2007

AMBYSTOMA: PERSPECTIVES ON ADAPTATION AND THE EVOLUTION OF VERTEBRATE GENOMES

Jeramiah James Smith
University of Kentucky, jjsmit3@uky.edu

ABSTRACT OF DISSERTATION

Jeramiah James Smith

The Graduate School
University of Kentucky
2007

AMBYSTOMA: PERSPECTIVES ON ADAPTATION AND THE EVOLUTION OF VERTEBRATE GENOMES

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Arts and Sciences
at the University of Kentucky

By

Jeramiah James Smith

Lexington, Kentucky

Director: Dr. S. Randal Voss, Associate Professor of Biology

Lexington, Kentucky

2007

ABSTRACT OF DISSERTATION

AMBYSTOMA: PERSPECTIVES ON ADAPTATION AND THE EVOLUTION OF VERTEBRATE GENOMES

Tiger salamanders, and especially the Mexican axolotl (*Ambystoma mexicanum*), are important model organisms in biological research. This dissertation describes new genomic resources and scientific results that greatly extend the utility of tiger salamanders. With respect to new resources, this dissertation describes the development of expressed sequence tags and assembled contigs, a comparative genome map, a web-portal that makes genomic information freely available to the scientific community, and a computer program that compares structure features of organism genomes. With respect to new scientific results, this dissertation describes a quantitative trait locus that is associated with ecologically and evolutionarily relevant variation in developmental timing, the evolutionary history of the tiger salamander genome in relation to other vertebrate genomes, the likely origin of amniote sex chromosomes, and the identification of the Mexican axolotl sex-determining locus. This dissertation is concluded with a brief outline of future research directions that can extend from the works that are presented here.

Keywords: *Ambystoma*, Genome, Metamorphosis, Evolution, Sex

AMBYSTOMA: PERSPECTIVES ON ADAPTATION AND THE EVOLUTION OF VERTEBRATE GENOMES

By

Jeramiah James Smith

S. Randal Voss
Director of Dissertation

Brian C. Rymond
Director of Graduate Studies

August 25, 2007

RULES FOR THE USE OF DISSERTATIONS

Unpublished dissertations submitted for the Doctor's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgments.

Extensive copying or publication of the dissertation in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

A library that borrows this dissertation for use by its patrons is expected to secure the signature of each user.

Name Date

_____

_____

_____

_____

_____

_____

_____

_____

_____

DISSERTATION

Jeramiah James Smith

The Graduate School

University of Kentucky

2007

AMBYSTOMA: PERSPECTIVES ON ADAPTATION AND THE EVOLUTION OF VERTEBRATE GENOMES

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Arts and Sciences
at the University of Kentucky

By
Jeramiah James Smith

Lexington, Kentucky

Director: Dr. S. Randal Voss, Professor of Biology

Lexington, Kentucky

2007

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FILES

File Name	Size
Supplementary_Table_3_1.xls	256 KB
Supplementary_Table_3_2.doc	28 KB
Supplementary_Table_3_3.xls	76 KB
Supplementary_Document_6_1.txt	540 KB
Supplementary_Figure_6_1.tif	648 KB
Supplementary_Figure_6_2.tif	656 KB
Supplementary_Figure_6_3.tif	624 KB
Supplementary_Table_6_1.xls	129 KB
Supplementary_Table_6_2.xls	296 KB
Supplementary_Table_6_3.doc	36 KB
Supplementary_Table_7_1.xls	764 KB
Supplementary_Table_7_2.xls	88 KB
Supplementary_Table_8_1.xls	24 KB
Supplementary_Table_8_2.xls	20 KB
Supplementary_Table_8_3.xls	1.9 MB
Supplementary_Table_8_4.xls	1.8 MB
Supplementary_Table_8_5.xls	28 KB
Supplementary_Table_8_6.xls	28 KB
Supplementary_Table_8_7.xls	3.4 MB
Supplementary_Table_8_8.xls	528 KB
Supplementary_Table_9_1.xls	392 KB
Supplementary_Table_9_2.xls	44 KB

CHAPTER 1: Introduction.

Genetic information from the salamander *Ambystoma mexicanum*, commonly known as the Mexican axolotl, provides unique on adaptation and the evolution of vertebrate genomes. The evolutionary history of *A. mexicanum* has suited it for these studies in two important ways. First, the species is a member of the class Amphibia. The amphibian lineage and the amniote lineage (collectively: mammals, birds, and other reptiles) shared a common ancestor approximately 350 million years ago. Comparisons between amphibian and amniote genomes can therefore reveal conserved features that are retained from the genome of their common ancestor. A second convenient feature of *A. mexicanum* is that it is a member of the tiger salamander species complex (*sensu* Shaffer, 1984a), which is distributed across the southern half of North America and consists of several closely related species that have adapted to exploit a variety of temporary and permanent aquatic habitats (Shaffer and McKnight, 1996). Coordinate with their diverse habitat use, tiger salamanders also exhibit extreme diversity in phenotype (Shaffer, 1984b). Despite this phenotypic diversity, crosses between tiger salamander species typically result in the production of viable and fertile hybrid offspring (reviewed by Voss and Shaffer, 1996). When these hybrids are mated, their offspring segregate genotypes that are inherited from their parental species and phenotypic variation that is associated with these genotypes.

Ambystoma mexicanum is by far the best-studied member of the tiger salamander species complex and has a long history as a model organism. It was originally brought into laboratory culture by Aguste Duméril in 1870 (Duméril, 1870; Smith, 1989). Since this time, the species has served as a model system for studying vertebrate development (Armstrong and Malacinski, 1989), hence its other handle – the laboratory axolotl. Until recently, however, extensive genetic analyses have been limited by several factors, including: an exceedingly large genome size of roughly 30 gigabases (Gb) (Licht and Lowcock, 1991), a relatively long generation time, and a relatively small number of assayable genetic markers. That is not to say that genetic analyses are impossible for the species. In fact, my predecessors have succeeded in identifying phenotypes that segregate as Mendelian (single gene) factors (reviewed by Armstrong and Malacinski, 1989). Most critical among these are the discoveries that: 1) sex segregates as a Mendelian locus,

under which the female phenotype is specified by a dominant locus (Humphrey, 1945; 1957); and 2) another Mendelian locus (*met*) controls the expression of the alternate developmental pathways: metamorphosis (the ancestral developmental pathway within the tiger salamander species complex) and paedomorphosis (a derived developmental pathway where in an individual forgoes metamorphosis and attains sexual maturity in it's larval form) (Voss, 1995; Voss and Shaffer 1997). Moreover, 10 of these Mendelian factors, including sex, have been tested for linkage to centromeric regions (Lindsley *et al*, 1956; Armstrong, 1984, Armstrong and Duhon, 1989). Recent advances in chemistries and computational methods have greatly facilitated the development of genetic markers. If not for these advances, the husbandry methods that were developed by my predecessors (*n.b.* Humphrey, 1967; Armstrong and Malacinski, 1989; Voss, 1995), and an interspecific cross (AxTg – Voss, 1995) the majority of the work that is described in this dissertation would have been impossible.

The remaining chapters of this dissertation describe advances in the development of genetic markers, use of these markers in elucidating the organization of the salamander genome, and the broader utility of these markers for studies of evolution and development. Chapter 3 describes the use of these markers in inferring the genomic location of three important developmental phenotypes: the pigment mutations *white* and *melanoid*, and a developmental phenotype known as paedomorphosis. Paedomorphosis is a presumably adaptive developmental strategy that is expressed by *A. mexicanum* and other members of the tiger salamander species complex. Under this strategy, an individual forgoes the metamorphic changes that are necessary for most amphibian species to transform into their adult form, rather attaining sexual maturity in its larval form (Gould, 1977). Chapter 5 describes how the genetic markers that were developed in Chapters 2-4, and previously by Dr. Voss (1997), were extended to additional crosses and used to further elucidate the developmental/genetic basis of paedomorphosis.

Chapters 6-8 focus on the deeper insights that are gained by understanding, for the first time, how protein-coding genes are distributed within an amphibian genome. First, Chapter 6 considers the relative location and distribution of genes within the genomes of *Ambystoma* and several other vertebrate species (mammals: human, mouse, rat and dog, a representative of the reptilian lineage: chicken; and fish: zebrafish and the freshwater

pufferfish). In particular, the analyses that are presented within Chapter 6 on the distribution of genes that presumably derive from a common ancestral gene that was located within the genome of the species common ancestor; such genes are known as orthologs. Comparing the orders of orthologs among these species reveals features of ancestral genomes that have been conserved through time and permits an estimation of the relative rates of genome rearrangement that have been experienced by different vertebrate lineages. In Chapter 7, the methods that were used to identify orthologs in Chapter 6 are improved somewhat and developed into a formal computer program called MapToGenome. Chapter 8 presents what is arguably the most controversial finding that has arisen from comparison of gene orders between *Ambystoma* and other vertebrates. This is that the sex chromosomes of birds and mammals (which were presumed to have evolved from independent ancestors, ***e.g.*** Fridolffson ***et al***, 1998; Nanda ***et al***, 1999; Ellegren, 2000; Nanda ***et al***, 2000; Graves ***et al***, 2002; Nanda ***et al***, 2002; Kohn ***et al***, 2004; Handley ***et al***, 2004; Khil and Camerini-Otero, 2005; Kohn ***et al***, 2006) share many orthologs with a single salamander linkage group. The distribution of orthologs on this linkage group and within the genomes of other vertebrate species is highly consistent with shared ancestry of the avian and mammalian sex chromosomes.

