The Dissertation Committee for Alisha Kay Holloway

Certifies that this is the approved version of the following dissertation:

POLYPLOID ORIGINS, EXPERIMENTAL EVOLUTION OF GENE DUPLICATES, AND DUPLICATION AND DIVERGENCE OF REPRODUCTIVE GENES

Committee:

_____
David M. Hillis, Supervisor

_____
James J. Bull, Co-Supervisor

_____
David C. Cannatella

_____
C. Randall Linder

_____
Ulrich G. Mueller

POLYPLOID ORIGINS, EXPERIMENTAL EVOLUTION OF GENE DUPLICATES, AND DUPLICATION AND DIVERGENCE OF REPRODUCTIVE GENES

by

Alisha Kay Holloway, B.A. M.S.

Dissertation

Presented to the Faculty of the Graduate School of

the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2004

POLYPLOID ORIGINS, EXPERIMENTAL EVOLUTION OF GENE DUPLICATES, AND DUPLICATION AND DIVERGENCE OF REPRODUCTIVE GENES

Publication No._____

Alisha Kay Holloway, Ph.D.

The University of Texas at Austin, 2004

Supervisors: David M. Hillis and James J. Bull

Gene and genome duplication are major sources of material for evolutionary innovations in eukaryotes that provide opportunities for novel function, increased complexity, and rapid speciation. Single gene duplications are prevalent even in eukaryotes with small genomes. Whole genome duplication plays an important role in the evolution of plants and some vertebrate groups.

I studied the origins of tetraploid gray tree frogs using molecular sequence data and advertisement calls. Molecular sequence data support multiple allopolyploid origins of tetraploids with a surprising twist. Apparently, tetraploids are composed of

a single interbreeding lineage that was created from at least three distinct diploid species, a paradigm of polyploid formation that has never been seen before.

The evolutionary fate of gene duplicates has been debated since the 1970s. Duplicate genes are maintained by purifying selection or evolve novel function. Drift, adaptive evolution of one duplicate, or subfunctionalization (the parsing of multiple ancestral functions between duplicates) leads to the diversification duplicates. Empirical evidence for all four of these models exists in nature, but the relative importance of each remains to be determined.

I developed an experimental model system in which the gene *β-lactamase* was duplicated. The system utilized the antibiotic resistance properties of the gene to test factors involved in maintenance and divergence of duplicates. Native *β-lactamase* confers resistance to ampicillin and to very low levels of cefotaxime. Mutations in *β-lactamase* allow it to confer increased resistance to cefotaxime, but decrease resistance to ampicillin, creating a tradeoff. I found that the mere existence of a tradeoff between old and new functions is insufficient for retention of the second copy; the environment must be such that the copy evolving novel function is no longer able to serve the original function.

Third, I examined the relative roles of novel function and subfunctionalization in evolution of duplicated male reproductive genes in *Drosophila*. Polymorphism and divergence data from these duplicated genes suggest that protein divergence between *D. melanogaster* and *D. simulans* is a result of adaptive evolution leading to novel

function. These data strengthen the conclusion that male reproductive genes may often be under directional selection in *Drosophila*.

Table of Contents

List of Figures……...…ix

List of Tables…….x

Chapter 1: Incidence and Evolution of Gene and Genome Duplications……..1

Chapter 2: Multiple Origins of a Single Tetraploid Species from Three Diploid Ancestral Species……..……..11

Abstract……11

Introduction……..11

Methods and Results……13

Discussion……16

Chapter 3: Experimental Evolution of Gene Duplicates in a Bacterial Plasmid Model……...20

Abstract……20

Introduction……..21

Materials and Methods…….…23

Culture Preparation……..23

Plate Selection……..24

Liquid Culture Selection……..24

Assays of Drug Resistance……...24

Sequence Analysis……..25

Competitive Fitness Assays……...……..25

Results……..26

Plate Selection……..26

Controls……26

Selection Experiments……27

Selection in Large Populations……27

Liquid Selection……...28

Controls……28

Selection Experiments…….28

Cost of Plasmids and Antibiotics…….29

Discussion……....30

Chapter 4: Molecular Evolution and Population Genetics of Duplicated Accessory Gland Protein Genes in *Drosophila*…….35

Abstract……35

Introduction……..35

Results and Discussion……36

Materials and Methods…….41

Figures……..43

Tables……...50

Appendix A……..56

References……69

Vita……...78

List of Figures

Figure 2.1. Distribution of: gray tree frogs, cytB and 65T allele lineages, and advertisement pulse rates……..…….…….…44

Figure 2.2. Bayesian inference of phylogenetic relationships of gray tree frogs and *H. avivoca*…….…..45

Figure 2.3. Model of tetraploid origins from diploid ancestors.....46

Figure 3.1. Tradeoff model of resistance to two antibiotics……...47

Figure 3.2. Empirical tradeoff in antibiotic resistance.....48

Figure 4.1. RT-PCR analysis of tissue-specific expression of putative duplicates....49

List of Tables

Table 1.1. Sexual polyploid amphibians……...…….……51

Table 3.1. Experimental design and resulting evolution…….…….…….....52

Table 3.2. Competitive fitness assays…….…….53

Table 4.1. Silent and replacement site heterozygosity and divergence for *Acp29AB* and *Acp53Ea* gene families in *D. melanogaster* and *D. simulans*…….…….54

Table 4.2. Silent and replacement variation for *Acp29AB* and *Acp53Ea* gene families in *D. melanogaster* and *D. simulans*……..55

Chapter 1: Incidence and Evolution of Gene and Genome Duplications

Gene and genome duplication have been recognized as major sources of material for evolutionary innovations in eukaryotes (OHNO 1970, 1999), providing opportunities for novel function, increased complexity, and rapid speciation. First, gene duplication provides raw material for the evolution of novel function by relaxation of constraints (OHNO 1970) or by directional selection (HUGHES 1994; LYNCH and CONERY 2000; WALSH 1995). Although gene loss is common after single gene duplications, it is estimated that half of genes evolved new functions following polyploidization in vertebrates (NADEAU and SANKOFF 1997). Second, duplications lead to increased complexity. In arthropods, tandem duplication of *Hox* genes led to increased segmental specialization (LEWIS 1978). In vertebrates, genome duplication enabled specialized tissues to evolve in concert with tissue-specific gene isoforms. These isoforms are virtually identical in structure and function, but are expressed in different tissues (KORTSCHAK *et al.* 2001; ONO-KOYANAGI *et al.* 2000). Third, duplication of genomes often leads to speciation due to incompatibilities with progenitors. Polyploidization often increases cell size which may influence morphology. Other incompatibilities include aberrant chromosomal segregation and lack of species recognition (BOGART 1980; KELLER and GERHARDT 2001; MULLER 1925; WHITE 1973). Speciation by duplication has also been extended to single gene duplications. Lynch and Force (LYNCH and FORCE 2000b) hypothesized that divergence of duplicates is an isolating

mechanism compatible with the idea of Bateson-Dobzhansky-Muller incompatibilities. Populations would become fixed for different alleles at duplicate loci, leading to hybrid incompatibilities.

Recent research has estimated the frequency of duplication events and the probability of maintaining duplicates. Lynch and Conery (2000) investigated the rates of duplication and frequency of novel function using genomic sequence data. They concluded that gene duplications occur at very high rates, but that most are silenced within a few million years. Long and Thornton (2001) responded to this claim with a different interpretation of the data and argued that a higher proportion of duplicate genes are maintained. Lack of genomic data presented serious impediments to empirical studies of the extent of gene duplication. The recent availability of completely sequenced genomes has facilitated deeper investigation of gene duplication by allowing genomes to be treated as searchable databases to identify putative duplicates. For example, recent analyses have revealed that even organisms with smaller genomes are highly duplicated. Roughly 20-50% of *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae* genes resulted from duplication events (GU *et al.* 2002; RUBIN *et al.* 2000).

Rather than gradual changes from single gene duplication, genome duplication may allow for saltational changes in the way organisms can respond to their environments. Polyploidization plays a major role in speciation and evolution of plants. Half of all plant species are either newly polyploid or derived

from polyploid ancestors. It has been estimated that 7% of new species in ferns and 2-4% of new species in angiosperms originate from polyploidization events (OTTO and WHITTON 2000). Contrary to the statement that "instances [of polyploidization] are so very extremely rare that they are quite insignificant in the general picture of animal evolution" (WHITE 1973), the volume of literature and current studies indicate that polyploidy is a rare, but important, evolutionary phenomenon. Ancient genome duplication events have been important in vertebrate lineages (OHNO 1970); however, *recent* polyploidization events in vertebrates are rare.

Because polyploidization in animals was thought to be an insignificant evolutionary process, studies have been biased towards understanding processes that select against polyploidy (MABLE 2004). Sex determination systems, complex developmental pathways, and spatial isolation of new polyploids may explain the paucity of polyploid vertebrates (MULLER 1925; ORR 1990; OTTO and WHITTON 2000; WHITE 1973). Molecular methods have reignited interest in mechanisms by which polyploidy arises and becomes established. Polyploid taxa are likely to be viable in cases of (1) asexual or hermaphroditic reproduction, (2) sex determination based on the presence of Y rather than on the X-to-autosome ratio, or (3) nondegenerate Y-chromosomes (ORR 1990; OTTO and WHITTON 2000). In addition, some taxa can bypass certain constraints; experimental *Xenopus* polyploids shifted to temperature-dependent sex determination (KOBEL and DU PASQUIER 1986).

Polyploids are concentrated in a few vertebrate groups. Except for one rodent, fishes and amphibians represent all known cases of sexual polyploids in vertebrates (OTTO and WHITTON 2000). In fishes, one clade (Actinopterygii, the ray-finned fishes), two families (Catostomidae and Salmonidae), and several lower taxonomic groups are polyploid (TURNER 1984). Among amphibians, sexual polyploid species are found in one salamander family, Sirenidae, and seven anuran families (DUELLMAN and TRUEB 1994; OTTO and WHITTON 2000) (Table 1.1). Two factors have facilitated the success of amphibian polyploids. First, sex chromosomes are weakly differentiated, which reduces problems due to dosage compensation (WHITE 1973). Second, polyploidization may produce instantaneous changes in advertisement calls that would provide immediate isolation from parental species (KELLER and GERHARDT 2001; UEDA 1993). In addition, certain temperature and pressure combinations cause incomplete meiosis in gametes, leading to increased rates of genome doubling (FERRIER and JAYLET 1978).

Polyploid organisms are formed either by duplication of a single genome (autopolyploidy) or fusion of two species' genomes (allopolyploidy). Questions remain as to which of these two processes of polyploid formation is more prevalent. Allopolyploidization may predominate in animals. For example, African clawed frogs (Pipidae: *Xenopus*), several species of loach (Cobitidae: *Cobitis*), and all salmon (Salmonidae) have allopolyploid origins (BORON and DANILKIEWICZ 1998; EVANS *et al.* 2004; MEZHZHERIN and CHUDAKOROVA 2002;

OTTO and WHITTON 2000; SLECHTOVA *et al.* 2000). In plants, autopolyploidization may be more common (RAMSEY and SCHEMSKE 1998). It also appears that repeated origins of the same allopolyploid species occur more often than single origins (OTTO and WHITTON 2000; SOLTIS and SOLTIS 1999; SOLTIS and SOLTIS 2000).

In Chapter 2, I examine the process of polyploid formation in gray tree frogs (*Hyla versicolor* complex), which have been a model system for speciation by polyploidization (MAXSON *et al.* 1977; PTACEK *et al.* 1994; RALIN *et al.* 1983; RALIN and SELANDER 1979; ROMANO *et al.* 1987). Two reproductively isolated advertisement call types were found to differ in ploidy level (WASSERMAN 1970); the name *Hyla versicolor* was assigned to the slow-trilling tetraploids and the name *H. chrysoscelis* to the fast-trilling diploids. These frogs are morphologically indistinguishable, but mating call pulse repetition rate for diploids is twice that for tetraploids. The difference in mating calls causes premating reproductive isolation (GERHARDT *et al.* 1994). Several data sets, including allozymes (RALIN *et al.* 1983; RALIN and SELANDER 1979; ROMANO *et al.* 1987), cytological evidence (WILEY and LITTLE 2000), immunological analysis (MAXSON *et al.* 1977), and mitochondrial DNA sequences (PTACEK *et al.* 1994) have been used to support both hypotheses of single autopolyploid and multiple allopolyploid origins of tetraploids from diploid ancestors. I studied the origins of tetraploid gray tree frogs using molecular sequence data. Molecular sequence data support multiple allopolyploid origins of tetraploids with a surprising twist. Apparently,

tetraploids are composed of a single interbreeding lineage that was created from at least three distinct diploid species. This mechanism of polyploid formation has never been seen before. One interesting area of future research would be to compare evolution of duplicate genes following different types of polyploidization, but researchers are just beginning to understand the fate of single gene duplications.

The evolutionary forces controlling the fate of gene duplicates have been debated since the 1970s. At least four classes of models have been proposed to explain the evolution of gene duplicates. First, Ohno (1970) deduced that purifying selection slows gene evolution, and that duplication events relax genetic constraints, allowing one copy of the gene to accumulate previously nonviable mutations and possibly acquire a new role (e.g. hemoglobin and opsin gene families). Second, Hughes (1994) argued that duplicates appear to be under strong purifying selection and that accumulation of deleterious mutations in one copy is unlikely. Data from *Xenopus* paralogs support his hypothesis; synonymous mutation rates exceed nonsynonymous rates in *Xenopus*, indicating that purifying selection is maintaining gene duplicates (HUGHES and HUGHES 1993). A third hypothesis is supported by evidence from multi-gene families and suggests that evolution of novel function is not the result of chance fixation but directional selection (HUGHES 1994). In agreement with this, Walsh (1995) contended that in large populations, the probability of an advantageous allele reaching fixation before a null allele is high. Fixation of an advantageous allele

may lead the two copies down diverging paths and to the evolution of novel function in one copy.

Finally, although neofunction and gene loss are commonly cited as fates of duplicates, theory suggests that another mechanism, subfunctionalization, contributes to the preservation of duplicate genes (LYNCH and FORCE 2000a). In subfunctionalization, complementary degenerate mutations occur in duplicates, leading each duplicate to perform a different portion of the original task (FORCE *et al.* 1999). Examples include: engrailed genes in zebrafish, ZAG1 and ZMM2 in maize, Hox genes in mouse (FORCE *et al.* 1999), and many genes in cotton (ADAMS *et al.* 2003).

Despite the evolutionary importance of gene duplication, we know little about pathways involved in the retention and divergence of gene duplicates (*e.g.* (WALSH 2003). In Chapter 3, I investigate the model that relies on new, beneficial mutations driving adaptive divergence of duplicates (HUGHES 1994). Divergence in function should lead to a selective advantage for retaining both genes because each gene cannot perform the functions of the other copy. Although the divergence in function may occur before or after the duplicate becomes fixed, this study focuses on divergence following fixation.

My study design in Chapter 3 includes development of an experimental model system to test factors involved in maintenance and divergence of duplicates. My model utilized the antibiotic resistance properties of the gene *β-lactamase*.

