont1.5 | 1,707,251 | 8,741 | 195 | 5.120 | 0.042 |
| bden_JEL423_supercont1.6 | 1,545,501 | 8,008 | 192 | 5.181 | 0.042 |
| bden_JEL423_supercont1.7 | 1,398,854 | 7,764 | 180 | 5.550 | 0.045 |
| bden_JEL423_supercont1.8 | 1,069,847 | 5,834 | 183 | 5.453 | 0.045 |
| bden_JEL423_supercont1.9 | 1,057,463 | 6,080 | 173 | 5.750 | 0.047 |
| bden_JEL423_supercont1.10 | 1,012,305 | 5,766 | 175 | 5.696 | 0.047 |
| bden_JEL423_supercont1.11 | 979,369 | 5,427 | 180 | 5.541 | 0.045 |
| bden_JEL423_supercont1.12 | 937,107 | 4,907 | 190 | 5.236 | 0.043 |
| bden_JEL423_supercont1.13 | 898,261 | 5,433 | 165 | 6.048 | 0.049 |
| bden_JEL423_supercont1.14 | 857,155 | 3,102 | 276 | 3.619 | 0.030 |
| bden_JEL423_supercont1.15 | 557,602 | 4,124 | 135 | 7.396 | 0.061 |
| bden_JEL423_supercont1.16 | 498,254 | 2,788 | 178 | 5.596 | 0.046 |
| bden_JEL423_supercont1.17 | 243,426 | 1,333 | 182 | 5.476 | 0.045 |
| bden_JEL423_supercont1.18 | 48,566 | 50 | 971 | 1.030 | 0.008 |
| bden_JEL423_supercont1.19 | 69,118 | 118 | 585 | 1.707 | 0.014 |
| bden_JEL423_supercont1.20 | 45,066 | 59 | 763 | 1.309 | 0.011 |
| bden_JEL423_supercont1.21 | 28,371 | 222 | 127 | 7.825 | 0.064 |
| **Total** | **23,339,511** | **122,220** | | **5.237** | |

**Supporting Information, Table 2.** Summary of potential functional relevance of nucleotide changes between two *Bd* isolates, JEL427-P9 and JEL427-P39, differing only in laboratory passage history. We assessed potential functional relevance of genes that had at least one nonsynonymous mutation in our dataset and were up-regulated on frog skin in Rosenblum *et al*. (2012) by conducting an over-representation test of Gene Ontology (GO) terms. We used the hypergeometric test in the GOstats R package (Falcon & Gentleman 2007) with the GO annotation assignments from Rosenblum *et al*. (2013). In all, 84 genes were both up-regulated and contained ≥1 nonsynonymous change. The hypergeometric test for over-representation of GO terms did not reveal any notable patterns, although the gene set contained one M36 metallopeptidase gene, a gene family that has previously been hypothesized to play a role in *Bd* virulence (Joneson *et al*. 2011).

| GOBPID | P-value | Odds ratio | ExpCount | Count | Size | Term | Gene Ids |
|---|---|---|---|---|---|---|---|
| GO:0032313 | 0.006 | 21.595 | 0.116 | 2 | 20 | regulation of Rab GTPase activity | BDEG_00189;BDEG_00594 |
| GO:0006369 | 0.006 | Inf | 0.006 | 1 | 1 | termination of RNA polymerase II transcription | BDEG_05286 |
| GO:0043144 | 0.006 | Inf | 0.006 | 1 | 1 | snoRNA processing | BDEG_05286 |
| GO:0016180 | 0.006 | Inf | 0.006 | 1 | 1 | snRNA processing | BDEG_05286 |
| GO:0043087 | 0.009 | 16.87 | 0.145 | 2 | 25 | regulation of GTPase activity | BDEG_00189;BDEG_00594 |
| GO:0033121 | 0.009 | 16.87 | 0.145 | 2 | 25 | regulation of purine nucleotide catabolic process | BDEG_00189;BDEG_00594 |
