AN ABSTRACT FOR THE DISSERTATION OF

Jacob A. Tennessen for the degree of Doctor of Philosophy in Zoology presented on February 27, 2009.

Title: Adaptive Diversity and Divergence at Frog Antimicrobial Peptide Loci.

Abstract approved:

_____

Michael S. Blouin

In this dissertation I study evolutionary patterns at genes encoding antimicrobial peptides (AMPs) in frogs. AMPs are short, amphipathic, cationic, secreted proteins that kill bacteria and other pathogens through a non-catalytic mechanism that involves binding to and disrupting the microbial cell membrane. In many animal taxa, positive selection is much more common at AMP genes than at most other types of genes, making them an ideal model for the study of non-neutral molecular evolution, as well as potentially important sites of adaptive differences in disease resistance between individuals, populations, or species. There is growing interest in the use of AMPs or their analogs in therapeutic applications, since they are functionally quite different from conventional antibiotics. However, evolutionary studies on AMPs or other immune effector molecules have been relatively rare compared to similar studies on other immunity genes. My research, which consists of both bioinformatic analyses of preexisting data and the generation of the first intraspecific polymorphism data at AMP loci in non-human vertebrates (leopard frogs; *Ranidae*: *Rana*: *Pantherana*), has uncovered several main results. First, I confirm that positive selection is common at frog AMP loci, indicative of coevolution with pathogens. Second, I demonstrate that a previoulsy proposed hypothesis of coordinated evolution between the mature AMP and the propiece is not supported by the data. Third, I show that substitution at synonymous sites is enhanced in the portion of the gene encoding the mature AMP; both the absence of an enhanced transversion/transition ratio and the low intraspecies polymorphism at silent sites argue against an enhanced mutation rate as the

explanation, and therefore selection on “silent” sites is probably responsible. Fourth, I demonstrate that AMP gene duplication has occurred many times in the evolution of leopard frog genomes, possibly because it has been favored by selection. Finally, I reveal a striking non-neutral pattern of either very low or very high intraspecific polymorphism at AMP loci in leopard frogs; the former is due to positive selective sweeps, and the latter occurs because several highly divergent alleles are maintained by balancing selection, which probably takes the form of fluctuating selection.

Adaptive Diversity and Divergence at Frog Antimicrobial Peptide Loci

by
Jacob A. Tennessen

A DISSERTATION
submitted to
Oregon State University

in partial fulfilment of
the requirements for the
degree of

Doctor of Philosophy

Presented February 27, 2009
Commencement June 2009

Doctor of Philosophy dissertation of Jacob A. Tennessen presented on February 27, 2009.

APPROVED:

_____

Major Professor, representing Zoology

_____

Chair of the Department of Zoology

_____

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

_____

Jacob A. Tennessen, Author

ACKNOWLEDGEMENTS

I have had an overwhelmingly positive graduate school experience, for which I owe deep gratitude to many people and institutions.

Mike Blouin has been an ideal advisor for me. He has given me great freedom to pursue my own interests, manage my own time, and choose my own research style. At the same time, his door has always been open, and he is always willing to drop what he is doing and chat with me. The Blouin lab is well managed and easy to work in, and the diversity of projects has allowed me to learn a great deal about various population genetic methods in many taxa. Most of all, Mike is a good human being, who treats others with respect, who if full of wisdom on how to live a full and satisfying life, and who I am pleased to have as both a mentor and a friend.

I am also thankful for the rest of my committee. Steve Arnold has taught me a lot about both evolutionary herpetology and scientific communication, has always generously let me uses his molecular lab, and acted as an unofficial in-house advisor while Mike was on sabbatical. Joey Spatafora has been a phylogenetics resource for me since I sought his advice on my undergraduate lizard phylogeography project right after I first arrived in Corvallis. Jerri Bartholomew was an indispensible guide as I investigated a possible project on salmonid parasite-resistance QTLs; even after I abandoned that research direction, she has been a useful expert on disease resistance. Dan Rockey has been a fair advocate as my grad rep.

In addition, I have relied on guidance from scientists outside of OSU. Kelly Zamudio, my undergraduate mentor from Cornell University, has continued to advise and collaborate with me up to this day. Louise Rollins-Smith, whose Vanderbilt University lab I visited in 2007, introduced me to the analysis of the actual peptides encoded by the genes I had been studying for years. Jianzhi "George" Zhang hosted me for a year at the University of Michigan and provided both stimulating discussions and ample lab space and resources for my research.

This dissertation would not have been possible without federal financial support. My first three years of graduate school were supported by a Graduate Research Fellowship from the National Science Foundation. The next three years were

supported by a Science To Achieve Results (STAR) Fellowship from the Environmental Protection Agency. I thank the American taxpayers for this invaluable aid.

Field and lab assistance came from many directions. Several people or institutions either provided tissue samples or helped me to collect them, including Eric Hoffman, Eric Tobin, Mike Redmer, Stephan Swanson, Mike Foy, Kim Fields, and the Museum of Vertebrate Zoology. Becky Cooper and Vindhya Amarasinghe kept the lab organized and well-stocked, and helped to troubleshoot many technical difficulties. The Nevada Genomics Center consistently provided a high-quality sequencing service though which all of my data were generated.

The Zoology Department has been a wonderful home. The administrative staff in the Zoology office are always helpful and kind. Several faculty members, beyond my committee, have been especially generous to me with their time and/or lab space, including Chris Bayne, Dee Denver, and Dave Lytle. The zoology graduate students and post-docs have been a circle of friends like I have never had before in my life. They have given me emotional support, intellectual stimulation, and countless fun and unforgettable experiences: costume parties, ski weekends, soccer games, jam sessions, camping trips, and many more. I would especially like to thank Hitoshi Araki, Kaitlin Bonner, Angela Brandt, Mark Christie, Charles Criscione, Leslie Dyall, Dafne Eerkes-Medrano, Deb Finn, Elise Granek, Margot Hessing-Lewis, Darren Johnson, Karen Kiemnec-Tyburczy, Laura McMullen, Rodrigo Miranda, Catherine Palmer, Santiago Perez, Laura Petes, Ivan Phillipsen, Erin Scheessele, Catherine Searle, Joe Tyburczy, Richard Watts, Mike Westphal, Román Vilas, and Asako Yamamuro.

Finally, I am eternally indebted to my family. My fiancée Monica Jacobson has been infinitely supportive and tolerant whenever my academic responsibilities have kept me away from her. She has been a beacon of inspiration and has constantly refreshed my spirit. My parents, Chuck and Karen Tennessen, raised me to be curious, creative, and self-confident, and to love nature and knowledge; these core values have brought me where I am today. My siblings, grandparents, and extended family have encouraged me to seek my dreams and are enthusiastic about my research, whether or not they entirely understand it.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

LIST OF APPENDIX TABLES

ADAPTIVE DIVERSITY AND DIVERGENCE AT FROG ANTIMICROBIAL PEPTIDE LOCI

CHAPTER 1
INTRODUCTION

The majority of genetic differences between individuals and species are likley to be neutral with respect to natural selection and therefore have equivalent effects on fitness (Ford 2002). Only certain molecular substitutions have non-neutral phenotypic effects that contribute to adaptation. Identifying and characterizing the distributions of these non-neutral variants is essential for the study of natural selection at the molecular level. Determining the genetic basis of adaptive variation is also relevant to conservation biology. In addition, genes with unique functional effects favored by selection may have useful biotechnological applications.

My dissertation focuses on the evolution of antimicrobial peptides (AMPs) in frogs. AMPs are short, amphipathic, cationic, secreted proteins that kill bacteria and other pathogens through a non-catalytic mechanism that involves binding to and disrupting the microbial cell membrane (Hancock 2001; Bowman 2003; Yeaman and Yount 2003). AMPs are excellent models for the study of adaptive molecular evolution because they evolve under positive selection more often than most other types of genes, because non-neutral patterns of evolution are concentrated in a relatively short and easily localized sequence region, and because their functions can feasibly be assessed *in vitro*. Although some immune system genes have been well characterized as important sites of adaptive genetic variation, evolutionary genetic research on immune effector molecules like AMPs has been limited. My research addresses questions like: How is adaptive variation in disease resistance maintained within species? What genes are responsible for adaptive divergence between species? What are the relative roles of natural selection, genetic drift, and mutation in the evolution of immune effectors?

Chapter 2 (Tennessen 2005a) is a review of the literature in which I show that positive selection on AMPs is common in many animal taxa from insects to amphibians to mammals. Evidence for positive selection on antimicrobial peptides includes an excess of nonsynonymous nucleotide substitutions, an excess of charge-

changing amino acid substitutions, non-neutral patterns of allelic variation, and functional assays *in vivo* and *in vitro* that show improved antimicrobial effects for derived sequence variants. The synthesis of these studies suggests that AMP loci are a good target for studies of natural selection at the molecular level.

Chapter 3 (Tennessen 2005b) consists of evolutionary bioinformatic analyses of preexisting data on AMPs. Using AMP gene sequences from frogs and fishes, I test whether AMPs evolve in a coordinated fashion with their propiece and whether substitution at synonymous nucleotide sites is enhanced in the mature peptide region. These hypotheses were suggested by previous researchers. Only the latter hypothesis was supported, suggesting either selection on synonymous nucleotide sites or an enhanced mutation rate in AMP genes, although the originally proposed mechanism of an error-prone polymerase that increases the transversion/transition ratio is not supported. I also confirm that positive selection is common on AMPs in both taxa.

Chapters 4 and 5 both concern AMP evolution in leopard frogs of the genus *Rana* (*Lithobates*), subgenus *Pantherana*. This work represents the first study of intraspecific variation at AMP loci in a non-human vertebrate. In order to examine genomic and population genetic patterns at AMP loci in these frogs, and in particular to determine what kinds of natural selection, if any, are acting on them, I sequence several AMP loci in multiple individuals of several closely related species. In Chapter 4 (Tennessen and Blouin 2007), I estimate the number of AMP sequences encoded in leopard frog genomes, demonstrate that positive selection has increased divergence among genomic paralogs and between orthologs in different species, and explore the effects of selection and mutation on synonymous nucleotide sites. In Chapter 5 (Tennessen and Blouin 2008), I sequence one particular AMP locus in over 500 frogs from multiple populations of three species, in order to assess population genetic and geographic patterns at this locus. I demonstrate that selection maintains several highly divergent alleles at this locus, and that the specific evolutionary mechanism is probably a form a fluctuating selection as opposed to stable balancing selection.

My work contributes in important ways to a scientific understanding of pathogen-imposed natural selection and its genomic effects. It confirms previous

hypotheses about common features of AMP evolution, including positive selection, extensive gene duplication, and enhanced substitution at synonymous sites. In addition, it implies that some patterns of non-neutral evolution, especially fluctuating selection, may be more common on immune effector molecules like AMPs than was previously known. It suggests that variation at AMP loci could be important for the ability of frogs to resist the emerging infectious diseases that are decimating many amphibian populations. It also provides strategies for finding functionally divergent AMPs with different potential uses in therapeutic applications.

MOLECULAR EVOLUTION OF ANIMAL ANTIMICROBIAL PEPTIDES: WIDESPREAD MODERATE POSITIVE SELECTION

Jacob A. Tennessen

Journal of Evolutionary Biology
Blackwell Publishing
2005. 18(6):1387–1394

CHAPTER 2

MOLECULAR EVOLUTION OF ANIMAL ANTIMICROBIAL PEPTIDES: WIDESPREAD MODERATE POSITIVE SELECTION

2.1 Abstract

An increasing number of studies in both vertebrates and invertebrates show that the evolution of antimicrobial peptides is driven by positive selection. Because these diverse molecules show potential for therapeutic applications, they are currently the targets of much structural and functional research, providing extensive background data for evolutionary studies. In this paper, patterns of molecular evolution in antimicrobial peptide genes are reviewed. Evidence for positive selection on antimicrobial peptides includes an excess of nonsynonymous nucleotide substitutions, an excess of charge-changing amino acid substitutions, nonneutral patterns of allelic variation, and functional assays in vivo and in vitro that show improved antimicrobial effects for derived sequence variants. Positive selection on antimicrobial peptides may be as common as, but perhaps weaker than, selection on the best-known example of adaptively evolving immunity genes, the major histocompatibility complex. Thus, antimicrobial peptides present a useful and underutilized model for the study of adaptive molecular evolution.

2.2 Introduction

Determining which genes underlie adaptation, how these genes change in response to natural selection, and what selective pressures drive their evolution are primary goals in the unification of evolutionary biology and genetics. Although many genes under positive natural selection have been identified in particular taxa (Ford 2002), there are few types of genes for which neutrality is commonly rejected in taxonomically widespread groups. Furthermore, there is often little biochemical knowledge about genes potentially under selection, so the causes of their molecular evolutionary patterns remain largely unknown. Systems of genes regularly under

positive selection in diverse species, for which the functional properties of their protein products are well understood, could allow fundamental questions about the genetic basis for Darwinian processes to be addressed. The identification and characterization of antimicrobial peptides (AMPs) from various taxa is a rapidly expanding field of biology, largely due to the potential for therapeutic applications (Bowman 2003; Yeaman and Yount 2003). These peptides are exceptionally diverse in sequence, structure, and function, even among closely related species. Because their diversity is so extensive, and because immune system proteins that interact directly with molecules of pathogens often evolve adaptively (Ford 2002; Garrigan and Hedrick 2003), it is likely that much of this variation results from positive selection on the ability to combat new or altered pathogens (Hedengren et al. 2000). Furthermore, because medical and pharmaceutical research is resulting in an increasingly comprehensive understanding of the biochemical properties of these peptides, AMPs may be an ideal system in which to study adaptive molecular evolution.

The AMPs are generally short (usually 15–45 amino acid residues), cationic and amphipathic (Bowman 2003; Yeaman and Yount 2003). The mature peptide is often cleaved off of a larger protein containing a signal sequence and a propiece (Bowman 2003). The occurrence, length and relative position of these three gene regions vary among AMP families (Fig. 2.1). As it is only the mature peptide that interacts with microbes, selective pressures are likely to differ between this region of the gene and the rest of the coding sequence. The structure of the mature peptide varies among gene families. Some AMPs form linear α-helices, while others form β-sheets, and some are composed of an unusually high proportion of particular amino acids (Andreu and Rivas 1998; Bowman 2003). Cysteine residues in AMPs often form disulfide bonds important for molecular structure (Fig. 2.1), thus cysteine codons are expected to be more conserved than other sequence regions. Other modifications such as amidation also occur in some peptides (Andreu and Rivas 1998).

The mature peptides are active against a wide range of microbes, including bacteria, fungi and viruses. Being positively charged, they bind to anionic microbial lipid membranes and disrupt them through a noncatalytic mechanism. The AMPs may

form membrane pores in which the hydrophobic surfaces of the peptides face the hydrophobic membrane core, while the charged surfaces face inward, lining the pore. According to other models, insertion into the hydrophobic membrane core does not necessarily occur, and AMPs simply carpet the membrane until it cannot function (Yeaman and Yount 2003). The AMPs may also inhibit intracellular functions including DNA or protein synthesis (Andreu and Rivas 1998; Yeaman and Yount 2003).

In comparison to the extensive literature describing activity spectra, mechanisms of action, and structural features of AMPs, evolutionary analyses have been rarer. These few studies, though, create a strong case for widespread nonneutral evolution in taxa as diverse as mammals, birds, amphibians, fish and insects (Tables 2.1 and 2.2). Most evolutionary change and variation in genes is neutral, with selection serving primarily as a conservative force, so genes at which positive selection is driving evolution are unusual and show unique signatures. In this paper, studies of the molecular evolution of AMPs are reviewed, with a focus on evidence for positive selection acting on these molecules.

2.3 Evolutionary Relationships

It is unknown whether all antimicrobial peptides are homologous. Neither sequence nor structure is conserved among gene families, and a reasonable hypothesis is that small peptides with antimicrobial abilities have evolved multiple times. Nevertheless, similar evolutionary pressures may be acting on unrelated families of AMPs, and they may show similar signatures of these forces.

Evolutionary relationships are not always clear within gene families, either. Any analysis of molecular evolution at AMP loci is complicated by the fact that these genes are frequently duplicated. For example, the mouse genome contains 49 β-defensin loci, in addition to other AMPs (Schutte et al. 2002). In humans, a cluster of three β-defensin genes varies in copy number from 2 to 12 repeats per diploid genome (Hollox et al. 2003). The frequent duplication of AMP genes may itself be adaptive,

allowing new copies to develop novel antimicrobial properties without constraint to retain old functions. In general, it can be assumed that a gene's most similar homolog in a different species is an ortholog, although the true ortholog may have been deleted or gone undetected. Conversely, homologous sequences from the same species may be reported to be paralogs (e.g. Duda et al. 2002), when they may in fact be different alleles of the same locus. An added complication is that all regions of a gene may not share the same evolutionary history due to exon shuffling (Froy and Gurevitz 2003). Confidence of orthology is especially important in tests of selection based on intraspecies variation, where, for example, unknowingly including sequences from multiple loci can lead to erroneous conclusions. Tests of selection based on substitution rate ratios do not depend on assumptions about orthology. However, the absolute divergences at both synonymous and nonsynonymous sites are dependent on the time since the sequences last shared a common ancestor, which can only be estimated with knowledge of orthology. In addition, identification of orthologs and paralogs allows one to address other issues, such as whether selection is strongest on recently duplicated genes.

2.4 Nonsynonymous and Synonymous Substitution Rates

If the ratio of nonsynonymous nucleotide differences (dn) to synonymous nucleotide differences (ds) between sequences is significantly greater than one, positive selection can be inferred (Hill and Hastie 1987; Hughes and Nei 1988). The dn/ds ratio is often greater than one for some AMP families in both vertebrates and invertebrates (Table 2.1). Most comparisons have been between paralogs, but a few apparent β-defensin orthologs show significantly high dn/ds between species of mouse (Morrison et al. 2003), and between species of primate (Boniotto et al. 2003b).

Such examples are noteworthy, but it is important to realize that for many pair-wise comparisons in these AMP families, dn/ds is not significantly different from, or is even less than, one. In addition, for several other families of AMPs, dn/ds is consistently less than one (e.g. β-defensin 1, Del Pero et al. 2002; β-defensin 3,

Boniotto et al. 2003a). Among fruit fly AMPs, dn/ds greater than one has only been observed in a small minority of comparisons from tephritid flies (Rosetto et al. 2003); in *Drosophila*, this ratio approaches, but has not been observed to exceed, one. A deficiency of nonsynonymous substitutions does not rule out selection, though, because the dn/ds test is very conservative. In almost any coding sequence, purifying selection is acting on many of the codons, which can lower the value of dn even in the face of positive selection on other codons. Thus, by using maximum likelihood methods that test dn and ds at individual codons (Yang and Bielawski 2000), significant positive selection on portions of AMP sequences has been detected in many gene families (Table 2.1). Cysteines are highly conserved and are not under positive selection, but other sites near them in the mature peptide region are frequently positively selected.

Several authors have noticed that dn/ds ratios in AMPs are highest when divergence is low (Hughes and Yeager 1997; Hughes 1999; Morrison et al. 2003; Nicolas et al. 2003; Semple et al. 2003). This pattern, in which selection appears the strongest between closely related genes, could be due to intense selection immediately after gene duplication that eventually diminishes after the duplicated gene has adapted to its new function. It is also possible that nonsynonymous sites at which positive selection tends to act are more saturated at greater genetic distances, limiting the observed dn in highly divergent sequences.

As noted above, the translated protein usually consists of a signal sequence, a propiece, and the mature peptide that is cleaved off. In vertebrates, the mature peptide evolves faster than the rest of the protein (Hughes and Yeager 1997; Duda et al. 2002; Maxwell et al. 2003; Morrison et al. 2003; Semple et al. 2003; Xiao et al. 2004). Although such rapid evolution may be due to a relaxation of purifying selection relative to the rest of the sequence, it implies that, if positive selection is indeed occurring, it is acting on the mature peptide. Some insect AMPs, such as termite termicins (Bulmer and Crozier 2004) and andropin in *Drosophila* (Date-Ito et al. 2002) follow the vertebrate pattern of greater amino acid divergence in the mature peptide than in the signal sequence. Other *Drosophila* AMPs, however, show the

opposite pattern: the mature peptide itself is highly conserved, and positive selection, if it is acting, is focused on the other parts of the protein or on regulatory sequences (Lazzaro and Clark 2001, 2003).

Interestingly, even silent sites at several AMP loci show unusually high divergence in the mature peptide region with respect to other regions of the same gene. For example, synonymous divergence among rabbit defensins is significantly higher in the propiece and mature peptide then in the signal sequence (Hughes and Yeager 1997). A similar trend has been recorded in amphibian AMPs (Nicolas et al. 2003; Vanhoye et al. 2003; Tennessen 2005b). Furthermore, ds in the flatfish pleurocidin mature peptide region is significantly higher than both ds in the signal sequence and divergence in the introns of the gene (Tennessen 2005b). The difference between synonymous divergence in the signal sequence and mature peptide is especially unexpected in the pleurocidins, because part of the mature peptide is on the same exon as the signal sequence, immediately adjacent to it. Thus, the mechanism causing enhanced synonymous divergence is extremely specific to the mature peptide region. All gene families exhibiting enhanced synonymous divergence in the mature peptide region also show evidence of positive selection on that part of the gene. This widespread trend may be caused by an increased mutation rate, suggesting that vertebrates promote variation in a gene region where it is adaptive to do so. Alternatively, selection on 'silent' sites, perhaps for translational accuracy or mRNA structure, may be the explanation.

In similar but not identical pattern, the *D. melanogaster* cecropin and attacin gene regions show significantly high levels of silent nucleotide variation relative to other loci in that species (Date et al. 1998; Lazzaro and Clark 2001). The authors hypothesized that intragenic recombination and/or introgression from a related species could have caused the excess of synonymous variation, but the hypotheses described above for the pattern seen in vertebrate AMPs could also be responsible. There are no reports of a difference between silent sites in the mature peptide and the rest of the coding region in *Drosophila*.

2.5 Patterns of Amino Acid Change

Nonsynonymous substitutions that replace an amino acid residue with a residue of a different charge can be defined as radical, and an excess of radical nonsynonymous differences between sequences can be taken as evidence for positive selection (Zhang 2000; Pupko et al. 2003). In the mature peptide, nonsynonymous changes that are radical with respect to charge are sometimes significantly more frequent than conservative nonsynonymous changes (Table 2.1). Some AMPs evolve conservatively with respect to charge, such as those from ranid frogs (Duda et al. 2002), but in general change in charge is surprisingly common. The termicins from termites in particular are noteworthy in that selection appears to have acted in several independent lineages to reduce net charge (Bulmer and Crozier 2004). Although radical nonsynonymous changes can also be defined as changes in polarity or volume, AMP evolution is rarely radical by these definitions. As positive charge is thought to be important in antimicrobial activity (Andreu and Rivas 1998), it is notable that charge is not a conserved feature. Rather, subtle differences in net charge or the location of charged residues apparently convey higher fitness on a frequent basis.

It has been proposed that the anionic propiece of the processed domain interacts with the cationic mature peptide and prevents it from damaging the host cells (Michaelson et al. 1992). In mammalian α-defensins, the propiece and the mature peptide evolve in a coordinated manner such that charge changes in the propiece tend to be accompanied by opposite charge changes in the mature peptide, resulting in little change in net charge (Hughes and Yeager 1997). A similar hypothesis was proposed for hylid frog AMPs (Duda et al. 2002) but an alternate analysis suggests that coordinated evolution does not occur in this group (Tennessen 2005b). No evidence of coordinated evolution was seen in ranid frog AMPs mammalian β-defensins, or flatfish pleurocidins (Duda et al. 2002; Morrison et al. 2003; Tennessen 2005b). Thus, coordinated evolution may be unique to mammalian α-defensins, and not a general feature of AMPs.

2.6 Allelic variation

Intraspecies variation at AMP loci has been assessed in *Drosophila* (Clark and Wang 1997; Date et al. 1998; Ramos-Onsins and Aguadé 1998; Lazzaro and Clark 2001, 2003; Date-Ito et al. 2002) and humans (Dörk and Stuhrmann 1998; Vatta et al. 2000; Jurevic et al. 2002) (Table 2.2). This variation often has phenotypic effects, which could be subject to selection. For example, polymorphisms in both coding and noncoding regions of the human β-defensin 1 gene have been correlated with susceptibility to diseases such as oral fungal infection, chronic obstructive pulmonary disease, and HIV-1 infection (Matsushita et al. 2002; Jurevic et al. 2003; Braida et al. 2004). Likewise, in *Drosophila*, variation at some AMP loci shows an association with phenotypic variability in pathogen resistance, and variation at several other loci affects disease resistance through epistatic interaction with intracellular signalling loci, probably due to variation in transcriptional regulation (Lazzaro et al. 2004) (Table 2.2). In some of these cases, the allele conveying higher fitness is the ancestral sequence, which is consistent with purifying, not positive, selection. In other cases, the opposite is true. For example, a derived variant of the β-defensin 1 5′-UTR in humans is significantly associated with lower oral fungal infection, which arguably could convey higher fitness and thus be under positive selection (Jurevic et al. 2003).

True tests of neutrality using intraspecies variation have only been reported in *Drosophila*. In some regards this variation behaves neutrally. For example, a significant excess of nonsynonymous variation has not been observed at any locus. McDonald and Kreitman (1991) tests on cecropin, andropin and diptericin loci have failed to reject neutrality (Clark and Wang 1997; Date-Ito et al. 2002). The cecropin family in particular shows no evidence of positive selection despite extensive study (Clark and Wang 1997; Date et al. 1998; Ramos-Onsins and Aguadé 1998; Lazzaro and Clark 2003).

Other tests show nonneutral patterns in *Drosophila* allelic variation (Table 2.2). For example, a significant excess of high-frequency derived variants, as demonstrated by Fay and Wu (2000) H-tests, has been noted at several loci (Lazzaro

and Clark 2001, 2003). A propiece variant of drosocin appears to "have substantially increased in frequency in only 2 years" (Lazzaro and Clark 2003). In an application of the HKA test (Hudson et al. 1987), andropin exhibits a significantly high ratio of divergence to polymorphism when compared to cecropin B, which conforms to neutral patterns (Clark and Wang 1997). Finally, Fu and Li (1993) tests gave a significantly negative result for two loci (Clark and Wang 1997), which is probably due to purifying selection but could potentially indicate recovery from a positive selective sweep.

2.7 Antimicrobial Activity and Selection

Because most studies of AMPs focus on their capacity to kill pathogens, there is a great deal of evidence that AMPs vary considerably in their antimicrobial activities (Tossi et al. 2000; Yeaman and Yount 2003). Diversity in antimicrobial activity among closely related AMPs is consistent with the hypothesis that their evolutionary radiation has not been merely a neutral drift to variants of equivalent fitness. Derived variants sometimes show improved antimicrobial activities, which further confirms the hypothesis of positive selection. For example, in the northern leopard frog *Rana pipiens*, the derived AMP brevinin-1Pc is nearly three times more efficient at killing *E. coli* than the more basal AMPs brevinin-1Pa and brevinin-1Pb (Goraya et al. 2000; Conlon et al. 2004). In addition, variation in antimicrobial activities within adaptively evolving gene families can provide insight into the selective pressures acting on these AMPs. For example, in a primate phylogeny, a significant excess of nonsynonymous change in the β-defensin 2 gene was detected along the branch leading to humans from the common ancestor of humans and macaques (Boniotto et al. 2003b). Counterintuitively, macaque β-defensin 2 appears more lethal towards some human pathogens than the positively selected human β-defensin 2 (Antcheva et al. 2004). This result suggests either that these microbes were not the ones driving selection, or that the microbes themselves have evolved a greater tolerance towards human AMPs.

2.8 Causes of Positive Selection

Although individual tests of selection on AMPs are sometimes merely suggestive, taken together these studies present convincing evidence that positive selection on AMPs is common and taxonomically widespread. Positive selection is not constant in all lineages at all times, as some evolutionary change in AMPs appears to be neutral. In addition, even where positive selection does occur, it may not be very strong. Nevertheless, there are few examples of gene types for which neutrality is rejected as frequently as for AMP genes, suggesting that these molecules are highly important to adaptive evolution. Clearly, the innate immune system is not merely a vestigial relic. Rather, new AMP variants are frequently able to provide higher fitness to both vertebrates and invertebrates.

Why does selection appear to act differently on different AMP loci, given that they all have similar functions? Is it simply an artifact of the statistical tests used? If not, are there fundamental differences in the roles of these molecules? Or, do most or all of them alternate between periods of neutral evolution and periods of positive adaptation, with researchers arbitrarily happening to observe them in one period or the other? The specific pathogens driving selection undoubtedly vary among hosts, which could result in different patterns of evolution. Some hosts might be in a constant co-evolutionary arms race with pathogens that are under selection to resist their defenses. A variety of resistance mechanisms to AMPs are known in microbes, some of which involve a single gene product (Andreu and Rivas 1998; Yeaman and Yount 2003). In some cases, derived microbial strains are more resistant than the wild type (Fernandez and Weiss 1996; Thevissen et al. 2004), which is consistent with the hypothesis that positive selection on microbial genomes can result in increased resistance to AMPs. Thus, it may be the case that resistance is easy to evolve and happens frequently. On the other hand, some AMPs might attack their targets in such a way that evolving resistance is not possible without coordinated changes at many microbial genes.

Selection on these AMPs would primarily occur when hosts enter new niches and are forced to adapt to completely different pathogen species not previously encountered.

2.9 Comparisons with MHC

Genes of the immune system, especially those that interact directly with molecules of pathogens, have been shown to evolve nonneutrally more often than many other types of genes (Ford 2002; Garrigan and Hedrick 2003). The best-studied example is the major histocompatibility complex (MHC), which encodes glycoproteins that bind foreign peptides for pathogen recognition. It has been shown to deviate from neutral expectations at population, species, and/or interspecies levels in multiple vertebrate taxa and is one of the clearest examples of a locus complex that frequently evolves adaptively (Hughes and Yeager 1998; Garrigan and Hedrick 2003). Having a similar but not identical function, AMPs might be expected a priori to evolve nonneutrally, albeit perhaps in different ways than genes such as MHC loci, and it is worth comparing signatures of selection on AMPs with those on other parts of the immune system.

Positive selection in both MHC and AMP loci is common, but not universal. As in AMPs, comparisons of MHC sequences among related species sometimes, but not always, show a dn/ds ratio significantly higher than one (Bernatchez and Landry 2003; Garrigan and Hedrick 2003). The AMPs are more taxonomically widespread than MHC, which only occurs in vertebrates. In the species that share both types of genes, an obvious question is whether nonneutral evolution is more frequent in MHC or in AMPs. Although MHC has received more attention, positive selection on AMPs may be equally common. If so, their occurrence in both vertebrates and invertebrates, as well as in plants, would make AMPs a more general model system for the study of adaptive molecular evolution.