In Chapter 9, I describe experiments that have identified the location of the major sex-determining factor of *A. mexicanum*. The fairly uniform patterns of recombination that are observed within sex-linked regions of the genome imply that this sex-determining factor evolved relatively recently, in contrast to the deep ancestry of avian and mammalian sex chromosomes. I conclude this dissertation with a final chapter that outlines some future avenues of research that can be extended from the works that are described in Chapters 2-9.

CHAPTER 2: From Biomedicine to Natural History Research: EST Resources for Ambystomatid Salamanders.

Introduction

Establishing genomic resources for closely related species will provide comparative insights that are crucial for understanding diversity and variability at multiple levels of biological organization. Expressed sequence tags (EST) are particularly useful genomic resources because they enable multiple lines of research and can be generated for any organism: ESTs allow the identification of molecular probes for developmental studies, provide clones for DNA microchip construction, reveal candidate genes for mutant phenotypes, and facilitate studies of genome structure and evolution. Furthermore, ESTs provide raw material from which strain-specific polymorphisms can be identified for use in population and quantitative genetic analyses. The utility of such resources can be tailored to target novel characteristics of organisms when ESTs are isolated from cell types and tissues that are actively being used by a particular research community, so as to bias the collection of sequences towards genes of special interest. Finally, EST resources produced for model organisms can greatly facilitate comparative and evolutionary studies when their uses are extended to other, closely related taxa.

Salamanders (urodele amphibians) are traditional model organisms whose popularity was unsurpassed early in the 20th century. At their pinnacle, salamanders were the primary model for early vertebrate development. Embryological studies in particular revealed many basic mechanisms of development, including organizer and inducer regions of developing embryos (Beetschen, 1996). Salamanders continue to be important vertebrate model organisms for regeneration because they have by far the greatest capacity to regenerate complex body parts in the adult phase. In contrast to mammals, which are not able to regenerate entire structures or organ systems upon injury or amputation, adult salamanders regenerate their limbs, tail, lens, retina, spinal cord, heart musculature, and jaw (Gardiner *et al*, 2002; Echeverri and Tanaka, 2002; Del Rio-Tsonis *et al*, 1997; Ikegami *et al*, 2002; Chernoff *et al*, 2003; Ferretti, 1996). In addition, salamanders are valuable models for several areas of research, including: vision, embryogenesis, heart development, olfaction, chromosome structure, evolution, ecology, science education, and conservation biology (Zhang and Wu 2003; Falck *et al*, 2002;

Zhang *et al*, 2003; Kauer, 2002; Voss *et al*, 2003; Riley *et al*, 2003; Borland *et al*, 2003; Jancovich *et al*, 2003). All of these disciplines were in need of genomic resources as fewer than 4100 salamander nucleotide sequences had been deposited in GenBank as of 3/10/04.

Here I describe results from an EST project for two ambystomatid salamanders: the Mexican axolotl, *Ambystoma mexicanum* and the eastern tiger salamander, *A. tigrinum tigrinum*. These two species are members of the Tiger Salamander Complex (Shaffer and McKnight, 1996), a group of closely related species and subspecies that are widely distributed in North America. Phylogenetic reconstruction suggests that these species probably arose from a common ancestor about 10–15 million years ago (Shaffer and McKnight, 1996). *Ambystoma mexicanum* has a long research history of over 100 years and is now principally supplied to the research community by the Axolotl Colony (http://www.ambystoma.org/AGTC), while *A. t. tigrinum* is obtained from natural populations in the eastern United States. Although closely related with equally large genomes ($32x10^9$ bp) (Straus, 1971; Licht and Lowcock, 1991), these two species and others of the Complex differ dramatically in life history: *A. mexicanum* is a paedomorphic species that retains many larval features and lives in water throughout it's life cycle while *A. t. tigrinum* undergoes a metamorphosis that is typical of many amphibians. Like many other traditional model organisms of the last century, interest in these two species declined during the rise of genetic models like the fly, zebrafish, and mouse (Davis, 2004). However, "early" model organisms such as salamanders are beginning to re-attract attention as genome resources can rapidly be developed to exploit the unique features that originally identified their utility for research. This point is made below by showing how the development of ESTs for these two species is enabling research in several areas. Furthermore, It is emphasized that the value of developing resources in model systems where the likelihood of information transfer to multiple, closely related taxa is high, thus simultaneously enabling both laboratory and natural history research programs.

Materials and Methods

cDNA library construction

Ten cDNA libraries were constructed for the project using various larval tissues of *A. mexicanum* and *A. t. tigrinum* (Table 2.1). Larval *A. mexicanum* were obtained from adult animals whose ancestry traces back to the Axolotl Colony (17). Larval *A. t. tigrinum* were obtained from Charles Sullivan Corp. The GARD and MATH *A. mexicanum* limb regeneration libraries were constructed using regenerating forelimb mesenchyme. Total RNAs were collected from anterior and posterior limbs amputated at the mid-stylopod level on 15 cm animals, and from the resulting regenerates at 12 h, 2 days, 5 days and early bud stages. One hundred microgram fractions of each were pooled together and polyA-selected to yield 5 μg that was utilized for directional library construction (Lambda Zap, Stratagene). The V1 (*A. mexicanum*), V2 (*A. t. tigrinum*), V4-5 (*A. t. tigrinum*), and V6-7 (*A. mexicanum*) libraries were made from an assortment of larval tissues (see Table 2.1) using the SMART cDNA cloning kits (Clontech). Total RNAs were isolated and reverse-transcribed to yield cDNAs that were amplified by long distance PCR and subsequently cloned into pTriplEX (Clontech). The V3 and AG libraries were constructed by commercial companies (BioS&T and Agencourt, respectively).

cDNA template preparation and sequencing

cDNA inserts were mass excised as phagemids, picked into microtitre plates, grown overnight in LB broth, and then diluted (1/20) to spike PCR reactions: 150 ng DNA, 50 ng each primer (pTriplEX 5'LD, pTriplEX 3'LD), 1.2 mM $MgCl_2$, 0.3 U Taq polymerase, 1x PCR buffer, 200 mM each of dATP, dCTP, dGTP, dTTP; thermal cycling 94°C for 2 min; then 30 cycles at 94°C for 45 sec, 58°C for 45°sec, and 72°C for 7 min. All successful amplifications with inserts larger than ~500 bp were sequenced (ABI Big Dye or Amersham Dye terminator chemistry and 5' universal primer - pTriplEX 5'LD). Sequencing and clean-up reactions was carried out according to manufacturers' protocols. ESTs were deposited into NCBI database under accession numbers BI817205-BI818091 and CN033008-CN045937 and CN045944-CN069430.

EST sequence processing and assembly

The PHRED base-calling program (Ewing and Green, 1998) was used to generate sequence and quality scores from trace files. PHRED files were then quality clipped and vector/contaminant screened. An in-house program called QUALSCREEN was used to quality clip the ends of sequence traces. Starting at the ends of sequence traces, this program uses a 20 bp sliding window to identify a continuous run of bases that has an average PHRED quality score of 15. Mitochondrial DNA sequences were identified by searching all ESTs against the complete mtDNA genome sequence of *A. mexicanum* (AJ584639). Finally, all sequences less than 100 bp were removed. The average length of the resulting ESTs was 629 bp. The resulting high quality ESTs were clustered initially using PaCE (Kalyanaraman, 2003) on the U.K. HP Superdome computer. Multi-sequence clusters were used as input sequence sets for assembly using CAP3 (Huang and Madan, 1999) with an 85% sequence similarity threshold. Clusters comprising single ESTs were assembled again using CAP3 with an 80% sequence similarity threshold to identify multi-EST contigs that were missed during the initial analysis. This procedure identified 550 additional contigs comprising 1150 ESTs.

Functional annotation

All contigs and singletons were searched against the human RefSeq database (Oct. 2003 release) using BLASTX. The subset of sequences that yielded no BLAST hit was searched against the non-redundant protein sequence database (Feb. 2004) using BLASTX. The remaining subset of sequences that yielded no BLAST hit was searched against *Xenopus laevis* and *X. tropicalis* UNIGENE ESTs (Mar. 2004) using TBLASTX. Zebrafish ESTs were downloaded from UNIGENE ESTs (May 2004). BLAST searches were performed using an E-value threshold of $E < 10^{-7}$ unless specified otherwise.

Sequence comparison of *A. mexicanum* and *A. t. tigrinum* assemblies

All low quality base calls within contigs were masked using a PHRED base quality threshold of 15. To identify polymorphisms for linkage mapping, contigs from *A. mexicanum* and *A. t. tigrinum* assemblies were joined into a single assembly using CAP3 and the following criteria: an assembly threshold of 12 bp to identify initial matches, a

minimum 100 bp match length, and 85% sequence identity. To identify putatively orthologous genes from *A. mexicanum* and *A. t. tigrinum* assemblies, and generate an estimate of gene sequence divergence, assemblies were compared using BLASTN with a threshold of E <10-20. Following BLAST, alignments were filtered to obtain reciprocal best BLAST hits.

Extending *A. mexicanum* / *A. t. tigrinum* sequence information to *A. ordinarium*

Polymorphic DNA marker loci were identified by locating single nucleotide polymorphisms (SNPs) in the joint *A. mexicanum* and *A. t. tigrinum* assembly. Polymerase chain reaction (PCR) primers were designed using Primer 3 (Rozen *et al*, 1999) to amplify 100 – 500 bp SNP-containing fragments from 123 different protein-coding loci (Table 2.2). DNA was isolated from salamander tail clips using SDS, RNAse and proteinase K treatment, followed by phenol-chloroform extraction. Fragments were amplified using 150 ng DNA, 75 ng each primer, 1.5 mM $MgCl_2$, 0.25 U Taq, and a 3-step profile (94°C for 4 min; 33 cycles of 94°C for 45 s, 60°C for 45 s, 72°C for 30 s; and 72°C for 7 min). DNA fragments were purified and sequenced using ABI Big Dye or Amersham Dye terminator chemistry. Single nucleotide polymorphisms were identified by eye from sequence alignments.