| GO:0009118 | 0.009 | 16.87 | 0.145 | 2 | 25 | regulation of nucleoside metabolic process | BDEG_00189;BDEG_00594 |
| GO:0006140 | 0.009 | 16.161 | 0.151 | 2 | 26 | regulation of nucleotide metabolic process | BDEG_00189;BDEG_00594 |
| GO:0031329 | 0.009 | 16.161 | 0.151 | 2 | 26 | regulation of cellular catabolic process | BDEG_00189;BDEG_00594 |
| GO:0006821 | 0.012 | 182.533 | 0.012 | 1 | 2 | chloride transport | BDEG_05538 |
| GO:0048583 | 0.014 | 7.034 | 0.523 | 3 | 90 | regulation of response to stimulus | BDEG_00189;BDEG_00452; BDEG_00594 |
| GO:0010646 | 0.014 | 7.034 | 0.523 | 3 | 90 | regulation of cell communication | BDEG_00189;BDEG_00452; BDEG_00594 |
| GO:0023051 | 0.014 | 7.034 | 0.523 | 3 | 90 | regulation of signaling | BDEG_00189;BDEG_00452; BDEG_00594 |
| GO:0006303 | 0.017 | 91.233 | 0.017 | 1 | 3 | double-strand break repair via nonhomologous end joining | BDEG_06507 |
| GO:0016575 | 0.017 | 91.233 | 0.017 | 1 | 3 | histone deacetylation | BDEG_00020 |
| GO:0006012 | 0.017 | 91.233 | 0.017 | 1 | 3 | galactose metabolic process | BDEG_04530 |
| GO:0006807 | 0.018 | 3.247 | 4.565 | 9 | 786 | nitrogen compound metabolic process | BDEG_00189;BDEG_00252; BDEG_00594;BDEG_03087; BDEG_03599;BDEG_04834; BDEG_05203;BDEG_05286; BDEG_06507 |
| GO:0050790 | 0.027 | 8.957 | 0.261 | 2 | 45 | regulation of catalytic activity | BDEG_00189;BDEG_00594 |
| GO:0051174 | 0.027 | 8.957 | 0.261 | 2 | 45 | regulation of phosphorus metabolic process | BDEG_00189;BDEG_00594 |
| GO:0046578 | 0.032 | 8.182 | 0.285 | 2 | 49 | regulation of Ras protein signal transduction | BDEG_00189;BDEG_00594 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| GO:0016569 | 0.034 | 36.453 | 0.035 | 1 | 6 | covalent chromatin modification | BDEG_00020 |
| GO:0035601 | 0.04 | 30.367 | 0.041 | 1 | 7 | protein deacylation | BDEG_00020 |
| GO:0052803 | 0.046 | 26.019 | 0.046 | 1 | 8 | imidazole-containing compound metabolic process | BDEG_04834 |
| GO:0000105 | 0.046 | 26.019 | 0.046 | 1 | 8 | histidine biosynthetic process | BDEG_04834 |
| GO:0006184 | 0.053 | 6.068 | 0.377 | 2 | 65 | GTP catabolic process | BDEG_00189;BDEG_00594 |
| GO:0006139 | 0.056 | 3.452 | 1.53 | 4 | 309 | nucleobase-containing compound metabolic process | BDEG_03087;BDEG_03599; BDEG_05203;BDEG_06507 |
| GO:0009146 | 0.056 | 5.877 | 0.389 | 2 | 67 | purine nucleoside triphosphate catabolic process | BDEG_00189;BDEG_00594 |
| GO:0009261 | 0.056 | 5.877 | 0.389 | 2 | 67 | ribonucleotide catabolic process | BDEG_00189;BDEG_00594 |
| GO:0009203 | 0.056 | 5.877 | 0.389 | 2 | 67 | ribonucleoside triphosphate catabolic process | BDEG_00189;BDEG_00594 |
| GO:0046130 | 0.056 | 5.877 | 0.389 | 2 | 67 | purine ribonucleoside catabolic process | BDEG_00189;BDEG_00594 |
| GO:0072523 | 0.056 | 5.877 | 0.389 | 2 | 67 | purine-containing compound catabolic process | BDEG_00189;BDEG_00594 |