Positive selection for new variants in MHC is often accompanied by the maintenance of multiple allelic lineages via extremely strong balancing selection. Evidence for balancing selection in MHC includes uniform allele frequencies, an

excess of nonsynonymous variation, and transspecies polymorphisms in which sequences are more similar to orthologous sequences in other species than they are to alleles at the same locus in the same species (Bernatchez and Landry 2003; Garrigan and Hedrick 2003). Like AMPs, MHC sometimes appears to have an elevated synonymous substitution rate relative to adjacent sequences (Hughes 2000). In this case, excess synonymous variation is attributed to selection: ancient MHC sequences have been maintained by balancing selection longer than neutral drift would allow, which has let many synonymous substitutions accumulate (Hughes 2000).

In contrast with MHC, there is little direct evidence for balancing selection on AMPs. Intraspecies variation, in the few AMP loci where it has been assessed, is not as excessive as in MHC (Clark and Wang 1997). Neither unusually uniform allele frequencies nor an excess of nonsynonymous variation over synonymous variation has been reported in these studies, either. However, such studies are rare, and include some of the AMPs that show the least evidence for selection at the interspecies level, such as β-defensin 1 in humans and cecropin in *Drosophila*. While transspecies polymorphisms of AMPs have not been shown to occur, either, sequences from the same species sometimes do not cluster together in evolutionary trees (e.g. Conlon et al. 2004). Although such a phylogenetic pattern is likely due to gene duplication, a clear understanding of orthology and paralogy is usually lacking. Balancing selection could enhance synonymous site diversity among alleles at a locus, and the subsequent separation of these divergent alleles during a speciation or gene duplication event would lead to the observed enhanced synonymous site divergence of the descendant sequences. However, for balancing selection to cause the pattern of enhanced synonymous divergence in the mature peptide region relative to the signal sequence, these two gene regions would have to be unlinked, which may not always be the case, especially for the pleurocidins. Thus, the possibility remains that balancing selection acts on at least some AMP loci, but lacking further evidence the null hypothesis of no balancing selection must be retained.

Whether selection on AMPs is balancing or simply directional, it may be generally weaker than the selection acting on MHC. It is difficult to evaluate this

hypothesis directly. Tests of MHC evolution do not always reject neutrality, and those that do reject it estimate selection coefficients ranging over several orders of magnitude (Satta et al. 1994; Garrigan and Hedrick 2001; Aguilar et al. 2004). The MHC evolution has been much more extensively studied than AMP evolution, making direct comparisons of selection estimates difficult. However, as AMPs appear to lack some of the extraordinary patterns exhibited by MHC, such as its extreme polymorphism, an appropriate null hypothesis is that AMPs evolve more neutrally than MHC. In addition, the fact that AMPs are not thought to recognize specific molecules the way that MHC does suggests that selection on AMPs should not be as strong as on MHC. Within certain motifs, all AMP variants might have equal antimicrobial ability, such that selection is weaker than if every difference mattered.

2.10 Future Directions

The AMPs provide an opportunity to study adaptation at the level of the molecules where it actually happens. Specific selective pressures driving evolution in AMPs could be investigated by identifying how positively selected changes affect antimicrobial activity. In addition, research could aim to elucidate both sides of a coevolutionary arms race between host and pathogen by identifying the changes in microbial genes that cause resistance to AMPs. Further comparisons between AMPs and MHC would also be interesting, particularly, of the strength and frequency of selection and whether selection at AMP genes is sufficient to overcome the effects of drift in small populations, as has been shown for MHC (Aguilar et al. 2004; Jarvi et al. 2004). More research at the population genetic level may be able to address many of the unresolved issues in AMP evolution, including specific selective pressures, strength of selection on AMPs vs. MHC, and the cause of enhanced synonymous divergence. Those searching for therapeutic uses of AMPs should seek to identify AMP families in which selection has likely produced high variation in antimicrobial properties, as well as AMP families that do not appear to be under constant selection to change, perhaps because pathogens cannot easily evolve resistance to them. On the

other hand, some proposed uses for AMPs, such as phylogenetic analysis (Conlon et al. 2004), may not be feasible if nonneutral evolution is common. Continued investigation into AMP evolution promises to be a highly fruitful and fascinating area of research.

Table 2.1

AMP families that show evidence of positive selection in some sequence comparisons (not necessarily over all available sequences).

AMP Family	**Taxon**	**Evidence of selection over all codons**	**Site heterogeneity**
α-defensin[a,b,q,r]	primates, rodents, rabbits	dn/ds >1*, radical/conservative change >1*	dn/ds>1* at certain codons
β-defensin[b,c,e,f,g,h,l,o]	primates, rodents, bovids	dn/ds >1*, radical/conservative change >1*	dn/ds>1* at certain codons
gallinacin[p,s]	chickens	dn/ds >1	dn/ds>1* at certain codons
frog AMP[d,i,j,m,t]	ranid and hylid frogs	dn/ds >1*, radical/conservative change >1*	not tested
pleurocidin[t]	flatfishes	dn/ds >1, radical/conservative change >1	dn/ds>1* at certain codons
termicin[n]	termites	dn/ds >1, radical/conservative change >1	dn/ds>1* at certain codons
ceratotoxin[k]	fruit flies	dn/ds >1 (very rarely)	not tested

*statistically significant

[a]Hughes and Yeager 1997, [b]Hughes 1999, [c]Del Pero et al. 2002, [d]Duda et al. 2002, [e]Boniotto et al. 2003a, [f]Boniotto et al. 2003b, [g]Maxwell et al. 2003, [h]Morrison et al. 2003, [i]Nicolas et al. 2003, [j]Pupko et al. 2003, [k]Rosetto et al. 2003, [l]Semple et al. 2003, [m]Vanhoye et al. 2003, [n]Bulmer and Crozier 2004, [o]Luenser and Ludwig 2004, [p]Lynn et al. 2004a, [q]Lynn et al. 2004b, [r]Patil et al. 2004, [s]Xiao et al. 2004, [t]Tennessen 2005b

Table 2.2

AMP families assessed for allelic variation in nucleotide sequences.

AMP Family	**Taxon**	**Evidence of non-neutrality**
β-defensin[c,f,h,i,j,l]	humans	disease resistance correlation
andropin[a,g]	fruit flies	HKA test, Fu and Li test
drosocin[k]	fruit flies	Fay and Wu's H, rapid allele frequency change
diptericin[a,k,m]	fruit flies	Fay and Wu's H, Fu and Li test, disease resistance correlation
attacin[e,k,m]	fruit flies	Fay and Wu's H, disease resistance correlation
metchnikowin[k,m]	fruit flies	Fay and Wu's H, epistatic disease resistance correlation
defensin[k,m]	fruit flies	epistatic disease resistance correlation
cecropin[a,b,d,k]	fruit flies	none

[a]Clark and Wang 1997, [b]Date et al. 1998, [c]Dörk and Stuhrmann 1998, [d]Ramos-Onsins and Aguadé 1998, [e]Lazzaro and Clark 2001, [f]Vatta et al. 2000, [g]Date-Ito et al. 2002, [h]Jurevic et al. 2002, [i]Matsushita et al. 2002, [j]Jurevic et al. 2003, [k]Lazzaro and Clark 2003, [l]Braida et al. 2004, [m]Lazzaro et al. 2004

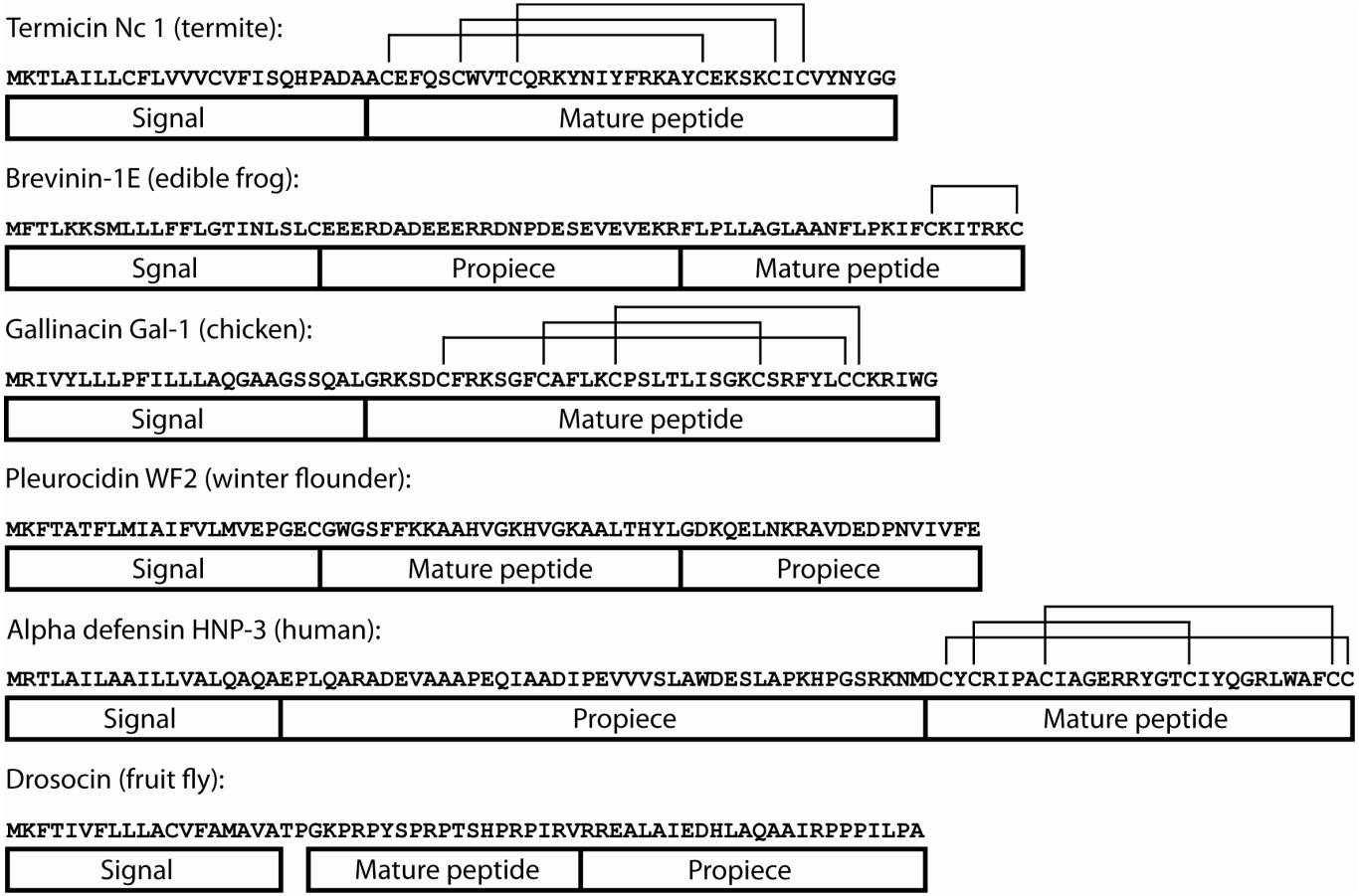

Figure 2.1

Examples of antimicrobial peptide genes from various taxa, highlighting structural organization. Disulfide bonds in the mature peptide are shown by connections between cysteines.

ENHANCED SYNONYMOUS SITE DIVERGENCE IN POSITIVELY SELECTED VERTEBRATE ANTIMICROBIAL PEPTIDE GENES

Jacob A. Tennessen

Journal of Molecular Evolution
Springer Publishing
2005. 61(4):445-455

CHAPTER 3

ENHANCED SYNONYMOUS SITE DIVERGENCE IN POSITIVELY SELECTED VERTEBRATE ANTIMICROBIAL PEPTIDE GENES

3.1 Abstract

Nonrandom patterns associated with adaptively evolving genes can shed light on how selection and mutation produce rapid changes in sequences. I examine such patterns in two independent families of antimicrobial peptide genes: those in frogs, which are known to have evolved under positive selection, and those in flatfishes, which I show have also evolved under positive selection. I address two recently proposed hypotheses about the molecular evolution of antimicrobial peptide genes. The first is that the mature peptide region is replicated by an error-prone polymerase that increases the mutation rate and the transversion/transition ratio compared to the signal sequence of the same genes. The second is that mature peptides evolve in a coordinated fashion with their propieces, such that a change in net charge in one molecular region prompts an opposite change in charge in the other region. I test these hypotheses using alternative methods that minimize alignment errors, correct for phylogenetic nonindependence, reduce sequence saturation, and account for differing selection pressures on different regions of the gene. In both gene families I show that divergence at both synonymous and nonsynonymous sites within the mature peptide region is enhanced. However, in neither gene family is there evidence of an increased mutational transversion/transition ratio or coordinated evolution. My observations are consistent with either an elevated mutation rate in an adaptively evolving gene region or widespread selection on ‘‘silent’’ sites. These hypotheses challenge the assumption that mutations are random and can be measured by the synonymous substitution rate.

3.2 Introduction

A thorough understanding of rapid adaptive molecular evolution requires knowledge of both its causes, including the relative importance of mutation and

selection, and its consequences, including the signatures it leaves in sequence data. Selection is often inferred by comparing patterns of amino acid divergence against the presumed neutral background of synonymous site divergence. Associations between synonymous substitution rates and adaptive evolution would obviously affect such estimates of selection, yet intragenic variation in synonymous site divergence is rarely considered in this context. Furthermore, hypotheses concerning the causes of rapid protein evolution can sometimes be evaluated with sequence data if more than just the ratio of nonsynonymous and synonymous substitutions is considered. These potential causes include mechanisms of enhanced mutation and selection for multiple coordinated changes.

Antimicrobial peptides have been identified and characterized in numerous taxa. These short, generally cationic peptides are often cleaved off of a larger protein containing a signal sequence and a propiece (Bowman 2003). The mature peptides are active against a wide range of microbes through a noncatalytic disruption of the microbial cell membrane and possibly of other cellular functions (Andreu and Rivas 1999; Yeaman and Yount 2003). Adaptive evolution is common in immune system genes, especially those that interact directly with molecules of pathogens, and thus antimicrobial peptides frequently exhibit signatures of positive selection. For example, in mature peptide sequences from frogs, the rate of nonsynonymous nucleotide substitutions per nonsynonymous site (dn) is sometimes significantly higher than the rate of synonymous nucleotide substitutions per synonymous site (ds), and nonsynonymous changes that are radical with respect to charge (pr) are sometimes significantly more frequent than conservative nonsynonymous changes (pc) (Duda et al. 2002; Pupko et al. 2003). In contrast, the signal sequence is highly conserved and never shows a dn/ds or pr/pc ratio higher than one.

While positive selection on antimicrobial peptides is intuitive and well documented, there may be other evolutionary patterns associated with these adaptively evolving molecules that are less well understood. In particular, two previously proposed hypotheses about the molecular evolution of antimicrobial peptides may explain observations in sequences from frogs: the elevated mutation rate hypothesis

and the coordinated evolution hypothesis (Duda et al. 2002; Vanhoye et al. 2003; Nicolas et al. 2003). In this paper, I examine these hypotheses in frog antimicrobial peptides using statistical methods that are different from those used in previously published studies. In addition, I explore whether these hypotheses apply to other similar genes by testing them in the pleurocidins, a family of flatfish antimicrobial peptides (Cole et al. 1997). The pleurocidins may not be homologous to frog antimicrobial peptides, although some similarities between these two gene families have been noted (Douglas et al. 2001). Regardless, equivalent evolutionary pressures may be acting on both of them, and they may show similar signatures of these forces. To confirm that pleurocidin evolution is appropriate to compare with frog antimicrobial peptide evolution, I test whether positive selection has also driven pleurocidin evolution, using dn/ds and pr/pc ratios.

3.2.1 *The Elevated Mutation Rate Hypothesis*

In frog antimicrobial peptides, even ds appears enhanced in the mature peptide region relative to the signal sequence (Vanhoye et al. 2003; Nicolas et al. 2003). In addition, the ratio of transversions to transitions (tv/ts) is much higher in the mature peptide region than in the signal sequence. In order to account for both of these observations, it has been hypothesized that an error-prone DNA polymerase similar to polymerase V in *Escherichia coli* targets the mature peptide region (Vanhoye et al. 2003; Nicolas et al. 2003). This hypothesis would also explain a similar trend of enhanced ds in antimicrobial peptides from rabbits, originally attributed to interlocus recombination events (Hughes and Yeager 1997).

This intriguing possibility demands further confirmation. Unfortunately, the many indels in frog mature peptides mean that it is difficult to align homologous codons with confidence (Fig. 3.1). Failure to do so could increase the estimate of ds and could also explain the observation of an apparent enhanced tv/ts ratio. In addition, this hypothesis is apparently based on pairwise calculation of ds and tv/ts among all sequences, without phylogenetic corrections. Such pairwise comparisons are not independent, as the ds of each branch is included multiple times (Felsenstein 1985). I

examine synonymous divergence and the tv/ts ratio in antimicrobial peptides from frogs and flatfishes while minimizing such potential errors.

3.2.2. *The Coordinated Evolution Hypothesis*

Many antimicrobial peptides have an overall positive charge, and this is thought to be important to their function. The propiece of the processed domain often carries a negative charge, and it has been proposed that this interacts with the cationic peptide and prevents it from damaging the host cells (Michaelson et al. 1992). It has been suggested that the propiece and the mature peptide undergo coordinated evolution such that charge changes within the propiece tend to be accompanied by opposite charge changes within the mature peptide, resulting in little change in net charge. Evidence for such coordinated evolution has been observed in some mammalian antimicrobial peptides (Hughes and Yeager 1997) but not in others (Morrison et al. 2003). However, a suggestion that this hypothesis applies to hylid frogs (Duda et al. 2002) was not based on phylogenetically independent contrasts (Felsenstein 1985). Peptides of similar charge are likely to be closely related to each other, and thus it is expected that their propieces will have similar charges as well. Coordinated evolution can only be conclusively detected if this effect of common ancestry is removed. While accounting for relatedness, I test for coordinated evolution in antimicrobial peptides from hylid frogs and flatfishes.

3.3 Methods

3.3.1. *Relative Synonymous Substitution Rates in Frog Antimicrobial Peptide Genes*

To determine whether the conclusion of an enhanced mutation rate in frog antimicrobial peptides was simply owing to artifact, I selected 20 frog antimicrobial peptides to analyze more closely (from *Agalychnis annae*, *Agalychnis callidryas*, *Litoria caerulea*, *Pachymedusa dacnicolor*, *Phyllomedusa bicolor*, *Phyllomedusa sauvagei*, *Rana pipiens*, *Rana esculenta*, *Rana rugosa*, and *Rana temporaria*; accession numbers AJ005183, AJ005189, AJ005191, AJ005192, AJ251566,

AJ251567, AJ251875, AJ312003, AJ427746, AJ564793, AJ564794, AY218776, AY218778, AY218780, AY218783, AY218785, U22393, X72387, X77831, and X77832 [Amiche et al. 1994, 2000; Simmaco et al. 1994; Park et al. 1995; Vouille et al. 1997; Simmaco et al. 1998; Wechselberger 1998; Chen et al. 2003a, b; Vanhoye et al. 2003]). Most, but not all, of these had been used in the published study (Vanhoye et al. 2003; Nicolas et al. 2003). I chose these sequences because they allowed for a phylogenetically independent calculation of ds between closely related, homologous codons.

These 20 peptides fell into seven sets of two to three sequences (Fig. 3.2). In each set, alignment of the mature peptides was unambiguous, either because the mature peptides were the same length or because they differed by a single amino acid and the location of the indel was obvious. By calculating ds only within these sets, and not between them, I could maximize confidence that all aligned codons were homologous. To ensure that such comparisons are phylogenetically independent, I first inferred the relationships among the signal sequences using a maximum likelihood analysis in PAUP* (version 4.0b10 [Swofford 2002]). I only used the signal sequences in phylogenetic analysis to minimize error due to nonneutral evolution or faulty alignment (all signal sequences are 66 base pairs [bp], while propieces and mature peptides vary in length). As the sequences in each set group together, comparisons within sets do not result in any branch being included more than once. I calculated dn and ds for signal sequence and mature peptide at 19 branches among the seven sets of genes, using PAML (version 3.13a [Yang 1997]). These represented all the branches within the seven sets, and none between sets. I compared the estimates of ds using a two-sided paired t-test. Because synonymous divergence has been correlated with GC content and the number of CpG dinucleotides at the third codon position (Tsunoyama et al. 2001; Lercher et al. 2004), I calculated GC content and third codon position CpG dinucleotides for the mature peptide region and signal sequence of the 20 frog genes.

3.3.2. *Tv/Ts Ratios in Frog Antimicrobial Peptide Genes*

Using this same set of frog sequences, I also examined whether the tv/ts ratio was higher in the mature peptide region than in the signal sequence, as has been reported. I calculated the uncorrected tv/ts ratio in the mature peptide region and the signal sequence, both overall and only for synonymous changes. To test whether there is a significant difference in ratios once the effects of amino acid level conservation have been removed, I compared the ratios at synonymous sites using a chi-square test.

3.3.3 *The Coordinated Evolution Hypothesis in Hylid Frog Antimicrobial Peptides*

I looked for a correlation between the net charges of propieces and mature peptides in hylid frogs using the program PHYLIP 3.61 (Felsenstein 2004) to implement the contrasts calculation of Felsenstein (1985). The tree presented in Vanhoye et al. (2003) includes all 18 hylid sequences examined in Duda et al. (2002), as well as 11 other hylid sequences. Using this more extensive tree of 29 hylid sequences, I calculated independent contrasts for the net charge of the mature peptide and the net charge of the propiece. I looked for coordinated evolution both by using the entire peptide and by excluding the final three amino acids, as practiced in Duda et al. (2002).

3.3.4 *Identical Tests in Pleurocidins*

To determine whether aspects of antimicrobial peptide molecular evolution observed in frogs are the same as in other vertebrates, I examined nucleotide sequences of the antimicrobial peptide pleurocidin from four species of flatfish, *Hippoglossoides platessoides*, *Limanda ferruginea*, *Pleuronectes americanus*, and *Hippoglossus hippoglossus* (Douglas et al. 2001, 2003; Patrzykat et al. 2003). I selected 11 genomic pleurocidin sequences to examine (accession numbers AF301512, AF301513, AF301515, AY273172, AY273173, AY273174, AY273175, AY273180, AY273181, and AY282498). All 11 sequences are formed from three exons, which vary in length. I grouped them into two sets in which all members have exons of similar length, and I minimized unintended comparisons between nonhomologous codons by only aligning sequences within each set (Fig. 3.3). Using

only the signal sequences, which are alignable across all 11 genes, I generated a maximum-likelihood tree in PAUP* to confirm that the two sets form distinct clades. Then, using the entire coding sequences, I generated maximum-likelihood phylogenies for each set separately.

Using the pleurocidin sequences, I performed the same tests I performed on the frog antimicrobial peptide sequences. Namely, I tested for a difference in ds and tv/ts between the mature peptide region and the signal sequence, I calculated GC content and frequency of CpG dinucleotides in these two gene regions, and I tested for coordinated evolution between mature peptide and propiece, using the methods described above. I also calculated divergences in the introns and compared this to synonymous divergences in the mature peptide region and in the signal sequence using two-sided paired t-tests. I used a Bonferroni correction to test the significance of the three pleurocidin ds comparisons (mature peptide versus signal sequence, mature peptide versus introns, and signal sequence versus introns).

Using the program DnaSP 4.0 (Rozas et al. 2003), I generated sliding window graphs of pairwise divergence at both synonymous and nonsynonymous sites along the coding region for both sets of pleurocidins (window length = 6, step size = 1). While nonphylogenetic pairwise comparisons are inappropriate for statistical analysis, they can provide a rough graphical representation of intragenic variation in divergence.

3.3.5 *Confirmation of Positive Selection on Pleurocidins*

In order to formulate hypotheses about the causes of evolutionary patterns shared by frog antimicrobial peptides and pleurocidins, it was necessary to confirm that the pleurocidins have also evolved under positive selection. I tested for positive selection on the pleurocidins, using PAML to calculate pairwise dn/ds among sequences, as well as overall dn/ds in both sets for the signal sequence and for the mature peptide region. I tested whether a model which assumes the observed estimate of dn/ds is significantly different from a model where dn/ds is 1. I tested for varying selection pressures among codons using the entire coding region in each set, including the signal sequence, propiece, and mature peptide. I ran PAML models M0 (one ratio),

M1 (neutral), M2 (selection), M3 (discrete), M7 (β), and M8 (β plus ω), which make varying assumptions about selection pressures among codons (Yang et al. 2000), and compared them. As recommended, I ran models M2 and M8 both with initial ω value <1 (0.5) and >1 (3) and selected the version with the highest likelihood.

Finally, I tested whether amino acid changes in mature pleurocidin peptides tend to be radical or conservative, another measure of positive selection. I reconstructed ancestral sequences using PAUP* and used the program Hon-new (Zhang 2000) to calculate the proportion of radical (pr) and conservative (pc) nonsynonymous differences along all branches of both phylogenies. I used three definitions of radical substitutions, as implemented by Honnew: radical with respect to charge, radical with respect to polarity, and radical with respect to polarity and volume. I compared these proportions using paired t-tests.

3.4 Results

3.4.1 *Substitution Rates in Frog Antimicrobial Peptide Genes*

The signal sequence phylogeny of the 20 frog antimicrobial peptides I analyzed is shown in Fig. 3.4. Sequences in each of the seven sets cluster together, so comparisons within sets do not include any branch more than once. Sequences that also occur in the previously published phylogeny (Vanhoye et al. 2003; Nicolas et al. 2003) group into the same sets in that tree, as well. In addition, overall similarities among mature peptides within sets, especially similar lengths, support the signal sequence-based phylogeny. Because sequences in each set are closely related, it is unlikely that sites have become saturated.

Synonymous substitution in the frog mature peptide region (16.7%) is significantly higher than in the signal sequence (4.6%) ($p < 0.05$, two-sided paired t-test) (Table 3.1). In the mature peptide region, overall mean dn/ds = 0.7, and in the signal sequence, overall mean dn/ds = 0.3 (Table 3.1). The difference in dn/ds between the mature peptide region and the signal sequence is not quite significant ($p = 0.08$, chi-square test), but this difference is obviously affected by the enhanced ds in

the mature peptide region; dn is an order of magnitude higher in the mature peptide region (11.4%) than in the signal sequence (1.2%) ($p < 0.01$, two-sided paired t-test). Selection on the mature peptide does not appear to be constant among the seven sets, because a model which assigns a unique dn/ds ratio to each set is significantly better than a model which assumes that dn/ds is the same across sets (likelihood ratio test, $p < 0.05$). According to this nonconstant selection model, three sets have dn/ds >1 in the mature peptide region: set 2, set 3, and set 4. In contrast, a model of nonconstant dn/ds in the signal sequence is not significantly favored over a model of constant dn/ds (likelihood ratio test, $p > 0.05$).

Overall GC content in frog antimicrobial peptide genes is very similar for the signal sequence and the mature peptide region, but the difference at third codon positions considered alone is twofold (Table 3.1). CpG dinucleotides with the C in the third codon position occur with the same frequency as those with the G in the third codon position, and the per-site proportion in the mature peptide region is higher than the per-site proportion for the signal sequence (Table 3.1).

3.4.2 *Tv/Ts Ratios in Frog Antimicrobial Peptide Genes*

Among these frog sequences, the uncorrected tv/ts ratio is significantly lower in the signal sequence (0.4) than in the mature peptide region (1.3) ($p < 0.05$, chi-square test), confirming the previously published observation (Table 3.1). However, because the genetic code is conservative, transversions are more likely to result in nonsynonymous or radical change than transitions, so the tv/ts substitution ratio should be lower in more highly conserved regions even if the relative frequencies of these two types of mutation are constant across the gene (Zhang 2000). The signal sequence is much more highly conserved than the mature peptide region (Table 3.1). The difference in the ratio of transversions to transitions at synonymous sites, which is lower than when nonsynonymous changes are included, is not significant($p > 0.1$, chi-square test) (Table 3.1). Thus, the overall difference in tv/ts between the two regions can be explained more parsimoniously by selection than by mode of mutation.

3.4.3 *The Coordinated Evolution Hypothesis in Hylid Frog Antimicrobial Peptides*

Using a phylogenetically independent method to test for coordinated evolution between frog mature peptide and propiece, there is a slight trend of an inverse relationship, but it is not significant. This result holds whether the entire mature peptide is used ($R^2 = 0.04$, $p > 0.1$) or the final three amino acids are excluded ($R^2 = 0.09$, $p > 0.1$).

3.4.4. *Substitution Rates in Pleurocidin Genes*

The signal sequences of all 11 pleurocidin genes are alignable, and a phylogeny of just the signal sequences clusters the two sets into separate clades. Thus, comparisons within the sets do not include any branches more than once. For all other analyses, I generated separate phylogenies of both sets using the entire coding sequence under maximum likelihood in PAUP* (Fig. 3.5).

For many pairwise comparisons of the mature pleurocidin region, dn/ds >1. To test whether evidence of positive selection is significant, I used maximum likelihood analysis to assign a dn/ds ratio to the mature peptide region alone using PAML (Yang 1997). Under a model of constant selection pressure, set 1 has dn/ds = 0.9, and set 2 has dn/ds = 1.5. For the signal sequence alone, the overall estimate of dn/ds is 0.8 for set 1 and 0.6 for set 2. None of these four are significantly better than a model where dn/ds is set at 1 ($p > 0.1$).

However, such an appearance of overall neutrality could also occur if some codons are under purifying selection, while others are under positive selection. Nonsynonymous divergence exceeds synonymous divergence through much of the mature peptide region, even if not through all of it (Fig. 3.6). In every appropriate comparison, the PAML model that assumes positive selection on some codons (model M2, M3, or M8) is significantly better than the model that does not (model M0, M1, or M7) for both sets (Table 3.2). The proportion of sites under positive selection varies among the models, from 13% to 45% in set 1 and from 16% to 72% in set 2 (Table 3.2). Such variance in the proportion of positively selected sites may reflect some false positives, but consistent strong rejection of null models in both independent sequence

sets suggests that positive selection on at least some codons is real. In both sets, all three positive selection models assumed a class of sites for which dn/ds >4. Model M3 also assumed an additional class of sites for which 2 > dn/ds > 1 in both sets of sequences (Table 3.2). Bayesian prediction of the specific sites under positive selection is not very accurate when a small number of closely related sequences is used (Anisimova et al. 2002), so these are not reported. Notably, though, most of the predicted positively selected sites are in the mature peptide region. The likelihood ratio tests that reject neutral models in favor of positive selection models are conservative when a small data set is used (Anisimova et al. 2001), and this is the important result in the present study.

The pr/pc ratio of pleurocidins is >1 with respect to charge (pr/pc = 1.08) but not with respect to polarity (pr/pc = 0.89) or polarity and volume (pr/pc = 0.82). However, under all three definitions of radical change, the difference between pr and pc is not significant (paired t-tests, $p > 0.1$ for all). Thus, one cannot conclude that radical changes are more frequent than would be expected under neutrality, although they may be more frequent than for most genes, which usually have pr/pc <1 (Zhang 2000).

