

## I. Background and Significance

### A. Rationale

A key focus in evolutionary physiology is the question of how organisms adapt to specific or novel environmental conditions. Ultimately, adaptations must result from evolutionary changes at the molecular level. However, in many cases the environmental stimulus is complex or ill-defined. In addition, it is often impossible to pinpoint the genes undergoing evolution, or to identify how specific changes in nucleotide sequence alter their function. **An ideal model system for studying adaptation would be one in which a specific environmental stimulus enhances selection for specific variants in identified genes of known function, where the adaptive benefit of the change is obvious and its functional effects can be measured.**

We propose to study the evolution of toxin resistance at the molecular level in vertebrates, focusing on a group of teleost fishes in which some species sequester and tolerate a potent neurotoxin in exceedingly high levels within their own tissues. This toxin poses a highly specific challenge to animal physiology, as it binds selectively and with high affinity to a single class of ion channel proteins, the voltage-dependent sodium channels. Sodium channels form a family of genes that are present in multiple copies in vertebrate genomes, and that differ in their expression throughout the body. Functional studies and experiments have identified specific amino acids in vertebrate ion channels that strongly affect the binding affinity of this potentially lethal channel blocker.

The system we present below provides a unique opportunity to study functional changes that have occurred in parallel among members of a gene family, both within an organism (paralogs) and across species (orthologs). If one were to study this problem using a single gene ortholog or in a single species alone, it would be impossible to reconstruct the history of these changes. However, because this is an evolutionary problem, we propose to use a comparative phylogenetic approach to study multiple sodium channel paralogs in a diverse sample of species that vary widely in their degree of toxin resistance, allowing us to reconstruct the evolutionary history of specific molecular changes that confer a clear adaptive function. An additional advantage of the proposed research is that it utilizes data from three new teleost fish genome projects in order to study the biology of organisms comparatively. To complement this, we believe that comparative studies that include genomic model species can provide a biological context for understanding and utilizing the information in genome databases.

### B. Sodium Channel Proteins and Toxins

Voltage-dependent sodium channels are fundamental to the basic functions of the nervous system in all animals and, additionally, of cardiac and skeletal muscle cells in vertebrates. In vertebrate genomes, sodium channel genes typically occur in multiple copies, each of which has limited expression in the cells of a particular tissue or organ. However, the number of gene copies and their patterns of somatic expression are variable among taxa. In a recent study co-authored by two of the PI's for this proposal, phylogenetic and comparative analyses of sodium channel genes from vertebrates supported an evolutionary history of this gene family characterized by multiple duplications and mutations along separate lineages, resulting in genomes containing paralogous gene copies that are expressed in different cell types (1).

A number of animals have evolved potent toxins that bind to and block sodium channels with nanomolar affinity. Scorpions and cone snails have independently evolved different sodium channel blocking peptides for immobilization of prey (2, 3). Amazingly, one sodium channel toxin, tetrodotoxin (TTX), has been found in a staggeringly diverse variety of

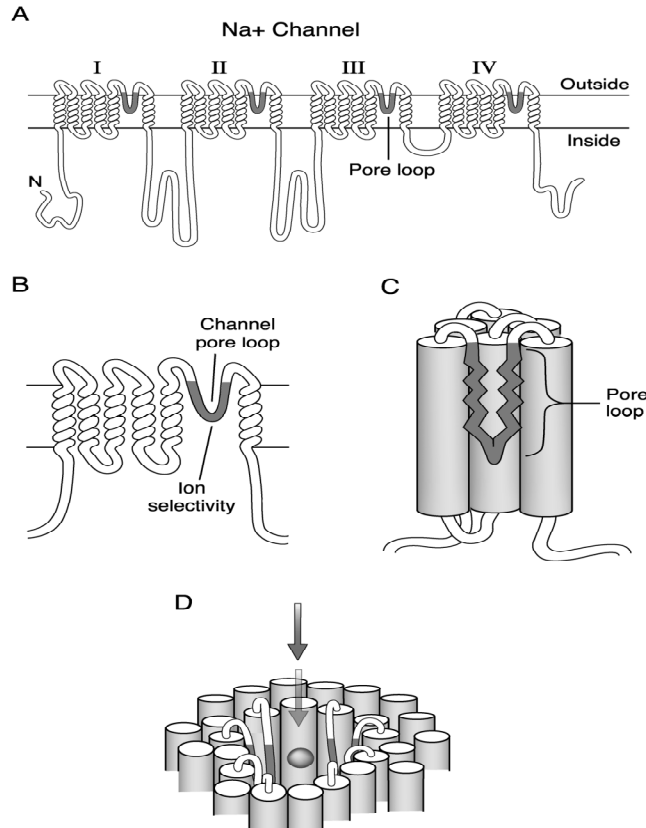
animals, including pufferfishes such as those in the genus *Fugu* (*Takifugu*) (Osteichthyes: Tetraodontiformes); related tetraodontiform and other fishes (4); some newts (e.g. *Taricha* spp. and *Cynops pyrrhogaster* (5), and frogs (6, 7) ; and a number of invertebrates including some species of ribbon worms (8), arrow worms (9), mollusks (including the blue-ringed octopus *Hapalochlaena* spp.) (10), starfish (11), and crabs (12). In some invertebrates, TTX is used actively as a venom for prey immobilization, while in vertebrates, its mere presence in an organism may deter potential predators. However, the biological origin of TTX is poorly understood, as it has not yet been shown to be actually produced by the animals that use or sequester it. A clue to the distribution of TTX among such disparate taxa came in the 1980's, when it was suggested that TTX in pufferfishes might be of bacterial origin, acquired through diet or perhaps resulting from a symbiosis (13). Some studies have suggested that the bacteria involved may be members of the genera *Pseudomonas* and/or *Vibrio* (14-16) but this topic will require further investigation outside the scope of our proposal.

**Most animals that use or come into regular contact with sodium channel toxins have also evolved mechanisms of preventing the toxins from affecting their own cells.** Venomous snails and scorpions produce and store peptide toxins in specialized glands so that the sodium channels in their own vital tissues are never exposed to the toxins. Although TTX in pufferfishes is often sequestered in high concentration in particular organs (e.g. the skin, liver, and ovaries of *Fugu* spp.), there are no specialized glands associated with TTX, and the toxin tends to be in general circulation (17, 18). In some pufferfish taxa, injection with TTX in concentrations manyfold higher than the lethal dose for mice or other teleost fishes has no effect (13, 19). Thus, the animals' own sodium channels must somehow be protected from TTX binding.