| GO:0009164 | 0.058 | 5.786 | 0.395 | 2 | 68 | nucleoside catabolic process | BDEG_00189;BDEG_00594 |
| GO:1901292 | 0.062 | 5.528 | 0.412 | 2 | 71 | nucleoside phosphate catabolic process | BDEG_00189;BDEG_00594 |
| GO:1902531 | 0.062 | 5.528 | 0.412 | 2 | 71 | regulation of intracellular signal transduction | BDEG_00189;BDEG_00594 |
| GO:1901068 | 0.062 | 5.528 | 0.412 | 2 | 71 | guanosine-containing compound metabolic process | BDEG_00189;BDEG_00594 |
| GO:0065007 | 0.063 | 2.788 | 2.259 | 5 | 389 | biological regulation | BDEG_00189;BDEG_00252; BDEG_00382;BDEG_00452; BDEG_00594 |
| GO:0006357 | 0.084 | 12.976 | 0.087 | 1 | 15 | regulation of transcription from RNA polymerase II promoter | BDEG_00252 |
| GO:0006364 | 0.089 | 12.107 | 0.093 | 1 | 16 | rRNA processing | BDEG_05286 |
| GO:0009968 | 0.1 | 10.675 | 0.105 | 1 | 18 | negative regulation of signal transduction | BDEG_00452 |
| GO:0038032 | 0.1 | 10.675 | 0.105 | 1 | 18 | termination of G-protein coupled receptor signaling pathway | BDEG_00452 |
| GO:0008277 | 0.1 | 10.675 | 0.105 | 1 | 18 | regulation of G-protein coupled receptor protein signaling pathway | BDEG_00452 |
| GO:0009205 | 0.103 | 4.065 | 0.552 | 2 | 95 | purine ribonucleoside triphosphate metabolic process | BDEG_00189;BDEG_00594 |
| GO:0009141 | 0.105 | 4.02 | 0.558 | 2 | 96 | nucleoside triphosphate metabolic process | BDEG_00189;BDEG_00594 |
| GO:0006325 | 0.116 | 9.063 | 0.122 | 1 | 21 | chromatin organization | BDEG_00020 |
| GO:0042278 | 0.12 | 3.693 | 0.604 | 2 | 104 | purine nucleoside metabolic process | BDEG_00189;BDEG_00594 |
| GO:0005996 | 0.121 | 8.629 | 0.128 | 1 | 22 | monosaccharide metabolic process | BDEG_04530 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| GO:0009150 | 0.128 | 3.549 | 0.627 | 2 | 108 | purine ribonucleotide metabolic process | BDEG_00189;BDEG_00594 |
| GO:0009119 | 0.13 | 3.514 | 0.633 | 2 | 109 | ribonucleoside metabolic process | BDEG_00189;BDEG_00594 |
| GO:0006820 | 0.131 | 7.872 | 0.139 | 1 | 24 | anion transport | BDEG_05538 |
| GO:0019693 | 0.132 | 3.48 | 0.639 | 2 | 110 | ribose phosphate metabolic process | BDEG_00189;BDEG_00594 |
| GO:0048523 | 0.141 | 7.237 | 0.151 | 1 | 26 | negative regulation of cellular process | BDEG_00452 |
| GO:0008033 | 0.146 | 6.956 | 0.157 | 1 | 27 | tRNA processing | BDEG_05286 |
| GO:1901657 | 0.148 | 3.23 | 0.685 | 2 | 118 | glycosyl compound metabolic process | BDEG_00189;BDEG_00594 |
| GO:0022613 | 0.176 | 5.64 | 0.192 | 1 | 33 | ribonucleoprotein complex biogenesis | BDEG_05286 |
| GO:0007166 | 0.19 | 5.15 | 0.209 | 1 | 36 | cell surface receptor signaling pathway | BDEG_00452 |
| GO:0071704 | 0.194 | 2.11 | 7.972 | 10 | 1481 | organic substance metabolic process | BDEG_00020;BDEG_00252; BDEG_03087;BDEG_03599; BDEG_04530;BDEG_04834; BDEG_05033;BDEG_05203; BDEG_05286;BDEG_06507 |