As with the frog peptides, ds is significantly higher at mature pleurocidin (15.9%) than at the signal sequence (3.9%) ($p < 0.05$, Bonferroni-corrected, two-sided paired t-test) (Table 3.1, Fig. 3.6). Divergence in the introns (4.6%; Table 3.1) is significantly lower than ds in the mature peptide ($p < 0.05$, Bonferroni-corrected, two-sided paired t-test) but not significantly different from ds in the signal sequence ($p > 0.1$, two-sided paired t-test with or without Bonferroni correction). In the mature peptide region, the highest mean ds occurs within the first exon, is sequentially lower in the second exon and the third exon, and increases again slightly in the propiece (Table 3.1).

Duplicating the pattern seen in frog sequences, overall GC content in pleurocidins is similar in the signal sequence and in the mature peptide region, but at the third codon positions of these regions GC content is higher in the signal sequence (Table 3.1). GC content in the introns, however, is lower than all of these estimates

(Table 3.1). The per-site proportions of CpG dinucleotides with the C in the third position and with the G in the third position are listed in Table 3.1 for the signal sequence and mature peptide region, as is the overall frequency of CpG dinucleotides in the introns.

3.4.5 *Tv/Ts Ratios in Pleurocidin Genes*

I examined the tv/ts ratio between signal sequence and mature pleurocidin peptides. Between both phylogenies, the uncorrected tv/ts ratio is lower in the signal sequence (0.8) than in the mature peptide region (1.5), although this difference is not significant ($p > 0.1$, chi-square test) (Table 3.1). As in frogs, this trend can be explained by the fact that the pleurocidin signal sequence is more highly conserved than the mature peptide region (Table 3.1). Also, the difference in dn/ds between these two gene regions is not as high as in the frog antimicrobial peptides, where the difference is more than twofold. Thus, it is not surprising that the difference in the tv/ts ratios between the pleurocidin gene regions is not significant, as it is in frog antimicrobial peptides. If only synonymous changes are counted in the pleurocidin genes, there is still no significant difference between tv/ts ratios ($p > 0.1$, chi-square test) (Table 3.1).

3.4.6 *The Coordinated Evolution Hypothesis in Pleurocidins*

To test for coordinated evolution between the propiece and the mature peptide in the pleurocidin family, I calculated independent contrasts of the charges as described above. Part of the propiece sequence was missing for four peptides, so I only analyzed the remaining seven sequences. No evidence of coordinated evolution in these genes was detected ($p > 0.1$), although this small sample size may have low statistical power.

3.5 Discussion

3.5.1 *Positive Selection*

My results confirm previous conclusions that, in frogs, some evolutionary changes in the mature peptide are driven by positive selection, while the signal sequence is not under positive selection (Duda et al. 2002). Positive selection also acts on at least some codons of flatfish pleurocidins. Presumably these genes have adapted to attack new pathogens or new strains of old pathogens that are coevolving with their host. It is noteworthy that dn/ds remains high in the mature peptide region in spite of the relatively high ds. Parallel signatures of positive selection suggest that molecular evolution is similar in both gene families, and thus both can be utilized to examine the coordinated evolution hypothesis and the elevated mutation rate hypothesis.

3.5.2 *The Coordinated Evolution Hypothesis*

There is no evidence that the mature peptide and the propiece evolve in a coordinated manner in either frogs or flatfishes. It is still likely that amino acid replacements occur in such a way that the overall net charge remains within a particular range. Indeed, this is probably true of almost all proteins. However, it does not appear that a charge change in the propiece is directly compensated with an opposite charge change in the mature peptide. The rapid evolution of these antimicrobial peptides is not driven by a selective pressure to conform to changes elsewhere in the gene.

3.5.3. *The Elevated Mutation Rate Hypothesis*

Two lines of evidence led to the original proposal of an elevated mutation rate in the mature peptide region: an enhanced tv/ts ratio, and enhanced synonymous site divergence (Vanhoye et al. 2003; Nicolas et al. 2003). My analysis of tv/ts ratios does not support this hypothesis. Although a significantly higher proportion of transversions is observed in the mature peptide region of frogs, and a similar nonsignificant trend occurs in pleurocidin sequences, variation in selective pressures alone explains this phenomenon. There is no evidence that transversional mutations occur more frequently in the mature peptide region, because there is no significant

difference in the tv/ts ratios at synonymous sites. Therefore, there is no evidence that these genes are replicated with a polymerase resembling polymerase V.

The elevated mutation rate hypothesis is, however, supported by the observation of greater synonymous site divergence in the mature peptide region than in the rest of the gene. This pattern is observed in frogs, flatfishes, and rabbits (Hughes and Yeager 1997; Vanhoye et al. 2003; Nicolas et al. 2003) and, thus, appears to be widespread among vertebrate antimicrobial peptides. Vertebrate ds is known to vary by more than 10-fold within genomes (Wolfe et al. 1989; Bernardi et al. 1993; Wolfe and Sharp 1993). Such variation in synonymous divergence would not necessarily imply significant differences between immediately adjacent sequences, but intragenic variation in ds has also been observed previously in vertebrates (Alvarez-Valin et al. 1998; Tsunoyama et al. 2001). Notably, several reports of enhanced synonymous divergence are in adaptively evolving genes (Pamilo and O'Neill 1997; Metz et al. 1998; Ogawa et al. 1999; Conticello et al. 2001). Thus, the trend in antimicrobial peptides might be part of a larger genomic phenomenon.

There is no known mutational mechanism that targets specific gene regions in animals, but accelerated mutation has been proposed to explain enhanced synonymous substitutions in other genes, especially those for rapidly evolving host-defense proteins (Ogawa et al. 1999; Conticello et al. 2001; Tani et al. 2002). In pleurocidins, the highest synonymous divergence occurs on the same exon as the signal sequence, immediately adjacent to it. A mutational mechanism would need to be remarkably specific to cause this pattern.

Methylated CpG dinucleotides are highly mutable in vertebrates, and thus sequences with high frequencies of these sites have higher substitution rates (Tsunoyama et al. 2001). These mutable dinucleotides could partially, but not entirely, explain the observations in antimicrobial peptides. CpG dinucleotide frequencies are generally higher in the mature peptide region third positions than in the signal sequence third positions or in the introns. However, after removing all codons in which a CpG overlaps the third position in at least one sequence, silent divergence is still significantly higher in the mature peptide region than in the signal sequence or

introns (two-sided paired t-tests, $p < 0.05$). Therefore, CpG dinucleotides are not the entire explanation.

Overall GC content has been correlated with ds (Lercher et al. 2004). However, there is no clear pattern relating base composition to ds in antimicrobial peptides. Although GC content is higher in signal sequence third codon positions than in mature peptide third codon positions, it is low in the pleurocidin introns, which also show low ds. It has been suggested that intermediate GC content causes enhanced synonymous divergence (Wolfe et al.1989). While this is consistent with the pleurocidin observations, in frogs GC content at third positions is much closer to 50% for signal sequences than for mature peptide sequences.

Antimicrobial peptide genes regularly show enhanced ds in the same part of the gene, the part under positive selection. Thus, the intragenic variation in divergence is not random. This observation is consistent with the hypothesis that accelerated mutation is adaptive, and serves to increase novelty and/or diversity in a gene region where new variants can provide higher fitness.

3.5.4 *Alternate Hypotheses*

An enhanced mutation rate in the mature peptide region is not the only explanation for my observations. Signal and mature peptide sequences from the same gene may not have always shared the same evolutionary history due to interlocus recombination or exon-shuffling in the formation of new genes, which is known to occur in some antimicrobial peptide families (Froy and Gurevitz 2003). However, it is unclear how these essentially random processes could consistently cause one gene region to appear more divergent than another, and different evolutionary histories seem especially unlikely when part of the mature peptide and the signal sequence share the same exon, as in the pleurocidins.

A second possible explanation is that synonymous sites and noncoding regions are not truly silent, and selection has acted on some of these sites. For example, stronger selection against translation errors in the signal sequence, leading to stronger codon bias, could explain the correlation between synonymous divergence and

nonsynonymous divergence (Alvarez- Valin et al. 1998). However, purifying selection on translational accuracy would not explain the low ds in the pleurocidin introns. Alternatively, conserved synonymous sites could be important in RNA folding and processing. Low ds at the start of a coding region, thought to be caused by selection on RNA structure, is common in vertebrate genes (Smith and Hurst 1999). Selection at the start of the coding region would explain the low ds in the signal sequences, which all consist of the first 22 codons of the sequence, but it is not as obvious why the introns would show lower divergence than the mature peptide region. Finally, positive selection on synonymous sites (DuMont et al. 2004) may have acted in the mature peptide region, but it is unclear what selective pressures could cause this. If selection on ‘‘silent’’ sites is common enough to cause significant intragenic variation in synonymous divergence, the synonymous substitution rate may not reflect the mutation rate as accurately as previously thought.

3.6 *Conclusions*

I have shown that the mutational tv/ts ratio is not enhanced in the mature peptide region, and the mature peptide does not undergo coordinated evolution with the propiece. Enhanced synonymous divergence in the mature peptide region, however, occurs in both frog antimicrobial peptides and flatfish pleurocidins, both of which are evolving under positive selection. There is no single preferred explanation for the observation of enhanced synonymous site divergence in the mature peptide region, but the two best hypotheses are an elevated mutation rate and selection on ‘‘silent’’ sites. These patterns of molecular evolution suggest that antimicrobial peptides present a useful system with which to study how mutation and selection cause rapid adaptive change in genes.

Table 3.1

Molecular evolutionary features of gene regions examined in this study.

	ds[a]	**dn/ds**[b]	**tv/ts**[c]	**s tv/ts**[d]	**GC**[e]	**GC3**[f]	**C3G1**[g]	**C2G3**[h]
Frog antimicrobial peptides								
Signal sequence	4.6%	0.3	0.4	0.3	41%	52%	0.7%	0.7%
Mature peptide region	16.7%	0.7	1.3	0.6	43%	26%	2.2%	2.2%
Flatfish pleurocidins								
Signal sequence	3.9%	0.7	0.8	0.4	50%	73%	5.4%	0.0%
Mature peptide region: entire	15.9%	1.2	1.5	0.7	46%	47%	5.3%	1.8%
Mature peptide region: exon 1	41.2%	0.7	n.c.[i]	n.c.[i]	n.c.[i]	n.c.[i]	n.c.[i]	n.c.[i]
Mature peptide region: exon 2	9.3%	2.5	n.c.[i]	n.c.[i]	n.c.[i]	n.c.[i]	n.c.[i]	n.c.[i]
Mature peptide region: exon 3	4.7%	2.5	n.c.[i]	n.c.[i]	n.c.[i]	n.c.[i]	n.c.[i]	n.c.[i]
Propiece	10.1%	0.7	n.c.[i]	n.c.[i]	n.c.[i]	n.c.[i]	n.c.[i]	n.c.[i]
Introns	4.6%	n/a	1.4	n/a	29%	n/a	1.0%	n/a

[a]mean divergence per branch at synonymous nucleotide sites, or overall for introns

[b]ratio of the mean divergence at nonsynonymous nucleotide sites per branch over ds

[c]uncorrected ratio of transversions over transitions

[d]tv/ts for synonymous changes only

[e]percentage of G and C nucleotides

[f]GC at third codon positions

Table 3.1 (Continued)

[g]per-site percentage of CpG dinucleotides with C in the third codon position, or overall for introns

[h]per-site percentage of CpG dinucleotides with G in the third codon position

[i]n.c. = not calculated

Table 3.2

Likelihood ratio tests to detect positive selection on pleurocidins.

Models	**2Δlog-likelihood**	**χ^2 value**	**df**	**p-value**	**Proportion of sites with dn/ds >1**	**dn/ds[a]**
Set 1						
M0 vs. M2	2|-725.4 - (-702.8)|	45.2	2	<0.001	0.13	5.5
M0 vs. M3	2|-725.4 - (-702.3)|	46.2	4	<0.001	0.45	1.9 or 8.7
M1 vs. M2	2|-710.8 - (-702.8)|	16	2	<0.001	0.13	5.5
M1 vs. M3	2|-710.8 - (-702.3)|	17	4	0.002	0.45	1.9 or 8.7
M7 vs. M8	2|-711.6 - (-703.2)|	16.8	2	<0.001	0.16	4.8
Set 2						
M0 vs. M2	2|-789.4 - (-772.3)|	34.2	2	<0.001	0.16	6.6
M0 vs. M3	2|-789.4 - (-771.7)|	35.4	4	<0.001	0.72	1.5 or 10.5
M1 vs. M2	2|-778.6 - (-772.3)|	12.6	2	0.002	0.16	6.6
M1 vs. M3	2|-778.6 - (-771.7)|	13.8	4	0.008	0.72	1.5 or 10.5
M7 vs. M8	2|-778.7 - (-772.4)|	12.6	2	0.002	0.17	6.2

[a]of positively selected sites

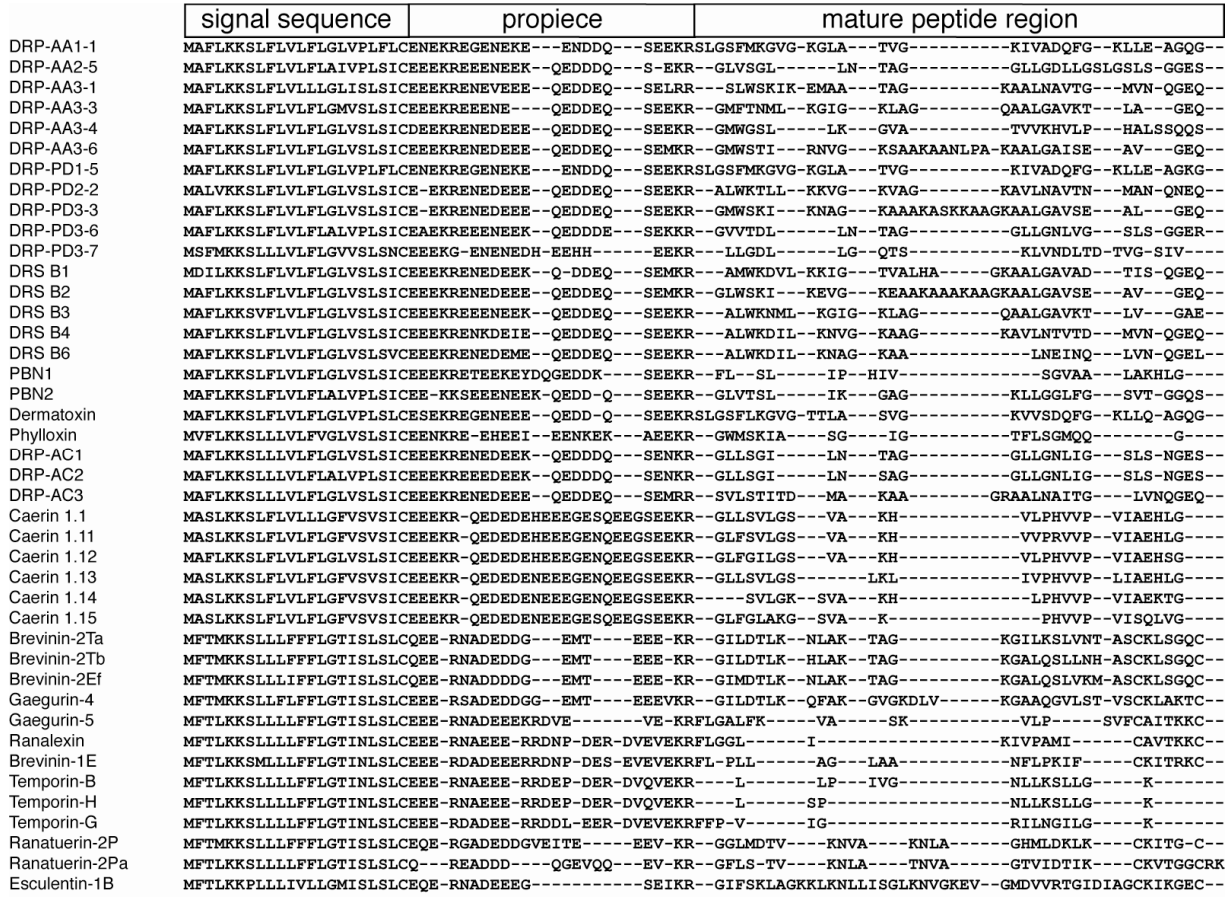

Figure 3.1

Alignment of frog antimicrobial peptide sequences used in the study in which the elevated mutation hypothesis was first proposed (based on Fig. 2 of Vanhoye et al. 2003). Multiple indels in the mature peptide region mean it is impossible to align homologous residues with certainty. Alignment errors could bias the estimate of ds and the tv/ts ratio.

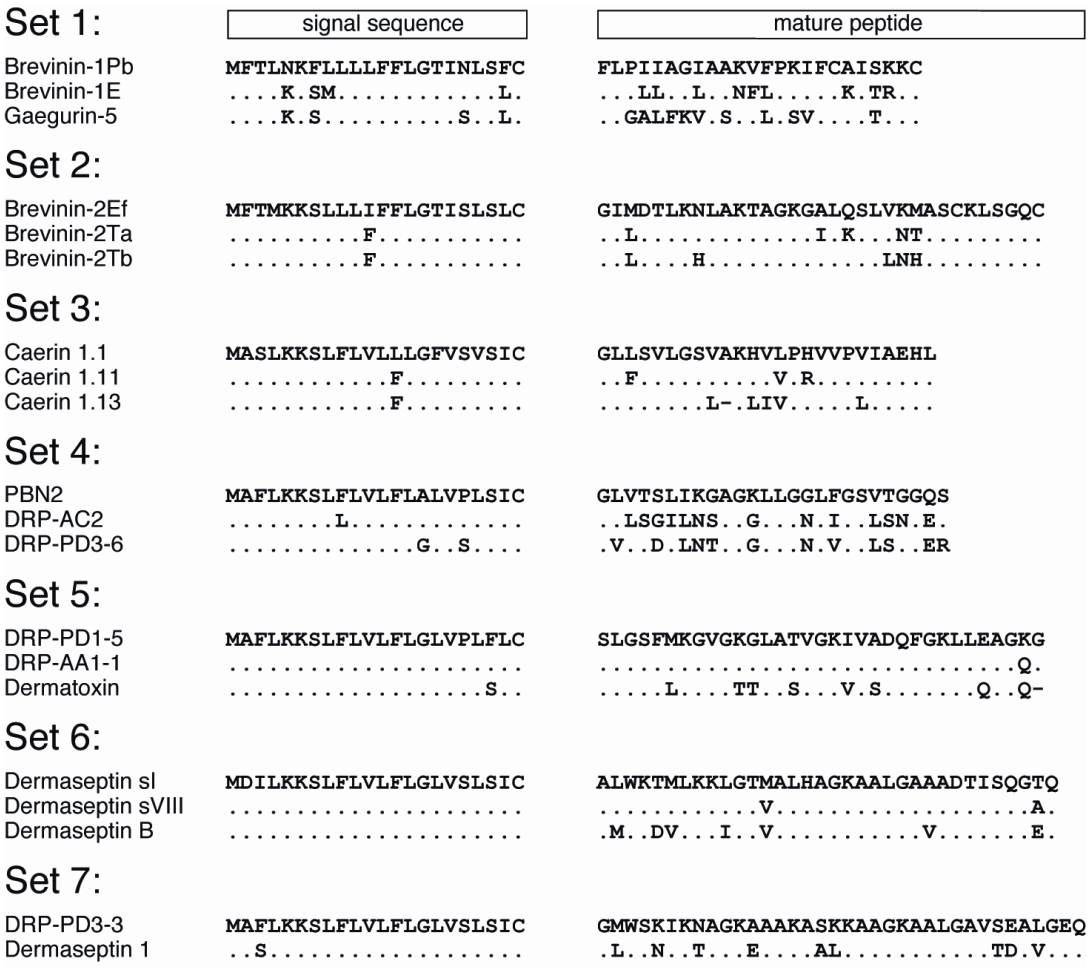

Figure 3.2

Alignments of 20 frog antimicrobial peptide sequences analyzed in this study. These sequences group into seven sets of similar length.

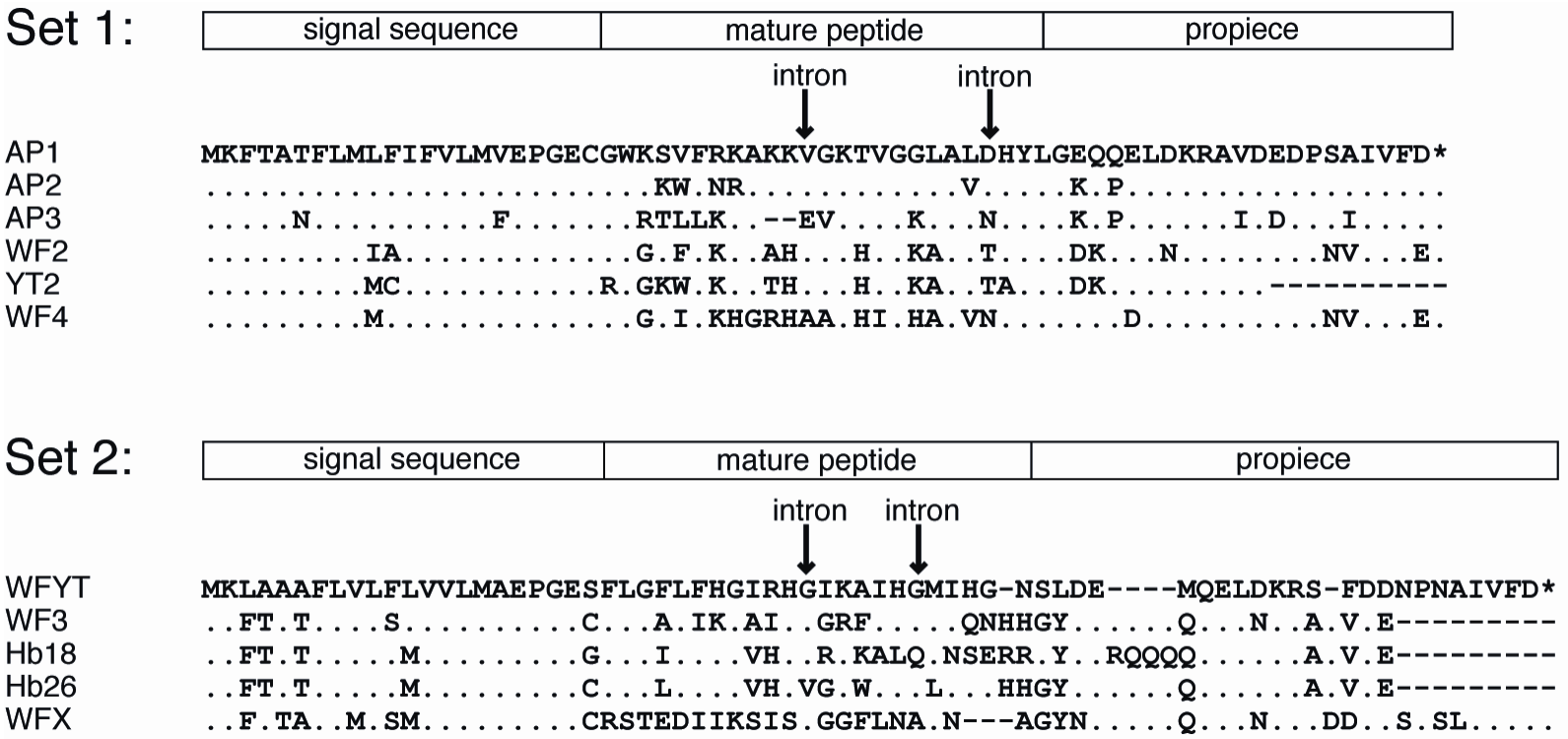

Figure 3.3

Alignments of 11 pleurocidin sequences analyzed in this study. These sequences group into two sets with exons of similar length.

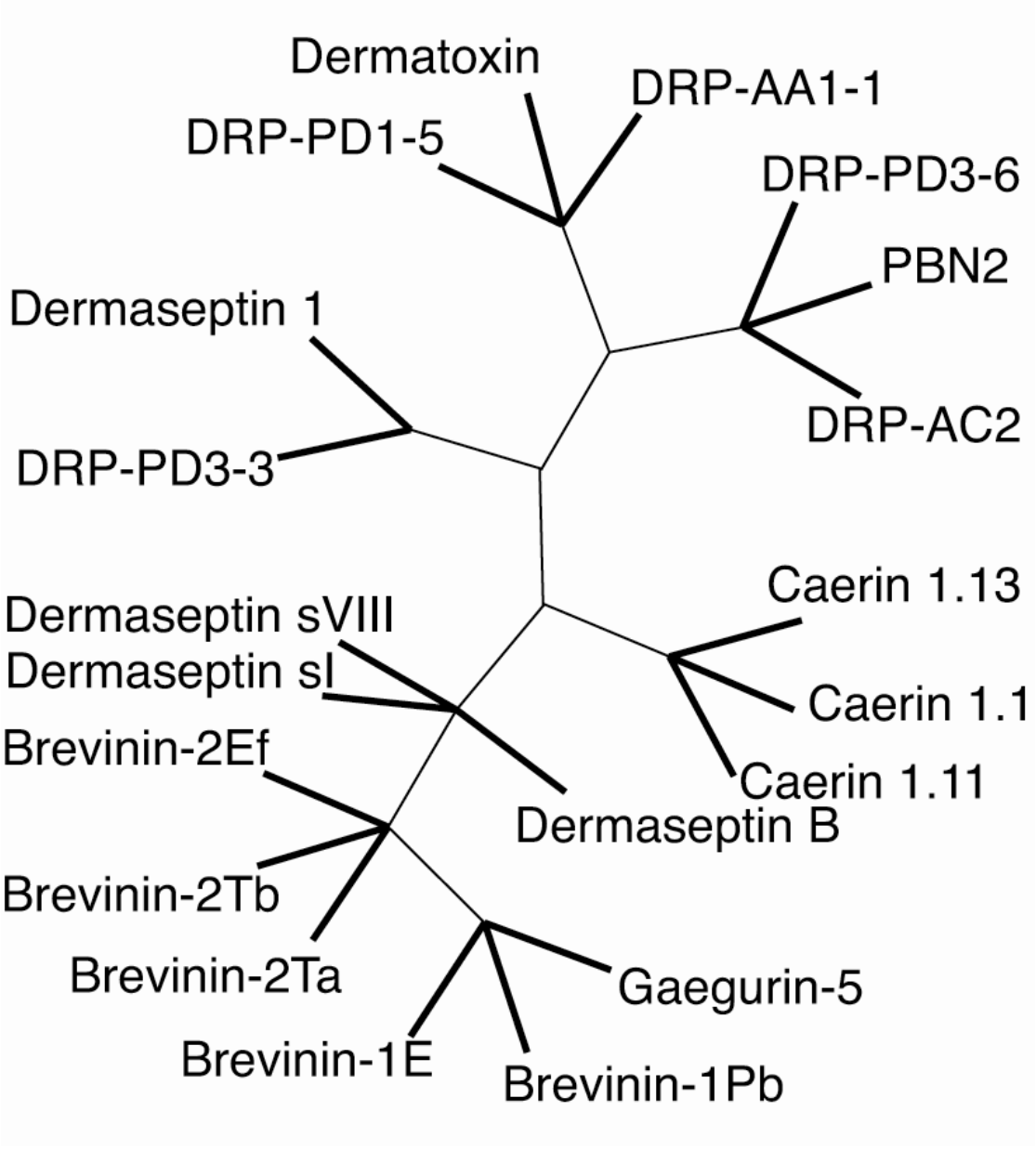

Figure 3.4

Maximum-likelihood phylogeny of the 20 frog antimicrobial peptides analyzed in this study. In this study, evolutionary changes have only been calculated along branches in boldface.

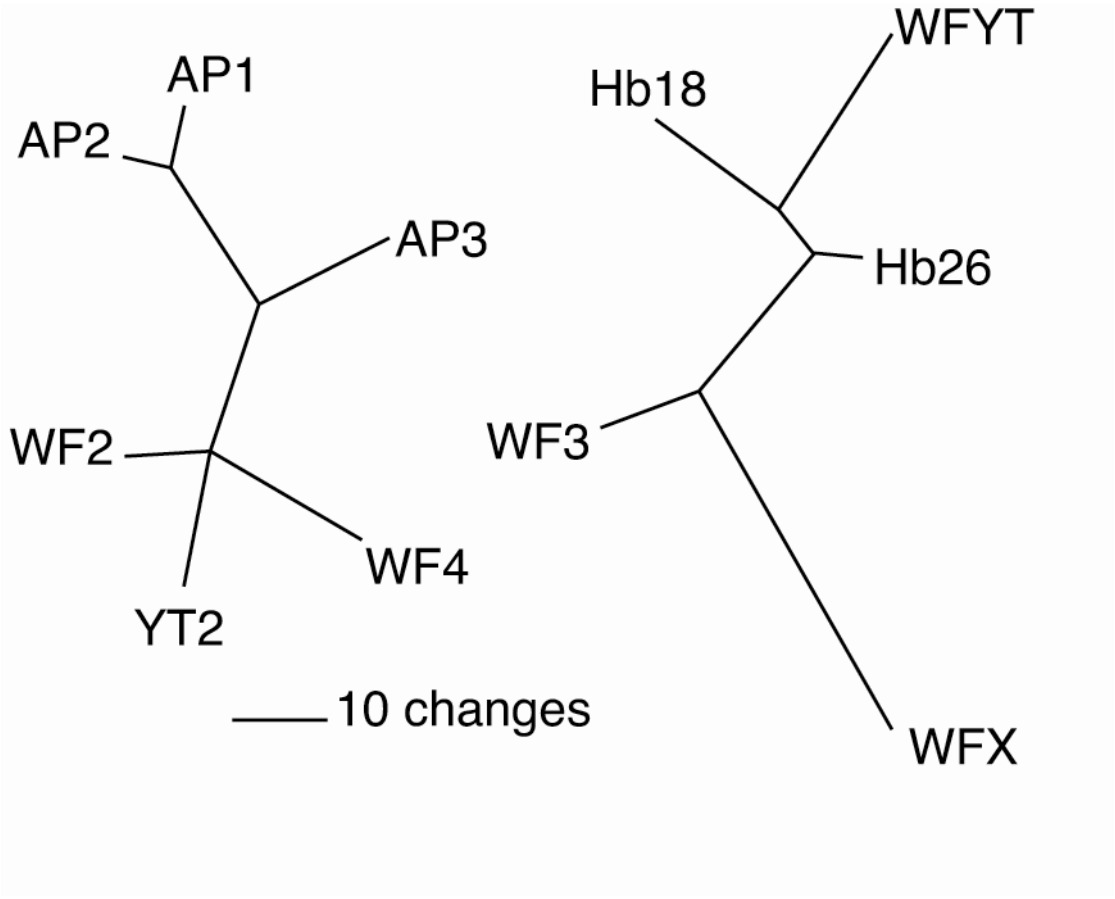

Figure 3.5

Maximum-likelihood phylogenies of 11 pleurocidin sequences analyzed in this study.