This gene confers resistance to ampicillin by catalyzing the hydrolysis of the antibiotic to inactive products (PALZKILL *et al.* 1994). Native *β-lactamase* confers resistance to high levels of ampicillin (30mg/mL) and to very low levels of the cephalosporin-type antibiotic, cefotaxime (0.05ug/mL). Mutations in *β-lactamase* allow it to confer increased resistance to cephalosporin antibiotics, but decrease resistance to ampicillin, creating a tradeoff (LONG-MCGIE *et al.* 2000; PALZKILL *et al.* 1994). This tradeoff prevents a single copy of the gene from achieving as high a level of resistance to both antibiotics as could be achieved by two, diverged copies. However, whether such a tradeoff actually leads to the maintenance of two diverged copies depends on the severity of the tradeoff as well as the selection imposed by the environment. I explored the landscape of the model by varying antibiotic concentrations to estimate the fate of duplicates subjected to diversifying selection.

In the final chapter I investigate the relationship between functional properties of genes and patterns of evolution. The relative importance of different evolutionary processes for the retention and subsequent divergence of gene duplicates is unclear (*e.g.*, WALSH 2003). For example, does directional selection or subfunctionalization represent the primary mechanism by which duplicate genes within a particular functional group evolve? Additionally, the interaction between the various population genetic parameters that may influence retention of duplicates remains to be determined. For example, retention and adaptive

divergence of duplicates may be facilitated in large populations (such as most *Drosophila*), the rationale being that there are more new beneficial mutations per generation in larger populations (WALSH 1995; WALSH 2003). However, *Drosophila* genomes appear to be biased towards deletion of non-functional DNA (PETROV and HARTL 1998) which would tend to rapidly eliminate new duplications conferring no fitness advantage to their carrier. The relative importance of these two countervailing forces for *Drosophila* genome evolution is uncertain. Alternatively, new duplicates from classes of proteins under chronic directional selection may have a higher probability of escaping elimination because a higher proportion of new mutations may be beneficial in such genes compared to genes primarily under mutation-selection balance.

Given the speculation that reproduction-related proteins may experience directional selection more frequently than other proteins (CIVETTA and SINGH 1998; NURMINSKY *et al.* 1998; RANZ *et al.* 2003; SWANSON and VACQUIER 2002), I investigated patterns of duplication and divergence in *Drosophila* accessory gland protein genes (*Acps*). *Acps* are male-specific seminal fluid proteins that affect multiple aspects of female physiology and behavior. They may reduce female receptivity to re-mating, decrease female attractiveness to males, increase egg laying rates, increase female mortality rates, and affect storage and utilization of sperm (for review see WOLFNER 1997). *Acps* are more polymorphic and evolve more rapidly than most other proteins (BEGUN *et al.* 2000; CIVETTA and SINGH 1998; COULTHART and SINGH 1988; SWANSON 2001; THOMAS and SINGH 1992).

Three *Acps*, *Acp29AB*, *Acp26Aa*, and *Acp36DE* show evidence of directional selection in *D. melanogaster* and *D. simulans* (AGUADE 1999; BEGUN *et al.* 2000; TSAUR *et al.* 1998). I present BLAST analyses of the *D. melanogaster* genome and gene expression studies to identify putative duplicated *Acp* genes. I then use molecular population genetic data from these duplicates in *D. melanogaster* and *D. simulans* to infer the evolutionary forces acting on these proteins.

The three sections of my dissertation shed light on different evolutionary questions relating to gene duplication and the evolutionary processes following duplication events. Investigation of tetraploid gray tree frog origins revealed a novel paradigm for the formation of polyploids. Exploration of an experimental model of gene duplications revealed the importance of the environment in maintaining duplicate genes. Finally, adaptive evolution often drives divergence between duplicated accessory gland protein genes in *Drosophila*.

Chapter 2. Multiple Origins of a Single Tetraploid Species from Three Diploid Ancestral Species

ABSTRACT

Polyploidization is one of the few mechanisms that can produce instantaneous speciation. Among animals, gray tree frogs (*Hyla versicolor* complex) have been the source of considerable debate regarding speciation by polyploidization. Molecular evidence and advertisement calls indicate that tetraploid gray tree frogs originated multiple times from extant diploid gray tree frogs and two lineages of extinct tree frogs. Tetraploid lineages then merged through interbreeding to result in a single lineage. Multiple origins of tetraploid lineages from the same two diploid progenitors are common, but this is the first known instance of a single tetraploid lineage that originated from at least three diploid ancestors.

INTRODUCTION

Speciation by polyploidy is common in some groups of organisms, but rare in others, such as vertebrates (OTTO and WHITTON 2000; WHITE 1973). Even when polyploidy is rare, genome duplication events have played an important role in diversification of major lineages (OHNO 1970). Polyploid species are formed either through genome duplication of a single diploid species (autopolyploidy) or through the fusion of two or more diploid genomes (allopolyploidy). Allopolyploidization events typically result in new, isolated polyploid species [e.g. ferns (GRANT 1981); *Brassica* (SOLTIS and SOLTIS 1993); salmonids

(TURNER 1984); *Xenopus* (EVANS *et al.* 2004)]. All known instances of tetraploid species with multiple origins were formed repeatedly from the same pair of diploid species (see: OTTO and WHITTON 2000; SOLTIS and SOLTIS 1999, 2000). Here we show that a single interbreeding lineage of polyploid gray tree frogs has resulted from multiple origins of allotetraploids from three different diploid ancestors, two of which are now extinct. Extinct lineages are inferred from the presence of allele lineages only found in tetraploids.

Gray tree frogs (*Hyla versicolor* complex) have been a central model system for speciation by polyploidization in animals (MAXSON *et al.* 1977; PTACEK *et al.* 1994; RALIN *et al.* 1983; RALIN and SELANDER 1979; ROMANO *et al.* 1987). Two reproductively isolated advertisement call types have been shown to differ in ploidy level (WASSERMAN 1970); the name *Hyla versicolor* has been assigned to the slow-trilling tetraploids and the name *H. chrysoscelis* to the fast-trilling diploids. Diploid and tetraploid gray tree frogs are morphologically indistinguishable, but mating call pulse repetition rate of diploids is twice that of tetraploids (GERHARDT 1974; GERHARDT 1994; JOHNSON 1966; LITTLEJOHN *et al.* 1960). Several data sets, including allozymes (RALIN *et al.* 1983; RALIN and SELANDER 1979; ROMANO *et al.* 1987), cytological evidence (WILEY and LITTLE 2000), immunological analysis (MAXSON *et al.* 1977), and mitochondrial DNA sequences (PTACEK *et al.* 1994) have been used to support a wide variety of hypotheses concerning both the number and the nature of the origins of polyploids in this group. Interpretations of the data conflict as to whether diploids represent a

single or multiple species, whether polyploids originated once or multiple times, and whether polyploids are autopolyploid or allopolyploid.

METHODS & RESULTS

We studied the origins of polyploids in the gray tree frog complex by sequencing three nuclear genes and one mitochondrial gene in 39 diploid and 32 tetraploid frogs from throughout their ranges (Fig. 2.1A), as well as from seven *H. avivoca*, a close relative. Phylogenetic relationships were inferred with maximum parsimony and model-based Bayesian methods of inference. Nuclear genes were PCR amplified and products were cloned into vectors. We directly sequenced PCR products of nuclear and mitochondrial markers and sequenced at least four colonies from cloning reactions for each nuclear marker in each individual. Detailed materials and methods are available as supporting material in Appendix A.

Four well-supported *cytochromeB* (*cytB*) haplotype lineages were found (Fig. 2.2A), supporting previous findings that suggest multiple allopolyploid origins (PTACEK *et al.* 1994). These haplotype lineages were geographically distinct (Fig. 2.1C). Geographical groupings indicate where the majority of a particular lineage was found. Eastern *H. chrysoscelis*, southwestern tetraploids, and the majority of *H. avivoca* shared one haplotype lineage. Western *H. chrysoscelis* and tetraploids shared a second haplotype lineage. A third lineage was present in only northwestern tetraploids, and not in any *H. chrysoscelis* or *H. avivoca*. Additional sampling from the northeastern part of the range that had not

previously been included revealed a fourth haplotype lineage that was also shared with *H. avivoca* and was not found in any *H. chrysoscelis*.

Four allele lineages of the nuclear fragment 65T were found (Fig. 2.2B), with some geographical distinction concordant with the *cytB* data (Fig. 2.1C & D). One allele lineage was present in western tetraploids, all *H. chrysoscelis*, and a single *H. avivoca* from Louisiana. Two additional allele lineages were found in tetraploids, but were absent in *H. chrysoscelis* and *H. avivoca*; one was primarily northeastern and the other was widespread throughout the range of tetraploids. A fourth allele was found in all *H. avivoca*, but was not present in any gray tree frogs. Data from the internal transcribed spacer 1 (ITS1) region and the gene fragment 11T from the cDNA library support findings from *cytB* and 65T. Data for ITS1 and 11T are presented in supporting material in Appendix A.

Evidence for three distinct genomes within polyploids indicates that *H. chrysoscelis* and two extinct diploid species contributed to forming the polyploid lineage. In three markers (*cytB*, 65T, and ITS1), tetraploid gray tree frogs have allele lineages not represented by any *H. chrysoscelis* or *H. avivoca*. Geographic distribution of these markers strongly suggests the ancestral presence of three distinct, diploid species, two of which are now extinct. We hypothesize that the two extinct diploid lineages formerly occupied the northeastern and northwestern part of the range, with subsequent replacement by the tetraploids. The model of hybridization among all possible pairs of diploids is presented in Figure 2.3. This hybridization led to the formation of polyploids multiple times from the three

diploid lineages, but these polyploid lineages are not reproductively isolated from one another (ESPINOZA and NOOR 2002). Different combinations of the genes from the various diploids are present in extant polyploid populations (Fig. 2.1C & D). Supporting evidence for the two extinct lineages as separate species includes the high level of sequence divergence between haplotype lineages of tetraploids relative to within other hylid species. For example, pairwise distances for *cytB* haplotype lineages was 2-3.5% between tetraploid haplotype lineages vs. <2% within species for two other Nearctic hylids, *H. andersonii* and *H. femoralis*.

Hyla avivoca shares some haplotype and allele lineages at all markers with gray tree frogs from ancestral polymorphisms or through hybridization or both. However, there was greater overall pairwise sequence divergence between gray tree frogs and *H. avivoca* than within either group. On average for all markers, pairwise distance was 4.1% between gray tree frogs and *H. avivoca*, 2.5% within gray tree frogs, and 1.2% within *H. avivoca*. Additionally, nuclear alleles for *H. avivoca* are primarily phylogenetically distinct from gray tree frog alleles. Given these two lines of support, we conclude that *H. avivoca* was not a key player in the formation of tetraploids.

Our data support multiple events of allopolyploidization with subsequent mixing of polyploid lineages. If tetraploids were originally autotetraploids of their individual diploid species, we would expect large geographic areas of homozygous genotypes among the tetraploids. As illustrated in Figure 2.1D, this

is not the case. Combinations of allelic lineages predominate among the tetraploids.

Geographic structure in tetraploid advertisement call pulse rates support our hypothesis of multiple origins from different diploid lineages. Tetraploids with southern *cytB* haplotype lineages differ significantly in corrected pulse rate from the northwestern and northeastern lineages (Fig. 2.1B; ANOVA results: Overall $F= 836.29$, $df=5$, $p<0.0001$; NE vs. S $F=26.02$, $p<0.0001$; NW vs. S $F=55.67$, $p<0.0001$; NE vs. NW $F=2.47$, $p=0.1168$). The disparity in pulse rates may reflect differences in the pulse rates of the diploid parental species. Nevertheless, these differences are too small to be discriminated against by tetraploid females and would not prevent interbreeding (GERHARDT 1994). Some populations with intermediate pulse rates were found in southwestern Missouri and western Tennessee. The geographic structure of pulse rates support conclusions of the genetic analysis that there is gene flow among lineages of tetraploids that have multiple origins.

DISCUSSION

Multiple independent origins of reproductively compatible tetraploids from more than two diploid ancestors has not been reported, even in groups such as plants in which speciation by polyploidy is common (GRANT 1981; OTTO and WHITTON 2000). Given that the polyploid gray tree frogs are reproductively isolated from the diploids (GERHARDT *et al.* 1994; JOHNSON 1963; LITTLEJOHN *et al.* 1960), recognition and subsequent mating among the polyploids lineages

seems counterintuitive. Premating reproductive isolation between diploids and tetraploids occurs because tetraploids call with a pulse rate approximately half that of the diploids and females do not recognize other call types as mates (GERHARDT 1974; GERHARDT 1994; JOHNSON 1966; LITTLEJOHN *et al.* 1960).

All polyploid gray tree frogs (regardless of origins) have a slow pulse rate, suggesting that a reduction in pulse rate may be a direct consequence of the ploidy level (KELLER and GERHARDT 2001). Polyploid cells have approximately twice the volume of diploid cells, even though the overall body size of diploids and polyploids is indistinguishable (CASH and BOGART 1978). This suggests a possible mechanism for mate recognition among different polyploid lineages: polyploidy may lead to a predictable and consistent change in the advertisement call (perhaps as a direct result of increased cell size) (KELLER and GERHARDT 2001), which would result in automatic isolation of the polyploids from the diploids, but identification of other tetraploids as possible mates.

Changes in cell sizes result in changes in pulse rate. Triploids created in the lab from diploids showed pulse rates that were about 13% lower than diploid controls but still higher than the intermediate values of triploid hybrids (2Nx4N offspring) (KELLER and GERHARDT 2001). Autotetraploids of *Hyla japonica* had pulse rates that were about 20% lower than diploid controls (UEDA 1993). Thus, the change in cell size appears to be insufficient to account for all of the variation between diploids and tetraploids. Changes in cell size coupled with reinforcement

against mating with frogs of a different ploidy level may have driven the differences between diploids and tetraploids to their current levels.

Our hypothesis for the origins of polyploids resolves the apparent conflicts among previous data sets on gray tree frogs. Although allozyme data have been used to support a single origin hypothesis of the polyploids, many polyploid populations contain alleles that are not found in any of the extant *H. chrysoscelis* populations (RALIN *et al.* 1983; RALIN and SELANDER 1979; ROMANO *et al.* 1987), which supports the hypothesis we have presented of extinct diploid species. Studies of both immunological and mitochondrial gene data also found evidence for tetraploid populations that were not closely related to any *H. chrysoscelis* sampled (MAXSON *et al.* 1977; PTACEK *et al.* 1994). Our study of both mitochondrial and nuclear loci from a comprehensive geographic sample demonstrates tetraploids originated from at least three diploid species, including the extant *H. chrysoscelis* and two extinct species. It is possible that *H. chrysoscelis* once contained all extant tetraploid alleles but that population extinction has reduced variability. However, the probability of finding multiple tetraploid-only alleles at every locus is extremely low. Regardless, our hypothesis that all pairwise combinations of three diploid lineages generated a single lineage of tetraploid gray tree frogs is the most straightforward and consistent with the data.

The multiple origins of a single, reproductively compatible polyploid species from different combinations of three diploid ancestors is unexpected (and

previously unreported). Many polyploids show evidence of recurrent formation, but these instances of multiple origins have always resulted from crossing of the same two diploid species, although sometimes from different populations (BELL 1982; OTTO and WHITTON 2000; SOLTIS and SOLTIS 1999, 2000). However, in organisms in which there is a direct relationship between polyploidy and critical aspects of the advertisement call (or other features that are important for reproductive recognition and isolation), automatic reproductive recognition and isolation associated with increased ploidy level may be expected. The connection between polyploidy and features associated with reproductive recognition and isolation may be part of the explanation for the increased frequency of polyploidy in frogs.