One form of protection is a TTX-binding protein, as occurs in some TTX-bearing crabs (12). This seems to be the crabs' sole means of protection, since their nervous systems are vulnerable to TTX when hemolymph containing the binding protein is washed away and TTX is applied in saline. While a protein of similar function has been reported in pufferfishes (20, 21), this is not their only protection; neurophysiological and toxin-binding assays show that the TTX binding affinity of *Fugu pardalis* muscle and brain cells is extraordinarily low (22). Similarly, in a neural channel of the newt *Cynops pyrrhogaster*, TTX has little physiological effect (23). This implies that **the Na<sup>+</sup> channels themselves in pufferfishes, newts, and perhaps some other TTX-bearing organisms, have convergently evolved insensitivity to TTX.**

### C. Structure of the Sodium Channel and TTX-binding

The Na<sup>+</sup> channel is a single large protein with four highly conserved membrane-spanning domains, connected by intracellular loops of variable length and less conserved sequence (Fig. 1). Each of these repeated membrane spanning domains has a highly conserved region called the P loop (for "pore" loop), and these are believed to face each other and dip down into the membrane to form the four sides of the mouth of an ion-permeating pore. About two dozen conserved amino acids in the P loop are responsible for critical features of the channel, such as its selectivity for sodium ions and channel toxin binding. Data and modeling of TTX-binding to the pore of the mammalian sodium channels indicate that **residues in each of the four domains are critical in anchoring the toxin in the mouth of the pore** (24, 25). Furthermore, **the binding affinity of TTX to the channel is differentially affected by mutations of different amino acids in the P region, varying from few-fold to thousand-fold decreases in affinity** (25). Thus, some kinds of substitutions are likely to provide greater TTX-protection than others.



**Figure 1**—Structure of the sodium channel. (A) The sodium channel is composed of four repeating domains, each of which possesses six membrane spanning regions and a pore loop (or P region) that dips into, but does not cross, the membrane. (B) A blow up of a single domain. (C) Each domain folds such that the P region faces the interior of the channel. (D) The four P regions face each other and line the mouth of the channel. **The amino acids of the P region confer sodium selectivity on the channel, and it is here that TTX binds, straddling all four P regions.** Figure modified from (26).

It is informative to compare the amino acid sites involved in TTX-binding – which have been determined by site-directed mutagenesis of mammalian sodium channels (24, 25, 27) -- with the same sites in the P regions of sodium channels in TTX resistant animals. For example, a comparison of amino acid sequences in the P loop of domain I of Na<sup>+</sup> channels reveals some striking differences between TTX-sensitive and TTX-resistant organisms. In the first domain of the muscle Na<sup>+</sup> channel in mouse, rat, human, and non-tetrodotoxic teleosts such as *Danio rerio* and *Sternopygus macrurus*, the P region contains the amino acid sequence D(Y/F)WEN (Fig. 2). In some Na<sup>+</sup> channel genes there is a Tyr (Y), in others a Phe (F): both aromatic amino acids. In the same site in the muscle Na<sup>+</sup> channel gene of the pufferfish *Fugu pardalis* there is an Asn (N), which is a non-aromatic amino acid (22). A comparison of the same site in a brain Na<sup>+</sup> channel gene in the rat and in the TTX-producing newt *Cynops pyrrhogaster* shows again Y or F (aromatic) in the rat, but Ala (A) (non-aromatic) in the newt (23). This suggests that evolution at this particular site has occurred independently in *Fugu* and *Cynops*, and that this mutation might be responsible for the observed decrease in TTX binding affinity. **Indeed, mutagenesis of this residue in the TTX-sensitive mammalian muscle Na<sup>+</sup> channel gene confers a much lower affinity for TTX (27-29).**

**A**

"muscle" Na <sup>+</sup> channel genes	pore, domain I	TTX-sensitive
	*	
	DYWEN	yes
	DYWEN	yes
	DFWEN	?
	DNWEN	no
Rat 1.4	DFWEN	yes

**B**

"brain" Na <sup>+</sup> channel genes	pore, domain I	TTX-sensitive
	*	
	DÄWEN	no
	DYWEN	yes
	DFWEN	yes
	DFWEN	yes
	DYWEN	yes

**Figure 2--**Sequences in the P region of domain I in two sodium channel genes in a number of species. (A) Domain I of the muscle sodium channel gene. (B) Domain I of a sodium channel gene expressed in the brain. The asterisk denotes a site that is a highly conserved **aromatic amino acid** in TTX-sensitive channel genes, but is **not an aromatic amino acid** in TTX-resistant genes. From (30).

Most site-directed mutagenesis work has been directed at determining how neutralizing or altering amino acid charge affects TTX binding, although it is recognized that TTX binding is not entirely dependent on electrostatic interactions (24). However, another mechanism may have evolved in TTX-resistant animals as well: amino acid substitutions that result in a consensus site for the attachment of a large molecule, such as a sugar, overhanging the mouth of the channel. This would provide an umbrella-like barrier to the TTX molecule but would allow smaller sodium ions to pass into the mouth. Such a substitution occurs in the acetylcholine receptor of the Egyptian cobra (*Naja haje*) and this is the basis for its insensitivity to its own acetylcholine receptor-blocking toxins (31).

#### D. Scenarios for Evolution of TTX Resistance by Sodium Channel Genes

Evolution of TTX resistance may be analogous to the rapid evolution of insecticide (pyrethroid and DDT) resistance by a single point mutation of the Na<sup>+</sup> channel gene in natural populations of flies, cockroaches, and other insects within a few decades of insecticide use (32, 33). It is easy to see how a spontaneous mutation that confers resistance against potentially lethal toxins (without altering normal function) would be strongly favored under natural selection.

However, the situation in vertebrates is more complicated (and interesting) since, unlike insects which have a single Na<sup>+</sup> channel gene (34), vertebrate genomes contain multiple Na<sup>+</sup> channel paralogs. Furthermore, some cell types (such as heart) express only a single sodium channel gene (35), while other cell types (such as skeletal muscle) express one sodium channel gene early in development but a different one throughout life (36) for mammals; A. Linares and A. Ribera, pers. comm. for *Danio rerio*) and still other cells (such as many types of neurons) permanently express multiple sodium channel genes (37, for mammals).

The total number of sodium channel genes has not been determined yet for any TTX-containing species; however, teleost fish have at least 6 distinct Na<sup>+</sup> channel genes (1), and amphibians are likely to have that number or more. **Thus, in addition to convergent origins of TTX resistance in diverse taxa, TTX-resistance is likely to have evolved by independent amino acid substitutions in multiple sodium channel paralogs during the evolutionary history of a single taxon.** Another possibility is that TTX resistance evolved in one gene (or a few genes) which subsequently became expressed more widely or with a changed

developmental sequence; this calls for changes at both the structural and transcriptional levels, and might ultimately be accompanied by secondary deletion of those sodium channel genes that had not acquired TTX-insensitivity.