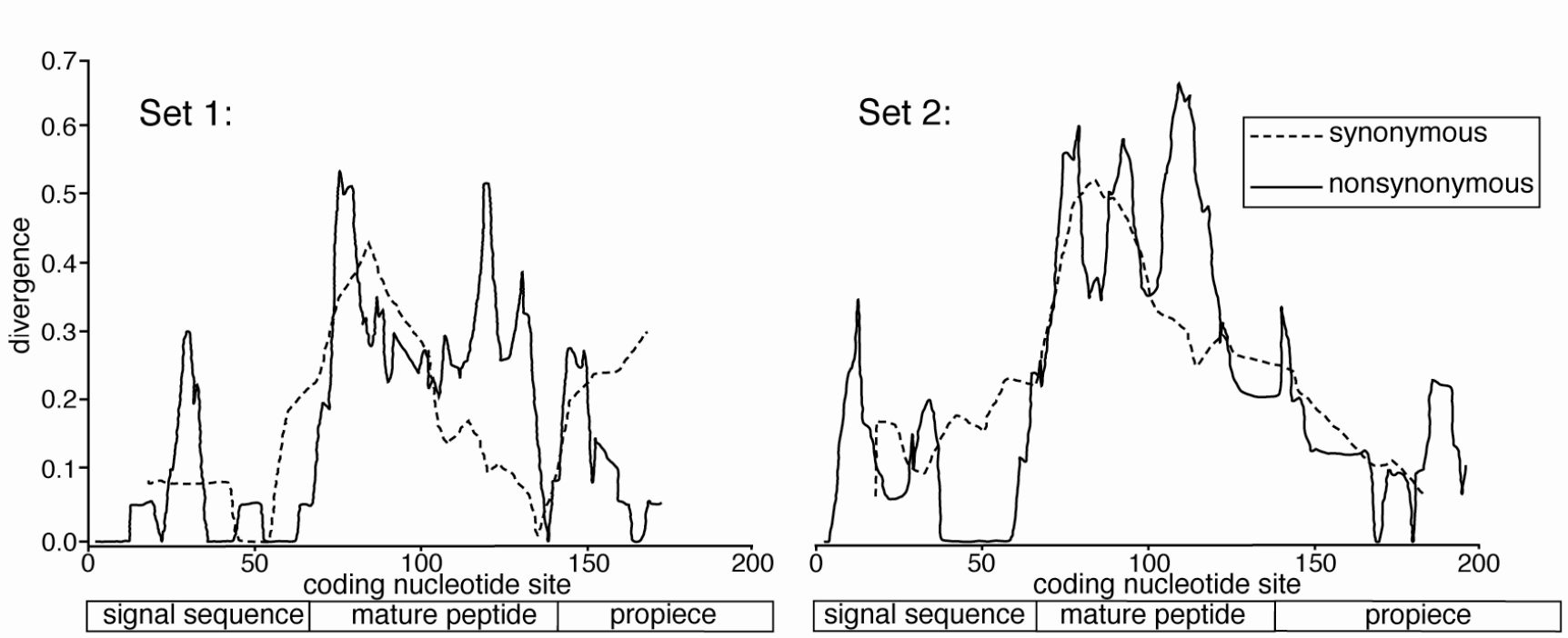

Figure 3.6

Sliding window representation of pairwise divergence at synonymous and nonsynonymous sites in two sets of pleurocidin sequences. Divergence at both synonymous and nonsynonymous sites is enhanced in the mature peptide region. Nonsynonymous divergence often exceeds synonymous divergence, especially in the mature peptide region.

SELECTION FOR ANTIMICROBIAL PEPTIDE DIVERSITY IN FROGS LEADS TO GENE DUPLICATION AND LOW ALLELIC VARIATION

Jacob A. Tennessen and Michael S. Blouin

Journal of Molecular Evolution
Springer Publishing
2007. 65(5):605-615

CHAPTER 4

SELECTION FOR ANTIMICROBIAL PEPTIDE DIVERSITY IN FROGS LEADS TO GENE DUPLICATION AND LOW ALLELIC VARIATION

4.1 Abstract

Antimicrobial peptides are highly diverse pathogen-killing molecules. In many taxa, their evolution is characterized by positive selection and frequent gene duplication. It has been proposed that genes encoding antimicrobial peptides might be subject to balancing selection and/or an enhanced mutation rate, but these hypotheses have not been well evaluated because allelic variation has rarely been studied at antimicrobial peptide loci. We present an evolutionary analysis of novel antimicrobial peptide genes from leopard frogs, *Rana*. Our results demonstrate that a single genome contains multiple homologous copies, among which there is an excess of nonsynonymous nucleotide site divergence relative to that expected from synonymous site divergence. Thus, we confirm the trends of recurrent duplication and positive selection. Allelic variation is quite low relative to interspecies divergence, indicating a recent positive selective sweep with no evidence of balancing selection. Repeated gene duplication, rather than a balanced maintenance of divergent allelic variants at individual loci, appears to be how frogs have responded to selection for a diverse suite of antimicrobial peptides. Our data also support a pattern of enhanced synonymous site substitution in the mature peptide region of the gene, but we cannot conclude that this is due to an elevated mutation rate.

4.2 Introduction

Unlike much of the genome, genes of the immune system are known to frequently evolve non-neutrally (Garrigan and Hedrick 2003). Thus, they permit the study of natural selection and adaptation at its fundamental molecular level. Studying immunity gene evolution is also important because changes driven by positive selection are likely to convey novel functional properties, and functionally divergent

disease resistance genes could lead to distinct practical therapeutic applications. Furthermore, it is particularly critical to understand how hosts adapt to pathogens, because emerging infectious diseases are exerting strong novel selective pressures on many species (Daszak et al. 2000; Altizer et al. 2003).

Antimicrobial peptides are short, cationic, amphipathic molecules that kill bacteria, viruses, and fungi through a noncatalytic disruption of the microbial membrane (Yeaman and Yount 2003; Tennessen 2005a). The mature peptide is often cleaved off a larger protein containing a signal sequence and a propiece (Tennessen 2005a). These molecules have been studied in many taxa. In ranid frogs, antimicrobial peptides include the brevinins and the ranatuerins, which consist of an α-helix with a conserved disulfide bridge forming a six- or seven-residue loop known as the ‘‘Rana box’’ at the C-terminal end (Park et al. 1994). Because they are easily obtained from the skin of live specimens, a large number of ranid frog antimicrobial peptides have been described (Conlon et al. 2004). They are diverse in functional properties, and many are active against human microbes and/or against pathogens causing emerging infectious diseases in amphibians (Rollins-Smith et al. 2002; Chinchar et al. 2004; Rollins-Smith and Conlon 2005). Much less is known about the genes encoding these peptides.

In this study, we sequence brevinin and ranatuerin loci from three closely related species: the northern leopard frog, *Rana pipiens*; the southern leopard frog, *R. sphenocephala*; and the Chiricahua leopard frog, *R. chiricahuensis* (Hillis and Wilcox 2005). We address four principal questions.

4.2.1 *Has Frequent Gene Duplication Resulted in Multiple Copies of Antimicrobial Peptide Genes Within Leopard Frog Genomes?*

All animal species appear to carry multiple duplicated copies of antimicrobial peptide genes, with vertebrate genomes typically containing tens of loci (Schutte et al. 2002; Lynn et al. 2004a). In most taxa that have been examined, several rounds of duplication have occurred relatively recently (Maxwell et al. 2003; Semple et al. 2003; Bulmer and Crozier 2004; Tennessen 2005a). Gene copy number has not been well

investigated in amphibians, though. Several similar peptides can often be isolated from the same frog species, suggesting multiple loci (Conlon et al. 2004). A suite of unique antimicrobial peptides would presumably maximize a host's ability to fight disease, and thus we hypothesize that selection has favored frequent duplication of these genes in leopard frogs.

4.2.2 *Has Positive Selection Driven the Evolution of Leopard Frog Antimicrobial Peptides?*

Positive natural selection on antimicrobial peptide genes occurs often in many animal taxa (Tennessen 2005a). The primary evidence for positive selection is that the ratio of the rate of nonsynonymous substitution per nonsynonymous nucleotide site to the rate of synonymous substitution per synonymous nucleotide site, or dn/ds, frequently exceeds one. The few available sequences of frog antimicrobial peptide genes suggest that these genes also evolve via positive natural selection, at least in some cases (Duda et al. 2002). Thus, antimicrobial peptides are potentially useful models for studying non-neutral molecular evolution. By testing the hypothesis that positive selection has occurred among the leopard frog sequences we examine, we add to the overall understanding of when and how antimicrobial peptides evolve adaptively. For example, does positive selection only occur immediately after gene duplication, or are there also adaptive differences between orthologs? Peptides that have diverged due to positive selection are also important to identify because they are more likely to differ in their functional properties.

4.2.3 *Is Allelic Variation at Antimicrobial Peptide Loci Affected by Non-Neutral Evolution?*

Allelic variation at antimicrobial peptide loci has never been systematically studied in a nonhuman vertebrate. Adaptive genetic variation is often high in other types of immunity genes, including the major histocompatibility complex (MHC) of vertebrates (Garrigan and Hedrick 2003; Piertney and Oliver 2006) and plant R-genes (Bergelson et al. 2001; Bakker et al. 2006). Multiple variants of antimicrobial peptides

would presumably provide an adaptive benefit, so it is possible that balancing selection maintains several alleles at individual loci via overdominance, frequency-dependent selection, and/or geographically heterogeneous selection due to distinct microbial communities affecting different populations (Tennessen 2005a). Alternatively, generalist immunity genes like antimicrobial peptide loci might harbor less diversity than specialist immunity genes (Tiffin et al. 2004). Under this scenario, recurrent gene duplication might satisfy the need for multiple antimicrobial peptide variants, while selective sweeps reduce allelic variation at individual loci. Distinct peptides are sometime found in frogs of the same species from different geographic locations (e.g., Won et al. 2004), but it is unclear whether these represent allelic variants or whether such variation is selectively neutral. We examine loci for allelic diversity, and we test whether variation and divergence at a particular locus in *R. pipiens* differs from neutral expectations by comparing it with other nuclear loci.

4.2.4 *To What Extent is the Synonymous Substitution Rate in the Mature Peptide Region Enhanced, and Why?*

Another widespread trend in antimicrobial peptide genes is a greater degree of synonymous site divergence in the mature peptide region, relative to the rest of the gene (Vanhoye et al. 2003; Tennessen 2005b). In frogs, this pattern has been shown with respect to the signal sequence, but it is unknown whether it holds with respect to noncoding sequence such as introns. The two most likely explanations for this phenomenon are an elevated mutation rate or natural selection acting on supposedly silent sites. The hypothesis of an elevated mutation rate predicts an excess of allelic variation relative to other loci. Because allelic variation is also influenced by natural selection, it cannot be used as a direct estimate of the relative mutation rate, but it can provide suggestive evidence as to whether the mutation rate is abnormally high.

4.3 Materials and Methods

4.3.1 *Samples*

Specimens of *R. pipiens* and *R. sphenocephala* were collected from multiple populations as described previously (Hoffman and Blouin 2004). Specimens of *R. chiricahuensis* consisted of toe clips collected September 2001 in Arizona, USA, and preserved by desiccation in 1.5 ml tubes filled with Drierite desiccant (W. A. Hammond Drierite Co., Xenia, OH). DNA was extracted as described previously (Hoffman and Blouin 2004).

Most polymerase chain reactions (PCRs) were carried out as 25 µl reactions using standard buffer conditions, 1.5 mM MgCl2, 0.2 mM each deoxyribonucleotide triphosphate, approximately 100 ng DNA, and 0.5 Units Taq DNA polymerase. The exceptions, noted below, employed either Pfu DNA polymerase (Promega, Madison, WI) or the TaqPlus Long PCR System (Stratagene, La Jolla, CA). For all sequences, the PCR product was visualized under ultraviolet light, purified with the MoBio Ultraclean PCR cleanup kit (Salana Beach, CA), and sequenced by the Nevada Genomics Center (Reno, NV).

4.3.2 *Brevinin Loci*

In order to estimate the minimum number of brevinin loci, to test for evidence of positive selection via the dn/ds ratio, and to compare synonymous site divergence with noncoding divergence, we cloned and sequenced brevinin genes. We designed primers for brevinin-1Pb (Brev1PF1: 5′-TCC TTC TAC TCC TTT TCT TC-3′; Brev1PF2: 5′-CAT CAA CTT ATC TTT CTG TG-3′; Brev1PR1: 5′-CCA ATT CAA GTT TCC AAA G-3′; Brev1PR2: 5′-TTT CAT CTG GCT CAT CTA TTC-3′) using an available cDNA sequence (Accession AJ427746; Chen et al. 2003). We amplified brevinin-1Pb from genomic *R. pipiens* DNA with the primers Brev1PF1 and Brev1PR1 using the TaqPlus Long PCR System under the following conditions: an initial denaturation step at 94°C for 5 min; 35 cycles of denaturation at 94°C for 45 sec, annealing at 51°C for 30 sec, and extension at 72°C for 8 min; and a final extension step at 72°C for 5 min. We then amplified the resultant PCR product with the nested primers Brev1PF2 and Brev1PR2 using the TaqPlus Long PCR System under the following conditions: an initial denaturation step at 94°C for 5 min; 35

cycles of denaturation at 94°C for 45 sec, annealing at 54°C for 30 sec, and extension at 72°C for 5.5 min; and a final extension step at 72°C for 30 min. We partially sequenced the resultant PCR product from the 3' end and used it to design the primer Brev1PF3 (5'-TTA GAA CAA TAT TAG ATG ACC-3').

We used the primers Brev1PF3 and Brev1PR1 in PCR reactions with the high-fidelity enzyme Pfu DNA polymerase and genomic DNA template from each of the following frogs: two *R. pipiens* individuals from the same population in Ontario, Canada (RpO1 and RpO2), a *R. pipiens* individual from British Columbia, Canada (RpBC1), and a *R. chiricahuensis* individual (Rc1). PCR amplification conditions consisted of an initial denaturation step at 94°C for 5 min; 35 cycles of denaturation at 94°C for 45 sec, annealing at 54°C for 30 sec, and extension at 72°C for 1 min; and a final single extension step at 72°C for 5 min. The resultant PCR products were incubated with Taq polymerase in order to add 30 adenines and cloned into *E. coli* using the TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA). The recombinant locus was amplified from screened colonies using the primers T3 and T7 under the following PCR conditions: an initial denaturation step at 94°C for 3 min; 35 cycles of denaturation at 94°C for 45 sec, annealing at 55°C for 30 sec, and extension at 72°C for 1 min; and a final single extension step at 72°C for 7 min. The resultant PCR products were sequenced with either T3 or T7. In addition, we used the primers Brev1PF3 and Brev1PR1 to amplify genomic DNA from two *R. sphenocephala* individuals under the aforementioned PCR conditions, and the resultant PCR products were sequenced without cloning.

4.3.3 *Ranatuerin Loci*

In order to estimate the minimum number of ranatuerin loci and to test for evidence of positive selection with the dn/ds ratio, we sequenced ranatuerin genes. We designed primers for the ranatuerin-2P gene (Ranat2PF1: 5'-ATG AAG AAA TCC CTG TTA CTC-3'; Ranat2PF2: 5'-ATC TCC TTA TCT CTC TGT G-3'; Ranat2PR1: 5'-CAC ATC AGA TGA CTT CCA ATT-3'; Ranat2PR2: 5'-TTC CAC ACC ATC GTC TCC-3') using an available cDNA sequence (Accession AJ427747; Chen et al.

2003). We amplified the ranatuerin-2P gene from genomic *R. pipiens* DNA with the primers Ranat2PF1 and Ranat2PR1 using the TaqPlus Long PCR System under the following conditions: an initial denaturation step at 94°C for 5 min; 35 cycles of denaturation at 94°C for 45 sec, annealing at 53°C for 30 sec, and extension at 72°C for 8 min; and a final extension step at 72°C for 30 min. We then amplified the resultant PCR product with the nested primers Ranat2PF2 and Ranat2PR2 using the TaqPlus Long PCR System under the following conditions: an initial denaturation step at 94°C for 5 min; 35 cycles of denaturation at 94°C for 45 sec, annealing at 55°C for 30 sec, and extension at 72°C for 5.5 min; and a final extension step at 72°C for 30 min. We partially sequenced the resultant PCR product and used it to design the primer Ranat2PF3 (5'-TCA ATG TTG TTT TAT GTA AGG-3'). We also designed the primer Ranat2PF4 (5'-AAG CCG ATG ACG ACC AAG-3') using an available cDNA sequence (AJ427748; Chen et al. 2003).

We used the primers Ranat2PF3 and Ranat2PR1 to amplify genomic DNA from *R. pipiens* and *R. chiricahuensis*. PCR amplification conditions consisted of an initial denaturation step at 94°C for 5 min; 35 cycles of denaturation at 94°C for 45 sec, annealing at 54°C for 30 sec, and extension at 72°C for 1 min; and a final single extension step at 72°C for 5 min. The resultant PCR product was sequenced. In addition, we used the primers Ranat2PF4 and Ranat2PR1 to amplify genomic DNA from *R. pipiens* and *R. chiricahuensis*. PCR amplification conditions consisted of an initial denaturation step at 94°C for 5 min; 35 cycles of denaturation at 94°C for 45 sec, annealing at 55°C for 30 sec, and extension at 72°C for 45 sec; and a final extension step at 72°C for 5 min. The resultant PCR product was sequenced.

In order to test for evidence of balancing selection and an enhanced mutation rate, we examined allelic variation at two particular ranatuerin loci. We amplified genomic DNA from 24 specimens of *R. pipiens* and 4 specimens of *R. chiricahuensis* with the primers Ranat2PF3 and Ranat2PR1. Each *R. pipiens* specimen was collected from a unique population from 1 of 13 states and provinces across North America, representing the major phylogeographic groups of this species as determined from mitochondrial DNA (Hoffman and Blouin 2004). We also amplified genomic DNA

from three specimens of *R. pipiens* and four specimens of *R. chiricahuensis* with the primers Ranat2PF4 and Ranat2PR1. The resultant PCR products were sequenced.

For the purpose of a Hudson–Kreitman–Aguadé (HKA) test (Hudson et al. 1987), we compared the gene coding for ranatuerin-2P with two putatively neutral loci coding for arcadlin and type-1 myosin heavy chain. Arcadlin is a cadherin involved in synaptic reorganization (Crump et al. 2002), and myosin is a muscle fiber protein (Lutz et al. 1998). We designed primers for the arcadlin gene (ArcadF: 5'-AAC AGA AGC AGC AGT CAC-3'; ArcadR: 5'-CGC TGT CAT TGA AGT CAC-3') using an available cDNA sequence (Accession AF279872; Crump et al. 2002). We also designed primers for the myosin gene (MyosinF: 5'-GAA GAA CCT TGA ACA GAC-3'; MyosinR: 5'-ATT CTT CTT GTC CTC CTC-3') using an available cDNA sequence (Accession AF013132; Lutz et al. 1998). We amplified genomic DNA from 13 *R. pipiens* individuals and 2 *R. chiricahuensis* individuals under the following conditions for both arcadlin and myosin primer pairs: an initial denaturation step at 94°C for 5 min; 35 cycles of denaturation at 94°C for 45 sec, annealing at 52°C for 30 sec, and extension at 72°C for 1 min; and a final extension step at 72°C for 5 min. The resultant PCR products were sequenced. To determine whether adaptive substitutions occurred in the *R. pipiens* or the *R. chiricahuensis* lineage, we inferred the common ancestral sequences for the genes coding for ranatuerin-2P, arcadlin, and myosin. Ancestral state reconstruction for the ranatuerin-2P gene employed published DNA (Accession AM113511, AM113509; Chen et al. 2006) and amino acid (Fig. 4.1) sequences from closely related outgroup species, while ancestral state reconstruction for the arcadlin and myosin genes employed sequences that we amplified from *R. sphenocephala*.

4.3.4 *Analysis*

All sequences have been deposited in GenBank under accession numbers DQ276956–DQ276977 and DQ923137–DQ923159. Heterozygotes were identified by the presence of double peaks. Evolutionary analyses were performed with PAML (vers. 3.13a, Yang 1997), DnaSP (vers. 4.0, Rozas et al. 2003), and PAUP* (vers.

4.0b10, Swofford 2002). We used the maxchi method (Maynard Smith 1992) to test for recombination, which could invalidate any phylogeny-based tests of selection. To test for positive selection with PAML, we ran two models: one in which dn/ds is estimated from the data, and one in which dn/ds is constrained to be one. If the estimated value of dn/ds is greater than one, and this model is significantly better than the constrained model in a log-likelihood test, we can conclude that positive selection has occurred. We did not use any PAML models that allow dn/ds to differ among codons or among branches. Although conserved regions indicate that brevinin and ranatuerin genes are homologous (Vanhoye et al. 2003), the mature peptide regions are too divergent to be aligned reliably, so we analyzed these two gene families separately.

4.4 Results

4.4.1 *Brevinin Loci*

Our initial brevinin locus amplification resulted in a PCR product of approximately 5,000 base pairs. This was partially sequenced to reveal part of an intron, followed by the C-terminal region of the propiece, immediately followed by the mature peptide region (Fig. 4.1). We designed a forward primer that annealed in the intron, Brev1PF3. Using Brev1PF3 and Brev1PR1, we obtained 227 cloned brevinin-1 sequences consisting of 63 unique haplotypes among the frogs RpO1 (110 sequences, 35 haplotypes), RpO2 (43 sequences, 9 haplotypes), RpBC1 (42 sequences, 10 haplotypes), and Rc1 (32 sequences, 12 haplotypes). However, many of these haplotypes likely represent PCR-induced errors such as point mutation or recombination (Cronn et al. 2002), which can occur even with high-fidelity polymerase. Of the 63 haplotypes, 42 were only observed once, and 1 was only observed twice; out of these 43, 24 could have been formed by a single recombination event between two other multiply observed haplotypes in that same individual. To minimize the possibility of PCR-induced errors and to conservatively estimate the number of brevinin-1 loci, we excluded from further analysis all haplotypes observed

only once or twice. Our final dataset of cloned brevinin-1 genes thus consisted of 20 unique haplotypes of 438 to 461 nucleotides, all of which were observed at least three times in the same individual. Nine were found in RpO1, five were found in RpOC2, four were found in RpBC1, and five were found in Rc1 (Fig. 4.2). We also directly sequenced 222 nucleotides of two additional brevinin-1 genes from *R. sphenocephala*. The 22 haplotypes from the three species each contained a single exon and encoded a total of 13 unique amino acid sequences (Fig. 4.1).

There is no evidence of recombination among the brevinin sequences within the mature peptide region (maxchi test, $p > 0.1$), so a phylogeny-based test for positive selection is appropriate. We used PAUP* to construct a maximum likelihood phylogeny of the brevinin-1 haplotypes (Fig. 4.2). A phylogeny constructed using just the mature peptide region is consistent with this phylogeny (not shown). We tested for positive selection in the mature peptide region with PAML. Along the branches of this phylogeny, dn/ds averages 2.6. A model that employs this estimated value of dn/ds is significantly better than a model which forces dn/ds to equal one ($p < 0.05$). This result demonstrates that positive natural selection has caused much of the evolutionary diversification of these peptides.

We compared overall divergence in the intron with synonymous divergence in the mature peptide region along the branches of the tree (Fig. 4.2). We only obtained a small portion of the intron in the two *R. sphenocephala* sequences, so these were excluded from this analysis. Although synonymous divergence in the mature peptide region (mean $ds = 0.012$) is more than twice as high as overall divergence in the introns (mean $d = 0.005$), the difference is not significant ($p > 0.1$; paired t-test). Thus, these results provide inconclusive support for an enhanced synonymous substitution rate in the mature peptide region with respect to noncoding sequence.

4.4.2 *Ranatuerin Loci*

Our initial amplification of the locus encoding ranatuerin-2P resulted in a PCR product of approximately 5,000 base pairs. This was partially sequenced to reveal part of an intron, followed by the C-terminal region of the propiece, immediately followed

by the mature peptide region. We designed a forward primer that annealed in the intron, Ranat2PF3. The primers Ranat2PF3 and Ranat2PR1 consistently amplified 210 nucleotides of a single locus in both *R. pipiens* and *R. chiricahuensis*, which we call the *Ranatuerin2* gene since it codes for the previously described peptide ranatuerin-2P in *R. pipiens* (Goraya et al. 2000) (Fig. 4.1). The primers Ranat2PF4 and Ranat2PR1 consistently amplified 160 to 169 nucleotides of another locus in these species, which we call the *Ranatuerin2b* gene because it is distinct from the previously described cDNA sequence of ranatuerin-2Pa in *R. pipiens* (AJ427748; Chen et al. 2003) (Fig. 4.1).

There is no evidence of recombination among the ranatuerin sequences within the mature peptide region (maxchi test, $p > 0.1$), so a phylogeny-based test for positive selection is appropriate. We used PAUP* to construct a maximum likelihood phylogeny of the ranatuerin mature peptide region (not shown), and tested for positive selection with PAML. Along the branches of this phylogeny, dn/ds averages 0.6. A model which employs this estimated value of dn/ds is not significantly better than a model which forces dn/ds to equal one ($p > 0.1$). Thus, the dn/ds ratio provides no support for the hypothesis of positive selection acting on ranatuerins.

We sequenced the *Ranatuerin2* locus in 24 *R. pipiens* individuals, each from a unique population, and in four *R. chiricahuensis* individuals from two populations. We sequenced the *Ranatuerin2b* locus in three *R. pipiens* individuals from three populations and in the same four *R. chiricahuensis* individuals. No intraspecific variation was observed at either ranatuerin locus, with the exception of a synonymous substitution found in a single individual of *R. chiricahuensis* at *Ranatuerin2b*. These results provide no support for either the hypothesis of balancing selection or the hypothesis of an enhanced mutation rate.

We tested for a recent selective sweep on the Ranatuerin2 locus by employing an HKA test (Hudson et al. 1987). We arbitrarily chose two loci, coding for arcadlin (*Arcadlin*) and type 1 myosin heavy chain (*Myosin*), as putatively neutrally evolving control sequences. The assumption of neutrality is justified because there is no reported evidence of non-neutral evolution at these loci, 41% of the sequenced

nucleotides in these two loci were in introns (Fig. 4.3), no nonsynonymous differences were observed within or between species, and Tajima's D is not significant for either locus ($p > 0.1$ for both). We sequenced these loci in 13 of the *R. pipiens* individuals and two of the *R. chiricahuensis* individuals from which we had sequenced *Ranatuerin2*. At the 139 nucleotides sequenced from *Ranatuerin2*, we observed no intraspecific segregating sites among 26 *R. pipiens* sequences (all encoding ranatuerin-2P), and five fixed differences (four nonsynonymous and one synonymous) with respect to the *R. chiricahuensis* sequences (all encoding ranatuerin-2CHa). At the 150 nucleotides sequenced from *Myosin*, we observed 7 intraspecific segregating sites in these same individuals, and a mean of 3.5 interspecific differences. Likewise, at the 265 nucleotides of *Arcadlin*, we observed 5 intraspecific segregating sites and a mean of 2.85 interspecific differences. Divergence at the *Ranatuerin2* locus is significantly greater than expected, given the amount of polymorphism in *R. pipiens* (chi-square test, $p < 0.05$), indicating a positive selective sweep (Fig. 4.3). The test is still significant if only one sequence per population per locus is used (chi-square test, $p < 0.05$), which minimizes the effect of population structure (Wakeley 1999; Ironside and Filatov 2005). It is also significant if *Ranatuerin2* is just compared with *Arcadlin* (chi-square test, $p < 0.05$) or just with *Myosin* (chi-square test, $p < 0.05$). In contrast, the ratio of polymorphism to divergence is not significantly different between *Arcadlin* and *Myosin* (chi-square test; $p > 0.1$).

To confirm that the selective sweep occurred in the *R. pipiens* lineage, we reconstructed the sequence of the three loci *Ranatuerin2*, *Arcadlin*, and *Myosin* in the common ancestor of *R. pipiens* and *R. chiricahuensis*. We amplified *Arcadlin* and *Myosin* in *R. sphenocephala* to infer the ancestral sequence at these loci. We were unable to amplify a *Ranatuerin2* ortholog in *R. sphenocephala*, so we used published DNA and amino acid sequences from other species to reconstruct the ancestral sequence (Fig. 4.1). We call the ancestral peptide ranatuerin-2STa, for *Stertirana*, the clade descended from the most recent common ancestor of *R. pipiens* and *R. chiricahuensis* (Hillis and Wilcox 2005). The HKA test was then performed on the ratio of polymorphism in *R. pipiens* to divergence between *R. pipiens* and the common

ancestor of *R. pipiens* and *R. chiricahuensis*. The most parsimonious reconstruction suggests that four of the five fixed interspecies differences at *Ranatuerin2* occurred in the *R. pipiens* lineage. The ratio of divergence to polymorphism at *Ranatuerin2* is still significantly higher than expected, even if only one sequence per population per locus is used (chi-square test, $p < 0.05$).

4.5 Discussion

4.5.1 *Gene Duplication*

Most animal species examined have multiple related copies of antimicrobial peptide genes, and thus a pattern of recurrent duplication might be common (Schutte et al. 2002; Maxwell et al. 2003; Semple et al. 2003; Bulmer and Crozier 2004; Lynn et al. 2004a; Tennessen 2005a). Our results support this trend. These duplications are likely favored by natural selection, allowing organisms to battle pathogens with a suite of distinct peptides.

There are at least two ranatuerin loci in *R. chiricahuensis*, and at least three in *R. pipiens*, if we assume that the published ranatuerin-2Pa sequence (Chen et al. 2003), which we did not observe, is encoded by a unique locus (Fig. 4.1). We observed no allelic variation at either ranatuerin locus in *R. pipiens*. Furthermore, peptides with the expected molecular weight of both ranatuerin-2P and ranatuerin-2Pb are expressed in the skin of *R. pipiens* from several different geographic locations (L. Rollins-Smith, personal communication), suggesting that these sequence harbor low allelic variation and are not pseudogenes. If ranatuerin-2Pa were an allele at one of these loci, it would be rare and unusually divergent from the common allele. Therefore, it is probably encoded by a unique locus.