Chapter 3. Experimental Evolution of Gene Duplicates in a Bacterial Plasmid Model

ABSTRACT

An experimental bacterial system was developed to explore the fate of gene duplicates when gene copies are subjected to diversifying selection. The wild-type sequence of the gene *β-lactamase* confers resistance to ampicillin and to extremely low levels of cefotaxime. However, a few specific point mutations in the gene increase cefotaxime resistance but compromise ampicillin resistance. A bacterium with two copies of *β-lactamase* was selected for increased cefotaxime resistance while maintaining selection for ampicillin resistance. Our expectation was that one copy should diverge to improve cefotaxime resistance, but the other copy should be maintained to serve the original function of ampicillin resistance. Two "copies" of *β-lactamase* were introduced that were carried on separate, compatible plasmids, pBR322 and pACYC177. Stepwise selection for cefotaxime resistance resulted in the expected mutations in liquid cultures, but not on plates. As expected, increased cefotaxime resistance conveyed a cost to ampicillin resistance. This cost was not extreme enough to favor maintenance of the wild-type plasmid, given the doses of ampicillin that could be feasibly applied and the high cost of carrying the second plasmid. The gene duplicate was therefore lost. This study highlights the importance of the quantitative nature of the trade-off in the evolution of gene duplication through functional divergence.

INTRODUCTION

Gene and genome duplication have been recognized as major sources of material for evolutionary innovations in eukaryotes (OHNO 1970, 1999). Recent analyses have revealed that even organisms with compact genomes have high levels of duplication. Roughly a fifth to a half of *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae* functional genes are the result of duplication events (GU *et al.* 2002; RUBIN *et al.* 2000). These duplications allow for increases in complexity. In arthropods, tandem duplication of *Hox* genes led to increased segmental specialization (LEWIS 1978). In vertebrates, genome duplication enabled specialized tissues to evolve in concert with tissue-specific gene isoforms (KORTSCHAK *et al.* 2001; ONO-KOYANAGI *et al.* 2000).

Despite the evolutionary importance of gene duplication, we know little about pathways involved in the retention and divergence of gene duplicates (*e.g.* WALSH 2003). Under one model, gene duplicates subfunctionalize (LYNCH and FORCE 2000a), with each duplicate performing part of the original function (e.g. differential tissue-specific expression). We investigated a second model that relies on new, beneficial mutations driving adaptive divergence of duplicates (HUGHES 1994). Under either model, divergence in function would lead to a selective advantage for retaining both genes because each gene cannot perform the functions of the other copy. Although the divergence in function may occur

before or after the duplicate's sojourn to fixation, this study focuses on divergence following fixation.

We developed an experimental model system to test factors involved in maintenance and divergence of duplicates. Our model utilized the antibiotic resistance properties of the gene *β-lactamase*. This gene confers resistance to ampicillin by catalyzing the hydrolysis of the antibiotic to inactive products (PALZKILL *et al.* 1994). Native *β-lactamase* confers resistance to high levels of ampicillin (30mg/mL) and to very low levels of the cephalosporin-type antibiotic, cefotaxime (0.05ug/mL). Mutations in *β-lactamase* allow it to confer increased resistance to cephalosporin antibiotics, but decrease resistance to ampicillin, creating a tradeoff (LONG-MCGIE *et al.* 2000; PALZKILL *et al.* 1994). This tradeoff prevents a single copy of the gene from achieving as high a level of resistance to both antibiotics as could be achieved by two, diverged copies (Fig. 3.1). However, whether such a tradeoff actually leads to the maintenance of two diverged copies depends on the shape of the tradeoff as well as the selection imposed by the environment (Fig. 3.1C).

The conceptual foundation for our design follows the fitness set methodology introduced by Levins for two environments (LEVINS 1968) (Fig. 3.1), with some simple modifications. A specific cell or plasmid type has an intrinsic level of resistance to each antibiotic, which we may represent as the maximal concentration at which the cell (containing the plasmid) is able to grow. The joint

maximal tolerated concentrations of two antibiotics may thus be plotted as a point in this space for gene copy. Here we explore the landscape of the model by varying antibiotic concentrations to estimate the fate of duplicates subjected to diversifying selection.

MATERIALS AND METHODS

Culture preparation: We transformed bacterial cells to carry two different low copy number (~10/cell) plasmids, pACYC177 and pBR322, both of which have the gene *β-lactamase*. We treated these different copies of *β-lactamase* that are carried on the two plasmids as a gene duplication event. pBR322 also carries the gene for tetracycline (*Tet*) resistance and pACYC177 carries the gene for kanamycin (*Kn*) resistance. These additional antibiotic resistance genes allowed us to detect the presence of different copies of the *β-lactamase* gene in cell cultures.

After transformation, cells containing both plasmids were isolated by plating 1 μl of suspended cells on LB-agar plates with tetracycline (12.5 μg/mL) and kanamycin (25 μg/mL). Individual colonies isolated on these double antibiotics were grown in LB with ampicillin (1mg/mL) to log-phase and frozen in 20% glycerol. These aliquots were used as starting stocks for all experiments.

The strain of *E. coli* cells, W3110, has slight native antibiotic resistance to ampicillin and cefotaxime (1 μg/mL and 5 ng/mL, respectively), orders of magnitude lower than the concentrations used for selection. Cells can evolve

some minimal amount of resistance to antibiotics by changing lipopolysaccharides on the outer membrane (LIVERMORE and PITT 1986). Therefore, we periodically transformed plasmids into naïve cells to eliminate host evolution as a factor for increasing resistance to antibiotics.

Plate selection: In four experiments (P1-P4) cells were plated on LB-agar plates with cefotaxime and/or ampicillin and were incubated at 37ºC for 24 h (Table 3.1). Following incubation, 1mL of LB was added to the plates, which were then scraped to make a concentrated solution of bacterial colonies. These concentrated solutions were mixed vigorously in Eppendorf tubes and then frozen at -80ºC with 20% glycerol added. Prior to freezing, an aliquot of cells was diluted and spread on a new plate for the next passage.

Liquid culture selection: Four experiments were conducted in liquid media (LC1-LC3, Table 3.1). Approximately 10^6 suspended cells were added to tubes containing 2 mL of LB with the appropriate antibiotics (see Table 3.1). Cells were grown at 37°C under aerobic conditions for 8 h (to log-phase). Every 8 h, approximately 10^6 cells were transferred to new tubes with the same antibiotic treatments. Aliquots of cultures at log-phase were frozen with 20% glycerol.

Assays of drug resistance: In experiments P3, P4, and LC3, parallel cultures were grown in two concentrations of cefotaxime: one that was known to permit growth and one 5-10-fold higher than that known permissible concentration. If the culture at the higher concentration grew, it was used to

continue the line. Periodically, plasmid DNA was isolated (Wizard® *Plus* SV Minipreps DNA Purification System, A1330, Promega, Inc.) and reinserted into fresh W3110 cells to check evolution of host resistance; passages were continued using these naïve cells. Increase in resistance was assayed by growing cells in several different concentrations of cefotaxime above the known resistance level. Plasmid genes were sequenced at each stage of detectable increase in resistance to cefotaxime that was not due to host evolution.

Sequence analysis: From PCR products, we sequenced *β-lactamase*, the ORI (origin of replication initiation) and ROP (replication of plasmid; only present in pBR322). The ORI and ROP were monitored to determine if copy number of the plasmid increased. Sequences were obtained from PCR products. Sequence files were manipulated with SeqMan and aligned in MegAlign (DNAStar, Inc. 1993-2000).

Competitive fitness assays: Three types of cell stocks were prepared: i) cells carrying just pACYC177, ii) cells carrying just pBR322, or iii) cells carrying both plasmids. These were grown as overnight cultures with ampicillin plus the other appropriate antibiotic(s). The overnight cultures were then transferred to new media with 0.2mg/mL ampicillin and grown separately for an additional three hours to equilibrate the cultures to the same media.

To measure the differential cost of each plasmid type in liquid cultures, the three stocks were mixed with 5% of cells containing pACYC177, 5% of cells containing pBR322, and 90% of cells containing both plasmids. We expected cells

with a single plasmid type to grow faster, therefore, we used lower concentrations of single plasmid cells in order to more reliably measure increases in frequency. This culture was grown in ampicillin (0.1 mg/mL) for 8 h at 37ºC. After incubation, a dilution of cells was immediately plated on ampicillin and grown for 24h at 37°C, colonies were then replica-plated on ampicillin, kanamycin, and tetracycline to calculate the proportion of the total culture containing each plasmid type.

The evolved copy of *β-lactamase* was also competed against the wild-type. The assay began with 50% of cells containing pACYC177 with wild-type *β-lactamase* and 50% of cells containing pBR322 whose *β-lactamase* carried a single amino acid substitution, G238S (resistant to 0.5 μg/mL cefotaxime). This culture was grown in ampicillin (0.1 mg/mL) and cefotaxime (0.05 μg/mL) for 8 h at 37ºC. Platings at the start and end of the eight hours were used to measure the change in frequency of each plasmid type. Fitness was determined by the equation: $w = \frac{P^0(1-P^1)}{P^1(1-P^0)}$ where w is fitness, P^0 is initial frequency, and P^1 is final frequency, and $\sqrt[t]{w}$ is the fitness/generation.

RESULTS

Plate Selection

Controls: Experiments P1 and P2 were conducted as controls. Antibiotics were maintained at starting levels, so there was no apparent selection to maintain both plasmids. Thus, it was expected that one plasmid would be lost in both of

these experiments. Three of four replicates lost one copy of *β-lactamase* by passage 30 in the presence of ampicillin alone (Table 3.1). In the fourth replicate, the number of cells with both copies decreased with each passage and only 20% cells actually had both copies at passage 35. All five replicates of experiment P2 lost one copy of *β-lactamase* by 30 passages (Table 3.1). No change in resistance to cefotaxime was observed nor were there any changes in the sequence of *β-lactamase* or ORI.

Selection experiment: In P3 concentrations of cefotaxime were increased as resistance increased. All replicates gained 10-fold resistance to cefotaxime by 18 passages and one replicate lost one copy of *β-lactamase* (Table 3.1). When plasmids were purified and retransformed into naïve cells resistance to cefotaxime was reduced to original levels (0.05ug/mL) indicating that host cells had evolved. No changes were observed in the sequence of *β-lactamase* or ORI. Only one of four replicates in P3 lost one of the plasmids (pACYC177).

Selection in a large population: To determine the effect of population size on the retention of duplicates, population size was increased 10-100 fold for experiment P4. As in P3, cefotaxime concentrations were increased as resistance increased. After only eight passages, cellular evolution accounted for a 10-fold increase in resistance and all four replicates lost one of the plasmids. There was no change in the sequence of *β-lactamase* or ORI. In all plate experiments, the plasmid lost was pBR322.

Liquid selection

Controls: Experiments LC1 and LC2 (one trial each) were conducted as controls with antibiotics maintained at starting levels. Both LC1 and LC2 lost pBR322 by 44 passages and had no changes in the sequence of pACYC177 *β-lactamase* or ORI (Table 3.1).

Selection experiment: In LC3, cefotaxime concentration was increased as resistance increased. Following each increase in resistance, *β-lactamase*, ORI, and ROP were sequenced. Plasmids transformed into naïve cells indicated that increased resistance to cefotaxime was due to evolution of *β-lactamase*. In all replicates, three nonsynonymous substitutions accumulated in a step-wise progression in pBR322 (Fig. 3.2). pACYC177 was lost between the second and third mutations on pBR322 and there were no sequence changes in pACYC177.

The first nonsynonymous change (residue 238) resulted in an 8-fold increase in resistance to cefotaxime (Fig. 2). In one replicate, the change was from glycine (GGT) to serine (AGT), but in the other three replicates, the change was to alanine (GCT). This amino acid substitution resulted in no apparent decrease in resistance to ampicillin. At this point all replicates contained both plasmids.

The second mutation (S268G), in conjunction with the first mutation, resulted in a 50-fold increase in resistance to cefotaxime and a concomitant 33% decrease in resistance to ampicillin (Fig. 3.2). All four replicates showed this

change at high frequency. Additionally, the three replicates that previously had a G238A were now fixed for serine at that position.

In all replicates, pACYC177 was lost between the second and third mutations of pBR322. The third mutation was E104K. All four culture lines were polymorphic for this change and for S268G, meaning that there were some pBR322 within cells which did not have these mutations in *β-lactamase*. The third mutation, in conjunction with the other two, resulted in a 100-fold increase in resistance to cefotaxime, but a 66% decrease in resistance to ampicillin compared to wild-type (Fig. 3.2). There were no changes in ORI or ROP. Although there was a decrease in resistance to ampicillin when *β-lactamase* gained resistance to cefotaxime, the level of ampicillin in the experiment was lower than the inhibitory level (*i.e*. in region I of Fig. 3.1B).

Cost of plasmids and antibiotics

Competitive fitness assays were used to measure the possible advantage of losing a plasmid and the possible fitness advantage of the evolved plasmid in the original media. pBR322 was typically lost if no mutations arose. Therefore, we tested whether there was a difference in cost of maintaining one of the two plasmids. The fitness of cells containing only pACYC177 was 30% higher than cells containing only pBR322 (Table 3.2). This result holds for cultures grown in ampicillin as well as in ampicillin and cefotaxime at basal resistance levels.

In the first competition assay, ampicillin only and ampicillin plus cefotaxime cultures were started with the same aliquot of cells and therefore had

the same concentration of cells. After eight hours each of the two cultures were diluted and plated on ampicillin. The total numbers of colonies on ampicillin were counted. Cultures with wild-type *β-lactamase* grew 3-fold faster in ampicillin only than in the combination of antibiotics. Therefore, in the second competition assay, we tested whether the single mutation, G238S, conferred a fitness advantage over wild-type *β-lactamase* when ampicillin and cefotaxime were present at basal levels. The fitness of cells containing *β-lactamase* with the single mutation (on pBR322) was three-fold higher than those with wild-type *β-lactamase* (on pACYC177; Table 3.2).

DISCUSSION

Ohno's (1970) model for the evolution of gene duplicates states that the two copies of a gene in the original environment are redundant, hence one copy is free to evolve new function. The design of our experiment fits the scenario presented by Ohno, with some modification. In our system, the second copy of the gene was redundant until we pushed the drug concentrations to new levels, effectively changing the environment. In principle, the second copy should have diverged in response to our changing environment, with an unaltered copy being retained to maintain ampicillin resistance. The high cost of maintaining two plasmids resulted in the loss of one plasmid in both plate and liquid culture experiments. In some cases, one copy of *β-lactamase* did evolve novel function; however, this evolved copy was still able to serve the ancestral function

(inactivating ampicillin) which resulted in the loss of the wild-type plasmid (Fig. 3.2).

The model presented in Figure 3.1 illustrates the tradeoff between evolution of novel function and maintenance of the ancestral function of a gene. Whether duplicates will be maintained depends on the environment (gray zones in Fig. 3.1) and the tradeoff between loss of old function as new function evolved. In the case of liquid culture experiment LC3, the level of ampicillin used was so low that the tradeoff did not prevent loss of one copy. That is, the evolved copy of *β-lactamase* was in a similar position as point A′ in zone I of Figure 3.1C and was able to combat the effects of both antibiotics, even at elevated levels of cefotaxime. If the evolved copy would have been in zone III (A″ in Fig. 3.1C), both copies would have been maintained. Our second competitive fitness assays shows that the fitness of *β-lactamase* with the single mutation (G238S) is over 10-fold higher than that of the wild-type. Therefore cells containing only the evolved copy were so much more fit than wild-type, that the second plasmid was lost before the system could accumulate additional mutations that pushed the gene far enough along the tradeoff to enter zone III.