As noted above, various mutations at the pore region can alter TTX sensitivity either slightly or greatly. While it is possible that a single favorable substitution occurred that immediately rendered a particular sodium channel TTX-resistant, it is also possible that multiple mutations that conferred weak TTX-resistance accumulated, endowing fish with the capacity to tolerate increasing amounts of TTX. In support of this notion, pyrethroid resistance is found in insects with a single and specific sodium channel mutation, while very high levels of resistance occurs in some insects with an additional mutation at another site (33). **Our BLAST searches of the *Fugu rubripes* and *Tetraodon nigroviridis* genome databases for the P region of domain I found sequences homologous to sodium channel genes in *Fugu pardalis* and other vertebrates. In both of these species, there were sequences with both aromatic and non-aromatic amino acids at the site described above (see Figure 2). Therefore, TTX insensitivity in some channel genes may involve other sites.** In this context, it is worth noting that in the P region of domain 4 of the pufferfish muscle sodium channel (which was described above as having a non-aromatic substitution in domain 1), there is a T instead of the G which is highly conserved across mammalian sodium channel genes (22). Because this mutation occurs at the pore region, it may be an additional site that helps provide TTX resistance in pufferfishes.

Understanding the evolution of TTX and TTX resistance in taxa as diverse as fishes, amphibians, and no fewer than five invertebrate phyla is a large and far-reaching project, and is complicated by the fact that TTX may not be produced by the animals themselves, but by bacteria acquired through symbiosis or diet. **Therefore, we propose to begin by studying this problem comparatively in a diverse sample of tetraodontiform fishes, including the two genomic species *Fugu rubripes* and *Tetraodon nigroviridis*. In particular we wish to ask: How have sodium channel genes in tetraodontiform fishes evolved TTX immunity? We will address this question by 1.) cloning and sequencing sodium channel gene paralogs from the taxa in our sample; 2.) using phylogenetic methods to reconstruct the history of this gene family in Tetraodontiformes; and 3.) comparing sequences of the known toxin binding regions of sodium channels in toxic and non-toxic species within the Tetraodontiformes, using the TTX-sensitive teleosts *Danio rerio* and *Sternopygus macrurus* as outgroups. These studies will be supported by a phylogenetic study of the Tetraodontiformes themselves, and experimental studies to independently assess TTX sensitivity, including TTX-binding assays on brain, heart, and skeletal muscle.**

## **E. Pufferfishes and other Tetraodontiformes as Model Organisms**

Tetraodontiform fishes are ideal organisms to carry out this work for five reasons:

1.) More is known about the biology of TTX in the pufferfish genus *Fugu* than in any other animal, and the distribution of TTX among other Tetraodontiformes is well documented (13, 17-22, 38).

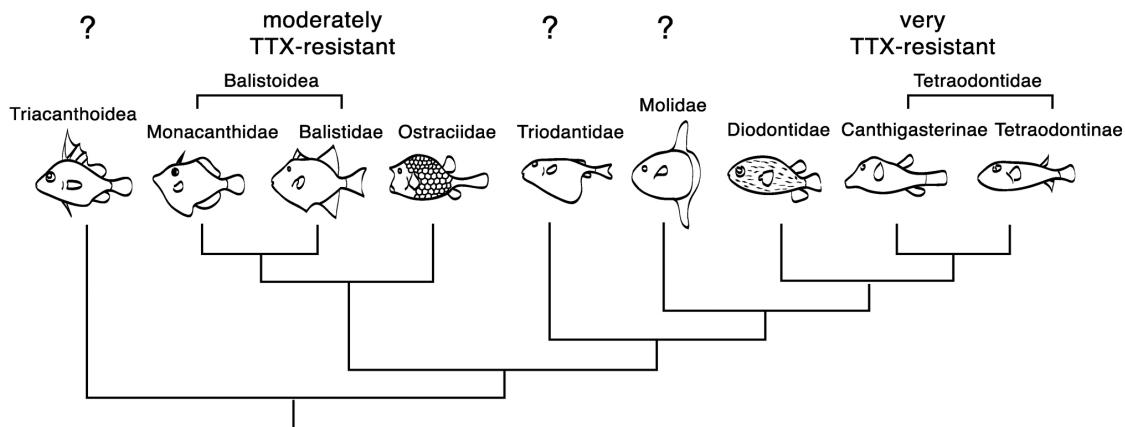
2.) Two sodium channel genes have already been cloned and studied in *Fugu*. **In addition, we have already identified a number of other sodium channel genes from the *Fugu* and *Tetraodon* genome databases (<http://fugu.hgmp.mrc.ac.uk/>), (<http://www.genoscope.cns.fr/externe/tetraodon/>). A variety of cDNA and genomic libraries for these species are available for further analysis. The increasing prominence of pufferfishes as genomic organisms offers other advantages for future studies, mainly the ability to quickly isolate other genes of interest such as those for other ion channels or genes for putative TTX-binding**

proteins. **We believe our utilization of these new databases to be a great strength of this proposal.**

3.) Family-level phylogenetic relationships of the Tetraodontiformes have been proposed based on morphology (39-41) (Fig. 3), which will be helpful in designing a taxon sample for our project. However, our study will use molecular data from three separate loci (not sodium channels) to further elaborate phylogeny within the Tetraodontiformes. This will be helpful for future comparative, systematic, and genomic studies.

4.) **Most importantly, there is great diversity in TTX-sensitivity among tetraodontiform fishes.** Whereas some Tetraodontiformes contain TTX and are TTX-resistant, other groups are not tetrodotoxigenic and show more normal sensitivity to applied TTX (19, 38) (Fig. 3). *Fugu* spp. (family Tetraodontidae) are about 500-fold less sensitive to TTX than non-toxin-producing fish, and even some representatives of non-toxic tetraodontiform families (e.g. Balistidae, Ostraciidae) may still be ten- to twenty-fold less sensitive to TTX than other teleosts, suggesting limited TTX-insensitivity (or pre-adaptations for TTX insensitivity) in these lineages (19). Even more intriguing is the fact that some families may have both TTX-sensitive and TTX-insensitive members(19, 38): for example, within the Tetraodontidae, the genera *Arothron*, *Fugu*, and *Canthigaster* are highly TTX-resistant whereas the genus *Lagocephalus* is only moderately resistant. **This diversity will be critical for investigating the evolutionary changes in candidate binding sites at the pore region of the sodium channel.**

5.) Specimens of a variety of species are readily available in the pet trade, in marketplaces in Asia, and/or by collection from natural marine and freshwater habitats.