Linkage mapping of human chromosome 17 orthologous genes

Putative salamander orthologs of genes on human chromosome 17 (HSA17) were identified by comparing the joint *A. mexicanum* and *A. t. tigrinum* assembly to sequences from the human RefSeq (NCBI) protein database, using BLASTX at threshold $E<10^{-7}$. Linkage distance and arrangement among markers was estimated using MapManager QTXb19 software (Meer *et al*, 2004) and the Kosambi mapping function at a threshold of p = 0.001. All markers were mapped using DNA from a previously described meiotic mapping panel (Voss *et al*, 2001). All PCR primers and primer extension probes were designed using Primer 3 (Rozen *et al*, 1999) and Array Designer2 (Premier Biosoft) software. Species-specific polymorphisms were assayed by allele specific amplification, restriction digestion, or primer extension, using the reagent and PCR conditions described above. Primer extension markers were genotyped using the AcycloPrime-FP SNP

detection assay (Perkin Elmer). See Table 2.3 for amplification and extension primer sequences, and information about genotyping methodology.

Results and Discussion

Selection of libraries for EST sequencing

Eleven cDNA libraries were constructed using a variety of tissues (Table 2.1). Pilot sequencing of randomly selected clones revealed that the majority of the non-normalized libraries were moderate to highly redundant for relatively few transcripts. For example, hemoglobin-like transcripts represented 15–25% of the sampled clones from cDNA libraries V1, V2, and V6. Accordingly, sequencing efforts were focused on the non-normalized MATH library as well as the normalized AG library, which had lower levels of redundancy (5.5 and 0.25% globins, respectively). By concentrating sequencing efforts on these two libraries the transcripts were obtained that derived primarily from regenerating larval tissues in *A. mexicanum* and several non-regenerating larval tissues in *A. t. tigrinum*.

EST sequencing and clustering

A total of 46,064 cDNA clones were sequenced, yielding 39,982 high quality sequences for *A. mexicanum* and *A. t. tigrinum* (Table 2.4). Of these, 3,745 corresponded to mtDNA and were removed from the dataset; complete mtDNA genome data for these and other ambystomatid species will be reported elsewhere. The remaining nuclear ESTs for each species were clustered and assembled separately. An additional 16,030 high quality ESTs that were generated recently for regenerating tail and neurula stage embryos (Habermann *et al*, 2004) were included in the *A. mexicanum* assembly. Thus, a total of 32,891 and 19,376 ESTs were clustered for *A. mexicanum* and *A. t. tigrinum*, respectively. Using PaCE clustering and CAP3 assembly, a similar number of EST clusters and contigs were identified for each species (Table 2.4). Overall contig totals were 11,190 and 9,901 *for A. mexicanum* and *A. t. tigrinum* respectively. Thus, although 13,515 more *A. mexicanum* ESTs were assembled, a roughly equivalent number of contigs were obtained for both species. This indicates that EST development was more efficient for *A. t. tigrinum*, presumably because ESTs were obtained primarily from the normalized AG library;

indeed, there were approximately twice as many ESTs on average per *A. mexicanum* contig (Table 2.4). Thus, this EST project yielded an approximately equivalent number of contigs for *A. mexicanum* and *A. t. tigrinum*, and overall > 21,000 different contigs were identified. Assuming that 20% of the contigs correspond to redundant loci, which has been found generally in large EST projects (Kawai *et al*, 2001), transcripts were identified for approximately 17,000 different ambystomatid loci. If ambystomatid salamanders have approximately the same number of loci as other vertebrates (*e.g.* Ewing and Green, 2000), these represent roughly half the expected number of genes in the genome.

Identification of vertebrate sequences similar to *Ambystoma* contigs

All contigs were searched against several vertebrate databases to identify sequences that exhibited significant sequence similarity. As the objective was to annotate reliably as many contigs as possible, these were first searched against 19,804 sequences in the NCBI human RefSeq database (Figure 2.1), which is actively reviewed and curated by biologists. This search revealed 5619 and 4973 "best hit" matches for the *A. mexicanum* and *A. t. tigrinum* EST datasets at a BLASTX threshold of $E = 10^{-7}$. The majority of contigs were supported at more stringent E-value thresholds (Table 2.5). Non-matching contigs were subsequently searched against the Non-Redundant (nr) Protein database and *Xenopus tropicalus* and *X. laevis* UNIGENE ESTs (Figure 2.1). These later two searches yielded a few hundred more 'best hit' matches, however a relatively large number of ESTs from both ambystomatid species were not similar to any sequences from the databases above. Presumably, these non-matching sequences were obtained from the non-coding regions of transcripts or they contain protein-coding sequences that are novel to salamander. Although the majority of these are probably of the former type, 3,273 sequences were identified from the non-matching set that had open reading frames (ORFs) of at least 200 bp, and 911 of these were greater than 300 bp.

The distribution of ESTs among contigs can provide perspective on gene expression when clones are randomly sequenced from non-normalized cDNA libraries. In general, frequently sampled transcripts may be expressed at higher levels. The 20 contigs from *A. mexicanum* and *A. t. tigrinum* that contained the most assembled ESTs are

presented in Tables 2.6 and 2.7, respectively. The largest *A. t. tigrinum* contigs contained fewer ESTs than the largest *A. mexicanum* contigs, probably because fewer overall *A. t. tigrinum* clones were sequenced, with the majority selected from a normalized library. However, it is notable that the contig with the most ESTs was identified for *A. t. tigrinum*: delta globin. In both species, transcripts corresponding to globin genes were sampled more frequently than all other loci. This may reflect the fact that amphibians, unlike mammals, have nucleated red blood cells that are transcriptionally active. In addition to globin transcripts, a few other house-keeping genes were identified in common from both species, however the majority of the contigs were unique to each list. Overall, the strategy of sequencing cDNAs from a diverse collection of tissues (from normalized and non-normalized libraries) yielded different sets of highly redundant contigs. Only 25% and 28% of the *A. mexicanum* and *A. t. tigrinum* contigs, respectively, were identified in common (Figure 2.2). Moreover, there are several hundred contigs were identified in common between *Xenopus* and *Ambystoma*; this will help facilitate comparative studies among these amphibian models.

Functional annotation

For the 10,592 contigs that showed significant similarity to sequences from the human RefSeq database, Gene Ontology (http://www.godatabase.org/dev/database/) information was obtained to describe ESTs in functional terms. Although there are hundreds of possible annotations, a list of descriptors for molecular and biological processes were selected that are of interest for research programs currently utilizing salamanders as model organisms (Table 2.8). In all searches, each match between a contig and a RefSeq sequence was counted as a different ambystomatid gene, even when different contigs matched the same RefSeq reference. In almost all cases, approximately the same number of matches was found per functional descriptor for both species. This was not simply because the same loci were being identified for both species, as only 20% of the total number of searched contigs shared sufficient identity (BLASTN; $E<10^{-80}$ or $E<10^{-20}$) to be potential homologues. In this sense, the sequencing effort between these two species was complementary in yielding a more diverse collection of ESTs that were highly similar to human gene sequences.

Informatic searches for regeneration probes

The value of a salamander model to regeneration research will ultimately rest on the ease in which data and results can be cross-referenced to other vertebrate models. For example, differences in the ability of mammals and salamanders to regenerate spinal cord may reflect differences in the way cells of the ependymal layer respond to injury. As is observed in salamanders, ependymal cells in adult mammals also proliferate and differentiate after spinal cord injury (SCI) (Adrian and Walker, 1962; Namiki and Tator, 1999); immediately after contusion injury in adult rat, ependymal cell numbers increase and proliferation continues for at least 4 days (Bruni and Anderson, 1985; but see Takahashi *et al*, 2003). Rat ependymal cells share some of the same gene expression and protein properties of embryonic stem cells (Yamamoto *et al*, 2001), however no new neurons have been observed to derive from these cells *in vivo* after SCI (Horner *et al*, 2000). Thus, although endogenous neural progenitors of the ependymal layer may have latent regenerative potential in adult mammals, this potential is not realized. Several recently completed microarray analyses of spinal cord injury in rat now make it possible to cross-reference information between amphibians and mammals. For example, all *Ambystoma* ESTs were searched against the complete list of significantly up and down regulated genes from Carmel *et al*. (2001) and Song *et al*. (2001). Based upon amino acid sequence similarity of translated ESTs (TBLASTX; $E<10^{-7}$), DNA sequences were identified that corresponded to 69 of these 164 SCI rat genes (Table 2.9). It is likely that there are sequences that correspond to other presumptive orthologs from hit list as many ESTs only contain a portion of the coding sequence or the untranslated regions (UTR), and in many cases searches identified closely related gene family members. Thus, many of the genes that show interesting expression patterns after SCI in rat can now be examined in salamander.