Fixed gene duplicates with neutral effects on fitness will be lost at a rate of $1-\frac{1}{Ne}$ due to drift; any fitness cost will increase this rate (WALSH 2003). In experiments with no plasmid evolution, pBR322 was always lost. Competitive fitness assays established that the two plasmids, our proxy for duplicate genes, had

differential fitness effects on cell growth. Cells containing pACYC177 were 30% more fit than those containing pBR322 in the presence of ampicillin and 46% more fit in a combination of ampicillin and cefotaxime. This asymmetry explains why pBR322 was lost. The exact mechanism for the difference in cost is unknown, but could be due to a difference in replication machinery between the two plasmids. Unequal costs of the duplicate present the most severe limitation of this two plasmid system of gene duplication; a limitation that could possibly be overcome by choosing plasmids with more similar effects on the fitness of the host. Nevertheless, a high symmetric cost of maintaining two different plasmids would still lead to rapid loss of one copy.

To determine the effect of population size on the retention of duplicates, population size was increased 10-100 fold for the plate experiment P4. In a large population a beneficial mutation has a much higher probability of occurring before the duplicate is lost. However, the high cost of maintaining the pBR322 plasmid outweighed this effect. In fact, large population sizes in plate experiment P4 increased the rate at which the costly duplicate (pBR322) was lost compared to plate experiment P3.

A mechanism by which the system could have evolved higher levels of resistance to cefotaxime (without mutations in *β-lactamase*) is to produce more of the enzyme. This could have been accomplished through two different processes that would effectively produce more mRNA. First, substitutions in regulatory regions of *β-lactamase* could elevate production of or stability of mRNA at the

level of the plasmid. Second, since both plasmids have relatively low copy number (~10 copies/cell), increased copy number would likewise produce more transcript. However, no changes in the sequence of the promoter region, ORI or ROP were observed in any experiment.

In liquid culture experiment LC3 some populations were polymorphic for the final two amino acid substitutions. This polymorphism could have been transient, as the new mutation increased in frequency. Conversely, it could also be stable because of density-dependent, frequency-dependent selection: because *β-lactamase* is diffusible, cells lacking the enzyme are somewhat protected by cells that produce this enzyme. In one scenario, a mutation that increases resistance to cefotaxime occurs and is at low frequency in the population. As the concentration of cefotaxime is increased, the mutated copy that confers greater resistance to cefotaxime increases in frequency in the population. The evolved *β-lactamase* enzyme leaks out of cells, so copies that do not produce the mutated enzyme may be protected from antibiotics and remain in the culture at low frequency.

This study has illustrated the feasibility and difficulties underlying one of the main models for the evolution of novel function through gene duplication and divergence. The mere existence of a tradeoff between old and new functions is not sufficient to lead to retention of the second copy. When the duplicate gene starts down the pathway toward acquisition of novel function, the environment must be such that the copy evolving novel function is no longer able to serve the

original function adequately. This would then provide some benefit to maintaining both copies of the gene. Yet for duplicate genes to be maintained during the period of redundant function, costs must be low. In this model system, the outcome of each experiment was driven by the cost of maintaining both duplicates. There is no empirical evidence in nature of the costs of gene duplicates in cases where the duplicate is incorporated into the genome.

Chapter 4. Molecular Evolution and Population Genetics of Duplicated Accessory Gland Protein Genes in *Drosophila*

ABSTRACT

To investigate the potential importance of gene duplication in *D. melanogaster* accessory gland protein (*Acp*) gene evolution we carried out a computational analysis comparing annotated *D. melanogaster Acp* genes to the entire *D. melanogaster* genome. We found that two known *Acp* genes are actually members of small multigene families. Polymorphism and divergence data from these duplicated genes suggest that in at least three cases, protein divergence between *D. melanogaster* and *D. simulans* is a result of directional selection. One putative *Acp* revealed by our computational analysis shows evidence of a recent selective sweep in a non-African population (but not in an African population). These data support the idea that selection on reproduction-related genes may drive divergence of populations within species, and strengthen the conclusion that *Acps* may often be under directional selection in *Drosophila*.

INTRODUCTION

At least three classes of models have been proposed to explain the evolutionary processes for the retention and subsequent divergence of gene duplicates. Lynch and Force (2000a) suggest that ancestral genes with multiple functions in different tissues or developmental stages may have high rates of retention of duplicates under mutation-selection balance. In this model,

degenerative mutations result in subfunctionalization, which favors retention and subsequent evolution of duplicates. A second class of models invokes fixation of duplications by genetic drift (WALSH 2003). Finally, a third class of models relies on new, beneficial mutations driving adaptive divergence (and thus retention) of duplicates (HUGHES 1994). One would expect new duplicates from classes of proteins under chronic directional selection to have unusually high fixation probabilities because a higher proportion of new mutations may be beneficial in such genes. For example, if reproduction-related proteins experience directional selection more frequently than other proteins (CIVETTA and SINGH 1998; NURMINSKY *et al.* 1998; RANZ *et al.* 2003; SWANSON and VACQUIER 2002), then perhaps a large number of duplicate reproduction-related genes spread through populations and diverge under directional selection.

RESULTS AND DISCUSSION

We investigated duplication and divergence in reproduction-related accessory gland proteins genes (*Acps*) in *Drosophila*. *Acps* are male-specific seminal fluid proteins that affect multiple aspects of female physiology and behavior (for review see WOLFNER 1997). We carried out BLAST comparisons of the 13 annotated *Acps* (see Methods) to the *D. melanogaster* reference sequence (FLYBASE 2003). These BLAST analyses suggested that two genes, *Acp29AB* and *Acp53Ea*, are members of small multigene families.

E-values returned from the tBLASTn search (default parameters) with *Acp29AB* as the query sequence were 1.5×10^{-47} and 2.6×10^{-35} for *Lectin29Ca*

and *Lectin30A*, respectively. Intraspecific paralogous protein divergence was, on average, 31% between *Acp29AB* and *Lectin29Ca*, 35.5% between *Acp29AB* and *Lectin30A*, and 38% between *Lectin29Ca* and *Lectin30A*. *Lectin29Ca* is 356 bases distal to *Acp29AB* and *Lectin30A* is approximately 1 Mbase distal to these tandem duplicates. *Acp29AB* is 234 amino acids, while *Lectin29Ca* and *Lectin30A* are 236 and 223 amino acids long, respectively. Each gene is composed of a single exon. Our analysis of *Lectin30A* and comparison to its paralogs suggested that the 5' end was incorrectly annotated. We confirmed this hypothesis by RACE, and used our annotation in all analyses. The three members of the *Acp29AB* family are predicted to be lectin galactose binding proteins (THEOPOLD *et al.* 1999) and to have a signal sequence, (SignalP v2.0, NIELSEN *et al.* 1997). The tBLASTn search returned several other more distantly related putative *Acp29AB* paralogs, primarily lectins (*Lectin 21Cb*, *Lectin 24Db*, *Lectin22C*, *Lectin 21Ca*, *Lectin24A*, *Lectin28C*, and CG15818). However, we will not present data from these genes in this report.

E-values returned from the tBLASTn search with *Acp53Ea* as the query sequence were 2.1×10^{-5} for CG8626 and 9.4×10^{-4} for CG15616. CG8626 and CG15616 will hereafter be referred to as *Acp53C14a* and *Acp53C14b*, respectively, based on putative function, genomic location and gene structure. Another more highly diverged putative duplicate identified by B. Wagstaff (pers. comm.) did not appear in our BLAST results. However, this gene (*Acp53C14c*) appears to be another tandem duplicate and shows male-limited expression (B.

Wagstaff, pers. comm.). Intraspecific paralogous protein divergence was 48.5% between *Acp53Ea* and *Acp53C14a*, 42.5% between *Acp53Ea* and *Acp53C14b*, 45% between *Acp53C14a* and *Acp53C14b*. The divergence of *Acp53C14c* from other putative *Acp53Ea* duplicates was >65%. These genes are tandem duplications, with *Acp53C14a* located 423 bp proximal to *Acp53C14b*, *Acp53Ea* 487 bases distal to *Acp53C14b*, and *Acp53C14c* 519 bp distal to *Acp53Ea*. *Acp53Ea, Acp53C14a*, *Acp53C14b*, *Acp53C14c* are predicted to be 120, 121, 132, and 124 amino acids long, respectively. Each is composed of two exons with a 50-60nt intron roughly 40 bases from the initiation codon. All genes are predicted to be peptide hormones and to have a signal sequence (SignalP v2.0, NIELSEN *et al.* 1997).

The high levels of silent and replacement divergence among putative paralogs suggest that the duplication events pre-date the split of *D. yakuba* from the *D. melanogaster*/*D. simulans* lineage. Nevertheless, the conserved genes structures, expression patterns, presence of predicted signal peptides, and for most cases, tandem organization, all indicate that we have correctly identified paralogous genes.

Acp29AB and *Acp53Ea* are expressed only in male accessory glands (WOLFNER *et al.* 1997). Our RT-PCR experiments showed that the only detectable expression of *Lectin29Ca*, *Acp53C14a*, and *Acp53C14b* is in accessory glands (Fig. 4.1). We were unable to detect an RT-PCR product from *Lectin30A*. However, given that our RACE products were derived from male cDNA, we are

certain the gene is expressed in males (perhaps at low levels). The fact that the *Acp* duplicates identified here share accessory-gland enriched expression further supports the inference of paralogy and suggests that subfunctionalization, at least with respect to gene expression, (*sensu* LYNCH and FORCE 2000a) cannot explain fixation of *Acp* duplicates. Levels of protein polymorphism and divergence for these *Acp* genes and putative duplicates (Table 4.1) were higher than those typically seen in *D. simulans* and *D. melanogaster* genes, as was the case for previous surveys of *Acp* variation (BEGUN *et al.* 2000; SWANSON 2001). There was, however, a major exception in the *Acp29AB* family.

Lectin29Ca had no silent polymorphisms and only a single replacement polymorphism in our US *D. melanogaster* sample (Table 4.2). This is highly unusual given the relatively high levels of variation in *D. melanogaster* generally, and in *Acp* genes specifically. Low levels of heterozygosity are even more surprising given high levels of silent and replacement divergence at this gene. We used the HKA test (HUDSON *et al.* 1987) to compare polymorphism and divergence data from *Lectin29Ca* and *vermilion* [we chose *vermilion* because of the availability of molecular population genetic data for both US and African samples and because there is no evidence that *vermilion* has been subject to directional or balancing selection (BEGUN and AQUADRO 1995)]. *Lectin29Ca* vs. the coding region of *vermilion* for the US sample showed a highly significant departure from neutrality ($p = 0.0022$), consistent with a hitchhiking event reducing variation in *Lectin29Ca*. In contrast, the African sample showed

considerably higher levels of polymorphism, but similar levels of divergence (Table 4.2). The HKA test of African variation at *Lectin29Ca* and *vermilion* showed no deviation from the neutral model ($p = 0.7875$). Given that the North American populations are thought to be recently derived from ancestral African populations (DAVID and CAPY 1988), the data support the idea that a selective sweep at *Lectin29Ca* occurred in the very recent past. Note that *Acp29AB* is only 356 bp upstream of *Lectin29Ca*, yet shows normal levels of heterozygosity in the California sample. This suggests that the window of reduced heterozygosity associated with *Lectin29Ca* may be quite small, though additional population genetic data from the other flanking region would be necessary to determine if this is indeed the case. It seems notable that of three solid cases of individual genes with evidence for recent selection in non-African *D. melanogaster* samples (*desat*: Takahashi et al. 2001, *Acp36DE*: Begun et al. 2000, *Lectin29Ca*: this report), two are *Acps*. Such observations support the notion that selection on sexual traits can cause rapid divergence of *Drosophila* populations (KNOWLES and MARKOW 2001; MILLER *et al.* 2003; PITNICK *et al.* 2003). Additional work will be required to precisely determine the extent of the "swept" region of *Lectin29Ca* in non-African populations and identify putative mutations that might be targets of selection.

Numbers of polymorphic and fixed, silent and replacement mutations were compared to predictions of the neutral model (MCDONALD and KREITMAN 1991). Of the six genes in our study, four (*Acp29AB*, *Lectin30A*, *Acp53C14b*, and *Acp53C14c*) reject the null hypothesis in a direction consistent with adaptive

protein evolution (Table 4.2). Moreover, our population genetic data support the notion that all three members of the *Acp29AB* family have been influenced by recent directional selection. Overall, our results put on firmer ground the conclusion that adaptive protein evolution is a major cause of divergence of *Acp* proteins in *D. melanogaster* and *D. simulans*.

MATERIALS AND METHODS

Amino acid sequences of 13 annotated *Acps* from *D. melanogaster* (*Acp26Aa, Acp26Ab, Acp29AB, Acp32CD, Acp33A, Acp36DE, Acp53Ea, Acp62F, Acp63F, Acp70A, Acp76A, Acp95EF, and Acp 98AB*) were compared to the *D. melanogaster* reference sequence (Genome Release 3.0, FLYBASE 2003). by tBLASTn searches using default parameters. Sequences were aligned by manual curation and molecular population genetics of putative duplicates with <50% nucleotide divergence from a known *Acp* was investigated. Nucleotide divergence (or uncorrected pairwise-distance) was calculated as a measure of the number of mismatched nucleotides/total number of nucleotides at 1st and 2nd codon positions. Duplicate genes are and considered putative *Acps*, but are referred to as *Acps* for convenience.

Candidate *Acp* duplicates were subjected to RT-PCR to determine whether their expression was restricted to accessory glands. mRNA was extracted from four tissues of *D. melanogaster*: male accessory glands, testes, male carcasses, and whole females. First strand synthesis was carried out using an oligo-dT primer and Superscript Reverse Transcriptase (Invitrogen). RT-PCR was carried

out using gene specific primers on RNA/DNA heteroduplex isolated from each tissue.

D. simulans population genetic data are from inbred lines established from flies collected at the Wolfskill Orchard in Winters CA (BEGUN and WHITLEY 2000). *D. melanogaster* population data are from isochromosomal lines derived from the Wolfskill Orchard or from isofemale lines from Malawi, Africa. *Drosophila yakuba* sequences are from an isofemale line. Some *D. simulans* *Acp29AB* sequences are from (BEGUN *et al.* 2000). In most cases DNA sequencing was carried out directly on PCR products. For cases in which inbred lines were not available, PCR products were cloned prior to sequencing.

Sequences were assembled using SeqMan (DNAstar, Inc.) and manually curated in MacClade 4.0 (MADDISON and MADDISON 2000). Alignments are available upon request from the authors. Summary statistics and tests of the neutral equilibrium model were carried out using DnaSP v. 3.53 (ROZAS and ROZAS 1999). SignalP Version 2.0 was used to predict presence/absence of signal peptides characteristic of *Acps* and other secreted proteins (NIELSEN *et al.* 1997). Sequences were submitted to GenBank under accession numbers AY635196-AY635290.