**Figure 3**—Phylogenetic tree of the tetraodontiform fishes (19; figure modified from 42). Question marks refer to families whose representatives have not yet been tested for TTX sensitivity.

## II. Specific Aims and Research Objectives

We will clone and sequence all four domains from multiple sodium channel genes, from a diverse sample of both TTX-containing and non-toxic species of Tetraodontiformes, using domain specific primers. Using phylogenetic analyses, we will then reconstruct the evolutionary history of this gene family in our taxon sample, thus identifying gene orthologs and polarizing the direction of change, using the teleosts *Sternopygus macrurus* and *Danio rerio* as outgroups. We will then compare the P regions of all 4 domains among our taxa in order to identify substitutions that may confer resistance to TTX. Because reconstructing ortholog

relationships requires knowledge of the evolutionary history of the animals, we will also sequence three loci to estimate a phylogeny of our study species: ND-2, rhodopsin, and cytochrome *b*, all of which have proven useful in prior studies for estimating relationships within teleost fishes (see Methods, below).

We will use these sequence data to reconstruct the evolutionary history of sodium channel genes and TTX resistance in Tetraodontiformes, and to test two specific evolutionary hypotheses. The first is that paralogous sodium channel genes in some tetraodontiform fishes have accumulated one or more substitutions in parallel that are predicted to make them insensitive to TTX. These changes may have also occurred independently in different taxa. Our second hypothesis is that gene loss and/or duplication may have occurred in some taxa once some TTX-sensitive sodium channels evolved. These hypotheses are not mutually exclusive, and our analyses may reveal that both kinds of evolutionary events have occurred in some taxa.

Finally, these studies will be supported by TTX-binding assays on preparations of heart, skeletal muscle, and brain tissues, to confirm that resistance comes as a result of reduced TTX binding to sodium channels and not to circulating binding proteins.

### **A. Have Paralogous Sodium Channel Genes Evolved TTX resistance in the Tetraodontiformes?**

We will approach this question by analyzing amino acid sequences at the pore regions of all four domains of at least three sodium channel genes. Because sodium channels are large proteins (~2,000 amino acids), cloning all of the sodium channel genes in a large number of species would be beyond the scope of this proposal. Instead, we plan to focus our analysis on a few specific genes that we have experience cloning using RT-PCR. Below we list the genes we have chosen to study, and our rationale for choosing them.

**Cardiac sodium channel:** This gene has not yet been cloned in the pufferfish, but we have identified its probable ortholog in a BLAST search of the *Fugu* database. Cardiac muscle is known to be TTX-sensitive in *Sternopygus* and *Danio*, but TTX-resistant in *F. pardalis*. This is the only sodium channel gene expressed by the heart. Primers previously used for cloning the cardiac sodium channel in *Sternopygus macrurus* and *Danio rerio* will be used to amplify the P regions of the cardiac sodium channel gene in all study taxa. In addition, the heart should provide enough tissue for a TTX binding assay.

**Muscle sodium channel:** This gene has been selected because it has already been cloned and sequenced in *Fugu pardalis*, and is known to have at least one specific amino acid substitution conferring TTX-resistance. This gene is likely to be the sole gene expressed in adult muscle, and we expect that it can be easily cloned from the taxa in our sample using RT-PCR based on sequence from *F. pardalis*. Furthermore, muscle can easily be tested for TTX-sensitivity with binding assays.

**Brain Sodium Channel (SterNa5/zNa<sub>v</sub>1.6):** Multiple sodium channel genes are expressed in the brain of fish and mammals. This gene is the most abundant neural sodium channel gene in mammals, and likely also in fish (43, 44). Orthologs have been cloned in pufferfish (*Fugu rubripes*), the weakly electric fish *Sternopygus macrurus*, and zebrafish *Danio rerio* (1, 44, 45). Because multiple sodium channel genes are expressed in the brain, and because their sequences are similar, it will be necessary to clone all of the gene except the C and N termini to ensure that the P regions belong to a single gene.

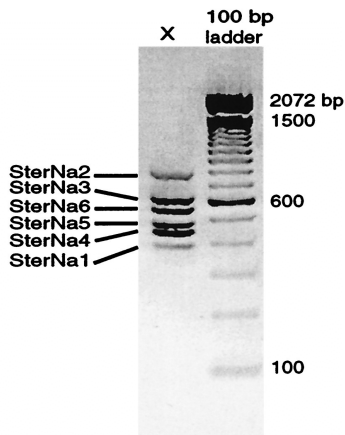
### **B. Has Evolution of TTX Resistance Been Accompanied by Gene Loss and/or Duplication?**

We plan to address this question by 1.) determining the number of sodium channel paralogs in the genomes of each taxon in our sample, and 2.) determining the orthologs of these genes in related taxa and in our outgroups. Given the propensity for some tetraodontiformes to

have small genomes, presumably due to high rates of deletions (42), it is possible that redundant (TTX-sensitive) sodium channel genes have become pseudogenes or lost.

For *Fugu rubripes*, and *Tetraodon nigroviridis* we will use the publicly accessible databases to screen for sodium channel gene orthologs. For the other species in our sample, we will identify them with PCR of genomic DNA. Using genomic PCR, we have already had success cloning six sodium channel genes from the weakly electric fish *Sternopygus macrurus* that we used to study the evolution of this gene family in vertebrates (1). (Following this study we identified a seventh gene and have adjusted our primers so that it can be isolated by this technique as well). We are now using this method successfully in another study to identify sodium channel genes in catfish, the sister group of the Gymnotiformes (H. Zakon, unpubl.)

We use primers that flank introns in highly conserved regions, on the premise that intron position is highly conserved (in sodium channel genes it is conserved from jellyfish to humans, 46-48) whereas intron length is highly variable. Thus, one set of primers reproducibly produces a set of bands, each corresponding to a different sodium channel gene (see Figure 4.) We design the primers to include a hundred or more bases of open reading frame on each side of the intron so that we can use this information for designing other primers for further PCR reactions. This method works successfully in teleost fishes since they typically have short introns, at least compared with mammals (1). This method has limitations if the targeted intron has been lost from a gene, it is too large to amplify across, or if the primer sequence is not optimal for some genes. To minimize these potential problems we repeat the process across a second intron.