Similar gene expression programs may underlie regeneration of vertebrate appendages such as fish fins and tetrapod limbs. Regeneration could depend on reiterative expression of genes that function in patterning, morphogenesis, and metabolism during normal development and homeostasis. Or, regeneration could depend in part on novel genes that function exclusively in this process. These alternatives were

investigated by searching *A. mexicanum* limb regeneration ESTs against UNIGENE zebrafish fin regeneration ESTs (Figure 2.3). This search identified 1357 significant BLAST hits (TBLASTX; $E<10^{-7}$) that corresponded to 1058 unique zebrafish ESTs. A search of the 1058 zebrafish ESTs against > 400,000 zebrafish ESTs that were sampled from non-regenerating tissues was then performed to assess whether any of these potential regeneration homologues were represented uniquely in limb and fin regeneration databases (and not in databases derived from other zebrafish tissues). These comparisons revealed 43 that were unique to the zebrafish regeneration database (Table 2.10). Conceivably, these 43 ESTs may represent transcripts important to appendage regeneration. For example, this search identified several genes (*e.g.* hspc128, pre-B-cell colony enhancing factor 1, galectin 4, galectin 8) that may be expressed in progenitor cells that proliferate and differentiate during appendage regeneration. Overall, these results suggest that regeneration is achieved largely through the reiterative expression of genes having additional functions in other developmental contexts, however a small number of genes may be expressed uniquely during appendage regeneration.

DNA sequence polymorphisms within and between *A. mexicanum* and *A. t. tigrinum*

The identification of single nucleotide polymorphisms (SNPs) within and between orthologous sequences of *A. mexicanum* and *A. t. tigrinum* is needed to develop DNA markers for genome mapping (Parichy *et al*, 1999), quantitative genetic analysis (Voss and Shaffer *et al*, 1997), and population genetics (Fitzpatrick and Shaffer, 2004). The frequency of within species polymorphism was estimated for both species by calculating the frequency of SNPs among ESTs within the 20 largest contigs (Tables 2.6 and 2.7). These analyses considered a total of 30,638 base positions for *A. mexicanum* and 18,765 base positions for *A. t. tigrinum*. Two classes of polymorphism were considered in this analysis: those occurring at moderate (identified in 10–30% of the EST sequences) and high frequencies (identified in at least 30% of the EST sequences). Within the *A. mexicanum* contigs, 0.49% and 0.06% of positions were polymorphic at moderate and high frequency, while higher levels of polymorphism were observed for *A. t. tigrinum* (1.41% and 0.20%). Higher levels of polymorphism are expected for *A. t. tigrinum* because they exist in larger, out-bred populations in nature.

To identify SNPs between species, it was necessary to first identify presumptive, interspecific orthologs. This was achieved by first performing BLASTN searches between the *A. mexicanum* and *A. t. tigrinum* assemblies, then filtering the resulting alignments were to retain only those alignments between sequences that were one another's reciprocal best BLAST hit. As expected, the number of reciprocal 'best hits' varied depending upon the E value threshold, although increasing the E threshold by several orders of magnitude had a disproportionately small effect on the overall total length of BLAST alignments. A threshold of $E<10^{-80}$ yielded 2414 alignments encompassing a total of 1.25 Mbp from each species, whereas a threshold of $E<10^{-20}$ yielded 2820 alignments encompassing a total of 1.32 Mbp. The percent sequence identity of alignments was very high among presumptive orthologs, ranging from 84–100% at the more stringent threshold of $E<10^{-80}$. On average, *A. mexicanum* and *A. t. tigrinum* transcripts are estimated to be 97% identical at the nucleotide level, including both protein coding and UTR sequence. This estimate for nuclear sequence identity is surprisingly similar to estimates obtained from complete mtDNA reference sequences for these species (96%, unpublished data), and to estimates for partial mtDNA sequence data obtained from multiple natural populations (Riley et al, 2003). These results are consistent with the idea that mitochondrial mutation rates are lower in cold *vs.* warm-blooded vertebrates (Martin, 1993). From a resource perspective, the high level of sequence identity observed between these species suggests that informatics will enable rapid development of probes between these and other species of the *A. tigrinum* complex.

Extending EST resources to other ambystomatid species

Relatively little DNA sequence has been obtained from species that are closely related to commonly used model organisms, and yet, such extensions would greatly facilitate genetic studies of natural phenotypes, population structures, species boundaries, and conservation and divergence of developmental mechanisms. Like many amphibian species that are threatened by extinction, many of the ambystomatid salamanders are currently in need of population genetic studies to inform conservation and management strategies (*e.g.* Riley *et al*, 2003). Primers were designed for orthologous *A. mexicanum* and *A. t. tigrinum* ESTs and extended to develop informative molecular markers for a

related species, *A. ordinarium*. *Ambystoma ordinarium* is a stream dwelling paedomorph endemic to high elevation habitats in central Mexico (Anderson and Worthington, 1971). This species is particularly interesting from an ecological and evolutionary standpoint because it harbors a high level of intraspecific mitochondrial variation, and as an independently derived stream paedomorph, is unique among the typically pond-breeding tiger salamanders. As a reference of molecular divergence, *A. ordinarium* shares approximately 98 and 97% mtDNA sequence identity with *A. mexicanum* and *A. t. tigrinum* respectively (Shaffer and McKnight, 1996).

To identify informative markers for *A. ordinarium*, *A. mexicanum* and *A. t. tigrinum* EST contigs were aligned to identify orthologous genes with species-specific sequence variations (SNPs or Insertion/Deletions = INDELs). Primer pairs corresponding to 123 ESTs (Table 2.2) were screened by PCR using a pool of DNA template made from individuals of 10 *A. ordinarium* populations, which span the geographic range of *A. ordinarium*. Seventy-nine percent (N = 97) of the primer pairs yielded amplification products that were approximately the same size as corresponding *A. mexicanum* and *A. t. tigrinum* fragments, using only a single set of PCR conditions. To estimate the frequency of intraspecific DNA sequence polymorphism among this set of DNA marker loci, 43 loci were sequenced using a single individual sampled randomly from each of the 10 populations. At least one polymorphic site was observed for 20 of the sequenced loci, with the frequency of polymorphisms being dependent upon the size of the DNA fragment amplified. These results suggest that the vast majority of primer sets designed for *A. mexicanum* / *A. t. tigrinum* EST orthologues can be used to amplify the corresponding sequence in a related *A. tigrinum* complex species, and for small DNA fragments in the range of 150–500 bp, approximately half are expected to have informative polymorphisms.

Comparative gene mapping

Salamanders occupy a pivotal phylogenetic position for reconstructing the ancestral tetrapod genome structure and for providing perspective on the extremely derived anuran *Xenopus* (Cannatella and De Sa, 1993) that is currently providing the bulk of amphibian genome information. To evaluate both the technical feasibility of mapping *Ambystoma*

ESTs and to assess likelihood that presumptive orthologues map to the same synteny group, assemblies were searched for presumptive HSA17 orthologs and a subset of these loci were developed for genetic linkage mapping. A region of conserved synteny that corresponds to human chromosome (HSA) 17q has been identified in several non-mammalian taxa including reptiles (Schmid *et al*, 2000) and fishes (Postlethwait *et al*, 2000). In a previous study, Voss *et al* (2001) identified a region of conserved synteny between *Ambystoma* and HSA17q that included collagen type 1 alpha 1 (Col1a1), thyroid hormone receptor alpha (Thra), homeo box b13 (Hoxb13), and distal-less 3 (Dlx3) (Figure 2.4). Using a joint assembly of *A. mexicanum* and *A. t. tigrinum* contigs, 97 HSA17 presumptive orthologs were identified. Fifteen genes were chosen from this list and PCR primers were designed to amplify a short DNA fragment containing 1 or more of the presumptive SNPs that were identified in the joint assembly (Table 2.3). All but two of these genes were mapped, indicating a high probability of mapping success using markers developed from the joint assembly of *A. mexicanum* and *A. t. tigrinum* contigs. All six ESTs that exhibited 'best hits' to loci within the previously defined human-*Ambystoma* synteny group mapped to this region (Hspc009, Sui1, Krt17, Krt24, Flj13855, and Rpl19). These results show that BLAST-based definitions of orthology can be informative between salamanders and human. All other presumptive HSA17 loci mapped to *Ambystoma* chromosomal regions outside of the previously defined synteny group. It is interesting to note that two of these loci mapped to the same ambystomatid linkage group (Cgi-125, Flj20345), but in human the presumptive orthologs are 50 Mb apart and distantly flank the syntenic loci in Figure 2.4. Assuming orthology has been correctly assigned for these loci, this suggests a dynamic history for some HSA17 orthologs during vertebrate evolution.

Conclusions

Approximately 40,000 cDNA sequences were isolated from a variety of tissues to develop expressed sequence tags for two model salamander species (*A. mexicanum* and *A. t. tigrinum*). An approximately equivalent number of contigs were identified for each species, with 21,091 unique contigs identified overall. The strategy to sequence cDNAs from a diverse collection of tissues from normalized and non-normalized libraries yielded

different sets of highly redundant contigs. Only 25% and 28% of the *A. mexicanum* and *A. t. tigrinum* contigs, respectively, were identified in common. To demonstrate the utility of these EST resources, databases were searched to identify new probes for regeneration research, characterized intra- and interspecific nucleotide polymorphism, saturated a human/*Ambystoma* synteny group with marker loci, and extended PCR primer sets designed for *A. mexicanum* / *A. t. tigrinum* orthologs to a related tiger salamander species. Over 100 new probes were identified for regeneration research using informatic approaches. With respect to comparative mapping, 13 of 15 EST markers were mapped successfully, and 6 EST markers were mapped to a previously defined synteny group in *Ambystoma*. These results indicate a high probability of mapping success using EST markers developed from the joint assembly of *A. mexicanum* and *A. t. tigrinum* contigs. Finally, it was demonstrated that primer sets designed for *A. mexicanum* / *A. t. tigrinum* EST orthologs can be used to amplify the corresponding sequence in a related species from the tiger salamander complex. Overall, the EST resources reported here will enable a diversity of new research areas using ambystomatid salamanders.