Figures

Figure 2.1. A) Distribution of *H. chrysoscelis* and tetraploid gray tree frogs, modified from (BLACKBURN *et al.* 2002; GERHARDT 1999). B) Population means of tetraploid pulse rates with colors representing *cytB* haplotype lineages. Distribution of: C) *cytB* haplotype lineages and D) nuclear marker 65T allele lineages for gray tree frogs and *H. avivoca*.

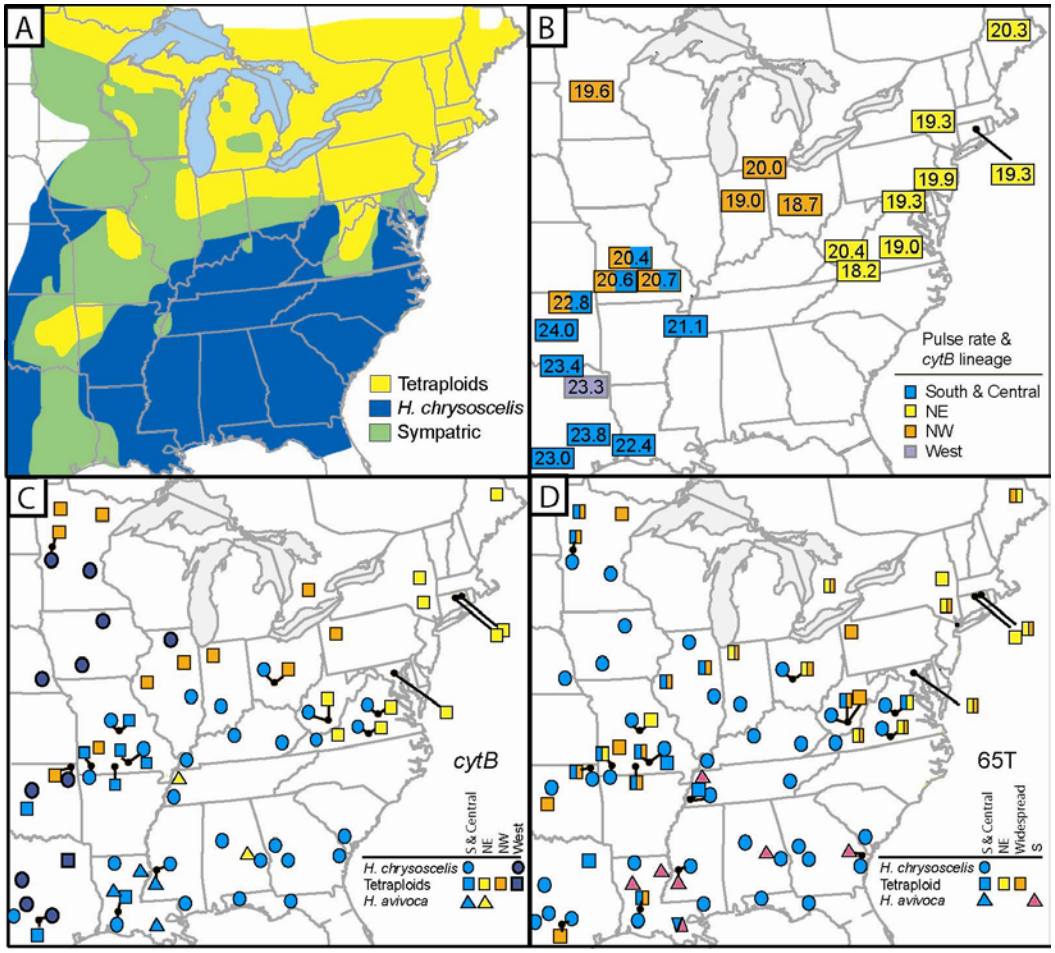


Figure 2.2. Bayesian inference of phylogenetic relationships of gray tree frogs and *H. avivoca*: A) *cytB* gene and B) nuclear marker 65T. Posterior probabilities next to branches (* represents >95% posterior probability), colors correspond to those in Fig. 1C and D.

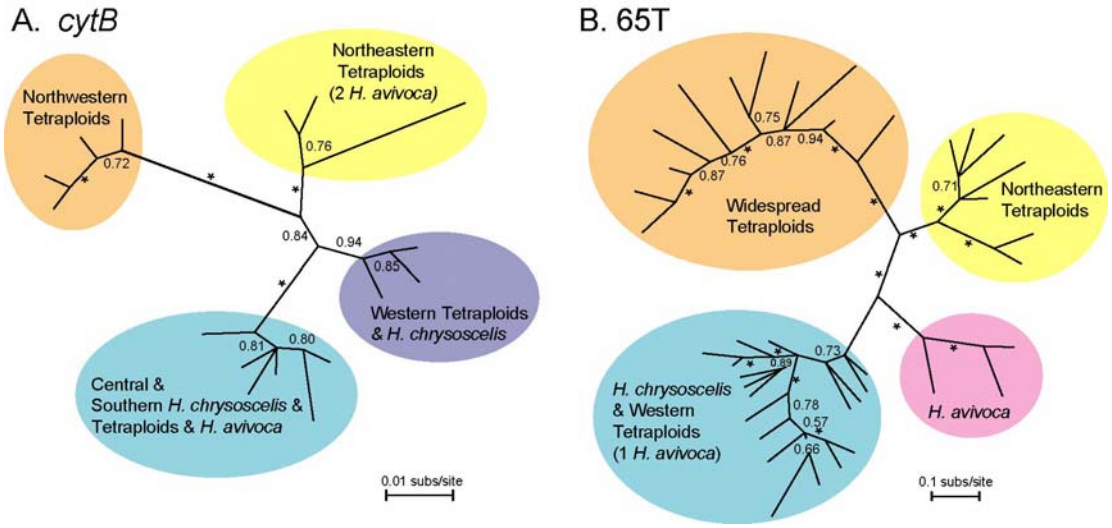


Figure 2.3. Model of tetraploid origins from diploid ancestors. The relationships of diploid gray tree frogs and *H. avivoca* are depicted in the tree at the bottom. Extinct diploids (2N), *H. sp. A* and *H. sp. B*, were inferred from tetraploid allele lineages. Tetraploid gray tree frogs, *H. versicolor* (4N), were formed multiple times from extinct diploid ancestors (AxB, AxC, and BxC). Tetraploid gray tree frogs now share a gene pool.

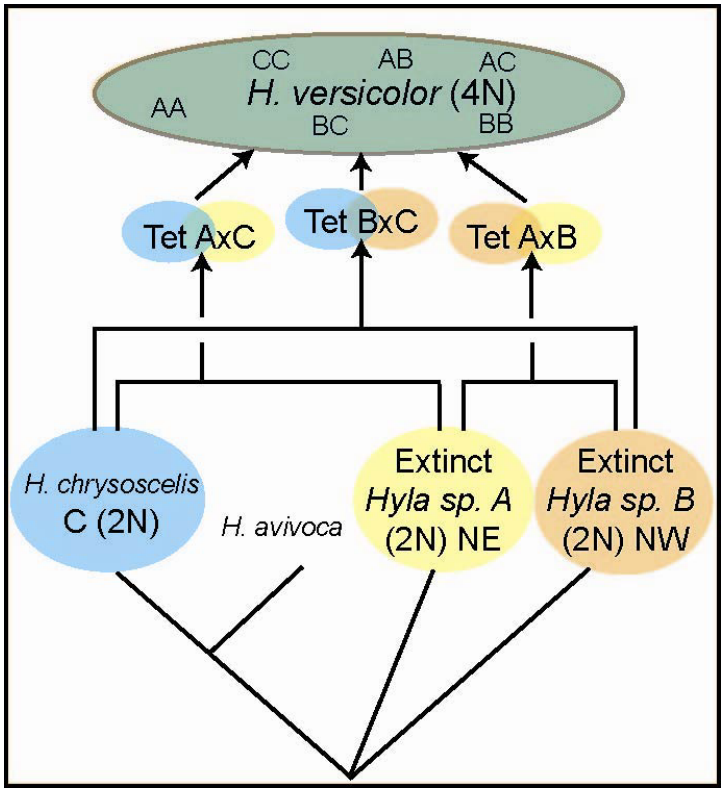


Figure 3.1. A) Tradeoff in drug resistance. Each point (A and A′) represents the joint level of resistance to each antibiotic separately. If A and A′ are different mutants of the same gene, they depict a tradeoff in drug resistance, because increases in resistance to one antibiotic causes a decrease in resistance to the other. B) Four zones in an environment containing fixed levels of the two antibiotics. Shaded areas represent concentrations of the two antibiotics in the environment where a specific plasmid would be inhibited by neither (I), one (II, III), or both (IV) antibiotics. In zones II and III an additional copy of the gene is necessary for host growth. For example, if the first copy is in zone II, a second copy of the gene with a joint level of resistance that resides in zone III would allow host growth. C) Conditional maintenance of a gene duplication. Starting at point A, if the drug resistance tradeoff follows the trajectory to A′, then a single copy confers resistance to both antibiotics in this environment. If the tradeoff boundary extends instead to A″, then no single copy will permit growth, and the cell must maintain two copies (A and A″) to persist in this environment.

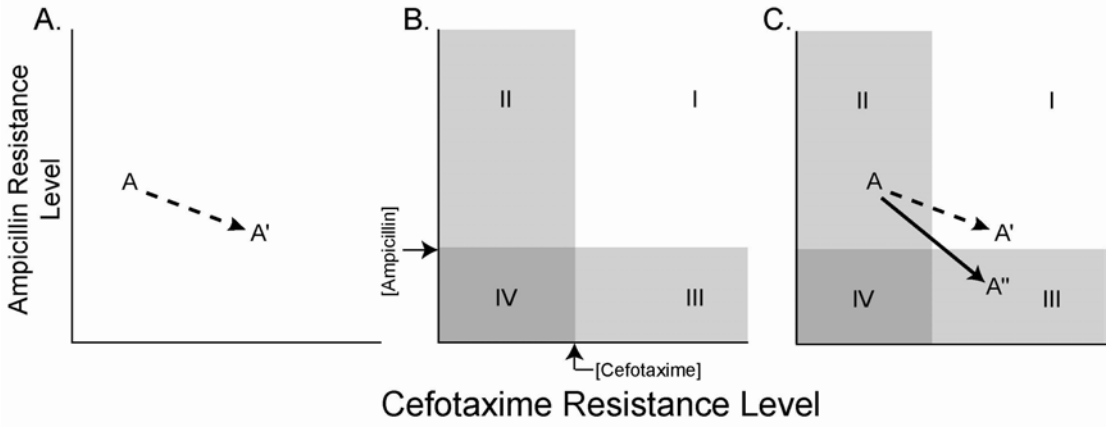


Figure 3.2. Resistance to ampicillin decreased as *β-lactamase* accumulated mutations that conferred a greater resistance to cefotaxime. Amino acid substitutions are noted along the line. Cultures were grown separately with each antibiotic to determine maximal levels that permit growth. Thin gray bar along the X-axis designates the concentration of ampicillin (0.1mg/mL), which was held constant during liquid culture experiment LC3. The level of ampicillin was too low to inhibit the mutated copy of *β-lactamase*. There was no tradeoff between accumulating mutations that increased resistance to cefotaxime and reduced resistance to ampicillin (*i.e.* not in area III from Fig 1).

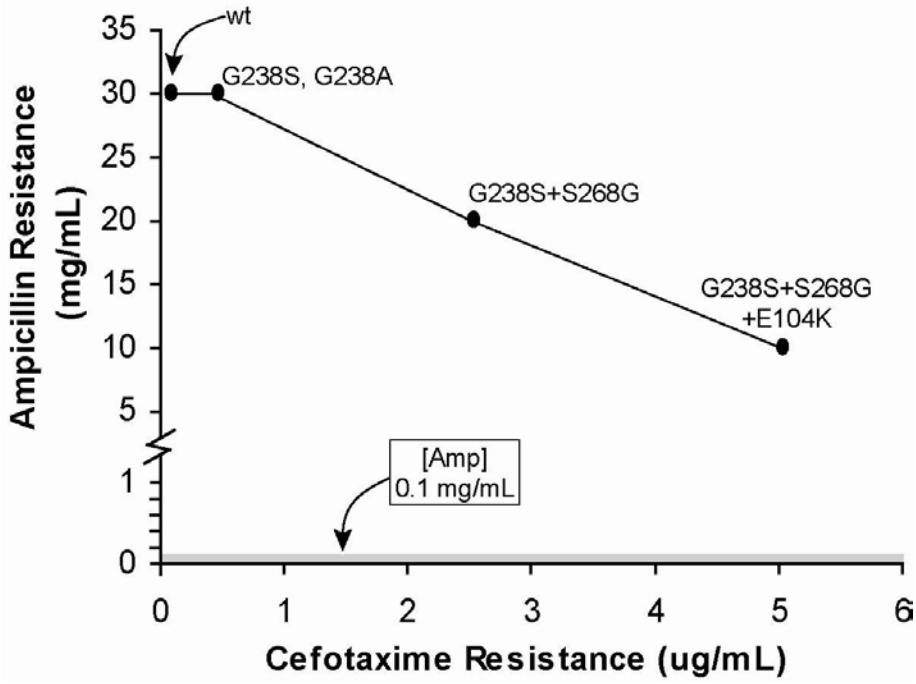
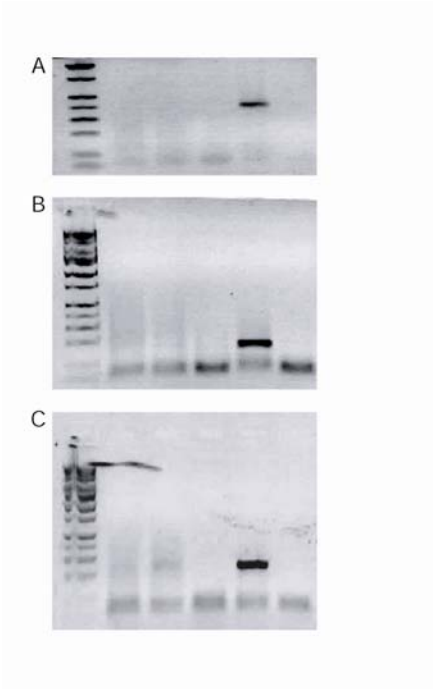


Figure 4.1. RT-PCR analysis of tissue-specific expression of putative duplicates. A) *Lectin29Ca*, B) *Acp53C14a*, and C) *Acp53C14b*. Lane assignments for each gel: 1) 1kB ladder, 2) whole females, 3) males without reproductive tracts, 4) testes, 5) accessory glands, and 6) negative control.



Tables

Table 1.1. Sexual polyploid amphibians
Bufonidae – *Bufo danatensis*, *B. viridis*, *Bufo* sp.
Hylidae – *Hyla versicolor*, *Phyllomedusa tetraploidea*
Leptodactylidae – *Ceratophrys aurita*, *C. ornata*, *Odontophrynus americanus*, *Eleutherodactylus binotatus*, *Pleurodema bibroni*, *P. kriegi*
Microhylidae – *Aphanthophryne* (*Cophixalus*) *pansus*, *Chiasmocleis leucosticta*
Myobatrachidae – *Neobatrachus sudelli*, *N. sutor*
Pipidae – *Xenopus* (common ancestor polyploid and multiple secondary events), *Silurana epitropicalis*
Ranidae – *Dicroglossus occipitalis*, *Pyxicephalus* (*Tomopterna*) *delalandii*
Sirenidae – *Siren lacertina, Pseudobranchus striatus*

Table 3.1 Experimental design and resulting evolution

Design					Outcome				
Experiment[a]	Pop. Size	[Amp.] mg/mL	[Cefo.][b] ug/mL	ΔT[c]	Plasmid[d] A	Plasmid[d] B	Plasmid[d] A+B	Final Cefo. Resist.[e]	*β-lactamase* Mutations
P1	10^2	0.1	–	24h	3	–	1	1x	–
P2	10^2	0.1	0.05	24h	5	–	–	1x	–
P3	10^2	0.1	0.05↑	24h	1	–	3	1x*	–
P4	$10^{3\text{-}4}$	0.1	0.05↑	24h	4	–	–	1x*	–
LC1	10^6	0.1	–	8h	1	–	–	1x	–
LC2	10^6	–	0.05	8h	1	–	–	1x	–
LC3	10^6	0.1	0.05↑	8h	–	4	–	100x	G238S+S268G+E104K

[a] P = plate; LC = liquid culture.