There are at least five brevinin-1 loci in *R. pipiens*, and at least three in *R. chiricahuensis*. Nine unique brevinin-1 haplotypes were each observed at least four times in RpO1. None of these nine could be formed from another multiply observed haplotype with a single point mutation. Furthermore, none of them could be formed by a single recombination event between two other multiply observed haplotypes. This

evidence strongly suggests that all nine are real and not PCR-induced artifacts. Therefore, since there cannot be more than two alleles at a locus, there are a minimum of five brevinin-1 loci, at least in of one of the *R. pipiens* genomes. There are several possible reasons why fewer haplotypes were found in the other frogs, including homozygosity, variation in gene copy number, polymorphism at the primer binding sites, and absence by chance of certain haplotypes from the sample of clones we sequenced. Several of the brevinin-1 haplotypes in *R. pipiens* cluster together in the phylogeny, to the exclusion of the haplotypes from other species (Fig. 4.2). This phylogenetic pattern suggests that several duplications have occurred since *R. pipiens* speciated. Thus, leopard frog antimicrobial peptides are not merely the survivors of ancient gene duplications, but continue to be produced by diversity-generating evolutionary processes. Indeed, if intraspecies variation in gene copy number is partially responsible for the different haplotype numbers observed among individual frogs (Fig. 4.2), locus number might still be increasing in living populations. The total of 22 brevinin-1 haplotypes from all frogs encodes 13 unique amino acid sequences including 12 unique putative antimicrobial peptides (Fig. 4.1). Three of these have previously been isolated from *R. pipiens* skin and named brevinin-1Pa, brevinin-1Pb, and brevinin-1Pe (Goraya et al. 2000), or alternatively called pipinins (Horikawa et al. 1985; Marenah et al. 2005). A fourth peptide, brevinin-1Sb, has been previously isolated from *R. sphenocephala* skin (Conlon et al. 1999). Five of the remaining peptides, hereby named brevinin-1Pf, brevinin- 1Pg, brevinin-1CHa, brevinin-1CHb, and brevinin-1Sd, closely resemble peptides which have been isolated from species in the *R. pipiens* complex (Fig. 4.1). The three most divergent putative peptides, brevinin-1Ph, brevinin-1Pi, and brevinin-1CHc, do not closely resemble any previously described sequence, although they are similar to each other and the genes encoding them are clearly homologous with the brevinin-1 genes. As they lack the conserved lysine-arginine motif where the mature peptide is cleaved from the propiece (Fig. 4.1), it is possible that these genes do not encode antimicrobial peptides. However, the strong signature of positive selection along the branch leading to these genes suggests that they are not pseudogenes, and they might have been selected to

perform another function. A six-residue deletion including one of the conserved cysteines occurs in Brevinin-1Pi, but the ‘‘Rana box’’ is reestablished with a nonhomologous cysteine that had previously evolved (Fig. 4.1). We did not observe genes coding for the peptides brevinin-1Pc and brevinin-1Pd, which have been detected in *R. pipiens* skin (Fig. 4.1) (Horikawa et al. 1985; Goraya et al. 2000). These might represent allelic variants or loci that do not amplify with our primers.

Leopard frogs also produce antimicrobial peptides from the temporin and esculentin families, which were not examined in this study (Goraya et al. 2000). Although the mature peptides are quite divergent, conserved gene regions demonstrate that brevinins, ranatuerins, esculentins, and temporins are homologous and have all descended from a single sequence in the ancestral ranid frog (Duda et al. 2002; Vanhoye et al. 2003). Our results demonstrate that multiple gene duplications have produced at least ten loci descended from this ancestral antimicrobial peptide gene.

4.5.2 *Positive Selection*

Antimicrobial peptide genes are thought to evolve via positive natural selection in many taxa, including frogs (Duda et al. 2002; Tennessen 2005a). Our results confirm this pattern. Evolution of the brevinin-1 family of genes is characterized by an excess of nonsynonymous substitutions in the mature peptide region, even when averaged over all codons and all branches. Although this is not true for the ranatuerin genes observed, we can infer a recent positive selective sweep on the *Ranatuerin2* locus (Fig. 4.3). Novel variants are probably favored by natural selection because they are more active against newly encountered pathogens or pathogens that have evolved resistance to the old antimicrobial peptide variant (Perron et al. 2006). Although we have not conclusively shown that the in vivo function of all the genes we observed is to encode peptides that kill pathogens, the hypothesis of an arms race with microbes would explain the high rate of adaptive evolution (Peschel and Sahl 2006).

A positive selective sweep has acted on the *Ranatuerin2* locus. No variation at this locus was observed in *R. pipiens*, but divergence is high between species, and even between *R. pipiens* and the inferred extinct sequence from the common ancestor

of *R. pipiens* and *R. chiricahuensis*. It is unclear whether population structure can bias the HKA test, but we minimized any effects by sampling a single allele per locus in each population (Wakeley 1999; Ironside and Filatov 2005). Thus, positive selection has enhanced substitution in the *R. pipiens* lineage, and fixed the ranatuerin-2P peptide in that species. Interestingly, ranatuerin-2P can kill the fungus *Candida albicans* and the Gram-positive bacterium *Staphylococcus aureus* (Goraya et al. 2000), whereas the tested peptides most similar in sequence to ranatuerin-2P cannot kill these microbes (e.g., ranatuerin-2PLb and -PLd from *R. palustris*, Basir et al. 2000; ranatuerin-2ARa from *R. areolata*, Ali et al. 2002; ranatuerin-2BYb from *R. boylii*, Conlon et al. 2003; Fig. 4.1). Although these human pathogens are not thought to infect frogs, it is possible that similar fungi or gram-positive bacteria were responsible for the observed selective sweep. The most likely target of the selective sweep is the histidine residue of ranatuerin-2P, where the inferred ancestral peptide had a glutamine residue (Fig. 4.1), since charge-changing substitutions are known to be important in antimicrobial peptide evolution (Tennessen 2005a). Selective sweeps like this one allow species with low gene flow, such as leopard frogs, to evolve collectively and maintain a cohesive species identity, even if migration is not sufficient to prevent divergence at neutral loci (Morjan and Rieseberg 2004).

Although there are several ranatuerin loci, the selective sweep at *Ranatuerin2* does not appear to have occurred immediately after gene duplication. Whereas a duplication event could have preserved ranatuerin-2STa at one locus, while allowing the other locus to evolve into ranatuerin-2P, it appears that ranatuerin-2STa has been lost. Without a sequenced genome, we cannot say conclusively that there is no gene encoding ranatuerin-2STa, but if that gene existed we would probably have amplified it with our primers. Thus, antimicrobial peptide evolution is not simply a model of recurrent duplication followed by positive selection. Rather, genes sometimes evolve adaptively without duplicating.

4.5.3 *Allelic Variation*

Intraspecific variation at antimicrobial peptide loci is not well understood, having been examined only in humans and *Drosophila* (Tennessen 2005a). We hypothesized that balancing selection might maintain multiple alleles at a locus. However, our HKA test suggests that the *Ranatuerin2* locus exhibits even less variation that expected, not more, and therefore positive but not balancing selection is acting (Fig. 4.3). In combination with similar results from *Drosophila* (Clark and Wang 1997), our study suggests that positive selective sweeps of new antimicrobial peptides might be common, in contrast with the highly polymorphic immunity genes such as MHC (Garrigan and Hedrick 2003; Piertney and Oliver 2006) and R-genes (Bergelson et al. 2001; Bakker et al. 2006). This trend is consistent with the hypothesis that balancing selection is more common at specialist than at generalist immunity loci (Tiffin et al. 2004). Since antimicrobial peptides interact with the microbial cell membrane instead of specific microbial gene products, it might be more difficult for microbes to evolve resistance to an effective antimicrobial peptide variant than to alleles of MHC or R-genes (Schroder 1999). Thus, an antimicrobial peptide allele can sweep to fixation before the pathogens it kills evolve resistance to it and it loses its fitness advantage. Of course, it must lose its fitness advantage eventually, or antimicrobial peptides would not continue to be under positive selection. At this point, the old allelic variant would possibly convey higher fitness, but it cannot increase in frequency again since it has been lost. Selection will then favor new derived alleles. Thus, because selection precludes saving all potentially useful sequence variants, either as allelic polymorphism or as duplicated loci, adaptive sequence divergence is enhanced.

Different combinations of brevinin-1 haplotypes were found in three *R. pipiens* individuals (Fig. 4.2). Some of the haplotypes we observed are likely to be allelic variants. In addition, there appears to be polymorphism in either gene copy number or at the primer binding sites, since there are at least five loci in RpO1 but only four haplotypes were found in RpBC1. The different suites of peptides among the three frogs are quite notable. For example, brevinin-1Pg from RpO1 resembles brevinin-1Sc from *R. sphenocephala*, but is too divergent to be allelic with any peptide found in

either RpO2 or RpBC1, unless extremely strong balancing selection or introgression is the cause (Fig 4.1; Fig. 4.2). Similarly, RpBC1 does not have any haplotype closely related to brevinin-1Ph or brevinin-1Pi (Fig. 4.1; Fig. 4.2). Although we sequenced more brevinin-1 clones from RpO1 than from RpO2 or RpBC1, this is unlikely to be the complete explanation for why we observed more haplotypes in RpO1, for the following reason. We excluded, as possible PCR artifacts, all haplotypes observed only once, as well as the only haplotype observed exactly twice (in RpO2). All of the excluded haplotypes in RpO2 and RpBC1 could be formed from a single point mutation and/or recombination between two multiply observed haplotypes in that individual. None of these haplotypes resembled the divergent haplotypes found in the other frog(s); for example, none of the excluded haplotypes in RpBC1 were anything like brevinin-1Pg, brevinin-1Ph, or brevinin-1Pi. Thus, even if we sequenced 110 clones each from RpO2 and RpBC1, we would be unlikely to observe our minimum criterion of 3 copies of anything seeming to be allelic with these missing haplotypes. Therefore, the variation in brevinin-1 haplotype numbers appears to be partly due to variation in gene copy number and/or polymorphism at the primer binding sites, at least between the Ontario and British Columbia populations. Mitochondrial DNA is distinct between *R. pipiens* from eastern and western North America, so geographic differences in brevinin-1 haplotypes are not unexpected (Hoffman and Blouin 2004).

It is unclear which of the nonsynonymous differences among brevinin-1 haplotypes represent allelic variants. Geographically distinct populations of *R. pipiens* express different combinations of brevinin-1 peptides, which could be due to nonsynonymous allelic variation (L. Rollins-Smith, personal communication). Our data suggest that brevinin-1Ph and brevinin-1Pi are especially likely to be allelic variants (Fig. 4.1, Fig. 4.2). If these two peptides and brevinin-1CHc are encoded by the same locus, then brevinin-1Pi is a derived allelic variant of brevinin-1Ph, at a high enough frequency that frog RpO2 was homozygous for it. It remains to be seen if the six-residue difference between brevinin-1Ph and brevinin-1Pi affects the fitness of frogs. Despite the strong signature of positive selection between loci, our data do not permit us to reject the hypothesis that all polymorphism within loci is neutral. If

individual antimicrobial peptide loci do not harbor much adaptive allelic variation, the alternate evolutionary strategy of frequent gene duplication appears to meet the need for a diversity of peptide variants.

4.5.4 *Synonymous Substitution*

Synonymous substitution in the mature peptide region is known to be enhanced relative to the signal sequence in frogs (Vanhoye et al. 2003; Tennessen 2005b). Enhanced synonymous divergence could be due to an increased mutation rate in the mature peptide region of the gene, or it could be that natural selection acts on synonymous sites, perhaps because of their effects on translation accuracy or mRNA folding. Our results do not allow us to evaluate these hypotheses explicitly, but it is interesting that no allelic variation was observed at *Ranantuerin2* (Fig. 4.3). Thus, if the mutation rate is indeed abnormally high, the selective sweep would have to have been quite recent, such that new mutations have not had time to accumulate. Alternatively, a normal mutation rate and a less recent selective sweep, coupled with selection on ‘‘silent’’ sites, would explain all observations.

Synonymous substitution might also be enhanced with respect to intron substitution, as is the case with flatfishes (Tennessen 2005b). The presence of an intron splitting the propiece is conserved among brevinins, ranatuerins, and the gaegurin-4 gene in *R. rugosa* (Kwon et al. 2000). Among the brevinin sequences examined, synonymous divergence in the mature peptide region is more than twice as high as overall divergence in the intron, but the difference is not significant. Thus, this study provides suggestive, but inconclusive, evidence that synonymous sites in the mature peptide region of frogs evolve faster than noncoding sites.

4.6 Conclusions

We have presented novel antimicrobial peptide gene sequences from leopard frogs, and shown that their evolution is characterized by frequent duplication and positive selection, which reduces allelic diversity. We found no evidence of balancing

selection. Our data support the trend of enhanced synonymous divergence in the mature peptide region. Thus, leopard frogs adapt to pathogens by endlessly revising their cocktail of antimicrobial peptides, resulting in a diverse but never perfect set of molecules within and among frogs. These evolutionary patterns imply that the peptides encoded by the genes discovered here might have unique functional properties, and therefore their synthesis and study would be useful.

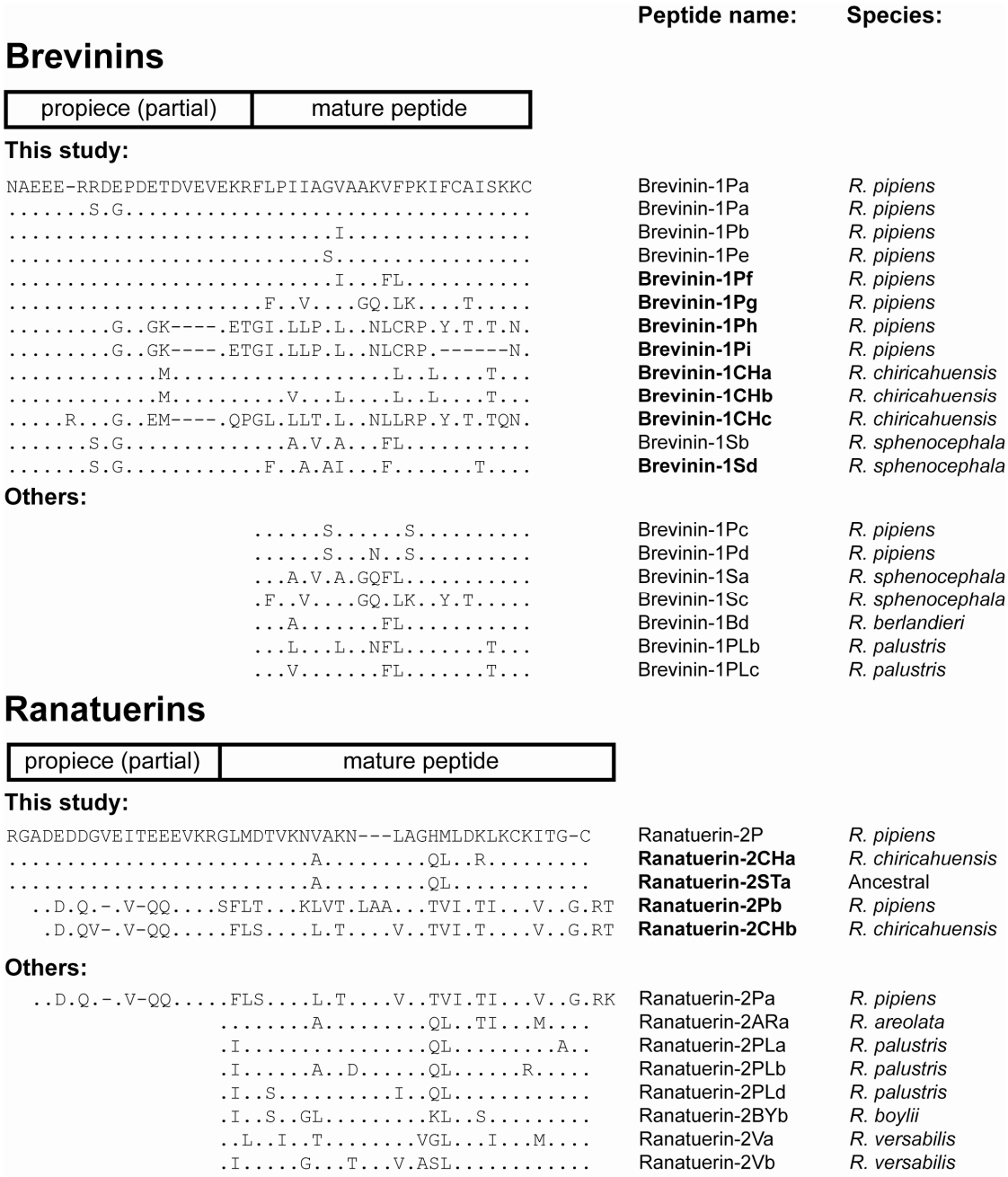

Figure 4.1

Amino acid sequences encoded by the antimicrobial peptide genes observed in this study, aligned with related peptides from the genus *Rana* (Conlon et al. 1999, 2003; Basir et al. 2000; Goraya et al. 2000; Ali et al. 2002; Chen et al. 2003, 2006). The names of previously undescribed mature peptides are presented in bold. A single exon encodes the C-terminal end of the propiece and the mature peptide region, and this is immediately preceded by an intron, which we also partially sequenced in most cases. Ranatuerin-2STa is a reconstructed extinct peptide from the most recent common ancestor of *R. pipiens* and *R. chiricahuensis*.

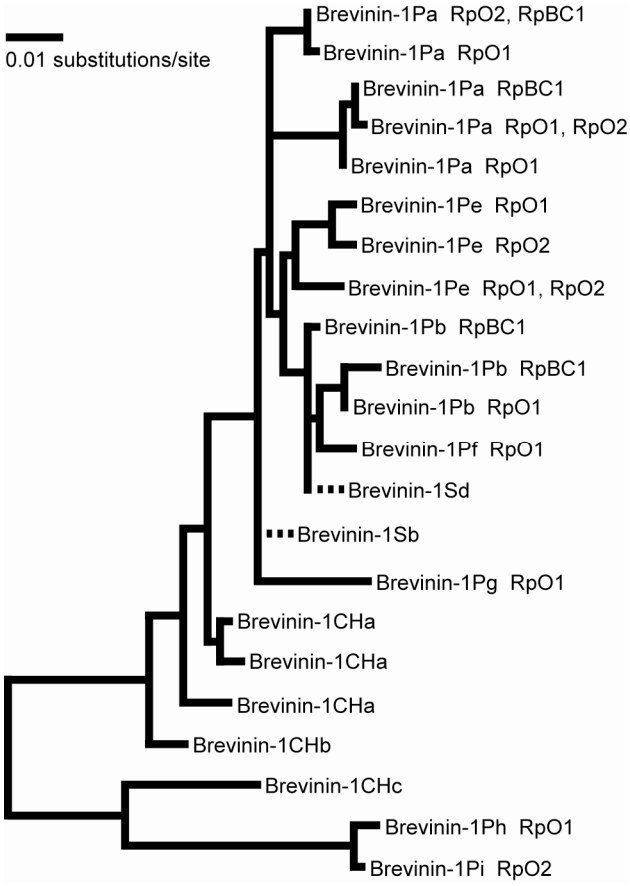

Figure 4.2

Maximum-likelihood phylogram of 22 brevinin-1 haplotypes obtained from *R. pipiens*, *R. chiricahuensis*, and *R. sphenocephala*, rooted by the midpoint. Haplotypes include part of an intron and an exon containing part of the propiece and the mature peptide region. Haplotypes are labeled according to the peptide that they encode. For *R. pipiens* haplotypes, the individual frog(s) from which the sequences were obtained are indicated (RpO1, RpO2, or RpBC1). Dotted branches leading to the *R. sphenocephala* haplotypes indicate that branch lengths are not proportional to the other branches, since only partial sequences were obtained. In the mature peptide region, dn/ds averages 2.6 along the branches of this phylogeny, indicating positive natural selection. Nine haplotypes obtained from a single frog (RpO1) indicate a minimum of five brevinin-1 loci.

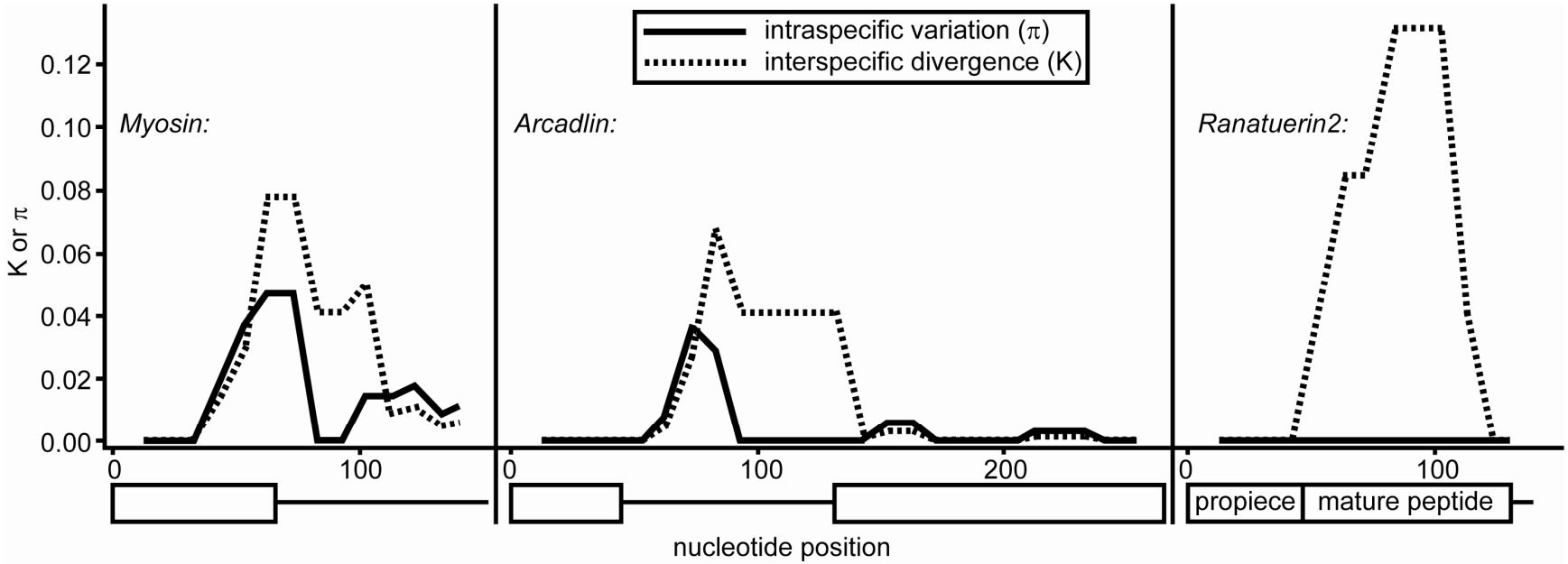

Figure 4.3

Variation and divergence at three leopard frog loci: *Myosin*, *Arcadlin* and *Ranatuerin2*. In *Rana pipiens*, intraspecific variation (π) is lowest at *Ranatuerin2*, where no polymorphism was observed, and higher at both *Myosin* and *Arcadlin*. However, divergence between *R. pipiens* and *R. chiricahuensis* (K) is higher at *Ranatuerin2* than at either *Myosin* or *Arcadlin*. Since the ratio of polymorphism to divergence should remain approximately equal throughout the genome under neutral evolution, this suggests a recent positive selective sweep on *Ranatuerin2*, if *Myosin* and *Arcadlin* are assumed to be evolving neutrally. The ratio of polymorphism to divergence is not significantly different between *Myosin* and *Arcadlin*. Window length = 25; step size = 10. White boxes beneath the graph indicate exons, and lines indicate noncoding sequence.

BALANCING SELECTION AT A FROG ANTIMICROBIAL PEPTIDE LOCUS: FLUCTUATING IMMUNE EFFECTOR ALLELES?

Jacob A. Tennessen and Michael S. Blouin

Molecular Biology and Evolution
Oxford Journals
2008. 25(12):2669-2680

CHAPTER 5

BALANCING SELECTION AT A FROG ANTIMICROBIAL PEPTIDE LOCUS: FLUCTUATING IMMUNE EFFECTOR ALLELES?

5.1 Abstract

Balancing selection is common on many defense genes, but it has rarely been reported for immune effector proteins such as antimicrobial peptides (AMPs). We describe genetic diversity at a brevinin-1 AMP locus in three species of leopard frogs (*Rana pipiens*, *Rana blairi*, and *Rana palustris*). Several highly divergent allelic lineages are segregating at this locus. That this unusual pattern results from balancing selection is demonstrated by multiple lines of evidence, including a ratio of nonsynonymous/synonymous polymorphism significantly higher than 1, the ZnS test, incongruence between the number of segregating sites and haplotype diversity, and significant Tajima's D values. Our data are more consistent with a model of fluctuating selection in which alleles change frequencies over time than with a model of stable balancing selection such as overdominance. Evidence for fluctuating selection includes skewed allele frequencies, low levels of synonymous variation, nonneutral values of Tajima's D within allelic lineages, an inverse relationship between the frequency of an allelic lineage and its degree of polymorphism, and divergent allele frequencies among populations. AMP loci could be important sites of adaptive genetic diversity, with consequences for host–pathogen coevolution and the ability of species to resist disease epidemics.

5.2 Introduction

Genes encoding immune system proteins often harbor adaptive variation maintained by balancing selection (Garrigan and Hedrick 2003). For example, patterns of genetic diversity consistent with balancing selection have frequently been observed at the vertebrate major histocompatibility complex (Garrigan and Hedrick 2003; Piertney and Oliver 2006), immunoglobulin genes (Su and Nei 1999), plant R-genes

(Stahl et al. 1999; Bakker et al. 2006), and other immunity genes (Bamshad et al. 2002; Jensen et al. 2008). Although extensive, these examples are not equally distributed among all classes of immunity genes. Nearly all immunogenetic adaptive variations have been found in detection and signaling proteins, not in the effector proteins that directly attack pathogens (Garrigan and Hedrick 2003; Lazzaro et al. 2004; Tiffin and Moeller 2006). The high polymorphism at many of these pathogen detection genes, such as loci of the vertebrate adaptive immune system and plant loci involved in gene-for-gene interactions, is associated with either high specificity or a cost of resistance (Garrigan and Hedrick 2003; Tian et al. 2003; Bakker et al. 2006). When immunity mechanisms are highly specific, pathogens can easily counteract them by altering the specific molecular target, so it benefits hosts to carry a diversity of immunity molecules which are effective against different variants of the molecular target (Stahl et al. 1999; Tiffin et al. 2004). Also, if alleles vary in overall effectiveness but there is a cost to resistance, both resistance and susceptibility alleles can be maintained at a locus (Tian et al. 2003). Some effector molecules show high structural and functional diversity across taxa, suggestive of specific immunological roles, and some could impose a cost by damaging host cells (Hancock 2001). Thus, we hypothesize that balancing selection might be relatively common at certain effector genes, especially those that show high interspecies divergence and/or are potentially costly. Testing this hypothesis would help to indicate the conditions under which balancing selection occurs and further illuminate host–pathogen coevolution at the molecular level.

Several mechanisms of balancing selection have been proposed (Charlesworth 2006). Under overdominant selection, heterozygotes are consistently fitter than homozygotes. Under instantaneous frequency-dependent selection or minority advantage, the fitness of an allele is directly and inversely proportional to its frequency (Takahata and Nei 1990). Instantaneous frequency-dependent selection can arise due to behavioral interactions among individuals in a single species, but it is not thought to occur in host–parasite interactions (Seger 1988). Under time-delayed frequency-dependent selection, including the ‘‘trench warfare’’ model of host–

pathogen coevolution, there is a lag between the change in host allele frequencies and the change in their fitnesses, owing to the need for the parasite to evolve first (Seger 1988; Stahl et al. 1999). Finally, under spatiotemporally varying selection, fitnesses of alleles vary over time and/or space due to variation in the presence or absence of pathogens which occurs independently of host allele frequencies (Hedrick 2002). These various mechanisms of balancing selection can be categorized according to whether or not allele frequencies change over time. Overdominance and instantaneous frequency-dependent selection are mathematically equivalent, and both are forms of stable balancing selection, predicting approximately constant intermediate allele frequencies over long periods of time (Takahata and Nei 1990). In contrast, the trench warfare model and spatiotemporally varying selection are forms of fluctuating selection, in which allele frequencies change frequency dynamically over time, becoming common when advantageous and rare when disadvantageous (Stahl et al. 1999; Tiffin et al. 2004).

Under stable balancing selection, allele frequencies should change more slowly over time than they would under neutral genetic drift, whereas under fluctuating selection, allele frequencies should change more quickly over time than they would under neutral genetic drift. Therefore, the two categories of balancing selection can be distinguished from each other by testing for evidence of substantial allele frequency change over time. When two or more allelic lineages are maintained by balancing selection, polymorphisms can be classified as either within-lineage variation or between-lineage variation (Innan and Tajima 1999). Under stable balancing selection, Tajima's D (Tajima 1989) for within-lineage variation should be close to 0. Under fluctuating selection, Tajima's D should be negative within lineages that have recently increased in frequency, which are analogous to growing populations, and positive for lineages that have recently decreased in frequency, which are analogous to populations going through a bottleneck. Similarly, stable balancing selection predicts highly similar allele frequencies among populations, high synonymous variation due to the antiquity of alleles at that locus, a correlation between the frequency of allelic lineages and the neutral variation they harbor, and a low probability of fixation for any allele.

In contrast, fluctuating selection permits divergent allele frequencies among populations and predicts low synonymous variation due to periodic bottlenecks for every allelic lineage, no correlation between the frequency of allelic lineages and the neutral variation they harbor, and the occasional fixation of particular alleles.

Antimicrobial peptides (AMPs) of the innate immune system have only rarely been observed to be under pathogen-driven balancing selection (Tennessen 2005a). These cationic, amphipathic mature peptides are cleaved off of a larger protein and then bind to the cell membranes of bacterial, fungal, and enveloped viral pathogens, killing them (Hancock 2001). Research on AMPs has been substantial in recent years, due in part to an interest in developing them for therapeutic application (Hancock 2001). Positive selection on AMP genes is very common and has resulted in an enormous functional diversity of these molecules among species and among loci in many taxa (Tennessen 2005a). Some human and mussel AMP loci appear to be under balancing selection (Hollox and Armour 2008; Pahdi and Verghese 2008), but other studies of intraspecies genetic diversity at AMP genes have revealed no evidence for balancing selection (Clark and Wang 1997; Lazzaro and Clark 2003; Tennessen and Blouin 2007).

The AMPs of leopard frogs (genus *Rana*; in this paper, we ignore the recent proposal by Frost et al. [2006] to revise the genus to *Lithobates*) are among the most well studied (Conlon et al. 2004). Most of them consist of an α-helix with a disulfide bridge forming a loop at the C-terminal end. They are functionally diverse, frequently with activity against both amphibian and human pathogens (Goraya et al. 2000; Chinchar et al. 2004; Rollins-Smith and Conlon 2005). Given the global crisis of amphibian population declines mediated by emerging infectious diseases and the immunological importance of AMPs, genetic diversity at AMP loci could be an important determinant of amphibian population stability (Daszak et al. 2003; Woodhams et al. 2006a,b). Of the four AMP families secreted by the northern leopard frog, *Rana pipiens* (brevinin-1, ranatuerin-2, temporin-1, and esculentin-2), the brevinin-1 family is both the most diverse and the most active against microbes (Goraya et al. 2000; Tennessen and Blouin 2007). Previously, we investigated allelic

variation at the *Ranatuerin2* AMP locus in *R. pipiens* and found that a single haplotype had been fixed in the species by a positive selective sweep (Tennessen and Blouin 2007). We also found substantial diversity among AMP sequences at five loci of the brevinin-1 family in *R. pipiens*, but we were unable to assess how much of this variation was within versus between loci (Tennessen and Blouin 2007). In this paper, we test whether some of the brevinin-1 variants are allelic by designing primers that amplify a single brevinin-1 locus. We examine patterns of genetic diversity in over 400 individuals of *R. pipiens*. In order to assess the generality of our observations across species, we also sequence this locus in a smaller number of plains leopard frogs (*Rana blairi*) and pickerel frogs (*Rana palustris*). Several highly divergent allelic lineages are segregating at this locus, and we present evidence that the balanced maintenance of these alleles is due to a dynamic process of fluctuating natural selection.