Table 2.1: Tissues selected to make cDNA libraries.

ID	Tissue	cDNAs sequenced
GARD	limb blastema	1029
MATH	limb blastema	16244
V1	tail blastema	1422
V2	brain	3196
V3	liver	792
V4	spleen	337
V5	heart	38
V6	gill	3039
V7	stage 22 embryo	96
AG	liver, gonad, lung, kidney, heart, gill	19871

Table 2.2: EST loci used in a population-level PCR amplification screen in *A. ordinarium*

Locus ID	Forward Primer 5' to 3'	Reverse Primer 5' to 3'
1F8	AAGAAGGTCGGGATTGTGGGTAA	CAGCCTTCCTCTTCATCTTTGTCTTG
1H3	GGCAAATGCTGGTCCCAACACAAA	GGACAACACTGCCAAATACCACAT
2C8	GCAAGCACCAGCCACATAAAG	GGCCACCATAACCACTCTGCT
3B10	TCAAAACGAATAAGGGAAGAGCGACTG	TTGCCCCCATAATAAGCCATCCATC
5E7	ACGCTTCGCTGGGGTTGACAT	CGGTAGGATTTCTGGTAGCGAGCAC
5F4	CCGAGATGAGATTTATAGAAGGAC	TAGGGGAAGTTAAACATAGATAGAA
6A3	GTTTATGAAGGCGAGAGGGCTATGACCA	ATCTTGTTCTCCTCGCCAGTGCTCTTGT
6B1	TGATGCTGGCGAGTACAAACCCCCTTCT	TTTACCATTCCTTCCCTTCGGCAGCACA
6B3	ACCACGTGCTGTCTTCCCATCCAT	ACGAAGCTCATTGTAGAAGGTGTG
6B4	CCCACGATGAATTGGAATTGGACAT	CTGCCTGCCAGACCTACAGACTATCGT
6C4	ATGGCGCCAAAGTGATGAGTA	GGGCCAGGCACACGACCACAAT
6D2	ATCAAGGCTGGCATGGTGGTCA	GGGGGTCGTTCTTGCTGTCA
6H8	GAAGAAGACAGAAACGCAGGAGAAAAAC	CGGGCGGGGGCGGGTCACAGTAAAAC
BL005B_A01	GACAGGTCATGAACTTTTGAAAATAA	AAAGTATATGTACCAAATGGGAGAGC
BL006A_G07	GATGTCCTCTCCACTATACAAGTGTG	GTTTGACTTGTCACCACTTTATCAAC
BL012D_F02	ACAGCCAGAAATAGAAACTTTGAACT	TGAAAGTATGTATTGTTTTCACAGGG
BL013C_E01	AGGATGAAATAATATGCTGTGCTTC	ACCGTGATAAACTCCATCCCTT
BL014D_B11	AGCAAAACTCCTCTATGAATCTCG	ATTGCACACTAAATAGGTGAATACGA
BL279A_G10	ATGGCAGGATGAAGAAAGACAT	ATGCACTTTGGACCCACTGAG
Et.fasta.Contig1023	TGTGGTTATTGGACTACTTCACTCTC	AAACGTCCATTTGACACTGTATTTTA
Et.fasta.Contig1166	GAATGAAGAGAAAATGTTTTGAAGGT	GCACAGTATTGGCTATGAGCAC
Et.fasta.Contig1311	AGAAAACTGTGTCAAGCTTATTTTCC	CAACTTAGTGTTCACATTTCTGAGGT
Et.fasta.Contig1335	CCACTTATGGTAGTTCCCACTTTTAT	GCTAAAGAATACCAAGAACCTTTGAC
Et.fasta.Contig1381	GTCACAGGTATAACATTGAAAGGATG	TAAATGAATCAAACATTGAAGAGAGC

Table 2.2 (continued)

Locus ID	Forward Primer 5' to 3'	Reverse Primer 5' to 3'
Et.fasta.Contig1459	ATAACAAGGACATGTTCTGCTGG	CTAGCAGAACCCTGTATAGCCTG
Et.fasta.Contig1506	AGGATATCCGCTCAGAAATATGAAG	CTGACCACTTGCAAAACTTACTACCT
Et.fasta.Contig1578	CCTAGAACATTACCAAAACAGACTCA	AATGAAGAAGTATTGCATGTGAGAAC
Et.fasta.Contig1647	GTACAACGTCAGGCAAAGCTATTCT	ATCTCCAACACCGTGGCTAAT
Et.fasta.Contig1717	GAACTTGTTGGCAGGTTTCTCTT	CTAGTGATAGGTTGGACATACCAGAG
Et.fasta.Contig1796	TGTGGGTATGTATATGGCTAACTTGT	AGATTTTATGTGCTACTGCATTTACG
Et.fasta.Contig1908	CTCATGACTTAATTGCTGTTCTTCG	ATAACCATTCTGAGGTTTTGAGTTG
Et.fasta.Contig1941	ATCTCCTGCTTCATCTCTTGATTTAT	TAACAGATTTAATAAACGTCCCCTTC
Et.fasta.Contig1943	AGTACGATGAATCTGGTCCTTCAAT	CCACAATACTGACATACTCTGGTCTT
Et.fasta.Contig325	GTGAAGTCAGTGAGTAAAGTCCATGT	CTAGGATACCAGTGGGAGAGTGTAAT
Et.fasta.Contig330	GTCATCACCTCCACTACTTCACAAG	TTTTGGCACTGTAAGATTCTATGAAC
Et.fasta.Contig536	CCTTAGGTAGAACAGACTGAAGCAG	GAAACATGAAACTGGACTTGTTTTAG
Et.fasta.Contig917	GGATGCAGATTCTTCCTATTTTACTC	CTGGTCACTTTACTTGTTTTCAGTGT
Et.fasta.Contig926	TTCATCACATTCTACTTCACAAATCA	CTAGGCAAGCAAGCTTTCTAATAGTT
Et.fasta.Contig93	GAATAAAAGCAACAATTGCAGAGTTA	CTCGACTCCTTCTACGATCTCTACTC
Et.fasta.Contig990	GTTTAGGTTAGTATGAAGGATCCCAA	TGCCAGTACTCACCAATTAGTAAAAG
G1-C12	CCCAAATCCAGGAGTTCAAA	TGGGACCTGGGGCTTCATT
G1-C13	TTGCCCGAGAAAAGGAAGGACATA	CAAGGGTGGGTGAGGGACATC
G1-C5	F-CACTGTTGACTTGGGTTATGTTATT	CTGCTCCTAGGGTTTGTGAAG
G1-C7	CCCGTGTGGCTGGCTTGTGC	TCGGCTACTTTGGTGTTTTTCTCCCTCAT
G1-C9	TGGTCCGGCAACAGCATCAGA	GCTTTTCGGTATTCAACGGCAGAGTG
G1-C9	TGGTCCGGCAACAGCATCAGA	GCTTTTCGGTATTCAACGGCAGAGTG
G1-D5	AGACCCTTGCTGTGTAACTGCT	GACTGGGACTGACTTCTATGACG
G1-D6	CAGCGTGCCCACCCGATAGAA	TCCCAAAAAGTAAAATGTGCAAAGAAAA

Table 2.2 (continued)

Locus ID	Forward Primer 5' to 3'	Reverse Primer 5' to 3'
G1-D7	CAGCGGTGGAAATGACAAACAGG	CCAAGACGACGAGGAACGGTATT
G1-E12	CAACCATGAGAGGAGGCCAGAGAAC	AAAACAGCACTACCTACAAAACCCTATT
G1-F1	TTAGTTTGGGTGCAGACAGGA	GGTGCTCAACAACAAATCAACT
G1-F20	TCCCCAACAACTCCAGCAGAT	GGAAACCACCTAGACGAAAAATG
G1-I18	CATGTTTGTGGGTGTGGTGAA	AAAAGCGGCATCTGGTAAGG
G1-I19	ACCCAGACCTGTCCACCTCA	GAACAGCTCTCCAATCCACAAG
G1-I21	CCAAGCGAAGGAGGCGTGTG	CATGTGGCTCTTTGTTTCTGGA
G1-I5	TAATCGTGTTTGGTGGCATCCTTGAGTC	AGCAGCAGTTCCATTTTCCCACACCA
G1-I8	ACCTGCAGTGGGCTAAGACC	ATGGAAATAATAAAATAAAATGTT
G1-J10	CGTTCGCTTTGCCTGCCACA	GGCTCTTCCCCGGTCGTCCAC
G1-J17	AGCGCCTTCTACACGGACAC	TATGCCCCAATTACTCTTCTGC
G1-J2	TACAGTAACTATGCCAAGATGAAATG	CAATATGGATAATGGCTGTAGACC
G1-J20	ATCCTCCAAGCTCACTACAACA	CCAGCCCCTTCCCAAACAG
G1-J9	CTGTCATTGCCTGCATCGGGGAGAAG	TGTTGAGGGGAAGCAGTTTTG
G1-K2	GCTTTCGCCTTTGACACCTC	GGCCGGACCATTGCTGAAGAAG
G1-L11	AAAGTGACCATCCAGTGCCCAAACCT	CCGGCCGAAACTGACGAGATACATTAG
G1-L13	TCAGCTGCACTAGGTTTGTC	CATTTTGATTTGCTCCATAA
G1-L19	GACAACCTTGAATCCTTTATG	AGATGTTGGTTGGTGACTTAT
G1-L20	TGGGCATAGATGGCAAGGAAAAA	CCCCCAGCATCTCGCATACAC
G1-L7	GTGCTACAGGAAGGAATGGATG	TAGCACAGGAACAGCCGACAATAA
G1-M14	CCGCTTGGACATGAGGAGAT	TGGCAAAGAAACAGAACACAACTA
G1-M19	GAGAAGTAGTGTCCCGGCAGAAAC	ATGGGTGAAAACTTAGGTGAAATG
G1-N9	GCGGGGCAATACATGACGTTCCACAG	GACCCCCATCTCCGTTTCCCATTCC
G1-O1	GGGGTAGAGCACAGTCCAGTT	TTGCAAGGCCGAAAAGGTG