[b] ↑ denotes increased concentration of cefotaxime over time during experiment.

[c] ΔT signifies the time between transfers to new media.

[d] Number of replicates with each plasmid present in final assay (A = pACYC177; B = pBR322).

[e] * indicates 10x increase in resistance due to host cell evolution.

Table 3.2. Competitive fitness assays

		Proportion in Culture			Fitness/Gen.	
		T_0	T_8 Amp.	T_8 Amp. + Cefo.	Amp.	Amp. + Cefo.
$pACYC177_{wt}$ vs. $pBR322_{wt}$	$pACYC177_{wt}$	0.049	0.063	0.049	1.017	1.024
	$pBR322_{wt}$	0.069	0.067	0.049	1.000	1.000
$pBR322_{wt}$ vs. $pBR322_{G238S}$	$pBR322_{wt}$	0.410	–	0.170	–	1.000
	$pBR322_{G238S}$	0.590	–	0.830	–	1.165

T_8 = 8 hours after experiment began.

Table 4.1. Silent and replacement site heterozygosity and divergence for *Acp29AB* and *Acp53Ea* gene families in *D. melanogaster* and *D. simulans*.

Gene	No. of sites		Sample	θ_{sil}	θ_{repl}	$Div^a_{.sil}$	$Div^a_{.repl}$
	Sil.	Repl.					
Acp29AB	148	554	*mel*	0.0209	0.0028	0.2368	0.0772
			sim	0.0359	0.0052		
Lectin29Ca	143	559	*mel*	0.0000	0.0007	0.1745	0.0807
			Af *mel*[b]	0.0142	0.0116	0.1841	0.0848
			sim	0.0286	0.0073		
Lectin30A	137	532	*mel*	0.0112	0.0029	0.1262	0.0579
			sim	0.0178	0.0015		
Acp53Ea	88	266	*mel*	0.0399	0.0066	0.1574	0.0390
			sim	0.0046	0.0061		
Acp53C14a	85	278	*mel*	0.0174	0.0013	0.1255	0.0068
			sim	0.0289	0.0029		
Acp53C14b	93	303	*mel*	0.0435	0.0024	0.1689	0.0352
			sim	0.0307	0.0027		
Acp53C14c	92	280	*mel*	0.0401	0.0058	0.1387	0.0823
			sim	0.0595	0.0156		
19 other	286	935	*sim*	0.0349	0.0013	0.1084	0.0107
Genes 3R[c]	(19)	(65)		(0.0044)	(0.0003)	(0.0097)	(0.0032)

Sil = silent sites in coding regions, Repl = replacement.

[a] Divergence (Div.) is between all pairs of *D. melanogaster* and *D. simulans* genes (Jukes-Cantor corrected).

[b] Malawi, Africa.

[c] Data are means with SE values below in parentheses for 19 unrelated proteins on chromosome arm 3R from Begun and Whitley (2000) and Begun (2002).

Table 4.2. Silent and replacement variation for *Acp29AB* and *Acp53Ea* gene families in *D. melanogaster* and *D. simulans*.

Gene		Polymorphic		Fixed		Prob.[a]
		Silent	Repl.	Silent	Repl.	
Acp29AB	*sim*	13	7			
	mel	8	4	24	33	**0.032**
Lectin29Ca	*sim*	10	10			
	mel	0	1	20	38	0.292
	Af *mel*[b]	15	26	21	36	0.979
Lectin30A	*sim*	6	2			
	mel	4	4	12	27	**0.030**
Acp53Ea	*sim*	1	4			
	mel	8	4	10	7	0.730
Acp53C14a	*sim*	6	2			
	mel	4	1	5	1	0.746
Acp53C14b	*sim*	7	2			
	mel	11	2	8	9	**0.022**
Acp53C14c	*sim*	10	8			
	mel	9	4	6	16	**0.025**

[a] Probability determined by G-test, numbers in bold indicate $p<0.05$.
[b] Malawi, Africa.

Appendix A

Materials and Methods

Source material. See Table S1 for localities of samples used in analyses. *H. avivoca*, found in the south central US, is sister to gray tree frogs to the exclusion of other Nearctic *Hyla* (Holloway, Hillis, and Cannatella, unpub. manuscript).

Species identification. *Hyla chrysoscelis* and tetraploid frogs were distinguished by mating calls. Advertisement call pulse rate, analyzed with custom-designed software by H.C. Gerhardt, was corrected for temperature to a mean of 20ºC, which is near the middle of the range of breeding temperatures. The corrected pulse rates are completely non-overlapping allowing for unequivocal distinction between *H. chrysoscelis* and tetraploids. The highest mean tetraploid pulse rate in any population was 24 pulses/sec, and the lowest mean *H. chrysoscelis* pulse rate was 35 pulses/sec [(GERHARDT 2001), and this study].

Library Construction & Preliminary Data. A cDNA library was created from testes tissue of gray tree frogs (Clontech, Inc. #PT3577-1, Palo Alto, CA). Messenger RNA was extracted (Clontech, Inc., #K3064-1, Palo Alto, CA) and reverse transcribed to produce double stranded cDNA that was then ligated into plasmids. Plasmids were transformed into bacteria to produce colonies containing single genes. Single-gene copies were amplified and sequenced using plasmid primers (see Sequencing Protocol below). These sequences were compared to gene sequences in GenBank (ALTSCHUL *et al.* 1990; BOGUSKI *et al.* 1993). Genes that were either very conserved or from the mitochondrial genome were discarded. For genes that either

had no match or were divergent from other sequences in the database gene-specific primers were designed and two individuals from each of four populations were sequenced to examine variation.

DNA Extraction and PCR Conditions. Total genomic DNA was extracted from preserved liver, muscle, or toe clips using the DNeasy extraction kit (Qiagen, Inc., #69506, Valencia, CA). Thermal conditions for 25 µl PCR reactions included initial denaturation at 95°C for 2 min, followed by 35 cycles of denaturation at 94°C for 30 sec, annealing at variable temperatures for 30-45 sec (see Table S2), and extension at 72°C for 1 min, and finally one extension at 72°C for 10 min. The nuclear ribosomal subunit, internal transcribed spacer region 1 (ITS1), was amplified using the Failsafe PCR Buffer G (Epicentre, Inc., #FSP995G, Madison, WI). All other genes were amplified using stock buffers supplied with *Taq* polymerase (NEB, Inc., #M0267L, Beverly, MA). For nuclear genes, a small portion of PCR products was used in cloning reactions. The remainder of the PCR product of nuclear genes and the entire PCR product from *CytochromeB* was cleaned using the Qiaquick PCR purification kit (Qiagen, Inc., #28104, Valencia, CA). Concentration of clean PCR products was determined by electrophoresis on a 1% agarose gel with 0.3µg/mL ethidium bromide.

Cloning of PCR Products. Four µl of PCR product were used in each ligation reaction. Protocols for ligation and transformation followed those described in the cloning manual (Invitrogen, Inc., #K4500-40, Carlsbad, CA). Ligations were carried out for the maximum suggested time (30 min.) and transformations were carried out

for the minimum suggested time (5 min.). Transformed products were grown overnight at 37ºC on LB plates containing X-Gal (64 µg/mL) for blue/white screening and kanamycin (50 µg/mL). White colonies were picked, added directly to PCR cocktails, and amplified using the plasmid primers M13F and M13R from Invitrogen with an annealing temperature of 55ºC. PCR and cleaning protocols were as described above.

Sequencing and Alignment. Ten µl cycle sequencing reactions with BigDye V3.1 (ABI, Inc., Foster City, CA) as the dye terminator and gene-specific or plasmid primers were carried out for 25 cycles, each with denaturation at 94°C for 10 sec, annealing at 50°C for 5 sec, and extension at 60°C for 4 min. Dye terminators were removed with Sephadex G-50 (Sigma, #S-6022, St. Louis, MO) in Centrisep columns (Princeton Separations #CS-901, Adelphia, NJ). Sequencing was performed on an ABI3100 sequencer (Applied Biosystems, Foster City, CA). Sequences were compiled in Sequencher v4.0 (GeneCodes, Inc., Ann Arbor, MI) and consensus sequences were exported to ClustalX (THOMPSON *et al.* 1997) for preliminary alignment. Final alignment was done by eye using MacClade v4.06 (MADDISON and MADDISON 2000). Genbank accession numbers are listed in Table A3.

Exclusion of Recombinant Sequences. During the initial PCR amplification before cloning, partially extended products may prime another allele in the next round of amplification. In order to minimize the number of partial amplicons, we designed primers that amplified less than 1kb of each gene. Any chimeric sequences were

identified by comparing the sequences of multiple copies of alleles from each individual; these sequences were excluded from analyses. Additionally, we tested for recombination in the remaining sequences using the following methods: pairwise scanning in Recombination Detection Program (RDP) v2b.08, (MARTIN and RYBICKI 2000), and the Bayesian multiple change point model in OhBrother, (SUCHARD *et al.* 2002). Both programs take into account multiple tests. The only gene with evidence for recombination was ITS1. Parsimony informative characters were also examined by eye on the most parsimonious ITS1 tree in MacClade v4.06 (MADDISON and MADDISON 2000). Putative recombinant sequences (sequences with homoplasies in >15% of sites and/or sequences that were identified in RDP) were excluded. Tests for recombination on the remaining sequences indicated that no other sequences were recombinants.

Phylogenetic Analyses. Phylogenetic relationships among aligned sequences were examined with maximum parsimony and Bayesian methods of inference. Prior to analysis, sequences with less than 1% pairwise sequence divergence were pruned from each dataset. Heuristic searches for the most parsimonious trees using unweighted data were conducted using 100 replicates of random addition of taxa with PAUP* V4.0b10 (SWOFFORD 2002). Support for relationships was estimated by nonparametric parsimony bootstrap analysis with 1000 replications, each with a single replication of random addition of taxa (FELSENSTEIN 1985). For Bayesian analyses, models of evolution were evaluated with likelihood ratio tests of tree scores from parsimony trees in PAUP* V4.0b10 (SWOFFORD 2002). Analyses were

performed in MrBayes 3.04b (HUELSENBECK and RONQUIST 2001). We used four Markov chains, one cold and four hot chains at default temperature settings, for each run in order to traverse a larger area of tree space and avoid entrapment in local topological optima. The chain was sampled every 100 generations for 10 million generations for ITS1 and 5 million generations for 65T, 11T, and *cytB*. Bayesian posterior probabilities were estimated as the proportion of trees sampled after burn-in that contained each of the observed bipartitions (HUELSENBECK and RONQUIST 2001; LARGET and SIMON 1999). Four separate analyses for each gene were run to examine convergence of bipartition posterior probabilities. Bipartition posterior probabilities differed by less than 5% between runs.

Results and Discussion of Nuclear Markers ITS1 and 11T

Four allele lineages of ITS1 were found (Fig. A1A,C). One allele lineage was present in all *H. chrysoscelis*, *H. avivoca*, and some tetraploids. A second allele lineage was found in a single *H. chrysoscelis*. A third allele lineage was composed of a single tetraploid and one *H. chrysoscelis*. A fourth allele lineage was composed entirely of tetraploids. Although the ITS1 region was less informative than *cytB* and 65T, the results support the conclusions that extinct diploid lineages contributed to tetraploid formation.

The vast majority of diploids and tetraploids as well as one *H. avivoca* shared a similar genotype for gene fragment 11T (Fig. A1B,D). A different genotype was present in all other *H. avivoca*, one diploid and one tetraploid. The data for marker

11T support our conclusions that gene flow between *H. avivoca* and gray tree frogs is minimal and that *H. avivoca* were not major players in tetraploid origins.

Figure A1. Distribution of: A) ITS1 allele lineages and B) nuclear marker 11T allele lineages for gray tree frogs and *H. avivoca*. Bayesian inference of phylogenetic relationships of gray tree frogs and *H. avivoca*: C) ITS1 and D) nuclear marker 11T. Posterior probabilities next to branches (* represents >95% posterior probability).

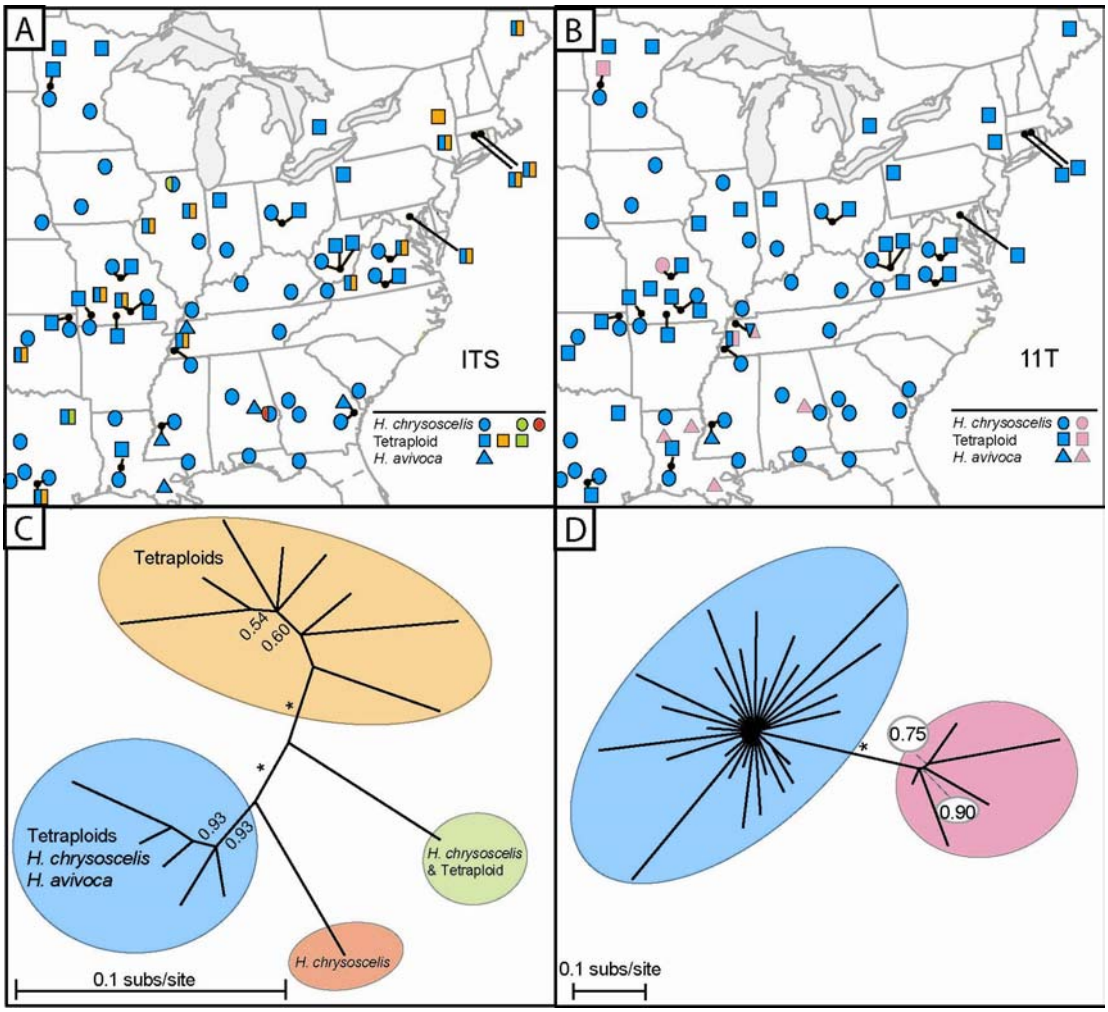


Table A1. Locality Information.