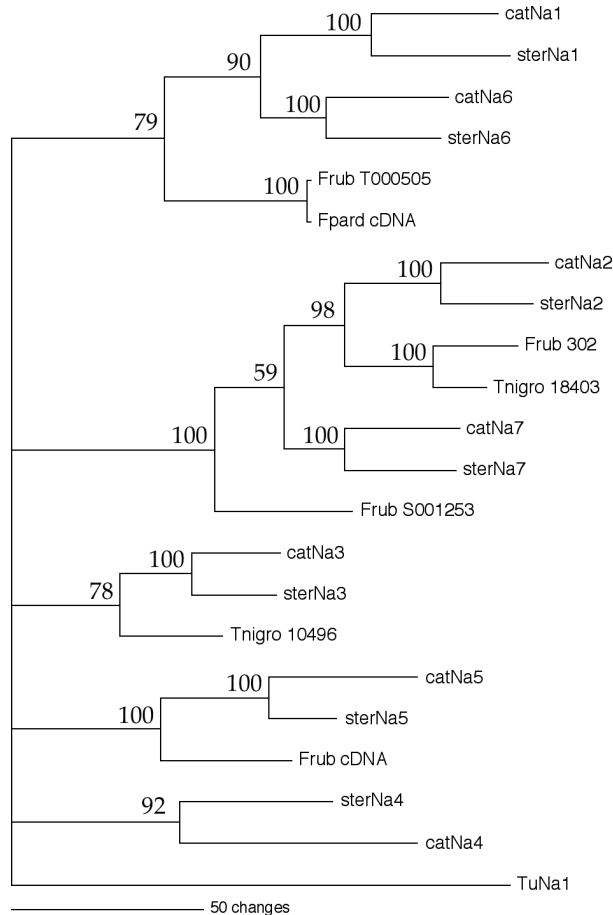


**Figure 4**—Gel of PCR fragments from genomic DNA of the weakly electric fish *Sternopygus macrurus*. Degenerate primers were designed to a highly conserved region of the sodium channel that contains an intron (which we designate as intron X.) As well as the introns (which are generally small in teleost fish) there are ~300 bp of the open reading frame to allow for targeting of specific primers for amplification of additional fragments. From (1).

We will then compare the exonic sequences we have obtained around these two introns with sequences available from sodium channel genes in other species to determine which gene orthologs are present in each species and which are absent. We have found that with as little as ~350 bps of sequence in this region, we can reconstruct the same gene tree found from analyses of more extensive amounts of sequence (~700 bp) in genes we have previously studied (H. Zakon, pers. observ.), thus identifying probable orthologs.

**Example:** We have already found a number of putative sodium channel sequences by BLAST-searching the *Fugu* and *Tetraodon* databases. In Figure 5, we show a preliminary Bayesian analysis of ~350 bp across domain 3 of sodium channel genes from *Sternopygus*, catfish, and some of the sequences we found for *Fugu* and *Tetraodon*. The result groups tetraodontid sodium channel sequences with putative orthologs in *Sternopygus* and catfish. We were unable to identify *Fugu* or *Tetraodon* matches for the sodium channel orthologs sterNa4 and catNa4. While there are putative sodium channel sequences in other *Fugu* and *Tetraodon* data scaffolds, it is also possible that this channel has been deleted in tetraodontid genomes, or that it is the product of a duplication event that occurred in a common ancestor to *Sternopygus*

and catfish after the divergence of Tetraodontiformes. This is the type of question we expect to be able to resolve using data from a wider sample of diverse taxa. In the proposed research, we will use primer sets across an additional domain, giving ~700 bp of sequence for each channel gene in order to make such analyses more robust (e.g., to obtain higher posterior probabilities at deep nodes).



**Figure 5** – Bayesian likelihood tree of ~350 bp of sodium channel sequences, including data from the *Fugu* and *Tetraodon* genome databases. (150,000 Markov chain Monte Carlo generations, 4 chains, burn-in at 6000). Numbers above branches indicate Bayesian posterior probabilities. sterNa 1-7 and catNa 1-7 indicate seven sodium channel genes cloned from *Sternopygus macrurus* and the catfish *Ictalurus punctatus* in the Zakon lab. Fpard cDNA is a *Fugu pardalis* muscle sodium channel known to be TTX-insensitive (GenBank accession AB030482). Sequence names beginning with Frub and Tnigro were retrieved by us from the *Fugu* and *Tetraodon* genome databases by BLAST searching for matches to sodium channel sequences for *Sternopygus*. The outgroup sequence TuNa1 is from the single sodium channel in the genome of the tunicate *Halocynthia roretzi*.

### III. Methods

**Taxon Sample** – Our taxon sample is designed to maximize phylogenetic diversity by sampling widely within all Tetraodontiform families. We will also select species as to maximize diversity in TTX toxicity and TTX sensitivity, based on prior reports and experiments in Tetraodontiformes (13, 19, 38). Two tetraodontiform families, Tetraodontidae and Diodontidae, contain the majority of tetrodotoxigenic species; our taxon sample will include as many representatives as we can obtain from the 16 tetraodontid and 2 diodontid genera in which TTX toxicity is known (Table 1) or suspected. Our sample will also include representatives of other tetraodontiform families, including Balistidae, Ostraciidae, Molidae, Triodontidae, and Monacanthidae, where TTX has been reported or suspected (Table 1), and where some groups are reported to show greater resistance to TTX than non-tetraodontiform species (but less than the most resistant tetraodontids – (13). Since TTX-resistant sodium channels need not be coupled with TTX toxicity of the organism itself, we will also sample species where toxicity is not known, including at least one representative of the Triacanthodidae (basal Tetraodontiform

family). Teleost outgroups will include the weakly electric gymnotiform *Sternopygus macrurus* for which 6 different sodium channel genes have been identified and sequenced, as well as the zebrafish *Danio rerio* in which a number of sodium channel sequences are known (44; A. Linares & A. Ribera, pers. comm.)

TETRAODONTIDAE			DIODONTIDAE		
<i>Canthigaster</i>	8	<i>Liosaccus</i>	1	<i>Chilomycterus</i>	7
<i>Amblyrhynchotes</i>	5	<i>Monotreta</i>	2	<i>Diodon</i>	3
<i>Arothron</i>	10	<i>Omegophora</i>	1	TRIODONTIDAE	
<i>Boesmanichthys</i>	1	<i>Sphaeroides</i>	6	<i>Triodon</i>	1
<i>Chelonodon</i>	3	<i>Fugu (Takifugu)</i>	13	MOLIDAE	
<i>Colomesus</i>	1	<i>Tetraodon</i>	1	<i>Mola</i>	1
<i>Ephippion</i>	1	<i>Torquigener</i>	6	<i>Ranzania</i>	1
<i>Lagocephalus</i>	6	<i>Xenopterus</i>	1	Other Tetraodontiformes	

**Table 1: Tetraodontiform genera in which tetrodotoxin (TTX) has been reported, with number of toxic species.** The taxon sample for the proposed research will include as many representatives from these genera as can be obtained, as they are the most likely candidates for possessing TTX-resistant sodium channels. However, additional taxa -- including non-toxic species of *Lagocephalus* and *Liosaccus*, as well as representatives of the tetraodontiform families Ostraciidae, Balistidae, and Monacanthidae, will be included in the sample. TTX toxicity information gathered from (4) and Fishbase (<http://www.fishbase.org>). For the organismic phylogeny (see below), our taxon sample will include much more diversity than indicated in this table, sampling as many genera as possible within each family to test family monophyly and to allow robust estimates of relationships in this group.