5.3 Materials and Methods

5.3.1 *Tissues*

Samples of *R. pipiens* were collected from 13 sites throughout its range across northern North America as described previously (n = 20–46 individuals per site; Appendix Table 1; Fig. 5.1; Hoffman et al. 2006). For one of the sites in New York, we also obtained a historical sample of *R. pipiens* collected in 1971 (approximately 15 generations earlier) as described previously (n = 25; Hoffman et al. 2006). We collected samples of *R. palustris* from two sites in Wisconsin and Michigan (n = 23 and 40, respectively) and samples of *R. blairi* from a single site in Illinois (n = 27; Appendix Table 1; Fig. 5.1). Further samples of *R. blairi* were obtained from the Museum of Vertebrate Zoology (MVZ Herp 240161–240184), having all been originally collected from a single site in Missouri (n = 24; Appendix Table 1; Fig. 5.1). The sample size at every site was at least 20 frogs (mean number of frogs per population = 28.8). All tissues consisted of toe clips preserved by desiccation in 1.5-ml tubes filled with Drierite desiccant (W. A. Hammond Drierite Co., Xenia, OH),

except for the MVZ samples, which had been frozen. DNA was extracted as described previously (Hoffman and Blouin 2004) or using DNeasy Blood and Tissue Kits (Qiagen Inc., Valencia, CA).

We carried out 25 μl polymerase chain reactions (PCRs) using standard buffer conditions, 1.5 mM $MgCl_2$, 0.2 mM each deoxyribonucleotide triphosphate, approximately 100 ng DNA, and 0.5 Units Taq DNA polymerase. We visualized PCR products under ultraviolet light, purified them with the MoBio Ultraclean PCR cleanup kit (Solana Beach, CA), and sequenced them through the Nevada Genomics Center (Reno, NV).

5.3.2 *Sequences*

Previously, we had cloned two brevinin-1 sequences from a single *R. pipiens* individual that were similar in the intron but quite divergent in the mature peptide region (GenBank accession numbers DQ276967 and DQ276968; Tennessen and Blouin 2007). In this study, we tested whether they were allelic by assessing whether the genotype frequencies were in Hardy–Weinberg equilibrium within populations. To do so, we designed a primer (Brev1PF4; 5′-GAT GAC CCA ATA ATA ATT TTT C-3′) that would bind to these sequences but not to any other known brevinin-1 gene in *R. pipiens*. We used the primers Brev1PF4 and Brev1PR1 (Tennessen and Blouin 2007), which bind outside of the coding region, to amplify genomic DNA from *R. pipiens*, *R. palustris*, and *R. blairi*. PCR amplification conditions consisted of an initial denaturation step at 94 °C for 5 min; 35 cycles of denaturation at 94 °C for 45 s, annealing at 51 °C for 30 s, and extension at 72 °C for 1 min; and a final single extension step at 72 °C for 5 min. We sequenced the resultant PCR products with an internal primer (Brev1PF5; 5′-GAA AGC TCT GTG CCA TAG-3′).

Due to difficulties amplifying some *R. blairi* individuals with this primer pair, we designed a forward primer specific to *R. blairi*. We sequenced the primer-binding site in *R. blairi* using a cloning procedure described previously (Tennessen and Blouin 2007). Briefly, we amplified *R. blairi* genomic DNA using the more distal and degenerate forward primer Brev1PF3 with Brev1PR1 in PCRs with the high fidelity

enzyme Pfu DNA polymerase (Promega, Madison, WI). The resultant PCR products were incubated with Taq polymerase in order to add 3′ adenines and cloned into *Escherichia coli* using the TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA). The recombinant locus was amplified from screened colonies using the primers T3 and T7, and the resultant PCR products were sequenced with T7. We used these sequences to design the primer Brev1BLF (5′-TAG ATG ACC TAA TAA TAA TTT TTC-3′), which binds in the intron. We amplified and sequenced DNA from all *R. blairi* individuals using both primer set (Brev1PF4 and Brev1PR1) and primer set (Brev1BLF and Brev1PR1).

In order to compare patterns of genetic diversity at the brevinin-1 locus with other, putatively neutral, nuclear loci, we sequenced 1,246 bp of four nuclear genes in a subset of individuals: *Arcadlin* (primers ArcadF and ArcadR, 178 bp coding, 87 bp intron; Tennessen and Blouin 2007), *Myosin* (primers MyosinF and MyosinR, 65 bp coding, 85 bp intron; Tennessen and Blouin 2007), *FIBI7* (primers FIBI7U and FIBI7L, 50 bp coding, 208 bp intron; Di Candia and Routman 2007), and *Tyrosinase* (primers Tyr1A and Tyr1G, 573 bp coding; Bossuyt and Milinkovitch 2000). These loci were chosen because we had preexisting protocols for amplifying them in *Rana* and because their polymorphisms are consistent with neutral evolution. We sequenced *Arcadlin*, *Myosin*, *FIBI7*, and *Tyrosinase* in all frogs from six populations: populations 1, 2, 14, 15, 16, and 17 (Fig. 5.1). These six populations represented all *R. palustris* and *R. blairi* individuals but only a fraction of all *R. pipiens* individuals (59 frogs). In order to obtain a more representative sample of *R. pipiens*, we randomly chose two frogs from each of the remaining 11 contemporary *R. pipiens* populations and we sequenced *Arcadlin*, *Myosin*, *FIBI7*, and *Tyrosinase* in these 22 frogs.

5.3.3 *Analysis*

All sequences have been deposited in GenBank under accession numbers EU407141–EU407149, EU407151–EU407176, and EU769510–EU769553. All loci were nuclear and diploid, so two alleles were obtained from all individuals. Heterozygotes were identified by the presence of double peaks in electropherograms.

To determine common haplotypes, we noted all homozygous genotypes and tested for recombination using the four-gamete test. Given evidence for minimal recombination, we determined the remaining haplotypes by subtracting out previously observed haplotypes from the heterozygous genotypes or, for some *R. blairi* genotypes, by using allele-specific PCR. To confirm that observed polymorphisms were allelic and not due to coamplification of duplicated loci, we tested whether each population was in Hardy–Weinberg equilibrium. It is highly unlikely that polymorphic sequences from multiple loci would conform to Hardy–Weinberg expectations.

We used PAUP* (version 4.0b10, Swofford 2002) to construct a maximum parsimony phylogeny of all brevinin-1 sequences. We used DnaSP (version 4.0, Rozas et al. 2003) to calculate standard statistics of selective neutrality, including ZnS (Kelly 1997), Tajima's D (Tajima 1989), haplotype diversity or expected heterozygosity (Hd), the number of haplotypes (h), genetic variation (π), and the number of segregating sites (S), and to test their significance using coalescent simulations (conditioned on S with no recombination and 1,000 replicates). To minimize the effects of population subdivision on these test statistics, we performed each test within individual populations. We used MEGA (version 2.1; Kumar et al. 2001) to test whether the ratio of nonsynonymous variation to synonymous variation ($\pi n/\pi s$) was significantly greater than 1. If observed πs was 0, we estimated $\pi n/\pi s$ to be greater than what the value would be if a single synonymous polymorphism had been observed in a single sequence. Neutrality tests were conducted on the mature peptide region only because that is the region where nonneutral evolution most frequently occurs (Tennessen 2005a). We represented relationships among alleles with a haplotype network, as opposed to a phylogenetic tree, because networks show allele frequencies and because there is little evidence for recombination at this locus.

To distinguish between fluctuating and stable balancing selection, we performed several tests of allelic stasis. A model of stable balancing selection predicts that the silent variation at the selected locus should be higher than elsewhere in the genome because allelic lineages have been maintained at intermediate frequency for a very long time and many mutations have been able to accumulate over this long

coalescence period (Charlesworth 2006). In contrast, under fluctuating selection each allelic lineage spends some time being rare when selection favors the other allelic lineage. Thus, the effective population size of the locus, estimated as the sum of the harmonic means of the effective sizes of all allelic lineages over time, is small. As a result, silent variation will be purged by genetic drift, making silent variation lower than silent variation at neutral loci (Tiffin et al. 2004). Thus, we calculated genetic variation (π) at silent sites (synonymous and noncoding) for *Arcadlin*, *Myosin*, *FIBI7*, *Tyrosinase*, and the brevinin-1 locus. We used equation 12.63 of Nei and Kumar (2000) to calculate the variance in π, and we used the square root of this variance as the standard error in t-tests comparing mean π values. Similarly, stable balancing selection predicts within-lineage Tajima's D values near 0, while fluctuating selection predicts high or low Tajima's D values for lineages that have been shrinking or growing, respectively. Thus, we calculated Tajima's D within each allelic lineage.

Our five easternmost populations of *R. pipiens* are the ''Eastern'' populations used by Hoffman et al. (2006) to generate a neutral distribution of the expected F_{ST} value, based on microsatellite markers and intersimple sequence repeat loci, using the method of Beaumont and Nichols (1996). Under stable balancing selection, F_{ST} would be lower than the neutral expectations. Under fluctuating selection, F_{ST} could be higher than the neutral expectation if selection coefficients varied substantially among populations or if the effective population size at the selected locus is substantially lower than at neutral loci. Therefore, we compared F_{ST} at the brevinin-1 locus with this distribution. Furthermore, we used the previously described genotype data from the *R. pipiens* individuals at nine microsatellite loci (Hoffman et al. 2006) to construct phylogenies of the *R. pipiens* populations east of the Mississippi River based on both the microsatellites and the brevinin-1 locus. We used the Fitch method based on the genetic distance of Reynolds et al. (1983) in PHYLIP (version 3.65; Felsenstein 1989). We only used the eastern populations because the high genetic divergence between eastern and western *R. pipiens* could result in substantial homoplasy at the microsatellite markers (Hoffman and Blouin 2004; Hoffman et al. 2006). Although microsatellites have a higher mutation rate than coding sequences, among closely

related populations genetic distances based on both marker types should be correlated under neutrality (Richard and Thorpe 2001), and therefore differences between the brevinin-1 phylogeny and the microsatellite phylogeny could be due to selection. To formally evaluate the difference between interpopulation divergences at the brevinin-1 locus and at the microsatellite loci, we performed a Mantel test of pairwise genetic distances in FSTAT version 2.9.3 (Goudet 1995).

Populations with low genetic diversity at the brevinin-1 locus could have recently experienced a selective sweep, or they could simply have low genome-wide variation due to a low effective population size. To distinguish between these two hypotheses, we calculated the mean expected heterozygosity (He) for each *R. pipiens* population using the previously described genotypes at the nine microsatellite loci (Hoffman et al. 2006). This analysis was restricted to *R. pipiens* because we lacked microsatellite data for the other species. We compared population He values at the brevinin-1 locus with the population He values at the microsatellite loci, which we assumed to be selectively neutral. If populations with low brevinin-1 diversity also have low microsatellite diversity, genetic drift is probably responsible in both cases. If not, the low brevinin-1 diversity could be due to a recent or ongoing selective sweep.

5.4 Results

5.4.1 *Brevinin-1 Locus*

We consistently obtained readable brevinin-1 sequences 236 bp in length, which consisted of a 98-bp partial intron followed by a 138-bp partial exon. The exon contained 61 bp of the C-terminal end of the propiece, the entire 72-bp mature peptide region, a stop codon, and 2 bp of postcoding sequence. Out of 517 frogs, 344 were homozygous. All common haplotypes (seen more than five times) were observed as homozygotes and showed no evidence of recombination. All remaining haplotypes (1% of the total) could easily be resolved, either because there was only one way to resolve the heterozygous genotypes such that at least one allele in every genotype matched an allele previously observed in that population or by using allele-specific

PCR. It remains possible that some of these nonhomozygous haplotypes were erroneously resolved, which could bias our results, but they are so rare that they are unlikely to affect our conclusions substantially. All populations were found to be in Hardy–Weinberg equilibrium, and there was no evidence of more than two alleles in any one individual, confirming that only a single locus was amplified, here named the *Brevinin1.1* locus.

Within *R. pipiens*, four common, highly divergent alleles were observed at the *Brevinin1.1* locus: alleles *Rp1*, *Rp2*, *Rp3*, and *Rp4* (Fig. 5.2). These four alleles accounted for 97% of the sequences; all remaining minor variants were one or two steps away from one of them (Fig. 5.3). Allele *Rp1* encoded the previously described peptide brevinin-1Pa, and allele *Rp2* encoded the previously described peptide brevinin-1Pg (Goraya et al. 2000; Tennessen and Blouin 2007). Alleles *Rp3* and *Rp4* both encoded a peptide, brevinin-1PLa, which previously had only been described in *R. palustris* (Basir et al. 2000). The distances among these four major *R. pipiens* alleles ranged from 1 to 13 substitutions (Fig. 5.3). The species overall showed an excess of homozygotes, probably owing to population subdivision, because every individual population was in Hardy–Weinberg equilibrium (Appendix Table 1) and because this species shows substantial population structure at neutral markers (Hoffman and Blouin 2004; Hoffman et al. 2006). We observed similar patterns of diversity in *R. blairi*. A single *R. blairi* sample, MVZ Herp 240161, was homozygous for unique alleles at both *Brevinin1.1* and *Arcadlin* and carried a unique allele at *Myosin*. Because this individual could be mislabeled or a migrant, we excluded it from all further analyses. Allele *Rb1* accounted for 86% of the *R. blairi* sequences at the *Brevinin1.1* locus (Figs. 5.2 and 5.3). The remaining *R. blairi* alleles differed from allele *Rb1* by 1–13 substitutions, including a single-nucleotide indel in the intron. In contrast to the other species, in *R. palustris*, we observed only two alleles which were separated by a single mutational step in the intron. Both coded for the peptide brevinin-1PLa (Figs. 5.2 and 5.3). The more common allele (allele *Rp4*) was identical to an allele observed in *R. pipiens*.

Haplotype networks of the *Brevinin1.1* locus alleles are shown in Fig. 5.3. There were three main allelic lineages, separated by substantial and mostly nonsynonymous divergence. There was no reciprocal monophyly between *R. pipiens* and either of the other two species, but *R. palustris* and *R. blairi* were reciprocally monophyletic with respect to each other. Although there were no shared *Brevinin1.1* alleles between *R. pipiens* and *R. blairi*, both species had alleles belonging to the same divergent allelic lineages, Lineage 1 and Lineage 2 (Table 5.1 and Fig. 5.3).

5.4.2 *Other Loci*

For all individuals, preexisting protocols for amplifying and sequencing *Arcadlin*, *Myosin*, *FIBI7*, and *Tyrosinase* (Bossuyt and Milinkovitch 2000; Di Candia and Routman 2007; Tennessen and Blouin 2007) were successful. In total, we obtained 1,246 bp of nuclear sequence unlinked to *Brevinin1.1*, including 866 bp of coding sequence and 380 bp of intronic sequence. These loci showed moderate levels of variation (Fig. 5.4 and Table 5.2; Appendix Table 2). In *R. pipiens* and *R. blairi*, patterns of variation did not deviate from neutral expectations. In *R. palustris*, several neutrality tests indicated an excess of divergent, intermediate frequency haplotypes. Because these patterns were observed across all loci, they probably have demographic causes such as population subdivision or recent migration from genetically distinct populations and are not due to selection. There were no fixed differences between *R. blairi* and *R. pipiens*. There were eight fixed differences between *R. palustris* and *R. blairi*; of these, four were also fixed between *R. palustris* and *R. pipiens*.

5.4.3 *Neutrality Tests*

To test whether the high allelic variation at the *Brevinin1.1* locus in *R. pipiens* and *R. blairi* was maintained by balancing selection, we conducted multiple tests of selective neutrality (Table 5.1; Appendix Table 1). Neutrality tests were not conducted on *R. palustris* because that species had no variation in the mature peptide region. The ratio of mature peptide region nonsynonymous variation to synonymous variation ($\pi n/\pi s$) was significantly greater than 1 in most populations as well as in *R. pipiens*

overall and *R. blairi* overall (Table 5.1; Appendix Table 1; Fig. 5.5). For both species, nonsynonymous variation in the mature peptide region was also significantly higher than synonymous variation at *Arcadlin*, *Myosin*, *FIBI7*, and *Tyrosinase* (Fig. 5.5). Coalescent simulations of several population genetic parameters in DnaSP demonstrated that polymorphisms were structured into divergent allelic lineages, with skewed frequencies, more so than would be expected under neutrality given the number of segregating sites (Table 5.1; Appendix Table 1; Fig. 5.3). Given the number of segregating sites, the number of haplotypes (h) was too low for all populations that had any variation and the haplotype diversity (Hd) was too low for nearly all populations. Linkage disequilibrium (ZnS) was significantly higher than neutral expectations for all populations that had any variation. In *R. pipiens*, the presence of several common, divergent alleles made pairwise variation (π) significantly higher than expected given the number of segregating sites, resulting in significantly positive values of Tajima's D. In contrast, Tajima's D was negative in *R. blairi*, owing to rare but divergent alleles.

Multiple tests rejected neutrality in both *R. pipiens* and *R. blairi*, providing strong evidence for balancing selection. In *R. pipiens*, neutrality was also rejected if the sequences from any one allelic lineage were removed (significant ZnS and h values, $\pi n/\pi s > 1$). This result suggested that all three allelic lineages are adaptively divergent *R. pipiens*. Similarly, in *R. blairi*, if Lineage 2 sequences were removed, multiple tests still rejected neutrality (significant ZnS and h values, $\pi n/\pi s > 1$). This result suggested that alleles *Rb1* and *Rb2*, while both in Lineage 1, might be adaptively divergent.

5.4.4 *Distinguishing Stable from Fluctuating Selection*

The first indication about the type of balancing selection acting was that allele frequencies were not intermediate but skewed. In *R. pipiens*, Lineage 1 alleles were 10 times more common than Lineage 3 alleles. In *R. blairi*, allele frequencies were even more skewed, with 86% of the sequences consisting of the same allele, and other alleles being quite rare. In *R. palustris*, only a single allelic lineage was observed. The

other allelic lineages could be so rare they were not sampled in *R. palustris* or they might not exist in this species. No species showed a pattern of equally frequent allelic lineages consistent with symmetrical models of stable balancing selection. The observed skew was more consistent with fluctuating selection or highly asymmetrical overdominance.

To formally distinguish between stable balancing selection and fluctuating selection, we compared silent variation at four non-AMP nuclear loci (*Arcadlin*, *Myosin*, *FIBI7*, and *Tyrosinase*) with silent variation at the *Brevinin1.1* locus (Fig. 5.5). Silent variation at the *Brevinin1.1* locus was lower than at all the non-AMP loci for both *R. pipiens* (*Brevinin1.1* $\pi s = 0.005$; $p = 0.08$) and *R. palustris* (*Brevinin1.1* $\pi s = 0.002$; $p = 0.02$). Silent variation was extremely low at all loci for *R. blairi* (*Brevinin1.1* $\pi s = 0.002$; $p > 0.1$). These results were consistent with fluctuating selection, which predicts low synonymous diversity, but not with stable balancing selection, which predicts high synonymous diversity. Although demographic processes appeared to have caused nonneutral patterns at the non-AMP loci *R. palustris*, these processes could not explain the difference in genetic diversity between *Brevinin1.1* and the other loci, which therefore is more likely to be due to selection.

We calculated values of Tajima's D within allelic lineages (Table 5.3). In *R. pipiens*, Tajima's D was significantly negative within the common Lineage 1 ($D = -1.84$; $p < 0.01$). Because we did not sequence non-AMP loci in all *R. pipiens* individuals, we cannot completely understand possible demographic factors that could affect Tajima's D, so this result should be interpreted with caution. In both *R. pipiens* and *R. blairi*, there was a nonsignificant trend of other common allelic lineages having negative Tajima's D values and rare allelic lineages having positive Tajima's D values consistent with fluctuating selection (Table 5.3). Because population subdivision can affect Tajima's D, we also calculated within-lineage Tajima's D values for individual populations of *R. pipiens* and *R. blairi*. For all populations, including those shown to have neutral Tajima's D values at non-AMP loci (populations 1, 2, 14, and 15; Appendix Table 2), the trend was the same as for the whole species (Table 5.3). Only

one allelic lineage with a single polymorphism occurred in *R. palustris*, so within-lineage Tajima's D could not be meaningfully estimated.

Under asymmetrical stable balancing selection, such as asymmetrical overdominance, the more favored allelic lineage will be more common and harbor most of the neutral genetic variation, owing to its larger effective size (Innan and Tajima 1999; Stahl et al. 1999). We therefore compared levels of silent variation among allelic lineages. Within *R. pipiens*, values of silent π were 0.0004 ± 0.0011 for Lineage 1, 0.0002 ± 0.0008 for Lineage 2, and 0.0038 ± 0.0037 for Lineage 3. Within *R. blairi*, values of silent π were 0.0018 ± 0.0018 for Lineage 1 and 0.0053 ± 0.0042 for Lineage 2. Thus, for both species, genetic variation was relatively low in the more common allelic lineages and much higher in the rarest allelic lineages.

We tested for excessive differentiation among populations, a signature of fluctuating selection. The F_{ST} value among the five easternmost populations of *R. pipiens* was 0.20, and the expected neutral value of F_{ST} at the same level of heterozygosity was approximately 0.06 (Fig. 2A in Hoffman et al. [2006]). Thus, the observed F_{ST} value at the *Brevinin1.1* locus was higher than the mean neutral expectation but not significantly so. The high F_{ST} value is primarily due to population 2, where allele *Rp1* predominates, in contrast to the other four populations, where allele *Rp2* predominates (Appendix Table 1). In addition, F_{ST} between populations 1 and 2 was 0.02 at the four non-AMP loci sequenced in this study and 0.04 at the microsatellites, but 0.34 at *Brevinin1.1*. The high divergence of population 2 at *Brevinin1.1* but not at microsatellites is shown in Figure 5.6. The matrix of pairwise microsatellite differences among eastern populations was not correlated with the corresponding matrix of *Brevinin1.1* pairwise differences (Mantel test; $p > 0.1$). Therefore, the high genetic divergence between population 2 and the other eastern populations is inconsistent with neutral expectations and could be due to spatially differing selective pressures.

The two *R. pipiens* populations with the lowest microsatellite He values were fixed for the allele *Rp1*, in contrast to the high genetic variation seen in other populations (Appendix Table 1). This result was consistent with the hypothesis that

these two populations had lower effective sizes than the others, in which case genetic drift would be more likely to fix alleles in these two populations. After these two fixed populations, the *R. pipiens* population with the lowest *Brevinin1.1* Hd was population 2 (Appendix Table 1), which had slightly higher than average microsatellite He (0.84), and therefore its effective population size did not appear to be substantially lower than the other populations. Because population 2 had a very different allele frequency pattern than its closest geographic neighboring populations (Fig. 5.6), and because it showed an unusually low *Brevinin1.1* Hd value for its microsatellite He value, population 2 was the population most likely to have undergone a recent shift in allele frequencies at the *Brevinin1.1* locus due to selection. However, allele frequencies were nearly identical between the 1971 sample and the 2001 sample, both at the *Brevinin1.1* locus (F_{ST} = 0.0; Appendix Table 1) and at the nine microsatellite loci (F_{ST} = 0.0).

5.5 Discussion

Patterns of genetic diversity at the *Brevinin1.1* locus are highly unusual in three species of leopard frogs, and multiple lines of evidence suggest that strong balancing selection is maintaining adaptive variation in the mature peptide region. Evidence for nonneutrality includes high ratios of nonsynonymous/synonymous variation, unusual values of Hd, h, and ZnS given the number of segregating sites, and significant values of Tajima's D (Table 5.1; Appendix Table 1). Because we did not sequence non-AMP loci from all *R. pipiens* individuals, we cannot rule out the possibility that demographic factors have contributed to some nonneutral patterns; however, in populations 1 and 2, neutrality is strongly rejected the *Brevinin1.1* locus but not at other loci (Appendix Table 1; Appendix Table 2), suggesting that our results are not an artifact of demography. Likewise, we cannot rule out the possibility that the high variation at the *Brevinin1.1* locus arose not merely through point mutations but also through gene conversion from other loci, which can contribute to a pattern of strikingly divergent alleles (Storz et al. 2007). Several Lineage 1 alleles, including allele *Rp1*, are identical in the mature peptide region to other brevinin-1 loci that exist

in leopard frog genomes (Tennessen and Blouin 2007). Lineage 2 and Lineage 3 sequences have not been observed at other brevinin-1 loci (Tennessen and Blouin 2007). However, even if demographic processes or gene conversion have enhanced the diversity at this locus, they must have acted in concert with selection as this is the only explanation for some nonneutral patterns such as an elevated $\pi n/\pi s$ ratio. The role played by these peptides in amphibian immunity indicates that pathogens are the most likely selective pressure acting on this locus. Perhaps this locus is specialized on a particular pathogen that is coevolving with leopard frogs, and alleles differ in their ability to defend against different strains. Alternatively, perhaps some alleles are more effective than others overall but there is a cost to resistance.

Stable balancing selection includes both overdominance and instantaneous frequency-dependent selection, while fluctuating selection can be caused by time-delayed frequency-dependent selection (Seger 1988) or spatiotemporally varying selection (Hedrick 2002). Although we cannot determine with absolute certainty which type of balancing selection is acting, our data support the fluctuating selection hypothesis more strongly than the stable balancing selection hypothesis. First, allele frequencies are skewed instead of intermediate in all three species. Second, Tajima's D for within-lineage variation is negative for common lineages and positive for rare lineages in both *R. blairi* and *R. pipiens* (Table 5.3). This result suggests that now common alleles were recently rare and have rapidly increased in frequency. Third, synonymous variation is lower at the *Brevinin1.1* locus than at unlinked, putatively neutral loci, the opposite of the prediction from stable balancing selection; for *R. palustris*, this difference is statistically significant, consistent with a recent selective sweep fixing Lineage 3 in this species (Fig. 5.5). This effect holds despite evidence that amphibian mature peptide regions feature an elevated mutation rate and/or positive selection on synonymous sites, which would tend to cause the opposite pattern (Tennessen 2005b). Fourth, the allelic lineages with the most variation are not the most common, again suggesting that allele frequencies were quite different in the recent past. Finally, genetic distance at *Brevinin1.1* is exceptionally high among some

R. pipiens populations, and stable balancing selection would predict a trend in the opposite direction (Fig. 5.6).

Our results provide unique insight into the evolution of immunity genes. Nearly all examples of pathogen-induced balancing selection are detection and signaling proteins, such as those of the vertebrate adaptive immune system (Garrigan and Hedrick 2003). The *Brevinin1.1* locus represents a rare example of an effector gene harboring an adaptive polymorphism. Loci encoding AMPs in mussels and humans also show high variation consistent with balancing selection (Hollox and Armour 2008; Pahdi and Verghese 2008). *Drosophila* AMPs are not under balancing selection (Clark and Wang 1997; Lazzaro and Clark 2003), possibly because fly AMPs, which also show little evidence for positive selection, may have less specific targets than AMPs in other taxa (Sackton et al. 2007). The AMP locus *Ranatuerin2* has undergone a selective sweep in *R. pipiens* and shows no evidence of balancing selection (Tennessen and Blouin 2007), but this pattern is also consistent with fluctuating selection if the other alleles are extremely rare or have recently been eliminated. Fluctuating selection in particular has only occasionally been convincingly demonstrated for any immunity locus, effector or otherwise. Many of the best examples of fluctuating selection on defense genes are from plants (Stahl et al. 1999; Tiffin et al. 2004; Tiffin and Moeller 2006), but a few animal examples also exist (Jensen et al. 2008). Even for well-studied loci under balancing selection like the genes of the major histocompatibility complex, the importance of stable versus fluctuating selection is not clear (Piertney and Oliver 2006); the universally observed high allelic diversity at these loci suggests that if fluctuating selection occurs, the fluctuations are nearly instantaneous and, therefore, closer to stable balancing selection than they are at the *Brevinin1.1* locus. Microbes can evolve resistance to detection mechanisms by simply substituting one or more amino acid residues on their surface proteins. In contrast, microbial resistance to AMPs involves changing the biochemistry of the cell membrane or producing AMP-degrading enzymes, adaptations which might involve several coordinated changes across multiple genes. Thus, we hypothesize that it is more difficult for microbes to evolve resistance to

AMPs than to detection mechanisms, resulting in longer lag periods between host evolution and pathogen evolution and therefore greater fluctuations in allele frequencies, including occasional allele fixation. Fluctuating selection may be difficult to distinguish from positive selection if divergent, low-frequency alleles are not sampled, which may be partly why documented examples of balancing selection on effector molecules are rare. Overall, our results help to illustrate when and how balancing selection can act on effector loci.

Allele frequency differences among populations are either due to genetic drift or different selective pressures in different habitats. In *R. pipiens*, allele frequencies differ substantially between populations east and west of the Mississippi River (Appendix Table 1), but this differentiation is consistent with genetic drift, given the high east–west mitochondrial DNA divergence also observed in this species (Hoffman and Blouin 2004). The high differentiation among the eastern populations at *Brevinin1.1*, despite low differentiation at neutral markers, is more likely to be due to selection (Fig. 5.6). Population 2 is the most likely candidate for a recent population-specific shift in allele frequencies because it is quite divergent from other populations nearby and because its haplotype diversity is low despite having normal levels of neutral genetic diversity. Allele frequencies in this population have not changed noticeably between 1971 and 2001, so if there is a difference in selective pressures between population 2 and its neighbors, it is more than a few decades old.

Variation at AMP loci is likely to be an important contributor to the ability of amphibian species to adapt to novel disease threats. Agents of emerging infectious diseases are causing precipitous declines, and in some cases extinction, in many amphibian species (Carey et al. 1999; Daszak et al. 2003). It is thought that AMPs are an important defense against these pathogens (Chinchar et al. 2004; Rollins-Smith and Conlon 2005; Woodhams et al. 2006a,b). Several epizootics of the fungal pathogen *Batrachochytrium dendrobatidis* and iridovirus have caused declines of *R. pipiens* populations in recent decades (Carey et al. 1999; Green et al. 2002; Greer et al. 2005). Although these epizootics could be selective agents on the *Brevinin1.1* locus, we did not observe any historic allele frequency change in population 2, and allelic lineages

are too divergent to have arisen within a few decades. Therefore, nonneutral patterns of genetic diversity are probably primarily due to long-term allelic fluctuations caused by unidentified native diseases, with recent disease outbreaks possibly shaping this variation.

Our study adds to the growing body of knowledge on diversifying selection among AMPs by demonstrating that AMPs encoded by alleles at the same locus can show evidence of major adaptive differences. We have shown that genes encoding immune effector proteins can harbor balanced polymorphisms comparable to many pathogen-detection genes. Researchers developing AMPs for therapeutic applications (Hancock 2001) should consider examining multiple individuals of the same species to encounter functionally novel allelic variants. Our data are consistent with a model of fluctuating selection, likely caused by pathogens. However, because the results of many of our tests were merely suggestive, we are reluctant to completely reject the hypothesis of stable balancing selection. Further study is required to determine the specific functional difference among these peptides, the particular pathogens that are driving selection, and the precise mechanism maintaining variation. We are currently examining the antimicrobial effects of these peptides in vitro (Tennessen JA, Woodhams DC, Reinert LK, Blouin MS, and Rollins-Smith LA, unpublished data). Loci encoding AMPs, such as *Brevinin1.1*, can be important sites of adaptive genetic diversity and major players in the coevolutionary arms race between hosts and pathogens.