Table 2.2 (continued)

Locus ID	Forward Primer 5' to 3'	Reverse Primer 5' to 3'
G1-O12	GGAATTCCGGGGCACTACT	TCGCGAGGACGGGGAAGAG
G1-O24	CGGCCTTCCTGCAGTACAACCATC	TCGGCAACGTGAAGACCATA
G2-A11	GCCCCTGGAAGCTGTTGTGA	GGGGTCCATCCGAGTCC
G2-A7	TTACCCCACAGACAAAATCAACACC	GGCGGCCCCTCATAGCAC
G2-B1	GGGCCTAGTCCTGCTGGTC	CAAAGAGTGCGGAGAAATGG
G2-B8	CAACATGCGACCACTATAGCCACTTCCT	CGCCACCGCCACCACCACA
G2-C2	TTTGCAGGAAGAGTCATAACACAG	GTCAACAACACCCTTTTCCCTTCCT
G2-D1	GCAGGTCGGCAAGAAGCTAAAGAAGGAA	AGGGTTGGTTTGAAAGGATGTGCTGGTAA
G2-E17	GGAGCACCAAATTCAAGTCAG	CGTCCCCGGTCAATCTCCAC
G2-E19	CCAGTTTGAGCCCCAGGAG	TCGCGGCAGTCAAGAGGTC
G2-F17	TATCCTCTTATTGCTGCATTCTCCTCAC	AGTACGGCCGTTCACCATCTCTG
G2-F2	CACACCACAGACGCATTGAC	TCCCCAGCCTGTGTAGAAC
G2-G13	GGGAGGGGAGAAGGCTACCA	ATACACGGCTTCCATGCTTCTTCTT
G2-G15	CCACGGCCCCACATCCAGC	TCCCGCAGAATTTCCGTATCCAT
G2-G21	TCCAAGAGGGTGTGAGGTGAAC	AAAGCCATGCGAAGCGGAAGAC
G2-G23	GGTTTGGTACTTCAGCGGATGT	CCAAAGCCTGTACTATGCGAAAAG
G2-G5	CGGTCCCTACTGTGGTCTATGGTTTTCA	GGCTCTGCATATCCTCGGTCACACTTCC
G2-G6	CCCATGGCTGCAAGGATTACG	CAGGGGTTGTTGGGAGGCAGTGT
G2-H18	TTGTCAAATGGGCGAGTTCA	TGTTTTGCACCCAGTTTTTG
G2-I18	GATCTCCTCAGGTCTCTTTCA	GATTATGGGCCGGTGTCTCT
G2-I23	TGACTTTCCCAATGTGAGCAGAC	CAGAGGTGGTGTTACAGCAGCAGTTT
G2-J12	CCTCTTGTCCCAGTGCCAGTG	TCCAGGGATCCGAAACAAAG
G2-J21	CCGCCTCAGCCTGTTTCTCTACTTTT	CTTTGAATTTCTGCTTTTGGTGCTCTGC
G2-K12	ACATTAGTCCTGGTTACGAGAGC	AAAGGGCAGTCCAGCATTGA

Table 2.2 (continued)

Locus ID	Forward Primer 5' to 3'	Reverse Primer 5' to 3'
G2-K2	CTGCCCAAGAAGACCGAGAGCCACAAG	AGCGCCCCCTGCACCAAAATCA
G2-L16	CCAAGGGTAGGAGAACAAGACA	ATGGCATGCTGGGAAATCA
G2-L21	GAATCTAGGTCCAAGCAGTCCCATCT	GACCATCACACCACTACCCACACTCA
G2-L3	TGAAAGAGGCCAGAAACAAGTAG	TTCCCAAGGTCTCCATAACAAT
G2-L4	TGGCCAAGAAGATGAAACAGGAAGAGGAG	TGGCAAAGGACACGACGCAGAG
G2-M14	CGGCCTCCTCGACGCATACG	CCAGGCCGGCCCATTGTTC
G2-M24	ACGGAGCACGGTCAGATTTCACG	CCCGGCTGGCTCTTCTTGCTCTT
G2-M3	CGATCCGCATTGAACGAGT	TGTGGCAGGAAGGAGAAGG
G2-N2	CGTGTTTTCCTCCTATGTCGACTTCTTTG	ACGTGCTCTGCCTTTCTTGATCTTGTGTT
G3-D7	AGGATTTCTTGGCCGGTGGAGTGG	GAAGTTGAGGGCCTGGGTGGGGAAGTA
NT001D_E08	AGAAGTTCCTAGATGAGTTGGAGGAG	AATTAATTTCCTAAACCAGGTGACAG
NT010B_E09	GAAGAGGTCCTAAAATATCAAGATGC	ATGATAGACTTCGTCCTTGTCATAGA
NT014D_E01	AAAGAAGTCCCGCATCTAACCT	ATTAAATATGAGAAGATGTGTGCAGG
V2_p1_b8	AGTCACTGTGTTACATTATCACCCAC	ATAATTATACACTGCGGTCTGCATCT
V2_p1_c5	AGTACCTGTTCGACAAGCACAC	TGAGAACATAGACAAGTTAACATACACC
V2_p1_d10	GAGATAGAAAGGCTGCATAAAGAAAT	TATGTTTCAACAATGTACAGGAAACC
V2_p1_d4	CACCAGAACAAGCTGTATTTTTATGT	TGGTTTGCATCATATATTAAAGGGTA
V2_p1_g7	GACTTCAAGCACATTGGGAAAC	ATTGTAAACTTGATAGGCTGGTGAG
V2_p2_g6	AGAATTCCCAATAGCACCTGAAAT	CACTTGGTAAATACATACACACAGCA
V2_p2_h2	CTTTTTGGCCTGGTCTTTTTG	AGATTCTTCAGACTCGTCCTTCTTT
V2_p3_a5	TTTACACAGAAACCTTGTTTATTTGGC	TTTAAGGATGCTTAGAGGCAAAGTATT
V2_p3_b1	AGTCACTGTGTTACATTATCACCCAC	TATACACTGCGGTCTGCATCTACT
V2_p5_b3	AATGGGATGAAGAGCGAGAAT	CTGCCCCATTGACATTTACCTA
V2_p5_h3	CCTTCAGACGAAAACAGCACTAAG	TACAGTGTATGAGAGCCCAATATTTC

Table 2.2 (continued)

Locus ID	Forward Primer 5' to 3'	Reverse Primer 5' to 3'
V2_p6_a4	AGAAATACATCAAATATCGGGTGG	AAAAAGGACAATGTTCAGCTCTCT
V3_p1_a21	ACCAAGTTCTTGGAAAGTGGTG	CTTAGTGTCTCCTGGGTTTGAATAG
V3_p1_b13	GTCTTGGTACTCAATGAAGGAGATG	TCAATCTGATGAAGAGTTTACATGTCT

Table 2.3: Presumptive human chromosome 17 loci that were mapped in *Ambystoma*

Marker ID	Primers[a]	Diagnosis[b]	LG[c]	Symbol[d]	RefSeq ID[e]	E-value[f]
E6E6	F-GAAAACCTGCTCAGCATTAGTGT	ASA	ul	PFN1	NP_005013	E-34
	R-TCTATTACCATAGCATTAATTGGCAG					
E5G5	F-CTATTTCATCTGAGTACCGTTGAATG	PE (A)	23	CGI-125	NP_057144	E-56
	R-TAATGTAGAACTAAATGGCATCCTTC					
	E-CCATGGTGCAGGAAGAGAGCCTATAT					
Rpl19	F-GTCTCATTATCCGCAAACCTGT	SP	1	RPL19	NP_000972	E-67
	R-ATTCTCATCCTCCTCATCCACGAC					
Krt24	F-CCTAGAACATTACCAAAACAGACTCA	RD (Dpn II)	1	KRT10	NP_061889	E-17
	R-AATGAAGAAGTATTGCATGTGAGAAC					
Krt17	F-GAACTTGTTGGCAGGTTTCTCTT	RD (AciI)	1	KRT17	NP_000413	E-146
	R-CTAGTGATAGGTTGGACATACCAGAG					
E10C4	F-CTCCACTATTTAAAGGACATGCTACA	PE (A)	1	SUI1	NP_005792	E-48
	R-TTAATATAGCACAACATTGCCTCATT					
	E-TGCTACATTAATGTAATAAACGGCATCATC					
E6E11	F-AAGAGAAGTTCCTAGATGAGTTGGAG	PE (A)	1	HSPC009	NP_054738	E-26
	R-TGAAGAGAGAACTCAAAGTGTCTGAT					
	E-TCATGTTTTGCTCTGCTGTGCAGT					
E9A10	F-TGATAGTTTCTGGATTAAGACGAGTG	PE (T)	1	FLJ13855	NP_075567	E-15
	R-CTTAGAGCCATTGTTACAAGATGTTC					
	E-GTGATCTAGTGGGATCAAACCCTAAAGACC					
E10C9	F-AAAGTGCCAAGAAGGAGATTAACTT	PE (T)	9	NME1	NP_000260	E-71
	R-GAGCTCAGAAAACAAGGCAGTAAC					
	E-AAATGGATCTACGAGTAGACCTTGACCC					