Initially, specimens for the organismic phylogeny and for genomic PCR of sodium channel genes can be obtained from a number of sources, including live or fresh die-off specimens from commercial aquarium suppliers, overseas fish merchants, and other researchers. For some species, live specimens for TTX binding experiments, RT-PCR, and the organismic phylogeny can be obtained from aquarium suppliers. However, to achieve the taxon diversity we have presented here, we will have to collect live specimens and/or tissues from natural marine and freshwater habitats. We have already received local sponsorship offers from Dr. Timothy Tricas at the University of Hawaii, and Dr. Shaun Collin at the University of Queensland, Brisbane Australia. All U.S. and international regulations will be adhered to when obtaining and transporting specimens, including the securing of all required collecting, export, and import permits. All specimens acquired live or reasonably fresh will be preserved in 100% ethanol after completion of tissue sampling, euthanasia (immersion in 1.0% 2-phenoxyethanol), and binding experiments. Ethanol effectively preserves tissues and nucleic acids as voucher specimens or as material for future genetic studies. All voucher specimens will be deposited in the Texas Memorial Museum collection of fishes, and tissue and DNA samples will be deposited in the Texas Memorial Museum frozen tissue collection. In addition, non-frozen ethanol-preserved tissue samples will be kept to permit future screening assays for TTX-producing bacterial symbionts using *in situ* taxon-specific probes of 16S rRNA (bacterial 16S probe library available from Dr. Michelle Nishiguchi at New Mexico State University). We do not propose to study the bacteria during the period of this proposal, but all specimens will be curated appropriately for future studies.

Live fish acquired from fish dealers or captured in the field will be housed at the University of Texas in aquarium facilities in the Zakon lab. We have a half-time technician who is responsible for fish care and feeding. In rare cases when fish are in ill health, we consult with fish veterinarians at Texas A&M University. Field-caught fish will be temporarily housed in aquaria at the University of Hawaii (Tricas lab) or the University of Queensland (Collin lab), depending on location of collection. Whenever possible, these fish will be brought back live to our Texas laboratory for tissue sampling; otherwise, tissue samples can be taken and prepared in the Tricas and Collin labs for processing back at the University of Texas.

**PCR and Sequencing** -- Sodium channel sequences will be cloned from fresh heart, muscle, and brain tissues using RT-PCR, with methods developed in the Zakon lab for this purpose in *Sternopygus*. If needed, custom primers will be developed for tetraodontiform fishes with the aid of sodium channel sequences in the *Fugu* and *Tetraodon* databases. PCR of total genomic extractions will also be used to obtain sequence for all copies of sodium channel genes in each species, for the purpose of counting these paralogs and giving data for phylogenetic analyses (see section IIB above). The cDNAs and amplified products from cloning and genomic PCR will be sequenced using the Big Dye version 3 sequencing system from Applied Biosystems, and read with an Applied Biosystems 3100 genetic analyzer. Forward and reverse chromatograms will be aligned and hand-checked in Sequencher version 4.1 by Gene Codes Corporation. Sodium channel sequences will be aligned to each other and to published protein sequences for *Fugu pardalis*, *Fugu rubripes*, *Sternopygus macrurus*, and *Danio rerio* using the multiple alignment algorithm in ClustalX 1.8 (49) and manual clean-up editing in MacClade 4.02 (50). The alignments will be compared to known sodium channel protein sequences to calculate correct codon positions, and will then be translated into amino acid sequences. The pore (P) regions of all 4 domains will be screened for candidate mutations which may block TTX activity (based on information from *Fugu pardalis* and mutagenesis studies of mammalian sodium channels), and for possible consensus sites for post-translational additions.

**Sequencing for Organismic Phylogeny** -- We will estimate the phylogenetic relationships of the species in our sample, as a framework for interpreting the evolution of sodium channel genes in the Tetraodontiformes. We regard this as a significant part of our project that will contribute greatly to future systematic and comparative studies of this group. Our taxon sample for this phase will also contain representatives from as many of the ~70 genera of Tetraodontiform fishes as we can obtain, to ensure sufficient sampling within major lineages to test their monophyly and relationships. Genomic extractions will follow a modified protocol for RNA extraction from (51) and used as templates to amplify three loci: ND-2, rhodopsin, and cytochrome *b*, through PCR using primers available from Dr. Thomas Near at the University of California at Davis. These loci have been found to contain varying levels of sequence variation which have been useful in prior studies for reconstructing phylogenetic relationships within and between teleost families (T. Near, pers. comm., see also 53-57).

Sequences for the organismic phylogeny will be generated and aligned using the same methods as for sodium channel sequences (above), and using previously published sequences for these three loci in teleosts to facilitate the alignments.

**Phylogenetic Analyses** -- Four separate data matrices will be produced from the nucleotide sequence alignments: one from the sodium channel gene sequences described in section II-B above, and three others from nucleotide sequence alignments for ND-2, rhodopsin, and cytochrome *b*. Alignment of these protein genes is not expected to be ambiguous, and alignments will be created using ClustalX (49) and previously published protein sequences for these genes. We will analyze each of these data sets using parsimony and Markov chain Monte Carlo (MCMC) Bayesian approaches. For the parsimony analyses, each matrix will be subjected to at least 500 heuristic search replicates (using different starting trees) in PAUP\*4.0

b8 (58). In the MCMC Bayesian analyses, we will run four Markov chains at different temperatures for well over 1,000,000 generations to thoroughly sample the solution space (using MrBayes; 59). The exact number of generations will be determined by examining the likelihood scores and model parameters to ensure that equilibrium has been reached before we begin to sample the chains. Once we have reached equilibrium, we will sample the chains once every 100 generations for 1,000,000 generations (to obtain a sample of 10,000 trees and associated model parameters). This sample will be used to compute the posterior probabilities of each inferred branch, as well as 95% confidence limits of branch lengths and model parameters. The number of model parameters to be estimated will be determined using ModelTest 3.06 (60) with approximations made from inferred parsimonious trees. For the parsimony analyses, clade robustness will be estimated by full heuristic bootstrap resampling (1000 replicates) and the calculation of Bremer indices using AutoDecay 4.0.2 (61). In optimal trees for sodium channel genes, it is expected that sequences will cluster with their orthologs from other taxa (as in Figure 5) and not with paralogs in the same taxa; paralog clustering within a taxon, if robust, will be interpreted as indicating a possible duplication event during the evolutionary history of that species (alternative hypotheses of paralogy and orthology will be tested through parametric bootstrapping, as described below). Further analyses will include published sodium channel sequences from mammals and other tetrapods, to place the results of our research into a broader evolutionary context for vertebrates.