| GO:0071840 | 0.205 | 2.593 | 0.842 | 2 | 145 | cellular component organization or biogenesis | BDEG_00020;BDEG_05286 |
| GO:0006351 | 0.235 | 2.349 | 0.923 | 2 | 159 | transcription, DNA-templated | BDEG_00252;BDEG_05286 |
| GO:1902589 | 0.237 | 3.991 | 0.267 | 1 | 46 | single-organism organelle organization | BDEG_00020 |
| GO:0006281 | 0.241 | 3.903 | 0.273 | 1 | 47 | DNA repair | BDEG_06507 |
| GO:0033554 | 0.25 | 3.738 | 0.285 | 1 | 49 | cellular response to stress | BDEG_06507 |
| GO:0055086 | 0.261 | 2.172 | 0.993 | 2 | 171 | nucleobase-containing small molecule metabolic process | BDEG_00189;BDEG_00594 |
| GO:1901575 | 0.286 | 2.031 | 1.057 | 2 | 182 | organic substance catabolic process | BDEG_00189;BDEG_00594 |
| GO:0051716 | 0.286 | 2.037 | 1.059 | 2 | 191 | cellular response to stimulus | BDEG_00382;BDEG_00452 |
| GO:0023052 | 0.306 | 1.927 | 1.109 | 2 | 191 | signaling | BDEG_00382;BDEG_00452 |
| GO:0008652 | 0.314 | 2.832 | 0.372 | 1 | 64 | cellular amino acid biosynthetic process | BDEG_04834 |
| GO:0044237 | 0.322 | 1.449 | 8.566 | 10 | 1475 | cellular metabolic process | BDEG_00020;BDEG_00189; BDEG_00252;BDEG_00594; BDEG_03087;BDEG_03599; BDEG_04834;BDEG_05203; BDEG_05286;BDEG_06507 |
| GO:1901605 | 0.326 | 2.7 | 0.389 | 1 | 67 | alpha-amino acid metabolic process | BDEG_04834 |
| GO:0034660 | 0.334 | 2.619 | 0.401 | 1 | 69 | ncRNA metabolic process | BDEG_05286 |
| GO:0035556 | 0.347 | 2.497 | 0.419 | 1 | 71 | intracellular signal transduction | BDEG_00382 |
| GO:0044710 | 0.377 | 1.502 | 2.183 | 3 | 451 | single-organism metabolic process | BDEG_00020;BDEG_04530; BDEG_04834 |
| GO:0050794 | 0.385 | 1.619 | 1.316 | 2 | 269 | regulation of cellular process | BDEG_00252;BDEG_00382 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| GO:0034654 | 0.391 | 1.573 | 1.336 | 2 | 230 | nucleobase-containing compound biosynthetic process | BDEG_00252;BDEG_05286 |
| GO:0016053 | 0.399 | 2.082 | 0.499 | 1 | 86 | organic acid biosynthetic process | BDEG_04834 |
| GO:0006259 | 0.429 | 1.891 | 0.548 | 1 | 99 | DNA metabolic process | BDEG_03087 |
| GO:0044763 | 0.434 | 1.311 | 3.316 | 4 | 689 | single-organism cellular process | BDEG_00382;BDEG_00452; BDEG_04834;BDEG_05538 |
| GO:0019219 | 0.439 | 1.839 | 0.564 | 1 | 110 | regulation of nucleobase-containing compound metabolic process | BDEG_00252 |
| GO:2001141 | 0.474 | 1.64 | 0.627 | 1 | 108 | regulation of RNA biosynthetic process | BDEG_00252 |
| GO:0006396 | 0.477 | 1.624 | 0.633 | 1 | 109 | RNA processing | BDEG_05286 |
| GO:0006629 | 0.486 | 1.578 | 0.65 | 1 | 112 | lipid metabolic process | BDEG_00382 |
| GO:0009889 | 0.492 | 1.549 | 0.662 | 1 | 114 | regulation of biosynthetic process | BDEG_00252 |
| GO:2000112 | 0.492 | 1.549 | 0.662 | 1 | 114 | regulation of cellular macromolecule biosynthetic process | BDEG_00252 |
| GO:0008152 | 0.498 | 1.942 | 0.616 | 1 | 157 | metabolic process | BDEG_08286 |