Species	Field #	Country	State	County
Hyla avivoca	MP710	USA:	AL:	Macon
Hyla avivoca	DCC3857	USA:	GA:	Chatham
Hyla avivoca	HCG21	USA:	LA:	Grant
Hyla avivoca	H146	USA:	LA:	Jefferson
Hyla avivoca	HCG84	USA:	LA:	Madison
Hyla avivoca	MP607	USA:	MS:	Hinds
Hyla avivoca	HCG28	USA:	TN:	Obion
Hyla chrysoscelis	DCC3883	USA:	AL:	Bibb
Hyla chrysoscelis	MP458	USA:	AL:	Russell
Hyla chrysoscelis	MP732	USA:	FL:	Leon
Hyla chrysoscelis	MP670	USA:	FL:	Okaloosa
Hyla chrysoscelis	MP002	USA:	GA:	Chatham
Hyla chrysoscelis	MP649	USA:	GA:	Houston
Hyla chrysoscelis	MP859	USA:	GA:	Upson
Hyla chrysoscelis	DCC3874	USA:	IA:	Butler
Hyla chrysoscelis	MP706	USA:	IA:	Clarke
Hyla chrysoscelis	INHS131T	USA:	IL:	Edgar
Hyla chrysoscelis	INHS924T	USA:	IL:	Ogle
Hyla chrysoscelis	DCC3751	USA:	IN:	Monroe
Hyla chrysoscelis	MP753	USA:	KY:	Carlisle
Hyla chrysoscelis	MP630	USA:	KY:	Laurel
Hyla chrysoscelis	MP452	USA:	KY:	Meade
Hyla chrysoscelis	MP195	USA:	LA:	Allen
Hyla chrysoscelis	MP847	USA:	LA:	Lincoln
Hyla chrysoscelis	MP632	USA:	MN:	Hennepin
Hyla chrysoscelis	MP802	USA:	MN:	Ottertail
Hyla chrysoscelis	MP773	USA:	MO:	Barry
Hyla chrysoscelis	MP386	USA:	MO:	Howell
Hyla chrysoscelis	MP350	USA:	MO:	Phelps
Hyla chrysoscelis	MP724	USA:	MS:	Hancock
Hyla chrysoscelis	MP249	USA:	MS:	Hinds
Hyla chrysoscelis	MP804	USA:	NE:	Otoe
Hyla chrysoscelis	DCC3829	USA:	OH:	Ross
Hyla chrysoscelis	MP329	USA:	OK:	Ottawa
Hyla chrysoscelis	MP327	USA:	OK:	Payne
Hyla chrysoscelis	MP205	USA:	SC:	Jasper
Hyla chrysoscelis	MP692	USA:	TN:	Monroe
Hyla chrysoscelis	MP696	USA:	TN:	Shelby

Table A1 continued.

Species	Field #	Country	State	County
Hyla chrysoscelis	MP723	USA:	TX:	Bastrop
Hyla chrysoscelis	MP686	USA:	TX:	Eastland
Hyla chrysoscelis	MP647	USA:	TX:	Gillespie
Hyla chrysoscelis	MP135	USA:	TX:	Travis
Hyla chrysoscelis	MP273	USA:	VA:	Goochland
Hyla chrysoscelis	MP816	USA:	VA:	Mecklenberg
Hyla chrysoscelis	MP701	USA:	VA:	Smyth
Hyla chrysoscelis	MP651	USA:	WV:	Summers
Hyla versicolor	MP625	Canada:	Ontario:	Guelph
Hyla versicolor	DCC3807	USA:	CT:	Tolland
Hyla versicolor	DCC3800	USA:	CT:	Windham
Hyla versicolor	INHS950T	USA:	IL:	Hancock
Hyla versicolor	INHS399T	USA:	IL:	Iroquois
Hyla versicolor	DCC3768	USA:	IN:	Porter
Hyla versicolor	MP162	USA:	LA:	Allen
Hyla versicolor	DCC3823	USA:	MD:	Ann Arundel
Hyla versicolor	MP576	USA:	ME:	Penobscot
Hyla versicolor	MP702	USA:	MN:	Clearwater
Hyla versicolor	MP795	USA:	MN:	Ottertail
Hyla versicolor	DCC3864	USA:	MN:	St. Louis
Hyla versicolor	MP759	USA:	MO:	Barry
Hyla versicolor	MP045	USA:	MO:	Greene
Hyla versicolor	MP409	USA:	MO:	Howell
Hyla versicolor	MP370	USA:	MO:	Oregon
Hyla versicolor	MP359	USA:	MO:	Ozark
Hyla versicolor	MP524	USA:	MO:	Phelps
Hyla versicolor	DCC3787	USA:	NY:	Renssalear
Hyla versicolor	DCC3795	USA:	NY:	Westchester
Hyla versicolor	DCC3828	USA:	OH:	Ross
Hyla versicolor	MP099	USA:	OK:	Cleveland
Hyla versicolor	MP793	USA:	OK:	Ottawa
Hyla versicolor	DCC3832	USA:	PA:	Crawford
Hyla versicolor	MP700	USA:	TN:	Shelby
Hyla versicolor	MP019	USA:	TX:	Bastrop
Hyla versicolor	MP717	USA:	TX:	Smith
Hyla versicolor	MP020	USA:	VA:	Giles
Hyla versicolor	MP296	USA:	VA:	Goochland
Hyla versicolor	MP809	USA:	VA:	Mecklenberg
Hyla versicolor	MP676	USA:	WV:	Summers
Hyla versicolor	MP678	USA:	WV:	Summers

Table A2. Gene product and PCR primer information.

Gene	Fragment Size (bp)	Forward Primers	Reverse Primers	Annealing Temperature
CytB	712	MVZ25-L* or MVZ15-L*	CytbAR-H*	52°C
ITS1	800	18b†	5.8c†	55°C
65T	650	65T_149F (5'–cccagggtaaattgtccgcagta–3')	65T_526R (5'–gttgggaaacactggtg–3')	58°C
11T	304	11T_84F (5'–tggagtacccctttaaatctgaat–3')	11T_388R (5'–ataaagtgcataagtaaaagtgaa–3')	60°C

*_CytB_ primers are from (GOEBEL *et al.* 1999)
†ITS1 primers are from (HILLIS and DIXON 1991).

Table A3. Genbank Accession Numbers.

Specimen	CYTB	G65T	ITS	G11T
MP002	AY830951	AY831028	AY831190-191	AY833135
MP019	AY830952	AY831029	AY831192-195	AY833136-37
MP020	AY830953	AY831030-31	AY831196-198	AY833138-41
MP045	AY830954	AY831032-33	AY831199-202	AY833142
MP099	AY830955	AY831034-37	AY831203-206	AY833143-44
MP135	AY830956	AY831038	AY831207-208	AY833145-46
MP162	AY830957	AY831039-41	AY831209-211	AY833147-48
MP195	AY830958	AY831042-43	AY831212-213	AY833149-50
MP205	AY830959	AY831044-45	AY831214-215	AY833151
MP249	AY830960	AY831046	AY831216-217	AY833152-53
MP273	AY830961	AY831047-48	AY831218-219	AY833154
MP296	AY830962	AY831049-51	AY831220-223	AY833155-56
MP327	AY830963	AY831052-53	AY831224-225	AY833157
MP329	AY830964	AY831054-55	AY831226-227	AY833158-59
MP350	AY830965	AY831056-57	AY831228-229	AY833160-61
MP359	AY830966	AY831058-60	AY831230-231	AY833162-63
MP370	AY830967	AY831061-62	AY831232-235	AY833164-65
MP386	AY830968	AY831063-64	AY831236	AY833166
MP409	AY830969	AY831065-66	AY831237-239	AY833167-68
MP452	AY830970	AY831067-68	AY831240-241	AY833169
MP458	AY830971	AY831069-70	AY831242-243	AY833170
MP524	AY830972	AY831071-72	AY831244-245	AY833171-73
MP576	AY830973	AY831073-75	AY831246-249	AY833174
MP625	AY830974	AY831076-78	AY831250-253	AY833175-76
MP630	AY830975	AY831079-80	AY831254-255	AY833177
MP632	AY830976	AY831081	AY831256-257	AY833178-79
MP647	AY830977	AY831082-83	AY831258-259	AY833180
MP649	AY830978	AY831084-85	AY831260-261	AY833181
MP651	AY830979	AY831086-87	AY831262-263	AY833182
MP670	AY830980	AY831088-89	AY831264-265	AY833183
MP676		AY831090	AY831266-268	AY833184-85
MP678	AY830981	AY831091-94	AY831269-270	AY833186
MP686	AY830982	AY831095	AY831271	AY833187-88
MP692		AY831096	AY831272-273	AY833189-90
MP696	AY830983	AY831097-98	AY831274-275	AY833191
MP700		AY831099-102	AY831276-279	AY833192-93
MP701	AY830984	AY831103-104	AY831280-281	AY833194-95
MP702	AY830985	AY831105-107	AY831282-283	AY833196-98
MP706	AY830986	AY831108-109	AY831284-286	AY8331990
MP717	AY830987	AY831110	AY831287-289	AY833201-3
MP723	AY830988	AY831111-112	AY831290-291	AY833204-5
MP724	AY830989	AY831113-114	AY831292-293	AY833206-7

Table A3. continued

Specimen	CYTB	G65T	ITS	G11T
MP732	AY830990	AY831115	AY831294-295	AY833208
MP753	AY830991	AY831116-117	AY831296-297	AY833209
MP759	AY830992	AY831118-119	AY831298-300	AY833210
MP773	AY830993	AY831120-121	AY831301-302	AY833211-12
MP793	AY830994	AY831122-124	AY831303-305	AY833213
MP795	AY830995	AY831125-126	AY831306-307	AY833214-16
MP802	AY830996	AY831127-128	AY831308-309	AY833217
MP804	AY830997	AY831129	AY831310-311	AY833218-19
MP809	AY830998	AY831130-131	AY831312-315	AY833220-21
MP816	AY830999	AY831132-133	AY831316-317	AY833222-23
MP847	AY831000	AY831134-135	AY831318-319	AY833224
MP859	AY831001	AY831136-137	AY831320-321	AY833225
DCC3751	AY831002	AY831138-139	AY831322-323	AY833226-27
DCC3768	AY831003	AY831140-141	AY831324-327	AY833228-29
DCC3787	AY831004	AY831142-143	AY831328-331	AY833230-31
DCC3795	AY831008	AY831144-146	AY831332-335	AY833232
DCC3800	AY831021	AY831147-149	AY831336-339	AY833233
DCC3807	AY831005	AY831150-151	AY831340-343	AY833234
DCC3823	AY831006	AY831152-153	AY831344-348	AY833235
DCC3828	AY831007	AY831154-157	AY831349-351	AY833236-37
DCC3829	AY831022	AY831158-159	AY831352-353	AY833238
DCC3832	AY831009	AY831160-163	AY831354-357	AY833239-40
DCC3857		AY831177-178	AY831377	
DCC3864	AY831010	AY831164	AY831358-360	AY833241-43
DCC3874	AY831011	AY831165-166	AY831361-362	AY833244-45
DCC3883	AY831012	AY831167	AY831363-364	AY833246
INHS131T	AY831013	AY831168-169	AY831365-366	AY833247
INHS399T	AY831014	AY831170-172	AY831367-369	AY833248-49
INHS924T	AY831015	AY831173-174	AY831370-371	AY833250-51
INHS950T	AY831016	AY831175-176	AY831372-376	AY833252
MP607	AY831017	AY831179-180	AY831378	AY833259
MP710	AY831020	AY831181	AY831379	AY833260-61
H146	AY831023	AY831182-183	AY831380	AY833253
HCG21	AY831018	AY831184		AY833254
HCG28	AY831024	AY831185-186	AY831381	AY833255-56
HCG84	AY831019	AY831187-188		AY833257-58
DCC3858	AY831025		AY831384	
WED54451	AY831026	AY831189		
DCC3043	AY831027		AY831382	
HCG2			AY831383	

References

ADAMS, K. L., R. CRONN, R. PERCIFIELD and J. F. WENDEL, 2003 Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc. Natl. Acad. Sci. U S A **100:** 4649-4654.

AGUADE, M., 1999 Positive selection drives the evolution of the *Acp29AB* accessory gland protein in *Drosophila*. Genetics **152:** 543-551.

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403-410.

BEGUN, D. J., and C. F. AQUADRO, 1995 Molecular variation at the *vermilion* locus in geographically diverse populations of *Drosophila melanogaster* and *D. simulans*. Genetics **140:** 1019-1032.

BEGUN, D. J., and P. WHITLEY, 2000 Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. Proc. Natl. Acad. Sci. U S A **97:** 5960-5965.

BEGUN, D. J., P. WHITLEY, B. L. TODD, H. M. WALDRIP-DAIL and A. G. CLARK, 2000 Molecular population genetics of male accessory gland proteins in *Drosophila*. Genetics **156:** 1879-1888.

BELL, G., 1982 *The Masterpiece of Nature*. University of California Press, Berkeley and Los Angeles.

BLACKBURN, L., P. NANJAPPA and M. LANNOO, 2002 Current Distributions of Selected Amphibians in the United States, http://nationalatlas.gov/amphibm.html

BOGART, J. P., 1980 Evolutionary implications of polyploidy in amphibians and reptiles, pp. 341-378 in *Polyploidy: Biological Relevance*, edited by W. H. LEWIS. Plenum Press, New York.

BOGUSKI, M., T. LOWE and C. TOLSTOSHEV, 1993 dbEST--database for "expressed sequence tags". Nature Genet. **4:** 332-333.

BORON, A., and Z. DANILKIEWICZ, 1998 Diploid-polyploid complex of spined loach *Cobitis taenia* sensu stricto and *Cobitis* sp. from the Bug River, Poland (Pisces, Cobitidae). Cytobios **96:** 13-22.

CASH, M. N., and J. P. BOGART, 1978 Cytological differentiation of the diploid-tetraploid species pair of North American treefrogs (Amphibia, Anura, Hylidae). J. Herpetol. **12:** 555-558.

CIVETTA, A., and R. S. SINGH, 1998 Sex-related genes, directional sexual selection, and speciation. Mol. Biol. Evol. **15:** 901-909.

COULTHART, M., and R. S. SINGH, 1988 High level of divergence of male-reproductive-tract proteins, between *Drosophila melanogaster* and its sibling species, *D. simulans*. Mol. Biol. Evol. **5:** 182-191.