To test specific hypotheses about gene duplication (including hypotheses of paralogy and orthology) and mutation events, we will use parametric bootstrapping analyses (62, 63). Using this approach, we will test our optimal solutions against the reasonable alternatives. For instance, if we find trees that support a paralogous relationship for sodium channel genes within a species or species group, we will test if the alternative arrangement of orthology (a tree that matches the species relationships based on the other sampled genes) can be rejected by our data. Although this approach will be computationally demanding (as it requires simulation and analysis of a large number of trees of many taxa), it will also allow explicit testing of paralogous and orthologous relationships among the sodium channel genes that we sample.

***TTX Binding Experiments*** -- **Fresh tissues will be used to assess and compare TTX binding activity in the study taxa.** These are intended as complementary studies to demonstrate actual TTX binding or insensitivity which can be associated with different sodium channel genotypes. Toxin binding experiments will abide by the National Institutes of Health and University of Texas guidelines for animal care, and procedures will be presented in advance to the University of Texas Institutional Animal Care and Use Committee. Following the methods described in (22) heart, brain, and skeletal muscle tissues will be dissected from fresh specimens, and used separately for comparative binding experiments with radiolabeled tetrodotoxin (11-3 H-TTX). In addition to providing experimental tests for hypotheses which attribute TTX resistance to amino acid substitutions resulting from point mutations, these binding experiments may also lead to the identification of presently unknown TTX binding sites.

Fish will be euthanized in 1% 2-phenoxyethanol following which tissues will be excised and membrane preparations made according to procedures described by (64, 65). Protein concentrations will be measured using the Lowrey method. Membrane fractions will be incubated with increasing concentrations of 3H-TTX (0.1 nM—50 nM) in 1 ml of buffered incubation medium. A second set of fractions will include an increasing concentration series with 1000x excess of unlabeled TTX to assess non-specific binding. Each sample will be run in triplicate. After 15 minutes of incubation at 1°C, the samples will be filtered and washed 3X to remove unbound TTX, and the bound radioactivity will be counted with a scintillation counter. Skatchard analysis will be used to measure the dissociation constants ( $K_d$ ) and number of binding sites ( $B_{max}$ ).

**Although TTX is a commonly used substance in neurological studies, it is recognized by the federal government as a weapons-grade toxin with potential use in chemical warfare.**

Following the events of Sept. 11, TTX is currently unavailable even to researchers due to recent anti-terrorism mandates precluding its sale and distribution. However, representatives at Perkin-Elmer (owner of New England Nuclear Labs) inform us that TTX should be available again before the end of 2002, when a TTX licensing system is expected to be in place. In the unlikely event that TTX continues to be unavailable during the 3-year course of this project, we wish to emphasize that the main comparative and analytical aims of this proposal can still be carried out successfully in absence of TTX binding experiments.

**Information Dissemination** -- Aside from submission of results to indexed journals, all data and phylogenetic results will be shared with relevant internet-accessible public biological databases. All sodium channel, ND-2, rhodopsin, and cytochrome *b* sequences will be submitted to GenBank (<http://www.ncbi.nlm.nih.gov/>) and their accession numbers provided in all printed and online journal submissions. Phylogenetic trees and data matrices for the organismic phylogeny will be submitted to TreeBase (<http://www.herbaria.harvard.edu/treebase/>) and to the Tree of Life web project (<http://phylogeny.arizona.edu/tree/phylogeny.html>). Experimental findings about species toxicity and TTX resistance will be shared with FishBase (<http://www.fishbase.org>).

#### **IV. Relationship of the Proposed Research to Ongoing Work in our Laboratories**

PI Zakon studies the neurobiological basis of behavior, especially the role of ion currents in shaping the electrical activity of neural circuits. He uses the weakly electric fish *Sternopygus* as his model system, and focuses on the function and regulation of sodium channels in their electric organs using biophysical, biochemical, and molecular methods. He began a collaboration with David Hillis to study the evolution of sodium channel genes in vertebrates as an outgrowth of this work. This work represents the fusion of concepts and techniques of molecular evolution with neurobiology. The project we are proposing here is a further outgrowth of that effort.

Co-PI Hillis studies many aspects of phylogenetics and its applications to biological problems. In experimental studies, he examines the effects of parallel selection on the efficacy of phylogenetic analysis. This study of sodium channel evolution (and the co-evolution of sodium channels with the acquisition of TTX) is a natural empirical extension of this work, and represents an excellent system to study parallel evolution in a phylogenetic framework. In addition, this system represents an excellent opportunity to develop algorithms and methodology for dealing with the problems of orthologous and paralogous genes.

Co-PI Jost studies the evolution of whole organisms using a phylogenetic and comparative approach, and has a special interest in parallel evolution of genes and phenotypes. Her dissertation research utilized molecular phylogenetics and statistical methods to test hypotheses on the evolution of functional morphology in animal communication. As the postdoctoral researcher on our project, the collaborative nature of this proposal provides her with a unique and exciting opportunity to merge knowledge and approaches from two fields which rarely overlap: molecular evolution/ systematics, and animal physiology at the cellular and molecular level. In addition, the proposed research has enormous potential for expansion in a number of directions that Dr. Jost is interested in exploring in the future, including gene family evolution in vertebrates, the evolution of TTX use and resistance through bacterial symbiosis, and the systematics of teleost fishes.