Table 5.1

Population genetic parameters for the *Brevinin1.1* mature peptide region (72 bp) in 3 frog species. For all statistical tests, * = $p < 0.05$ and ** = $p < 0.01$.

Species	**2N[a]**	**Lineage 1[b]**	**Lineage 2[b]**	**Lineage 3[b]**	**S[c]**	**h[d]**	**Hd[e]**	**π[f]**	**π_s[g]**	**ZnS[h]**	**π_n/π_s[i]**	**Tajima's D[j]**
R. pipiens	848	60%	34%	6%	16	8*	0.55	0.06*	0.009	0.32*	8.67**	2.23*
R. blairi	100	92%	8%	0%	12	4**	0.24**	0.02	0.001	0.48*	25.00**	-0.81
R. palustris	86	0%	0%	100%	0	1	0.00	0.00	0.000	0.00	-	-

[a]2N = number of alleles = twice the number of individuals

[b]Lineages 1, 2, and 3 refer to the percentage of alleles in each species belonging to each of these 3 allelic lineages

[c]S = number of segregating sites

[d]h = number of haplotypes; tested whether h is significantly low given S

[e]Hd = haplotype diversity, equivalent to expected heterozygosity; tested whether Hd is significantly low given S

[f]π = mean number of pairwise differences among sequences; tested whether π is significantly extreme given S

[g]π_s = mean number of pairwise differences among sequences at silent sites; tested whether π_s is significantly extreme given silent S

[h]ZnS = linkage disequilibrium statistic of Kelly (1997); tested whether ZnS is significantly high given S

[i]π_n/π_s = ratio of nonsynonymous nucleotide variation to synonymous nucleotide variation; tested whether significantly different than 1

[j]Tajima's D = statistic of Tajima (1989); tested whether significantly different than 0.

Table 5.2

Population genetic parameters for 4 putatively neutral nuclear loci unlinked to *Brevinin1.1* (1246 bp) in 3 frog species (*Arcadlin*, *Myosin*, *FIBI7*, and *Tyrosinase*). For all statistical tests, * = $p < 0.05$ and ** = $p < 0.01$.

Species	**2N**[a]	**Locus**	**Length**[b]	**S**[c]	**h**[d]	**Hd**[e]	**π**[f]	**π_s**[g]	**ZnS**[h]	**π_n/π_s**[i]	**Tajima's D**[j]
R. pipiens	162	*Arcadlin*	265	6	9	0.75	0.01	0.010	0.06	0.04**	0.54
		Myosin	150	7	10	0.76	0.01	0.020	0.12	0.00*	1.27
		FIBI7	258	12	9	0.57	0.01	0.010	0.24	0.00	0.12
		Tyrosinase	573	17	17	0.84	0.00	0.016	0.06	0.08*	-0.41
		Mean ±	312 ±	10.5 ±	11.3 ±	0.73 ±	0.01 ±	0.014 ±	0.12 ±	0.03**	0.38 ±
		St. Dev.	182	5.1	3.9	0.11	0.00	0.005	0.08	-	0.71
R. blairi	100	*Arcadlin*	265	2	3	0.46	0.00	0.004	0.02	0.00*	0.35
		Myosin	150	2	3	0.50	0.00	0.006	0.04	0.00	0.66
		FIBI7	258	2	3	0.10	0.00	0.000	0.00	0.00	-1.15
		Tyrosinase	573	3	4	0.56	0.00	0.000	0.02	>8.7	0.07
		Mean ±	312 ±	2.3 ±	3.3 ±	0.40 ±	0.00 ±	0.002 ±	0.02 ±	0.14	-0.02 ±
		St. Dev.	182	0.5	0.5	0.21	0.00	0.003	0.02	-	0.79
R. palustris	86	*Arcadlin*	265	4	4	0.56	0.00	0.009	0.33	0.00**	1.09
		Myosin	150	4	2*	0.34	0.01	0.010	1.00*	0.64	1.39
		FIBI7	258	14	7	0.72	0.02**	0.024**	0.37*	0.00*	2.53**
		Tyrosinase	573	7	5	0.75	0.00	0.011	0.38	0.07**	0.56

Table 5.2 (Continued)

Mean ±	312 ±	7.3 ±	4.5 ±	0.59 ±	0.01 ±	0.014 ±	0.52 ±	0.13**	1.39 ±
St. Dev.	182	4.7	2.1**	0.19	0.01**	0.007**	0.32**	-	0.83**

[a]2N = number of alleles = twice the number of individuals

[b]Length = locus length in base pairs

[c]S = total number of segregating sites among the 4 loci

[d]h = mean number of haplotypes per locus; tested whether h is significantly low given S

[e]Hd = mean haplotype diversity per locus, equivalent to expected heterozygosity; tested whether Hd is significantly low given S

[f]π = mean number of pairwise differences among sequences; tested whether π is significantly extreme given S

[g]π_s = mean number of pairwise differences among sequences at silent sites; tested whether π_s is significantly extreme given silent S

[h]ZnS = mean linkage disequilibrium statistic of Kelly (1997) per locus; tested whether ZnS is significantly high given S

[i]π_n/π_s = ratio of nonsynonymous nucleotide variation to synonymous nucleotide variation; tested whether significantly different than 1

[j]Tajima's D = statistic of Tajima (1989); tested whether significantly different than 0.

Table 5.3

Values of Tajima's D (Tajima 1989) within allelic lineages within species. Only lineages displaying variation for the species considered are shown. In both *R. pipiens* and *R. blairi*, common lineages have negative Tajima's D values and rare lineages have positive Tajima's D values, consistent with fluctuating selection. This trend is statistically significant for Lineage 1 in *R. pipiens*. Only a single polymorphism occurs in *R. palustris*, so Tajima's D cannot be meaningfully estimated.

Species	**Lineage**	**D[a]**	**frequency[b]**	**N[c]**	**population D (mean)[d]**	**population D (range)[e]**
R. pipiens	1	-1.84**	60%	7	-1.11	-1.69 to -0.53
	2	-0.79	34%	3	-0.95	-1.09 to -0.84
	3	0.62	6%	1	1.45	1.45
R. blairi	1	-1.26	87%	1	-0.42	-0.42
	2	0.16	8%	0	-	-

[a]D = within-lineage Tajima's D; tested whether significantly different than 0; ** = $p < 0.01$

[b]frequency = frequency of each allelic lineage in each species

[c]N = number of populations in each species displaying sufficient variation in that allelic lineage such that within-lineage Tajima's D could be calculated

[d]population D (mean) = mean within-lineage Tajima's D value among individual populations

[e]population D (range) = range of within-lineage Tajima's D values among individual populations

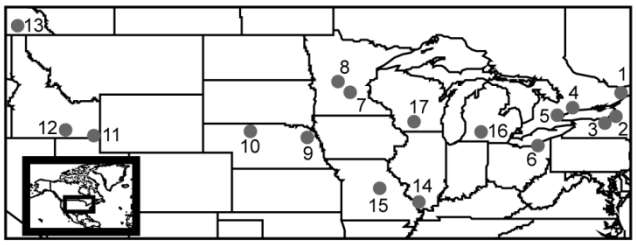

Figure 5.1

Map of sampling localities in the United States and Canada. Populations 1–13 are *Rana pipiens*, populations 14 and 15 are *Rana blairi*, and populations 16 and 17 are *Rana palustris*. Exact population locations are in Appendix Table 1.

propiece (partial)	mature peptide	Peptide name	Frequency	Alleles	Lineage
***Rana pipiens*:**					
AEEERRDEPDETDVEVEKRFLPIIAGVAAKVFPKIFCAISKKC		Brevinin-1Pa	58.6%	*Rp1*	1
.....F..V....GQ.LK....T.....		Brevinin-1Pg	34.1%	*Rp2*	2
.....F.NV.S.PGQ.LK..........		Brevinin-1PLa	6.0%	*Rp3*, *Rp4*	3
.....I.....		Brevinin-1Pb	0.5%		1
.....F.NV.S.PGQ.LR..........		Brevinin-1Pj	0.4%		3
.....T.....		Brevinin-1Pk	0.4%		1
.....S.....		Brevinin-1Pe	0.1%		1
***Rana blairi*:**					
.....I...FL.....T.....		Brevinin-1BLa	87.0%	*Rb1*	1
T.....M...LL..........		Brevinin-1BLb	5.0%	*Rb2*	1
.....F..V..M.GQ.LK....T.....		Brevinin-1BLc	6.0%	*Rb3*	2
.....S.G.....F..V....GQ.LK..Y.T.....		Brevinin-1BLd	2.0%		2
***Rana palustris*:**					
.....F.NV.S.PGQ.LK..........		Brevinin-1PLa	100.0%	*Rp4*	3

Figure 5.2

Alignment of *Brevinin1.1* locus amino acid sequences from all species, the frequencies of these sequences in each species, the alleles encoding them (some rare alleles were not named), and the allelic lineage.

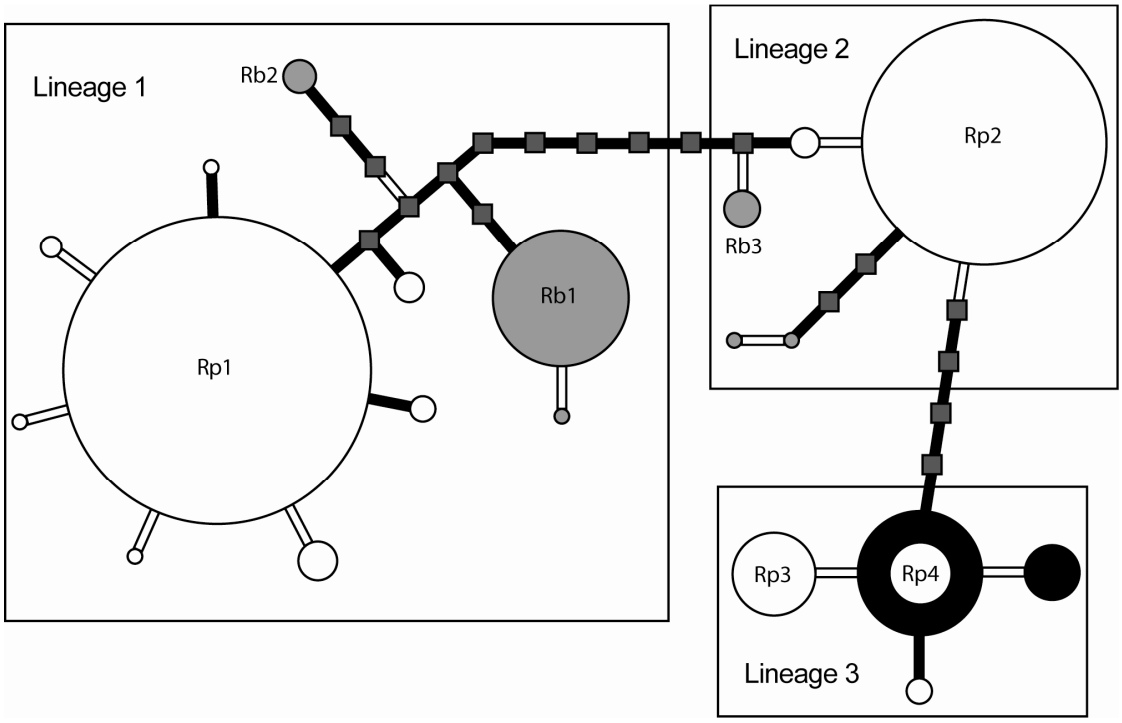

Figure 5.3

Maximum parsimony haplotype network for a 236-bp segment of the *Brevinin1.1* locus. White circles represent *Rana pipiens* haplotypes, gray circles represent *Rana blairi* haplotypes, and black circles represent *Rana palustris* haplotypes. Gray squares represent inferred nodes that were not observed. The size of the circles indicates the frequency of each allele in the respective species. Nonsynonymous substitutions are indicated by small black rectangular bars, and synonymous substitutions are indicated by small white rectangular bars; thus, branch lengths represent the total number of substitutions. Common alleles are labeled. Allelic lineages are surrounded by large rectangles. Highly divergent alleles, separated mostly by nonsynonymous substitutions, occur in both *R. pipiens* and *R. blairi*, causing several tests of selective neutrality to be significant (Table 5.1). A haplotype is shared between *R. pipiens* and *R. palustris*, indicated by a white circle superimposed on a black circle. A pattern of both very common and very rare alleles can be seen, especially for the Lineage 1 lineage in *R. pipiens*, suggestive of a recent rapid increase in allele frequency.

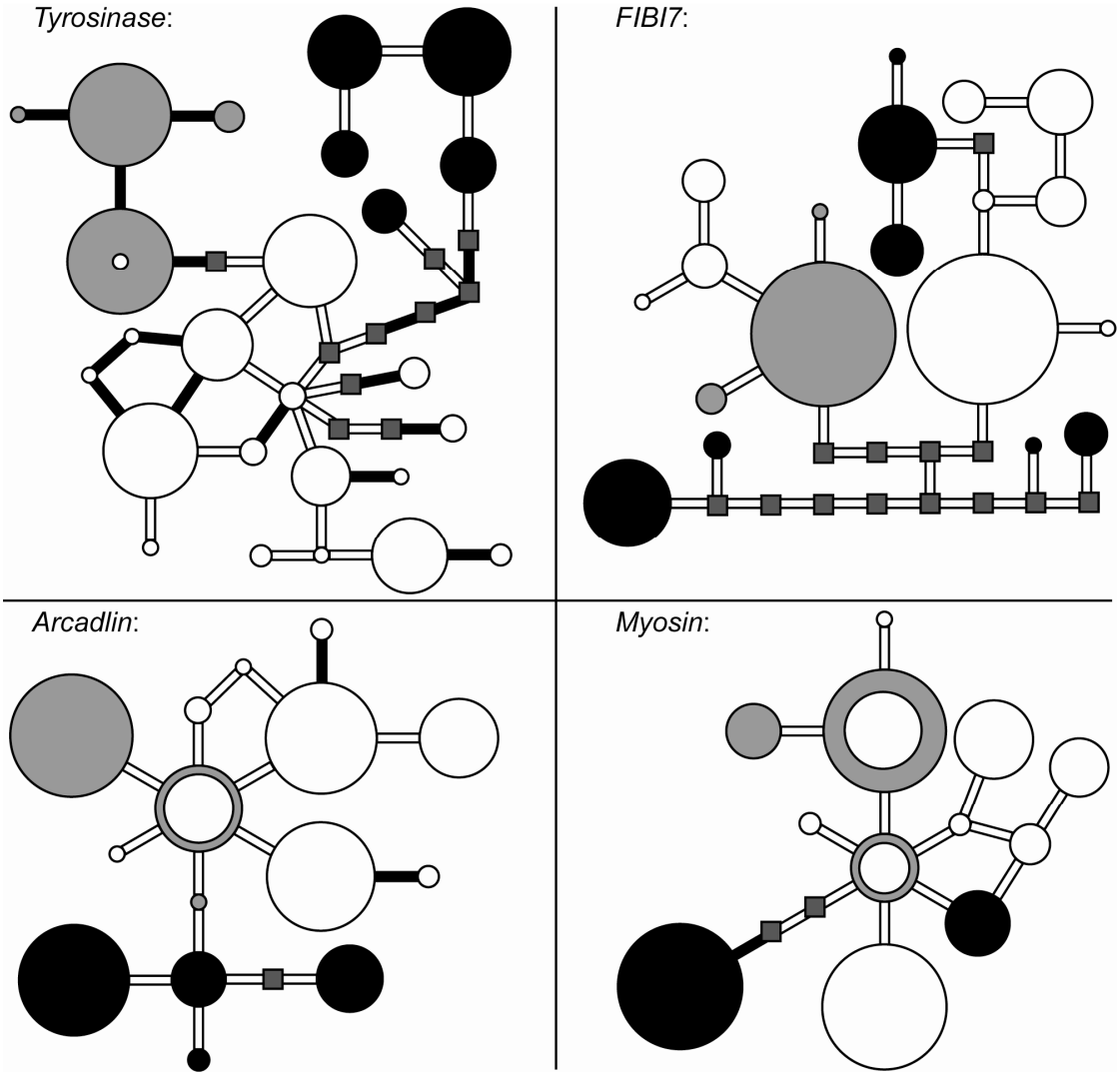

Figure 5.4

Maximum parsimony haplotype networks for four putatively neutral non-AMP nuclear loci unlinked to *Brevinin1.1*: *Arcadlin*, *Myosin*, *FIBI7*, and *Tyrosinase*. Symbols are the same as in Fig. 5.3. In contrast with the *Brevinin1.1* locus (Fig. 5.3), nonsynonymous substitutions (small black rectangular bars) are rarer than synonymous substitutions (small white rectangular bars). In both *Rana pipiens* (white circles) and *Rana blairi* (gray circles), non-AMP loci show intermediate frequency alleles and few missing transitional haplotypes (gray squares), in contrast with the *Brevinin1.1* locus (Fig. 5.3), which shows high- and low-frequency alleles forming divergent lineages separated by many missing transitional haplotypes. In *Rana palustris* (black circles), non-AMP loci show diverse, moderately divergent alleles, in contrast with the *Brevinin1.1* locus (Fig. 5.3), which shows very low diversity.

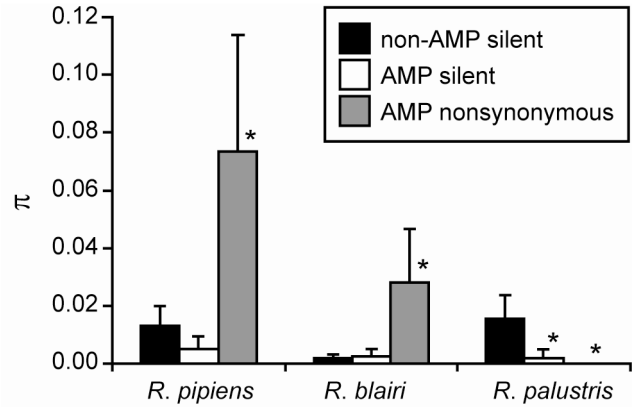

Figure 5.5

Genetic diversity (π) at the *Brevinin1.1* locus compared with neutral π for three frog species. ‘‘Non-AMP silent’’ is synonymous and noncoding π at four putatively neutral nuclear loci: *Arcadlin*, *Myosin*, *FIBI7*, and *Tyrosinase*. ‘‘AMP silent’’ is synonymous and noncoding π at *Brevinin1.1*. ‘‘AMP nonsynonymous’’ is nonsynonymous π in the mature peptide region of *Brevinin1.1*. Asterisks represent values that are significantly different from the corresponding non-AMP silent value ($p < 0.05$). For *Rana pipiens* and *Rana palustris*, *Brevinin1.1* silent sites show lower variation than is seen at non-AMP silent sites. For *R. pipiens* and *Rana blairi*, *Brevinin1.1* shows higher nonsynonymous variation than is seen at silent sites; *R. palustris* has no nonsynonymous variation at *Brevinin1.1*.

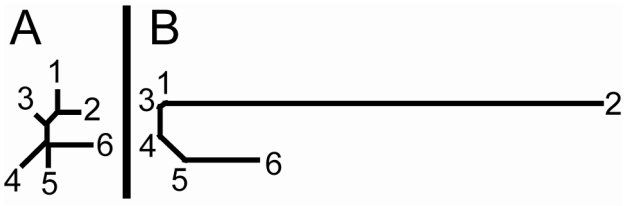

Figure 5.6

Distance-based phylogenies of contemporary *Rana pipiens* populations east of the Mississippi River, populations 1–6 (Fig. 5.1; Appendix Table 1). (A) Phylogeny based on genetic distance at nine microsatellite loci. All populations show intermediate divergence from each other. (B) Phylogeny based on the *Brevinin1.1* locus. All populations show intermediate divergence from each other except for population 2, which has highly different allele frequencies. Very little branching is apparent because almost all alleles at the *Brevinin1.1* locus in these populations are either allele *Rp1* or allele *Rp2*, so populations are effectively differentiated from each other in a single dimension (i.e., frequency of allele *Rp1*). Because the high divergence between population 2 and the other populations is only apparent at *Brevinin1.1*, not at the microsatellites, it is likely adaptive.

CHAPTER 6
CONCLUSIONS

I have presented a detailed examination of AMP evolution in frogs. This study in molecular natural history bridges the gap between molecular ecology, which typically employs neutral markers to estimate general evolutionary parameters in non-model species, and functional genetics, which typically examines phenotypically relevant genetic variants in model species. A thorough understanding of adaptive molecular biodiversity requires project like this one, which investigates variants that matter phenotypically at loci which are interesting in and of themselves in wild species.

Positive selection is clearly more common at AMP loci than at most other loci. As reviewed in Chapter 2, this trend was identified in individual AMP families by many researchers before me. However, Chapter 2 is the first published synthesis of these taxonomically disparate studies and recognition of this pattern as a general feature of AMP evolution. In addition, with population genetic data I have identified more subtle and recent examples of positive selection that would not have been detected with the dn/ds ratio, such as the selective sweep at the *Ranatuerin2* locus (Chapter 4); these results suggest that positive selection may be even more prevalent than is indicated by sequence comparisons. A fundamental goal in biology is to determine the major differences between species and the genetic and evolutionary basis for this differentiation. Due to positive selection, the cryptic differences at AMP loci are greater and more ecologically important than the differences at most other loci, on par with or surpassing any recognizable phenotypic trait.

In addition to widespread positive selection, I have uncovered evidence of other evolutionary trends at AMP loci. Gene duplication is common and greatly contributes to AMP diversity (Chapter 4). Balancing selection has rarely been described at AMP loci, but my results imply that AMP loci might be important sites of intraspecies variation in disease resistance (Chapter 5). Evidence for fluctuating selection in particular is rare at all types of loci, not just AMPs, and the results of

Chapter 5 will serve as an important example of probable fluctuating selection in the wild. The unexpected trend of selection on silent sites is also suggested by my results, with numerous consequences since the assumption of neutrality at these sites underlies a wide variety of evolutionary genetic analyses (Chapter 3). Coordinated evolution, on the other hand, may not be as common at AMP loci as previously hypothesized (Chapter 3).

The link between emerging infectious diseases and amphibian population declines has been well documented (Carey et al. 1999; Daszak et al. 2003). A better understanding of the genetics basis of disease resistance in amphibians could allow managers to predict the outcomes of disease outbreaks and potential intervention strategies. My research suggests that AMP genetic diversity could affect the ability of species to adapt to novel diseases. Populations and species may vary in immunity based on differences at AMP loci. Patterns of positive selective sweeps and fluctuating selection suggest that potentially useful variants are periodically lost when one allele is favored for a substantially long time, as can happen during an epizootic or if the natural microbial flora are diminished, and this could reduce the ability of a species to adapt to a subsequent disease threat. The study of how variation at disease resistance genes affects the long-term survival of a species, and the application of this knowledge to the conservation of biodiversity, can be called conservation immunogenetics. Proposals to manage for immunogenetic diversity are not new (Hughes 1991), but this approach may finally become practical as more genomic data in wild species are generated.

Due in part to rapid microbial evolution of resistance to conventional antibiotics, there is interest in developing AMPs for therapeutic applications in medicine or agriculture (Hancock 2001). Many technical hurdles, some perhaps insurmountable, must be overcome before AMPs or analogous molecules could be put to practical use. Nevertheless, my research suggests several strategies for the pursuit of this goal. First, the strong signature of positive selection indicates that microbes frequently evolve resistance to AMPs in nature, so AMPs shouldn't be thought of as an unbeatable weapon. Second, high functional divergence can potentially be found

between even closely related AMPs, including allelic AMPs from the same species, so the functional properties of one peptide do not necessarily represent the other peptides in the same family. Finally, even if AMPs are not used directly, understanding the coevolutionary dynamics between pathogens and AMPs in nature could lead to better predictive models of antibiotic resistance evolution.

My research has laid the groundwork for fruitful future research on AMPs. Major questions that remain to be addressed include: how do factors such as gene duplication, or host taxonomy, ecology, or effective population size, correlate with the probability and strength of positive selection on AMP loci? Which pathogen species are coevolving with AMP loci and through what genetic mechanisms? How do positively selected substitutions affect antimicrobial activity? The use of AMPs as molecular models for evolutionary studies can provide fruitful research for years to come.

BIBLIOGRAPHY

Aguilar A, Roemer G, Debenham S, Binns M, Garcelon D, Wayne RK (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. Proc Natl Acad Sci USA 101:3490–3494

Ali MF, Lips KR, Knoop FC, Fritzsch B, Miller C, Conlon JM (2002) Antimicrobial peptides and protease inhibitors in the skin secretions of the crawfish frog, *Rana areolata*. Biochim Biophys Acta 1601:55–63

Altizer S, Harvell D, Friedle E (2003) Rapid evolutionary dynamics and disease threats to biodiversity. Trends in Ecology and Evolution 18:589–596

Alvarez-Valin F, Jabbari K, Bernardi G (1998) Synonymous and nonsynonymous substitutions in mammalian genes: intragenic correlations. J Mol Evol 46:37–44

Amiche M, Ducancel F, Mor A, Boulain JC, Menez A, Nicolas P (1994) Precursors of vertebrate peptide antibiotics dermaseptin b and adenoregulin have extensive sequence identities with precursors of opioid peptides dermorphin, dermenkephalin, and deltorphins. J Biol Chem 269:17847–17852

Amiche M, Seon AA, Wroblewski H, Nicolas P (2000) Isolation of dermatoxin from frog skin, an antibacterial peptide encoded by a novel member of the dermaseptin genes family. Eur J Biochem 267:4583–4592

Andreu D, Rivas L (1998) Animal antimicrobial peptides: an overview. Biopolymers 47:415–433

Anisimova M, Bielawski JP, Yang Z (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol 18:1585–1592

Anisimova M, Bielawski JP, Yang Z (2002) Accuracy and power of Bayes prediction of amino acid sites under positive selection. Mol Biol Evol 19:950–958

Antcheva N, Boniotto M, Zelezetsky I, Pacor S, Verga Falzacappa MV, Crovella S, Tossi A (2004) Effects of positively selected sequence variations in humans and *Macaca fascicularis* β-defensins 2 on antimicrobial activity. Antimicrob Agents Chemother 48:685–688

Bakker EG, Toomajian C, Kreitman M, Bergelson J (2006) A genome-wide survey of R gene polymorphisms in Arabidopsis. Plant Cell 18:1803–1818

Bamshad MJ, Mummidi S, Gonzalez E, et al. (11 co-authors) (2002) A strong signature of balancing selection in the 5′ cis-regulatory region of *CCR5*. Proc Natl Acad Sci USA 99:10539–10544

Basir YJ, Knoop FC, Dulka J, Conlon JM (2000) Multiple antimicrobial peptides and peptides related to bradykinin and neuromedin N isolated from skin secretions of the pickerel frog, *Rana palustris*. Biochim Biophys Acta 1543:95–105

Beaumont MA, Nichols RA, (1996) Evaluating loci for use in genetic analysis of population structure. Proc R Soc Lond B Biol Sci 263:1619–1626

Bergelson J, Kreitman M, Stahl EA, Tian D (2001) Evolutionary dynamics of plant R-genes. Science 292:2281–2285

Bernardi G, Mouchiroud D, Gautier C (1993) Silent substitutions in mammalian genomes and their evolutionary implications. J Mol Evol 37:583–589

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J Evol Biol 16:363–377

Boniotto M, Antcheva N, Zelezetsky I, Tossi A, Palumbo V, Verga Falzacappa MV, Sgubin S, Braida L, Amoroso A, Crovella S (2003a) A study of host defense peptide β-defensin 3 in primates. Biochem J 374:707–714

Boniotto M, Tossi A, Del Pero M, Sgubin S, Antcheva N, Santon D, Masters J, Crovella S (2003b) Evolution of the beta defensin 2 gene in primates. Genes Immun 4:251–257

Bossuyt F, Milinkovitch MC (2000) Convergent adaptive radiations in Madagascan and Asian ranid frogs reveal covariation between larval and adult traits. Proc Natl Acad Sci USA 97:6585–6590

Bowman HG (2003) Antibacterial peptides: basic facts and emerging concepts. J Intern Med 254:197–215

Braida L, Boniotto M, Pontillo A, Tovo PA, Amoroso A, Crovella S (2004) Single-nucleotide polymorphism in the human beta-defensin 1 gene is associated with HIV-1 infection in Italian children. AIDS 18:1598–1600

Bulmer MS, Crozier RH (2004) Duplication and diversifying selection among termite antifungal peptides. Mol Biol Evol 21:2256–2264

Carey C, Cohen N, Rollins-Smith L (1999) Amphibian declines: an immunological perspective. Dev Comp Immunol 23:459–472

Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet 2:379–384

Chen T, Farragher S, Bjourson AJ, Orr DF, Rao P, Shaw C (2003a) Granular gland transcriptomes in stimulated amphibian skin secretions. Biochem J 371:125–130

Chen T, Tang L, Shaw C (2003b) Identification of three novel *Phyllomedusa sauvagei* dermaseptins (sVI–sVIII) by cloning from a skin secretion-derived cDNA library. Regul Pept 116:139–146

Chen T, Zhou M, Rao P, Walker B, Shaw C (2006) The Chinese bamboo leaf odorous frog (*Rana* (*Odorrana*) *versabilis*) and North American *Rana* frogs share the same families of skin antimicrobial peptides. Peptides 27:1738–1744

Chinchar VG, Bryan L, Silphadaung U, Noga E, Wade D, Rollins-Smith L (2004) Inactivation of viruses infecting ectothermic animals by amphibian and piscine antimicrobial peptides. Virology 323:268–275

Clark AG, Wang L (1997) Molecular population genetics of *Drosophila* immune system genes. Genetics 147:713–724

Cole AM, Weis P, Diamond G (1997) Isolation and characterization of pleurocidin, an antimicrobial peptide in the skin secretions of winter flounder. J Biol Chem 272:12008–12013

Conlon JM, Halverson T, Dulka J, Platz JE, Knoop FC (1999) Peptides with antimicrobial activity of the brevinin-1 family isolated from skin secretions of the southern leopard frog, *Rana sphenocephala*. J Peptide Res 54:522–527

Conlon JM, Sonnevend Á, Patel M, Davidson C, Nielsen PF, Pál T, Rollins-Smith LA (2003) Isolation of peptides of the brevinin-1 family with potent candidacidal activity from the skin secretions of the frog *Rana boylii*. J Peptide Res 62:207–213

Conlon JM, Kolodziejek J, Nowotny N (2004) Antimicrobial peptides from ranid frogs: taxonomic and phylogenetic markers and a potential source of new therapeutic agents. Biochim Biophys Acta 1696:1–14

Conticello SG, Gilad Y, Avidan N, Ben-Asher E, Levy Z, Fainzilber M (2001) Mechansims for evolving hypervariability: the case of conopeptides. Mol Biol Evol 18:120–131

Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF (2002) PCR-mediated recombination in amplification products derived from polyploidy cotton. Theor Appl Genet 104:482–489

Crump D, Lean D, Trudeau VL (2002) Octylphenol and UV-B radiation alter larval development and hypothalamic gene expression in the leopard frog (*Rana pipiens*). Environ Health Perspect 110:277–284

Daszak P, Cunningham AA, Hyatt AD (2000) Emerging infectious diseases of wildlife – threats to biodiversity and human health. Science 287:443–449

Date A, Satta Y, Takahata N, Chigusa SI (1998) Evolutionary history and mechanism of the *Drosophila Cecropin* gene family. Immunogenetics 47:417–429

Date-Ito A, Kasahara K, Sawai H, Chigusa SI (2002) Rapid evolution of the male-specific antibacterial protein *Andropin* gene in *Drosophila*. J Mol Evol 54:665–670

Del Pero M, Boniotto M, Zuccon D, Cervella P, Spanò A, Amoroso A, Crovella S (2002) β-defensin 1 gene variability among non-human primates. Immunogenetics 53:907–913

Di Candia MR, Routman EJ (2007) Cytonuclear discordance across a leopard frog contact zone. Mol Phylogenet Evol 45:564–575

Dörk T, Stuhrmann M (1998) Polymorphisms of the human β-defensin-1 gene. Mol Cell Probes 12:171–173

Douglas SE, Gallant JW, Gong Z, Hew C (2001) Cloning and developmental expression of a family of pleurocidin-like antimicrobial peptides from winter flounder, *Pleuronectes americanus* (Walbaum). Dev Comp Immunol 25:137–147

Douglas SE, Patrzykat A, Pytyck J, Gallant JW (2003) Identification, structure and differential expression of novel pleurocidins clustered on the genome of the winter flounder, *Pseudopleuronectes americanus* (Walbaum). Eur J Biochem 270:3720–3730

Duda TF Jr, Vanhoye D, Nicolas P (2002) Roles of diversifying selection and coordinated evolution in the evolution of amphibian antimicrobial peptides. Mol Biol Evol 19:858–864

DuMont VB, Fay JC, Calabrese PP, Aquadro CF (2004) DNA variability and divergence at the Notch locus in *Drosophila melanogaster* and *D. simulans*: a case of accelerated synonymous site divergence. Genetics 167:171–185

Fay JC, Wu C-I (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413

Felsenstein J (1985) Phylogenies and the comparative method. Am Nat 125:1–15

Felsenstein J (1989) PHYLIP—phylogeny inference package (version 3.2). Cladistics 5:164–166

Felsenstein J (2004) PHYLIP: (Phylogeny Inference Package), version 3.61. Department of Genetics, University of Washington, Seattle

Fernandez RC, Weiss AA (1996) Susceptibilities of *Bordetella pertussis* strains to antimicrobial peptides. Antimicrob Agents Chemother 40:1041–1043

Ford MJ (2002) Applications of selective neutrality tests to molecular ecology. Mol Ecol 11:1245–1262

Frost DR, Grant T, Faivovich J, et al. (19 co-authors) (2006) The amphibian tree of life. Bull Am Mus Nat Hist 297:1–370

Froy O, Gurevitz M (2003) Arthropod and mollusk defensins—evolution by exon-shuffling. Trends Genet 19:684–687

Fu Y, Li W-H (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Garrigan D, Hedrick PW (2001) Class I MHC polymorphism and evolution in endangered California Chinook and other Pacific salmon. Immunogenetics 53:483–489

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

Goraya J, Wang Y, Li Z, O'Flaherty M, Knoop FC, Platz JE, Conlon JM (2000) Peptides with antimicrobial activity from four different families isolated from the skins of the North American frogs *Rana luteiventris*, *Rana berlandieri*, and *Rana pipiens*. Eur J Biochem 267:894–900

Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. J Hered 86:485–486

Green DE, Converse KA, Schrader AK (2002) Epizootiology of sixty-four amphibian morbidity and mortality events in the USA, 1996–2001. Ann N Y Acad Sci 969:323–339

Greer AL, Berrill M, Wilson PJ (2005) Five amphibian mortality events associated with ranavirus infection in south central Ontario, Canada. Dis Aquat Org 67:9–14

Hancock REW (2001) Cationic peptides: effectors in innate immunity and novel antimicrobials. Lancet Infect Dis 1:156–164

Hedengren M, Borge K, Hultmark D (2000) Expression and evolution of the *Drosophila Attacin/Diptericin* gene family. Biochem Biophys Res Comm 279:574–581

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

Hill RE, Hastie ND (1987) Accelerated evolution in the reactive centre regions of serine protease inhibitors. Nature 326:96–99

Hillis DM, Wilcox TP (2005) Phylogeny of the New World true frogs (*Rana*). Mol Phylogenet and Evol 34:299–314

Hoffman EA, Blouin MS (2004) Evolutionary history of the northern leopard frog: reconstruction of phylogeny, phylogeography, and historical changes in population demography from mitochondrial DNA. Evolution 58:145–159

Hoffman EA, Schueler FW, Jones AG, Blouin MS (2006) An analysis of selection on a colour polymorphism in the northern leopard frog. Mol Ecol 15:2627–2641

Hollox EJ, Armour JAL, Barber JCK (2003) Extensive normal copy number variation of a β-defensin antimicrobial gene cluster. Am J Hum Genet 73:591–600

Hollox EJ, Armour JAL (2008) Directional and balancing selection in human beta-defensins. BMC Evol Biol 8:113

Horikawa R, Parka DS, Herring PL, Pisano JJ (1985) Pipinins: new mast cell degranulating peptides from *Rana pipiens*. Fed Proc 44:695

Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159

Hughes AL (1991) MHC polymorphism and the design of captive breeding programs. Cons Biol 5:249-251

Hughes AL (1999) Evolutionary diversification of the mammalian defensins. Cell Mol Life Sci. 56:94–103

Hughes AL (2000) Evolution of introns and exons of class II major histocompatibility complex genes of vertebrates. Immunogenetics 51:473–486

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I locus reveals overdominant selection. Nature 335: 167–170.

Hughes AL, Yeager M (1997) Coordinated amino acid changes in the evolution of mammalian defensins. J Mol Evol 44:675–682

Hughes AL, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates. Annu Rev Genet 32:415–435

Innan H, Tajima F (1999) The effect of selection on the amount of nucleotide variation within and between allelic classes. Genet Res [Camb] 73:15–28

Ironside JE, Filatov DA (2005) Extreme population structure and high interspecific divergence of the silene Y chromosome. Genetics 171:705–713

Jarvi SJ, Tarr CL, McIntosh CE, Atkinson CT, Fleischer RC (2004) Natural selection of the major histocompatibility complex in Hawaiian honeycreepers (Drepanidinae). Mol Ecol 13:2157–2168

Jensen LF, Hansen MM, Mensberg K-LD, Loeschcke V (2008) Spatially and temporally fluctuating selection at non-MHC immune genes: evidence from TAP polymorphism in populations of brown trout (*Salmo trutta*, L.). Heredity 100:79–91

Jurevic RJ, Chrisman P, Mancl L, Livingston R, Dale BA (2002) Single-nucleotide polymorphisms and haplotype analysis in β-defensin genes in different ethnic populations. Genet Test 6:261–269

Jurevic RJ, Bai M, Chadwick RB, White TC, Dale BA (2003) Single-nucleotide polymorphisms (SNPs) in human β-defensin 1: high-throughput SNP assays and association with Candida carriage in type I diabetic and nondiabetic controls. J Clin Microbiol 41:90–96

Kelly JK (1997) A test of neutrality based on interlocus associations. Genetics 146:1179–1206

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1245

Kwon SY, Carlson BA, Park JM, Lee BJ (2000) Structural organization and expression of the gaegurin 4 gene of *Rana rugosa*. Biochim Biophys Acta 1492:185–190

Lazzaro BP, Clark AG (2001) Evidence for recurrent paralogous gene conversion and exceptional allelic divergence in the *Attacin* genes of *Drosophila melanogaster*. Genetics 159:659–671

Lazzaro BP, Clark AG (2003) Molecular population genetics of inducible antibacterial peptide genes in *Drosophila melanogaster*. Mol Biol Evol 20:914–923

Lazzaro BP, Sceurman BK, Clark AG (2004) Genetic basis of natural variation in *D. melanogaster* antibacterial immunity. Science 303:1873–1876

Lercher MJ, Chamary J-V, Hurst LD (2004) Genomic regionality in rates of evolution is not explained by clustering of genes of comparable expression profile. Genome Res 14:1002–1013

Luenser K, Ludwig A (2005) Variability and evolution of bovine β-defensin genes. Genes Immun. 6:115–122

Lutz GJ, Cuizon DB, Ryan AF, Lieber RL (1998) Four novel myosin heavy chain transcripts define a molecular basis for muscle fibre types in *Rana pipiens*. J Physiol 508:667–80

Lynn DJ, Higgs R, Gaines S, Tierney J, James T, Lloyd AT, Fares MA, Mulcahy G, O'Farrelly C (2004a) Bioinformatic discovery and initial characterisation of nine novel antimicrobial peptide genes in the chicken. Immunogenetics 56:170–177

Lynn DJ, Lloyd AT, Fares MA, O'Farrelly C (2004b) Evidence of positively selected sites in mammalian α-defensins. Mol Biol Evol 21:819–827

Marenah L, Flatt PR, Orr DF, Shaw C, Abdel-Wahab YHA (2005) Characterization of naturally occurring peptides in the skin secretion of *Rana pipiens* frog reveal pipinin-1 as the novel insulin-releasing agent. J Pept Res 66:204–210

Matsushita I, Hasegawa K, Nakata K, Yasuda K, Tokunaga K, Keicho N (2002) Genetic variants of human β-defensin-1 and chronic obstructive pulmonary disease. Biochem Biophys Res Comm 291:17–22

Maxwell AI, Morrison GM, Dorin JR (2003) Rapid sequence divergence in mammalian β-defensins by adaptive evolution. Mol Immunol 40:413–421

Maynard Smith JM (1992) Analyzing the mosaic structure of genes. J Mol Evol 34:126–129

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 354:114–116

Metz EC, Robles-Sikisaka R, Vacquier VD (1998) Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. Proc Natl Acad Sci USA 95:10676–10681

Michaelson D, Rayner J, Conto M, Ganz T (1992) Cationic defensins arise from charge-neutralized propeptides: a mechanism for avoiding leukocyte autotoxicity? J Leukoc Biol 51:632–639

Morjan CL, Rieseberg LH (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. Mol Ecol 13:1341–1356

Morrison GM, Semple CAM, Kilanowski FM, Hill RE, Dorin JR (2003) Signal sequence conservation and mature peptide divergence within subgroups of the murine β-defensin gene family. Mol Biol Evol 20:460–470

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford: Oxford University Press.

Nicolas P, Vanhoye D, Amiche M (2003) Molecular strategies in biological evolution of antimicrobial peptides. Peptides 24:1669–1680

Ogawa T, Ishii C, Kagawa D, Muramoto K, Kamiya H (1999) Accelerated evolution in the protein-coding region of galectin cDNAs, congerin I and congerin II, from skin mucus of conger eel (*Conger myriaster*). Biosci Biotechnol Biochem 63:1203–1208

Pahdi A, Verghese B (2008) Molecular diversity and evolution of myticin-C antimicrobial peptide variants in the Mediterranean mussel, *Mytilus galloprovincialis*. Peptides. 29:1094–1101.

Pamilo P, O'Neill RJW (1997) Evolution of the Sry genes. Mol Biol Evol 14:49–55

Park JM, Jung J-E, Lee BJ (1994) Antimicrobial peptides from the skin of a Korean frog, *Rana rugosa*. Biochem Biophys Res Commun 205:948–954

Park JM, Lee JY, Moon HM, Lee BJ (1995) Molecular cloning of cDNAs encoding precursors of frog skin antimicrobial peptides from *Rana rugosa*. Biochim Biophys Acta 1264:23–25

Patil A, Hughes AL, Zhang G (2004) Rapid evolution and diversification of mammalian α-defensins as revealed by comparative analysis of rodent and primate genes. Physiol Genomics 20:1–11

Patrzykat A, Gallant JW, Seo J-K, Pytyck J, Douglas SE (2003) Novel antimicrobial peptides derived from flatfish genes. Antimicrob Agents Chemother 47:2464–2470

Perron GG, Zasloff M, Bell G (2006) Experimental evolution of resistance to an antimicrobial peptide. Proc Biol Sci 273:251–256

Peschel A, Sahl H-G (2006) The co-evolution of host cationic antimicrobial peptides and microbial resistance. Nat Rev Microbiol 4:529–536

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

Pupko T, Sharan R, Hasegawa M, Shamir R, Graur D (2003) Detecting excess radical replacements in phylogenetic trees. Gene 319:127–135

Ramos-Onsins S, Aguadé M (1998) Molecular evolution of the *Cecropin* multigene family in *Drosophila*: functional genes vs. pseudogenes. Genetics 150:157–171

Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short term genetic distance. Genetics 105:767–779

Richard M, Thorpe RS (2001) Can microsatellites be used to infer phylogenies? Evidence from population affinities of the western Canary Island lizard (*Gallotia galloti*). Mol Phylogenet Evol 20:351–360

Rollins-Smith LA, Conlon JM (2005) Antimicrobial peptide defenses against chytridiomycosis, an emerging infectious disease of amphibian populations. Dev Comp Immunol 29:589–598

Rollins-Smith LA, Carey C, Longcore J, Doersam JK, Boutte A, Bruzgal JE, Conlon JM (2002) Activity of antimicrobial skin peptides from ranid frogs against *Batrachochytrium dendrobatidis*, the chytrid fungus associated with global amphibian declines. Dev Comp Immunol 26:471–479

Rosetto M, Marchini D, de Filippis T, Ciolfi S, Frati F, Quilici S, Dallai R (2003) The ceratotoxin gene family in the medfly *Ceratitis capitata* and the Natal fruit fly *Ceratitis rosa* (Diptera: Tephritidae). Heredity 90:382–389

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG (2007) Dynamic evolution of the innate immune system in *Drosophila*. Nat Genet 39:1461–1468

Satta Y, O'hUigin C, Takahata N, Klein J (1994) Intensity of natural selection at the major histocompatibility complex locus. Proc Natl Acad Sci USA 91:7184–7188

Schroder JM (1999) Epithelial peptide antibiotics. Biochem Pharmacol 57:121–134

Schutte BC, Mitros JP, Bartlett JA, Walters JD, Jia HP, Welsh MJ, Casavant TL, McCray PB Jr (2002) Discovery of five conserved β-defensin gene clusters using a computational search strategy. Proc Natl Acad Sci USA 99:2129–2133

Seger J (1988) Dynamics of some simple host-parasite models with more than two genotypes in each species. Philos Trans R Soc Lond B Biol Sci 319:541–555

Semple CAM, Rolfe M, Dorin JR (2003) Duplication and selection in the evolution of primate β-defensin genes. Genome Biol 4:R31

Simmaco M, Mignogna G, Barra D, Bossa F (1994) Antimicrobial peptides from skin secretions of *Rana esculenta*. Molecular cloning of cDNAs encoding esculentin and brevinins and isolation of new active peptides. J Biol Chem 269:11956–11961

Simmaco M, Mignogna G, Barra D (1998) Antimicrobial peptides from amphibian skin: What do they tell us? Biopolymers 47:435–450

Smith NGC, Hurst LD (1999) The effect of tandem substitutions on the correlation between synonymous and nonsynonymous rates in rodents. Genetics 153:1395–1402

Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999) Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. Nature 400:667–671

Storz JF, Baze M, Waite JL, Hoffmann FG, Opazo JC, Hayes JP (2007) Complex signatures of selection and gene conversion in the duplicated globin genes of house mice. Genetics 177:481–500

Su C, Nei M (1999) Fifty-million-year-old polymorphism at an immunoglobulin variable region gene locus in the rabbit evolutionary lineage. Proc Natl Acad Sci USA 96:9710–9715

Swofford DL (2002) PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sinauer Associates, Sunderland, MA

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics 124:967–978

Tani A, Ogawa T, Nose T, Nikandrov NN, Deshimaru M, Chijiwa T, Chang C-C, Fukumaki Y, Ohno M (2002) Characterization, primary structure and molecular evolution of anticoagulant protein from *Agkistrodon actus* venom. Toxicon 40:803–813

Tennessen JA (2005a) Molecular evolution of animal antimicrobial peptides: widespread moderate positive selection. J Evol Biol 18:1387–1394

Tennessen JA (2005b) Enhanced synonymous site divergence in positively selected vertebrate antimicrobial peptide genes. J Mol Evol 61:445–455

Tennessen JA, Blouin MS (2007) Selection for antimicrobial peptide diversity in frogs leads to gene duplication and low allelic variation. J Mol Evol 65:605–615

Tennessen JA, Blouin MS (2008) Balancing selection at a frog antimicrobial peptide locus: fluctuating immune effector alleles? Mol Biol Evol 25:2669-2680

Thevissen K, Warnecke DC, François IEJA, Leipelt M, Heinz E, Ott C, Zähringer U, Thomma BPHJ, Ferket KKA, Cammue BPA (2004) Defensins from insects and plants interact with fungal glucosylceramides. J Biol Chem 279: 3900–3905

Tian D, Traw MB, Chen JQ, Kreitman M, Bergelson J (2003) Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. Nature 423:74–77

Tiffin P, Hacker R, Gaut BS (2004) Population genetic evidence for rapid changes in intraspecific diversity and allelic cycling of a specialist defense gene in *Zea*. Genetics 168:425–434

Tiffin P, Moeller DA (2006) Molecular evolution of plant immune system genes. Trends Genet 22:662–670

Tossi A, Sandri L, Giangaspero A (2000) Amphipathic, α-helical antimicrobial peptides. Biopolymers 55:4–30

Tsunoyama K, Bellgard MI, Gojobori T (2001) Intragenic variation of synonymous substitution rates is caused by nonrandom mutations at methylated CpG. J Mol Evol 53:456–464

Vanhoye D, Bruston F, Nicolas P, Amiche M (2003) Antimicrobial peptides from hylid and ranin frogs originated from a 150-million year old ancestral precursor with a conserved signal peptide but a hypermutable antimicrobial domain. Eur J Biochem 270:2068–2081

Vatta S, Boniotto M, Bevilacqua E, Belgrano A, Pirulli D, Crovella S, Amoroso A (2000) Human beta defensin 1 gene: six new variants. Hum Mutat 15: 582–583

Vouille V, Amiche M, Nicolas P (1997) Structure of genes for dermaseptins B, antimicrobial peptides from frog skin. Exon 1-encoded prepropeptide is conserved in genes for peptides of highly different structures and activities. FEBS Lett 414:27–32

Xiao Y, Hughes AL, Ando J, Matsuda Y, Cheng J-F, Skinner-Noble D Zhang G (2004) A genome-wide screen identifies a single beta-defensin cluster in the chicken: implications for the origin and evolution of mammalian defensins. BMC Genomics 5:56

Wakeley J (1999) Nonequilibrium migration in human history. Genetics 153:1863–1871

Wechselberger C (1998) Cloning of cDNAs encoding new peptides of the dermaseptin-family. Biochim Biophys Acta 1388:279–283

Wolfe KH, Sharp PM, Li W-H (1989) Mutation rates differ among regions of the mammalian genome. Nature 337:283–285

Wolfe KH, Sharp PM (1993) Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. J Mol Evol 37:441–456

Won H-S, Kim SS, Jung S-J, Son W-S, Lee B, Lee B-J (2004) Structure-activity relationships of antimicrobial peptides from the skin of *Rana esculenta* inhabiting Korea. Mol Cells 17:469– 476

Woodhams DC, Rollins-Smith LA, Carey C, Reinert L, Tyler MJ, Alford RA (2006a) Population trends associated with skin peptide defenses against chytridiomycosis in Australian frogs. Oecologia 146:531–540

Woodhams DC, Voyles J, Lips KR, Carey C, Rollins-Smith LA (2006b) Predicted disease susceptibility in a Panamanian amphibian assemblage based on skin peptide defenses. J Wildl Dis 42:207–218

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555–556

Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. TREE 15:496–503.

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

Yeaman MR, Yount NY (2003) Mechanisms of antimicrobial peptide action and resistance. Pharmacol Rev 55:27–55

Zhang J (2000) Rates of conserved and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J Mol Evol 50:56–68

APPENDIX

Appendix Table 1

Collection localities and population genetic parameters for all populations discussed in Chapter 5. Population sites are numbered as in Fig. 5.1. All genetic statistics refer to the *Brevinin1.1* locus mature peptide region (72 bp) except for the final column. For all statistical tests, * = $p < 0.05$ and ** = $p < 0.01$.

Site	Species	Location	Year	Latitude	Longitude	2N[a]	Lineage[b]			HW[c]	S[d]	h[e]	Hd[f]	π[g]	π_s[h]	ZnS[i]	π_n/π_s[j]	Tajima's	μsat
							1	2	3									D[k]	He[l]
1	*R. pipiens*	Ontario	2001	45.07	-75.65	74	39%	54%	7%	y	12	4**	0.57	0.06*	0.008	0.51*	9.79**	2.18*	0.87
2	*R. pipiens*	New York	2001	43.47	-76.01	44	89%	9%	2%	y	12	3**	0.21**	0.03	0.003	0.57**	11.15**	-1.06	0.84
2b	*R. pipiens*	New York	1971	43.47	-76.01	50	86%	12%	2%	y	12	3**	0.25**	0.03	0.003	0.55*	14.95**	-0.64	0.86
3	*R. pipiens*	New York	2001	42.99	-76.77	92	40%	60%	0%	y	8	2**	0.49	0.05**	0.000	1.00**	>50**	3.63**	0.89
4	*R. pipiens*	Ontario	2001	44.04	-78.98	80	26%	66%	8%	y	13	4**	0.50*	0.05	0.010	0.48*	7.05*	1.26	0.86
5	*R. pipiens*	Ontario	2001	43.52	-79.00	84	17%	82%	1%	y	12	4**	0.31**	0.03	0.002	0.49**	28.41**	-0.02	0.87
6	*R. pipiens*	Ohio	2000	41.59	-81.06	46	4%	78%	17%	y	12	3**	0.36**	0.03	0.019	0.47*	1.67	-0.73	0.84
7	*R. pipiens*	Minnesota	2000	45.13	-94.03	74	78%	22%	0%	y	11	4**	0.43*	0.04	0.005	0.53**	10.36**	0.79	0.84
8	*R. pipiens*	Minnesota	2000	45.71	-94.89	64	94%	6%	0%	y	9	3**	0.23**	0.02	0.007	0.78**	2.30*	-1.17	0.83
9	*R. pipiens*	Nebraska	1999	42.00	-97.01	52	85%	12%	4%	y	13	4**	0.31**	0.03	0.007	0.49*	5.84**	-0.49	0.82
10	*R. pipiens*	Nebraska	1999	41.95	-102.43	50	100%	0%	0%	y	0	1	0.00	0.00	0.000	0.00	-	-	0.57
11	*R. pipiens*	Idaho	1999	42.62	-113.28	48	62%	0%	38%	y	13	3**	0.53*	0.08**	0.034	0.73**	2.69*	2.58*	0.62
12	*R. pipiens*	Idaho	1999	42.14	-111.26	40	70%	0%	30%	y	11	2**	0.43*	0.07**	0.028	1.00**	2.86*	2.53*	0.65
13	*R. pipiens*	British Columbia	1999, 2000	49.05	-116.50	50	100%	0%	0%	y	0	1	0.00	0.00	0.000	0.00	-	-	0.44
14	*R. blairi*	Illinois	2007	37.71	-89.45	54	96%	4%	0%	y	10	2**	0.07**	0.01**	0.002	1.00**	5.19**	-1.89*	-
15	*R. blairi*	Missouri	1993	38.59	-92.07	46	87%	13%	0%	y	10	3**	0.40*	0.04	0.000	0.59*	>15**	0.29	-
16	*R. palustris*	Michigan	2007	42.30	-85.33	40	0%	0%	100%	y	0	1	0.00	0.00	0.000	0.00	-	-	-
17	*R. palustris*	Wisconsin	2007	43.06	-89.78	46	0%	0%	100%	y	0	1	0.00	0.00	0.000	0.00	-	-	-

[a]2N = number of alleles = twice the number of individuals

Appendix Table 1 (Continued)

[b]Lineage 1, Lineage 2, and Lineage 3 refer to the percentage of alleles in each population belonging to each of these 3 allelic groups

[c]HW = in Hardy-Weinberg equilibrium; y = yes

[d]S = number of segregating sites in mature peptide region

[e]h = number of haplotypes in mature peptide region; tested whether h is significantly low given S

[f]Hd = haplotype diversity in mature peptide region, equivalent to expected heterozygosity; tested whether Hd is significantly low given S

[g]π = mean number of pairwise differences among mature peptide region sequences; tested whether π is significantly extreme given S

[h]π_s = mean number of pairwise differences among sequences at silent sites; tested whether π_s is significantly extreme given silent S

[i]ZnS = linkage disequilibrium statistic of Kelly (1997); tested whether ZnS is significantly high given S

[j]π_n/π_s = ratio of nonsynonymous nucleotide variation to synonymous nucleotide variation; tested whether significantly different than 1

[k]Tajima's D = statistic of Tajima (1989); tested whether significantly different than 0

[l]μsat He = mean expected heterozygosity at 9 microsatellite loci

Appendix Table 2

Population genetic parameters for 4 putatively neutral nuclear loci unlinked to *Brevinin1.1* (1246 bp) in 3 frog species (*Arcadlin*, *Myosin*, *FIBI7*, and *Tyrosinase*). Population sites are numbered as in Fig. 5.1. For all statistical tests, * = $p < 0.05$ and ** = $p < 0.01$.

Site	Species	2N[a]	Locus	Length[b]	HW[c]	S[d]	H[e]	Hd[f]	π[g]	π_s[h]	ZnS[i]	π_n/π_s[j]	Tajima's D[k]
1	*R. pipiens*	74	*Arcadlin*	265	y	4	6	0.69	0.005	0.011	0.24	0.05**	1.60
			Myosin	150	y	6	6	0.61	0.011	0.016	0.22	0.00*	0.69
			FIBI7	258	y	10	5	0.33	0.004	0.005	0.30	0.00	-1.28
			Tyrosinase	573	y	9	8	0.78	0.005	0.017	0.21	0.07**	1.11
			Mean ±	312 ±	y	7.3 ±	6.3 ±	0.60 ±	0.005 ±	0.012 ±	0.24 ±	0.03**	0.53 ±
			St. Dev.	182		2.8	1.3	0.19	0.003	0.006	0.04	-	1.26
2	*R. pipiens*	44	*Arcadlin*	265	y	4	5	0.68	0.005	0.009	0.19	0.04**	0.80
			Myosin	150	y	6	5	0.67	0.012	0.019	0.23	0.05*	0.84
			FIBI7	258	y	4	3	0.32	0.004	0.005	0.61*	0.00	0.21
			Tyrosinase	573	y	9	7	0.73	0.004	0.013	0.21	0.00**	0.32
			Mean ±	312 ±	y	5.8 ±	5.0 ±	0.60 ±	0.005 ±	0.011 ±	0.31 ±	0.04**	0.54 ±
			St. Dev.	182		2.4	1.6	0.19	0.004	0.006	0.20	-	0.32
14	*R. blairi*	54	*Arcadlin*	265	y	2	3	0.50	0.002	0.004	0.03	0.00*	0.33
			Myosin	150	y	2	3	0.46	0.003	0.005	0.03	0.00	0.25
			FIBI7	258	y	1	2	0.04	0.000	0.000	-	0.00	-1.10
			Tyrosinase	573	y	2	3	0.51	0.001	0.000	0.01	>4.1	0.33

Appendix Table 2 (Continued)

			Mean ±	312 ±	y	1.8 ±	2.8 ±	0.38 ±	0.001 ±	0.002 ±	0.02 ±	0.13	-0.05 ±
			St. Dev.	182		0.5	0.5	0.23	0.001	0.003	0.01	-	0.70
15	*R. blairi*	46	*Arcadlin*	265	y	1	2	0.41	0.002	0.003	-	0.00	1.13
			Myosin	150	y	2	3	0.55	0.004	0.006	0.05	0.00	0.64
			FIBI7	258	y	2	3	0.17	0.001	0.001	0.00	0.00	-1.15
			Tyrosinase	573	y	3	4	0.54	0.001	0.000	0.05	>4.4	-0.08
			Mean ±	312 ±	y	2 ±	3.0 ±	0.42 ±	0.001 ±	0.003 ±	0.03 ±	0.14	0.13 ±
			St. Dev.	182		0.8	0.8	0.18	0.002	0.003	0.03	-	0.99
16	*R. palustris*	40	*Arcadlin*	265	y	4	4	0.67	0.006	0.012	0.32	0.00**	1.55
			Myosin	150	y	4	2*	0.51	0.014*	0.016*	1.00*	0.64	2.77**
			FIBI7	258	y	13	6	0.53*	0.015	0.017	0.36	0.00*	0.71
			Tyrosinase	573	y	5	3	0.57	0.002	0.007	0.66*	0.12**	0.14
			Mean ±	312 ±	y	6.5 ±	3.8 ±	0.57 ±	0.007 ±	0.013 ±	0.58 ±	0.21**	1.29 ±
			St. Dev.	182		4.4	1.7*	0.07	0.006	0.005	0.32*	-	1.14**
17	*R. palustris*	46	*Arcadlin*	265	y	1	2	0.20	0.001	0.002	-	0.00	-0.18
			Myosin	150	y	0	0	0.00	0.000	0.000	-	0.00	-
			FIBI7	258	y	13	3**	0.48*	0.018*	0.021*	0.56*	0.00*	1.62
			Tyrosinase	573	y	7	4	0.67	0.003	0.010	0.54*	0.08**	-0.04
			Mean ±	312 ±	y	5.3 ±	2.3 ±	0.34 ±	0.005 ±	0.008 ±	0.55 ±	0.02**	0.47 ±

Appendix Table 2 (Continued)

St. Dev.	182	6.0	1.7**	0.30	0.008	0.010	0.02*	-	1.00

[a]2N = number of alleles = twice the number of individuals

[b]Length = locus length in base pairs

[c]HW = in Hardy-Weinberg equilibrium; y = yes

[d]S = total number of segregating sites among the 4 loci

[e]h = mean number of haplotypes per locus; tested whether h is significantly low given S

[f]Hd = mean haplotype diversity per locus, equivalent to expected heterozygosity; tested whether Hd is significantly low given S

[g]π = mean number of pairwise differences among sequences; tested whether π is significantly extreme given S

[h]π_s = mean number of pairwise differences among sequences at silent sites; tested whether π_s is significantly extreme given silent S

[i]ZnS = mean linkage disequilibrium statistic of Kelly (1997) per locus; tested whether ZnS is significantly high given S

[j]π_n/π_s = ratio of nonsynonymous nucleotide variation to synonymous nucleotide variation; tested whether significantly different than 1

[k]Tajima's D = statistic of Tajima (1989); tested whether significantly different than 0.