DAVID, J., and P. CAPY, 1988 Genetic variation of *Drosophila melanogaster* natural populations. Trends Genet. **4:** 106-111.

DUELLMAN, W., and L. TRUEB, 1994 *Biology of Amphibians*. Johns Hopkins University Press, London.

ESPINOZA, N. R., and M. A. NOOR, 2002 Population genetics of a polyploid: is there hybridization between lineages of *Hyla versicolor*? J. Hered. **93:** 81-85.

EVANS, B., D. KELLEY, R. TINSLEY, D. MELNICK and D. CANNATELLA, 2004 A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. Mol. Phylogenet. Evol. **33:** 197-213.

FELSENSTEIN, J., 1985 Confidence-limits on phylogenies - an approach using the bootstrap. Evolution **39:** 783-791.

FERRIER, V. and A. JAYLET, 1978 Producing tetraploids with temperature shock and pressure. Chromosoma **69:** 47-63.

FLYBASE CONSORTIUM, 2003 The FlyBase database of the *Drosophila* genome projects and community literature. Nucleic Acids Res. **31:** 172-175. http://www.flybase.org.

FORCE, A., M. LYNCH, F. B. PICKETT, A. AMORES, Y.-L. YAN *et al.*, 1999 Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151:** 1531-1545.

GERHARDT, H. C., 1974 Mating call differences between eastern and western populations of the treefrog *Hyla chrysoscelis*. Copeia **1974:** 534-536.

GERHARDT, H. C., 1994 Reproductive character displacement of female mate choice in the gray treefrog, *Hyla-chrysoscelis*. Anim. Behav. **47:** 959-969.

GERHARDT, H. C., 1999 Reproductive character displacement and other sources of selection on acoustic communication systems, pp. 515-534 in *The design of animal communication*, edited by M. HAUSER and M. KONISHI. MIT Press, Cambridge.

GERHARDT, H. C., 2001 Acoustic communication in two groups of closely related treefrogs. Adv. Study Behav. **30:** 99-167.

GERHARDT, H. C., M. B. PTACEK, L. BARNETT and K. G. TORKE, 1994 Hybridization in the diploid-tetraploid treefrogs *Hyla chrysoscelis* and *Hyla versicolor*. Copeia **1994:** 51-59.

GOEBEL, A. M., J. M. DONNELLY and M. E. ATZ, 1999 PCR primers and amplification methods for 12S ribosomal DNA, the control region, cytochrome oxidase I, and cytochrome *b* in bufonids and other frogs, and an overview of PCR primers which have amplified DNA in amphibians successfully. Mol. Phylogenet. Evol. **11:** 163-199.

GRANT, V., 1981 *Plant Speciation*. Columbia University Press, New York.

GU, X., Y. WANG and J. GU, 2002 Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. Nature Genet. **31:** 205-209.

HILLIS, D. M., and M. T. DIXON, 1991 Ribosomal DNA: molecular evolution and phylogenetic inference. Q. Rev. Biol. **66:** 411-453.

HUDSON, R. R., M. KREITMAN and M. AGUADE, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153-159.

HUELSENBECK, J., and F. RONQUIST, 2001 MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:** 754-755.

HUGHES, A. L., 1994 The evolution of functionally novel proteins after gene duplication. Proc. R. Soc. Lond. [Biol] **256:** 119-124.

HUGHES, M. K., and A. L. HUGHES, 1993 Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*. Mol. Biol. Evol. **10:** 1360-1369.

JOHNSON, C., 1963 Additional evidence of sterility between call-types in the *Hyla versicolor* complex. Copeia **1963:** 139-143.

JOHNSON, C., 1966 Species recognition in the *Hyla versicolor* complex. Tex. J. Sci. **18:** 361-364.

KELLER, M. J., and H. C. GERHARDT, 2001 Polyploidy alters advertisement call structure in gray treefrogs. Proc. R. Soc. Lond. [Biol] **268:** 341-345.

KNOWLES, L. L., and T. A. MARKOW, 2001 Sexually antagonistic coevolution of a postmating-prezygotic reproductive character in desert *Drosophila*. Proc. Natl. Acad. Sci. U S A **98:** 8692-8696.

KOBEL, H., and L. DU PASQUIER, 1986 Genetics of polyploid *Xenopus*. Trends Genet. **2:** 310-315.

KORTSCHAK, R., R. TAMME and M. LARDELLI, 2001 Evolutionary analysis of vertebrate *Notch* genes. Dev. Genes Evol. **211:** 350-354.

LARGET, B., and D. SIMON, 1999 Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. **16:** 750-759.

LEVINS, R., 1968 *Evolution in Changing Environments: Some Theoretical Explorations*. Princeton University Press, Princeton, NJ.

LEWIS, E. B., 1978 A gene complex controlling segmentation in *Drosophila*. Nature **279:** 565-570.

LITTLEJOHN, M. J., M. J. FOUQUETTE, JR. and C. JOHNSON, 1960 Call discrimination by female frogs of the *Hyla versicolor* complex. Copeia **1960:** 47-49.

LIVERMORE, D. M., and T. L. PITT, 1986 Dissociation of surface properties and "intrinsic" resistance to beta lactams in *Pseudomonas aeruginosa*. J. Med. Microbiol. **22:** 217-224.

LONG, M., and K. THORNTON, 2001 Gene duplication and evolution. Science **293:** 1551a.

LONG-MCGIE, J., A. D. LIU and V. SCHELLENBERGER, 2000 Rapid *in vivo* evolution of a *β-lactamase* using phagemids. Biotechnol. Bioeng. **68:** 121-125.

LYNCH, M., and J. S. CONERY, 2000 The evolutionary fate and consequences of duplicate genes. Science **290:** 1151-1155.

LYNCH, M., and A. FORCE, 2000a The probability of duplicate gene preservation by subfunctionalization. Genetics **154:** 459-473.

LYNCH, M., and A. G. FORCE, 2000b The origin of interspecific genomic incompatibility via gene duplication. Am. Nat. **156:** 590-605.

MABLE, B., 2004 "Why polyploidy is rarer in animals than in plants": myths and mechanisms, pp. 453-466 in *Biological Relevance of Polyploidy: Ecology to Genomics*, edited by A. LEITCH, D. SOLTIS, P. SOLTIS, I. LEITCH and J. PIRES. Biol. J. Linn. Soc.

MADDISON, D. R., and W. P. MADDISON, 2000 MacClade 4: Analysis of phylogeny and character evolution, Sinauer Associates, Sunderland, MA.

MARTIN, D., and E. RYBICKI, 2000 RDP: detection of recombination amongst aligned sequences. Bioinformatics **16:** 562-563.

MAXSON, L., E. PEPPER and R. D. MAXSON, 1977 Immunological resolution of a diploid-tetraploid species complex of tree frogs. Science **197:** 1012-1013.

MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the Adh locus in *Drosophila*. Nature **351:** 652-654.

MEZHZHERIN, S. V., and T. Y. CHUDAKOROVA, 2002 The genetic structure of the diploid-polyploid complex of the spined loach *Cobitis taenia* (Cypriniformes : Cobitidae) from the middle Dnieper basin. Russian J. Genet. **38:** 70-75.

MILLER, G. T., W. T. STARMER and S. PITNICK, 2003 Quantitative genetic analysis of among-population variation in sperm and female sperm-storage organ length in *Drosophila mojavensis*. Genet. Res. **81:** 213-220.

MULLER, H. J., 1925 Why polyploidy is rarer in animals than in plants. Am. Nat. **59:** 346-353.

NADEAU, J., and D. SANKOFF, 1997 Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution. Genetics **147**.

NIELSEN, H., J. ENGELBRECHT, S. BRUNAK and G. VON HEIJNE, 1997 Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10:** 1-6.

NURMINSKY, D. I., M. V. NURMINSKAYA, D. DE AGUIAR and D. L. HARTL, 1998 Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. Nature **396:** 572-575.

OHNO, S., 1970 *Evolution by Gene Duplication*. Springer-Verlag, Heidelberg.

OHNO, S., 1999 Gene duplication and the uniqueness of vertebrate genomes circa 1970-1999. Cell Dev. Biol. **10:** 517-522.

ONO-KOYANAGI, K., H. SUGA, K. KATOH and T. MIYATA, 2000 Protein tyrosine phosphatases from amphioxus, hagfish, and ray: Divergence of tissue-specific isoform genes in the early evolution of vertebrates. J. Mol. Evol. **50:** 302-311.

ORR, H. A., 1990 Why polyploidy is rarer in animals than in plants: revisited. Am. Nat. **136:** 759-770.

OTTO, S. P., and J. WHITTON, 2000 Polyploid incidence and evolution. Annu. Rev. Genet. **34:** 401-437.

PALZKILL, T., Q.-Q. LE, K. V. VENKATACHALAM, M. LAROCCO and H. OCERA, 1994 Evolution of antibiotic resistance: several different amino acid substitutions in an active site loop alter the substrate profile of *β-lactamase*. Mol. Microbiol. **12:** 217-229.

PETROV, D. A., and D. L. HARTL, 1998 High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. Mol. Biol. Evol. **15:** 293-302.

PITNICK, S., G. T. MILLER, K. SCHNEIDER and T. A. MARKOW, 2003 Ejaculate-female coevolution in *Drosophila mojavensis*. Proc. R. Soc. Lond. [Biol] **270:** 1507-1512.

PTACEK, M. B., H. C. GERHARDT and R. D. SAGE, 1994 Speciation by polyploidy in treefrogs: multiple origins of the tetraploid, *Hyla versicolor*. Evolution **48:** 898-908.

RALIN, D. B., M. A. ROMANO and C. W. KILPATRICK, 1983 The tetraploid treefrog *Hyla versicolor*: evidence for a single origin from the diploid *H. chrysoscelis*. Herpetologica **39:** 212-225.

RALIN, D. B., and R. K. SELANDER, 1979 Evolutionary genetics of diploid-tetraploid species of treefrogs of the genus *Hyla*. Evolution **33:** 595-608.

RAMSEY, J., and D. W. SCHEMSKE, 1998 Pathways, mechanisms, and rates of polyploid formation in flowering plants. Ann. Rev. Ecol. Syst. **29:** 467-501.

RANZ, J. M., C. I. CASTILLO-DAVIS, C. D. MEIKLEJOHN and D. L. HARTL, 2003 Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. Science **300:** 1742-1745.

ROMANO, M. A., D. B. RALIN, S. I. GUTTMAN and J. H. SKILLINGS, 1987 Parallel electromorph variation in the diploid-tetraploid gray treefrog complex. Am. Nat. **130:** 864-878.

ROZAS, J., and R. ROZAS, 1999 DnaSP 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174-175.

RUBIN, G. M., M. D. YANDELL, J. R. WORTMAN, G. L. G. MIKLOS, C. R. NELSON *et al.*, 2000 Comparative genomics of the eukaryotes. Science **287:** 2204-2215.

SLECHTOVA, V., V. LUSKOVA, V. SLECHTA, S. LUSK, K. HALACKA *et al.*, 2000 Genetic differentiation of two diploid-polyploid complexes of spined leach, genus *Cobitis* (Cobitidae), in the Czech Republic, involving *C-taenia*, *C-elongatoides*, and *C. spp.*: Allozyme interpopulation and interspecific differences. Folia Zoologica **49:** 67-78.

SOLTIS, D. E., and P. S. SOLTIS, 1993 Molecular-data and the dynamic nature of polyploidy. Crit. Rev. Plant Sci. **12:** 243-273.

SOLTIS, D. E., and P. S. SOLTIS, 1999 Polyploidy: recurrent formation and genome evolution. Trends Ecol. Evol. **14:** 348-352.

SOLTIS, P. S., and D. E. SOLTIS, 2000 The role of genetic and genomic attributes in the success of polyploids. Proc. Natl. Acad. Sci. USA **97:** 7051-7057.

SUCHARD, M. A., R. E. WEISS, K. S. DORMAN and J. S. SINSHEIMER, 2002 Oh brother, where art thou? A Bayes factor test for recombination with uncertain heritage. Syst. Biol. **51:** 715-728.

SWANSON, W. J., 2001 Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc. Natl. Acad. Sci. USA **98:** 7375-7379.

SWANSON, W. J., and V. D. VACQUIER, 2002 The rapid evolution of reproductive proteins. Nature Rev. Genet. **3:** 137-144.

SWOFFORD, D. L., 2002 *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.* Sinauer Associates, Sunderland, MA.

THEOPOLD, U., M. RISSLER, M. FABBRI, O. SCHMIDT and S. NATORI, 1999 Insect glycobiology: a lectin multigene family in *Drosophila melanogaster*. Biochem. Biophys. Res. Comm. **261:** 923-927.

THOMAS, S., and R. S. SINGH, 1992 A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. VII. Varying rates of genic divergence as revealed by 2-dimensional electrophoresis. Mol. Biol. Evol. **9:** 507-525.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN and D. G. HIGGINS, 1997 The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **24:** 4876-4882.

TSAUR, S.-C., C. TING and C.-I. WU, 1998 Positive selection driving the evolution of a gene of male reproduction, *Acp26Aa*, of *Drosophila*. Mol. Biol. Evol. **15:** 1040-1046.

TURNER, B., 1984 *Evolutionary Genetics of Fishes*. Plenum Press, New York.

UEDA, H., 1993 Mating calls of autotriploid and autotetraploid males in *Hyla japonica*. Sci. Rep. Lab. Amphibian Biol., Hiroshima Univ. **12:** 177-189.

WALSH, J., 1995 How often do duplicated genes evolve new functions? Genetics **139:** 421-428.

WALSH, J. B., 2003 Population-genetic models of the fates of duplicate genes. Genetica **118:** 279-294.

WASSERMAN, A. O., 1970 Polyploidy in the common tree toad *Hyla versicolor* Le Conte. Science **167:** 385-386.

WHITE, M. J. D., 1973 *Animal Cytology and Evolution*. Cambridge University Press, London.

WILEY, J. E., and M. L. LITTLE, 2000 Replication banding patterns of the diploid-tetraploid treefrogs *Hyla chrysoscelis* and *H. versicolor*. Cytogenet. Cell Genet. **88:** 11-14.

WOLFNER, M. F., 1997 Tokens of love: functions and regulation of *Drosophila* male accessory gland products. Insect Biochem. Mol. Biol. **27:** 179-192.

WOLFNER, M. F., H. A. HARADA, M. J. BERTRAM, T. J. STELICK, K. W. KRAUS *et al.*, 1997 New genes for male accessory gland proteins in *Drosophila melanogaster*. Insect Biochem. Mol. Biol. **27:** 825-834.

VITA

Alisha K. Holloway was born in Oklahoma City, Oklahoma on 31 May 1972, the daughter of S. Joe and Linda K. Holloway. After completing her work at Crescent High School, Crescent, Oklahoma, in 1990, she was awarded a scholarship for her undergraduate studies at the University of Oklahoma. Alisha worked for two years at the Oklahoma Biological Survey and received her Bachelor of Science in Zoology in 1995. Alisha began graduate school at Colorado State University, Ft. Collins, Colorado in August 1995 and was awarded a Master of Science in Biology in December 1997. After graduation, she accepted a position as a research assistant at the University of Texas in Houston. In January of 2000, Alisha entered the Graduate School at the University of Texas in the Section of Integrative Biology.

Permanent Address: 18706 Candlechase Drive, Spring, Texas 77388.

This dissertation was typed by the author.