#### **V. Future Directions**

A project as cross-disciplinary as this has the potential to develop in a number of directions beyond the scope of this proposal. Although the biological problems we are

interested in are significant in their own right, we also suggest that current national defense strategies against biological terrorism would benefit from any study that further illuminates the natural sources and activity of TTX, as well as mechanisms of biological resistance to this weapons-grade toxin. One interesting avenue for research would be to identify toxin-producing bacterial species and study the symbioses with their animal hosts. This would include both phylogenetic analyses of hosts and putative symbionts as well as physiological studies of their interactions. Another direction would be to study the evolution of sodium channel paralogs in other vertebrates that are tetrodotoxic (such as amphibians) and/or show TTX-resistance in one or more Na<sup>+</sup> channel genes (such as some Na<sup>+</sup> channels in mammals), to determine the wider applicability of our findings. A third possibility would be a functional analysis of the amino acid changes that have been proposed to confer TTX-insensitivity. Certain amino acid substitutions are known to be critical for TTX binding from site-directed mutagenesis studies of mammalian channels; however, it is possible that we will identify amino acid substitutions at sites not previously linked with TTX binding. In this case it would be informative to introduce these mutations into comparable residues in mammalian orthologs, or to clone these genes from cDNA libraries available from the *Fugu* genome project, to study the effects of these additional sites. Finally, TTX and the structurally related toxin saxitoxin (STX) require further studies regarding their roles in marine ecology. STX, produced by dinoflagellates, is responsible for red tides (16) while TTX, (probably) produced by bacteria, has been implicated in large scale sea urchin die offs (66) which indirectly impact the health of coral reefs. In fact, we are interested in the hypothesis that periodic exposure to red tides (and STX) may have been an initial selection pressure on marine organisms to evolve insensitivity to channel-blocking toxins, thus potentially providing a pre-adaptation to coping with TTX.

## **VI. Results from Prior NSF Support: David Hillis**

### **1995-1998: National Science Foundation, DEB 9508987: "Experimental Phylogenies for Testing Phylogenetic Methods" (D. M. Hillis, J. Bull, and I. Mollineux) \$135,000.**

This project was designed to test the accuracy of major methods of phylogenetic analysis by creating laboratory phylogenies of viruses. Estimates of these known phylogenies were then produced by sequencing whole and partial viral genomes and subjecting the sequence data to a variety of phylogenetic analyses. The best estimates of phylogenies were produced by incorporating detailed models of sequence evolution into the phylogenetic analyses. Methods were developed to incorporate more biological information into the models, and to test alternative hypotheses in a phylogenetic framework. In particular, a likelihood-ratio test was developed and shown to be highly effective for testing phylogenetic hypotheses. The extent of parallel molecular evolution in response to parallel selection was assessed, and its effect on phylogenetic analyses was studied. Parallel molecular changes were shown to occur, including relatively complex traits such as major insertion/deletion events and associated substitutions. However, parallel molecular changes had an adverse effect on phylogenetic analyses only when real historical structure was absent from the experimental phylogenies. Phylogenetic estimates were shown to be accurate over a wide range of conditions, as long as taxonomic and character sampling were adequate for the problem addressed.

Publications supported in whole or part from this award:

- 1) Hillis, D. M. 1995. Approaches for assessing phylogenetic accuracy. *Systematic Biology* 44:3-16.

- 2) Huelsenbeck, J. P., D. M. Hillis, and R. Jones. 1996. Parametric bootstrapping in molecular phylogenetics: Applications and performance. Pp. 19-45 *in* *Molecular Zoology: Strategies and Protocols* (J. D. Ferraris and S. Palumbi, eds.). Wiley-Liss, New York.
- 3) Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic inference. Pp. 407-514 *in* *Molecular Systematics*, 2nd. ed. (D. M. Hillis, B. K. Mable, and C. Moritz, eds.). Sinauer, Sunderland, Massachusetts.
- 4) Hillis, D. M., B. K. Mable, and C. Moritz. 1996. Applications of molecular systematics and the future of the field. Pp. 515-543 *in* *Molecular Systematics*, 2nd. ed. (D. M. Hillis, B. K. Mable, and C. Moritz, eds.). Sinauer, Sunderland, Massachusetts.
- 5) Hillis, D. M. 1996. Inferring complex phylogenies. *Nature* 383:130-131.
- 6) Huelsenbeck, J. P., D. M. Hillis, and R. Nielsen. 1996. A likelihood-ratio test of monophyly. *Systematic Biology* 45:546-558.
- 7) Cunningham, C. W., K. Jeng, J. Husti, M. Badgett, I. J. Molineux, D. M. Hillis, and J. J. Bull. 1997. Parallel molecular evolution of deletions and nonsense mutations in bacteriophage T7. *Molecular Biology and Evolution* 14:113-116.
- 8) Hillis, D. M. 1997. Primer: Phylogenetic analysis. *Current Biology* 7:R129-R131.
- 9) Hillis, D. M. 1997. Are big trees indeed easy? *Trends in Ecology and Evolution* 12:358.
- 10) Bull, J. J., M. R. Badgett, H. A. Wichman, J. P. Huelsenbeck, D. M. Hillis, A. Gulati, C. Ho, and I. J. Molineux. 1997. Exceptional convergent evolution in a virus. *Genetics* 147:1497-1507.
- 11) Hillis, D. M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Systematic Biology* 48:3-8.
- 12) Cunningham, C. W., H. Zhu, and D. M. Hillis. 1998. Best-fit maximum likelihood models for phylogenetic inference: Empirical tests with known phylogenies. *Evolution* 52:978-987.

## VII. Results from Prior NSF Support: Manda C. Jost

### **2000-2002: National Science Foundation, DEB 0073187: "Dissertation Research: Heterochrony in Insects: Phylogeny and Evolution of Acoustic Characters in the True Crickets (Orthoptera: Grylloidea)" (N. Pierce, K. Shaw, and M. Jost) \$9989.**

Manda Jost's doctoral dissertation estimated phylogenetic relationships in the Ensifera (Hexapoda: Orthoptera) and the Grylloidea (Orthoptera: Ensifera) using sequence data from 3 ribosomal loci, and both maximum parsimony and Bayesian inference methods. Phylogenetic estimates were used to test evolutionary hypotheses by character evolution simulations, relative rates tests, and phylogeny-based comparative methods including independent contrasts and concentrated changes. The results showed that specialized acoustic morphology has been lost multiple times in parallel, most frequently together with a total reduction or loss of wings and auditory organs. Independent contrasts analyses for ontogenetically-ordered character states found strong evidence for developmental constraints between multiple morphological characters, consistent with the evolutionary process of paedomorphosis. This result is also supported by additional morphological, ontogenetic, phylogenetic, and ecological lines of evidence. The significance of this project was that it used new phylogenetic trees to test novel hypotheses in a well-studied group, finding good evidence for an important evolutionary process (heterochrony) that is rarely invoked in studies of insects. It also established a new link between the evolution of acoustic morphology and the evolution of flightlessness and aptery, the latter of which has been studied extensively by other researchers using both genetic and endocrinological approaches.

Six manuscripts are in preparation from this research, including two papers on phylogenetic relationships of Ensifera and Grylloidea (co-authored by Kerry Shaw); two papers on the evolution of acoustic characters in Grylloidea; and two taxonomy papers that will describe in total eleven new species discovered during the course of the project.