Table 2.3 (continued)

Marker ID	Primers[a]	Diagnosis[b]	LG[c]	Symbol[d]	RefSeq ID[e]	E-value[f]
E9C9	F-GAGTCTCCTTTAGGATTGACGTATCT	PE (T)	23	FLJ20345	NP_060247	E-17
	R-GCTATGTGAGCAGAGATAAAAGTCAG					
	E-GTTACAGCATCAGTGGGATGTGGTATGT					
E8C9	F-AGGATACCAACCTCTGTGCTATACAT	PE (C)	15	H3F3B	NP_005315	E-66
	R-TAAATGTATTTACAAACCGAAAGCAA					
	E-CGTGGCGAGCGTGCCTAGT					
E9C4	F-GTGGTTATTTGTAACATTTCGTTGAC	PE (A)	8	SFRS2	NP_003007	E-40
	R-AATTACATTTGGGCTTCTCAATTTAC					
	E-TTTTTAAACGCGTAAAAATGTTAACAGA					
E6C5	F-CCGTAAATGTTTCTAAATGACAGTTG	PE (G)	2	ACTG1	NP_001605	0
	R-GGAAAGAAAGTACAATCAAGTCCTTC					
	E-GATTGAAAACTGGAACCGAAAGAAGATAAA					

a – Sequences are 5' amplification primers, 3' amplification primers, or primer extension probes, and are preceded by F-, R-, and E- respectively. b – Genotyping methods are abbreviated: allele specific amplification (ASA), size polymorphism (SP), restriction digestion (RD), primer extension (PE). Diagnostic restriction enzymes and diagnostic extension bases are provided in parentheses. c – *Ambystoma* linkage group ID. "ul" designates markers that are unlinked. d – Official gene symbols as defined by the Human Genome Organization Gene Nomenclature Committee http://www.gene.ucl.ac.uk/nomenclature/. e – Best BLASTX hit (highest e-value) from the human RefSeq database using the contig from which each marker was designed as a query sequence. f – Highest E-value statistic obtained by searching contigs, from which EST markers were designed, against the human RefSeq database.

Table 2.4: EST summary and assembly results.

	A. mexicanum	*A. t. tigrinum*
cDNA clones sequenced	21830	24234
high-quality sequences	19383	20599
mt DNA sequence	2522	1223
seqs submitted to NCBI	16861	19376
sequences assembled	32891a	19376
PaCE clusters	11381	10226
ESTs in contigs	25457	12676
contigs	3756	3201
singlets	7434	6700
putative transcripts	11190	9901

[a]Includes 16,030 ESTs from (Habermann *et al*, 2004).

Table 2.5: *Ambystoma* contigs search of NCBI human RefSeq, nr, and *Xenopus* Unigene databases.

# BLASTX Best Matches	*A. mexicanum*	A. t. tigrinum
Total	6283	5545
< E-100	630	870
< E-50 > E-100	2015	1990
< E-20 > E-50	2153	1595
< E-10 > E-20	967	745
< E-7 > E-10	518	345

Table 2.6: Top 20 contigs with the most assembled ESTs. (*A. mexicanum*)

Contig ID	# ESTs	Best Human Match	E-value
MexCluster_4615_Contig1	415	(NM_000519) delta globin	E-39
MexCluster_600_Contig1	354	(NM_182985) ring finger protein 36 isoform a	E-110
MexCluster_6279_Contig1	337	(NM_000559) A-gamma globin	E-32
MexCluster_10867_Contig1	320	(NM_000558) alpha 1 globin	E-38
MexCluster_5357_Contig1	307	(NM_000558) alpha 1 globin	E-37
MexCluster_9285_Contig3	285	(NM_001614) actin, gamma 1 propeptide	0
MexCluster_7987_Contig3	252	(NM_001402) eukaryotic translation elongation f1	0
MexCluster_9285_Contig1	240	(NM_001101) beta actin; beta cytoskeletal actin	0
MexCluster_9279_Contig3	218	(NM_000223) keratin 12	E-113
MexCluster_11203_Contig1	181	(NM_002032) ferritin, heavy polypeptide 1	E-70
MexCluster_8737_Contig2	152	(NM_058242) keratin 6C	E-131
MexCluster_3193_Contig1	145	(NM_004499) heterogeneous nuclear ribonucleoprotein	E-90
MexCluster_8737_Contig7	134	(NM_058242) keratin 6C	E-131
MexCluster_5005_Contig3	132	(NM_031263) heterogeneous nuclear ribonucleoprotein	E-124
MexCluster_6225_Contig1	125	(NM_001152) solute carrier family 25, member 5	E-151
MexCluster_1066_Contig1	122	[31015660] IMAGE:6953586	E-16
MexCluster_8737_Contig4	114	(NM_058242) keratin 6C; keratin, epidermal type II	E-132
MexCluster_8187_Contig2	113	(NM_005507) cofilin 1 (non-muscle)	E-65
MexCluster_2761_Contig1	109	(NM_001961) eukaryotic translation elongation factor2	0
MexCluster_9187_Contig1	105	(NM_007355) heat shock 90 kDa protein 1, beta	0

Table 2.7: Top 20 contigs with the most assembled ESTs. (*A. t. tigrinum*)

Contig ID	# ESTs	Best Human Match	E-value
TigCluster_6298_Contig1	654	(NM_000519) delta globin	E-38
TigCluster_10099_Contig2	193	(NM_001614) actin, gamma 1 propeptide	0
TigCluster_6470_Contig1	167	(NM_000558) alpha 1 globin	E-39
TigCluster_9728_Contig2	142	(NM_000477) albumin precursor	E-140
TigCluster_6594_Contig1	117	(NM_001402) eukaryotic translation elongation f1	0
TigCluster_5960_Contig1	91	(NM_001101) beta actin; beta cytoskeletal actin	0
TigCluster_7383_Contig1	77	(NM_001614) actin, gamma 1 propeptide	0
TigCluster_6645_Contig1	76	(NM_001063) transferrin	0
TigCluster_7226_Contig4	74	(NM_006009) tubulin, alpha 3	E-160
TigCluster_7191_Contig1	67	(NM_019016) keratin 24	E-89
TigCluster_10121_Contig1	64	(NM_005141) fibrinogen, beta chain preproprotein	0
TigCluster_6705_Contig1	63	(NM_000558) alpha 1 globin	E-39
TigCluster_7854_Contig1	62	(NM_021870) fibrinogen, gamma chain isoform	E-121
TigCluster_6139_Contig1	52	(NM_001404) eukaryotic translation elongation f1	0
TigCluster_7226_Contig2	51	(NM_006009) tubulin, alpha 3	0
TigCluster_10231_Contig1	44	(NM_003018) surfactant, pulmonary-associated prot.	E-08
TigCluster_6619_Contig1	36	(NM_000041) apolipoprotein E	E-38
TigCluster_7232_Contig2	35	(NM_003651) cold shock domain protein A	E-46
TigCluster_5768_Contig1	34	(NM_003380) vimentin	E-177
TigCluster_9784_Contig3	32	|XP_218445.1| similar to RIKEN cDNA 1810065E05	E-15

Table 2.8: Functional annotation of contigs

	A. mexicanum	A. t. tigrinum
Molecular Function (0016209)		
antioxidant (0016209)	25	29
binding (0005488)	3117	2578
chaparone (0003754)	100	85
enzyme regulation (003023)	193	223
motor (0003774)	73	75
signal transduction (0004871)	344	375
structural protein (0005198)	501	411
transcriptional reg. (0030528)	296	221
translational reg. (0045182)	94	59
bone remodeling (0046849)	8	8
circulation (0008015)	23	78
immune response (000695)	182	263
respiratory ex. (0009605)	254	288
respiratory in. (0009719)	72	58
stress (0006950)	263	320
Biological Process (0008150)		
Cellular (0009987)		
activation (0001775)	4	6
aging and death (0008219)	158	148
communication (0007154)	701	696
differentiation (0030154)	31	20
extracellular mat. (0043062)	4	4
growth and main. (0008151)	1731	1445
migration (0016477)	8	14
motility (0006928)	163	154
Developmental (0007275)		
aging (0007568)	32	21
embryonic (0009790)	6	1
growth (0040007)	2	2
morphogenesis (0009653)	350	272
pigment (0048066)	13	26
post embryonic (0009791)	8	13
reproduction (0000003)	42	27

Table 2.8 (continued)

	A. mexicanum	*A. t. tigrinum*
Physiological (0007582)		
coagulation (0050817)	22	73
death and aging (0016265)	159	148
homeostasis (0042592)	22	27
metabolism (0008152)	3059	2513
secretion (0046903)	9	16
sex differentiation (0007548)	3	2

Numbers in parentheses reference GO numbers (Gene Ontology Consortium).

Table 2.9: *Ambystoma* contigs that show sequence similarity to rat spinal cord injury genes.

Ambystoma Contig ID	RAT cDNA clone	E-value
MexCluster_7440_Contig1	gi|1150557|c-myc, exon 2	E-29
MexCluster_4624_Contig1	gi|1468968| brain acyl-CoA synthtase II	E-09
TigCluster_4083_Contig1		E-09
TigSingletonClusters_Salamander_4_G20_ab1	gi|1552375| SKR6 gene, a CB1 cannabinoid recept.	E-08
MexSingletonClusters_NT009B_B04	gi|17352488| cyclin ania-6a	E-46
TigCluster_3719_Contig1		E-114
TigCluster_8423_Contig1	gi|1778068| binding zyginI	E-102
TigCluster_7064_Contig1	gi|1836160| Ca2+/calmodulin-dependent	E-20
MexCluster_3225_Contig1	gi|1906612| Rattus norvegicus CXC chemokine	E-68
TigSingletonClusters_Salamander_13_F03_ab1		E-38
MexSingletonClusters_BL285B_A06	gi|203042| (Na+, K+)-ATPase-beta-2 subunit	E-63
TigCluster_6994_Contig1		E-65
MexSingletonClusters_BL014B_F12	gi|203048| plasma membrane Ca2+ ATPase-isoform 2	E-112
TigSingletonClusters_Salamander_5_F07_ab1		E-92
MexCluster_1251_Contig1	gi|203167| GTP-binding protein (G-alpha-i1)	E-110
TigSingletonClusters_Salamander_3_P14_ab1		E-152
TigSingletonClusters_Salamander_22_B01_ab1	gi|203336| catechol-O-methyltransferase	E-47
TigSingletonClusters_Salamander_17_N04_ab1	gi|203467| voltage-gated K+ channel protein (RK5)	E-08
MexSingletonClusters_v1_p8_c16_triplex5ld_	gi|203583| cytosolic retinol-binding protein (CRBP)	E-77
TigCluster_6321_Contig1		E-18
MexCluster_5399_Contig1	gi|204647| heme oxygenase gene	E-67
TigCluster_2577_Contig1		E-67
MexCluster_4647_Contig1	gi|204664| heat shock protein 27 (Hsp27)	E-83
TigSingletonClusters_Salamander_12_M05_ab1		E-51
MexSingletonClusters_BL285C_F02	gi|205404| metabotropic glutamate receptor 3	E-41

Table 2.9 (continued)

Ambystoma Contig ID	RAT cDNA clone	E-value
TigSingletonClusters_Salamander_2_B24_ab1	gi|205508| myelin/oligodendrocyte glycoprotein	E-26
TigCluster_5740_V2_p10_M20_TriplEx5ld_	gi|205531| metallothionein-2 and metallothionein 1	E-08
TigSingletonClusters_V2_p5_A2_TriplEx5ld_	gi|205537| microtubule-associated protein 1A	E-59
MexCluster_1645_Contig1	gi|205633| Na, K-ATPase alpha-2 subunit	E-149
TigSingletonClusters_Contig328		0
TigSingletonClusters_Contig45	gi|205683| smallest neurofilament protein (NF-L)	E-63
MexSingletonClusters_NT016A_A09	gi|205693| nerve growth factor-induced (NGFI-A)	E-95
TigSingletonClusters_I09_Ag2_p9_K24_M13R		E-24
MexSingletonClusters_NT007A_E07	gi|205754| neuronal protein (NP25)	E-64
TigCluster_7148_Contig1		E-57
MexCluster_9504_Contig1	gi|206161| peripheral-type benzodiazepine receptor	E-73
MexSingletonClusters_BL016B_B02	gi|206166| protein kinase C type III	E-36
TigCluster_981_Contig1		E-27
MexSingletonClusters_nm_19_k3_t3_	gi|206170| brain type II Ca2+/calmodulin-dependent	E-117
MexSingletonClusters_v11_p42_j20_t3_049_ab1	gi|207138| norvegicus syntaxin B	E-79
MexSingletonClusters_nm_14_h19_t3_	gi|207473| neural receptor protein-tyrosine kinase	E-40
TigSingletonClusters_Contig336		E-34
TigSingletonClusters_E10_Ag2_p18_O19_M13	gi|2116627| SNAP-25A	E-123
MexCluster_211_Contig1	gi|220713| calcineurin A alpha	E-63
TigSingletonClusters_Salamander_7_K14_ab1		E-87
MexSingletonClusters_NT014A_G03	gi|220839| platelet-derived growth factor A chain	E-21
TigSingletonClusters_Salamander_9_M15_ab1		E-56
TigSingletonClusters_Salamander_19_M06_ab1	gi|2501807| brain digoxin carrier protein	E-55
MexSingletonClusters_Contig100	gi|2746069| MAP-kinase phosphatase (cpg21)	E-108

TABLE 2.9 (continued)

Ambystoma Contig ID	RAT cDNA clone	E-value
TigSingletonClusters_Salamander_11_A16_ab1		E-70
MexCluster_8345_Contig1	gi|2832312| survival motor neuron (smn)	E-40
TigCluster_8032_Contig1		E-49
MexCluster_3580_Contig1	gi|294567| heat shock protein 70 (HSP70)	0
TigCluster_8592_Contig2		E-161
TigSingletonClusters_Salamander_17_N08_ab1	gi|2961528| carboxyl-terminal PDZ	E-10
MexSingletonClusters_BL286C_D09	gi|298325| sodium-dependent neurotransmitter tran.	E-12
TigSingletonClusters_Contig95		E-22
MexSingletonClusters_Contig461	gi|2996031| brain finger protein (BFP)	E-08
TigSingletonClusters_Salamander_11_O19_ab1		E-23
TigSingletonClusters_E16_Ag2_p8_O20_M13R	gi|3135196| Ca2+/calmodulin-dependent	E-33
MexSingletonClusters_Contig188	gi|3252500| CC chemokine receptor protein	E-15
MexCluster_6961_Contig1	gi|3319323| suppressor of cytokine signaling-3	E-08
MexSingletonClusters_nm_14_p15_t3_	gi|349552| P-selectin	E-16
TigCluster_218_Contig2		E-99
MexSingletonClusters_Contig506	gi|3707306| Normalized rat embryo, cDNA clone	E-14
TigSingletonClusters_I16_Ag2_p5_N7_M13R	gi|3711670| Normalized rat muscle, cDNA clone	E-35
MexSingletonClusters_V1_p1_a10_Triplex5Ld	gi|3727094| Normalized rat ovary, cDNA clone	E-15
TigSingletonClusters_v2_p1_D20_triplex5ld		E-16
MexSingletonClusters_NT005B_F02	gi|3811504| Normalized rat brain, cDNA clone	E-35
TigSingletonClusters_Salamander_22_I04_ab1		E-34
TigSingletonClusters_Ag2_p34_N23_M13R	gi|405556| adenylyl cyclase-activated serotonin	E-17
TigSingletonClusters_Salamander_1_H02_ab1	gi|4103371| putative potassium channel TWIK	E-22

Table 2.9 (continued)

Ambystoma Contig ID	RAT cDNA clone	E-value
MexCluster_4589_Contig1	gi|4135567| Normalized rat embryo, cDNA clone	E-32
TigSingletonClusters_Contig220		E-09
TigCluster_4093_Contig1	gi|4228395| cDNA clone UI-R-A0-bc-h-02-0-UI	E-104
MexSingletonClusters_nm_21_2_m7_t3_	gi|425471| nuclear factor kappa B p105 subunit	E-22
TigCluster_8535_Contig1		E-11
MexSingletonClusters_v6_p1_j6_triplex5_1ld_	gi|430718| Sprague Dawley inducible nitric oxide	E-13
TigSingletonClusters_Salamander_15_D22_ab1		E-41
MexCluster_3498_Contig1	gi|436934| Sprague Dawley protein kinase C rec.	0
TigCluster_6648_Contig1		0
MexSingletonClusters_BL279A_B12	gi|464196| phosphodiesterase I	E-49
TigSingletonClusters_Salamander_25_P03_ab1		E-75
MexCluster_8708_Contig1	gi|466438| 40kDa ribosomal protein	E-168
TigCluster_5877_Contig1		E-168
MexSingletonClusters_nm_14_a9_t3_	gi|493208| stress activated protein kinase alpha II	E-51
TigSingletonClusters_Salamander_11_A13_ab1	gi|517393| tau microtubule-associated protein	E-44
TigSingletonClusters_Salamander_12_J14_ab1	gi|55933| c-fos	E-26
MexSingletonClusters_nm_21_2_l13_t3_	gi|56822| major synaptic vesicel protein p38	E-39
TigCluster_2065_Contig1		E-50
MexCluster_10965_Contig1	gi|56828| nuclear oncoprotein p53	E-75
TigCluster_5315_Contig1		E-66
MexCluster_4245_Contig1	gi|56909| pJunB gene	E-50
TigSingletonClusters_G05_Ag2_p9_G8_M13R		E-09
MexSingletonClusters_NT013D_C12	gi|56919| region fragment for protein kinase C	E-33

Table 2.9 (continued)

Ambystoma Contig ID	RAT cDNA clone	E-value
TigSingletonClusters_Salamander_21_H19_ab1		E-24
MexCluster_9585_Contig1	gi|57007| ras-related mRNA rab3	E-61
TigCluster_4885_Contig1		E-63
TigSingletonClusters_Salamander_1_M03_ab1	gi|57238| silencer factor B	E-13
MexSingletonClusters_NT008B_D05	gi|57341| transforming growth factor-beta 1	E-13
TigSingletonClusters_Salamander_24_I16_ab1		E-20
MexCluster_9533_Contig1	gi|57479| vimentin	0
TigCluster_5768_Contig1		0
MexSingletonClusters_BL283B_A11	gi|596053| immediate early gene transcription	E-12
TigSingletonClusters_Salamander_13_J19_ab1		E-16
MexSingletonClusters_v6_p4_j2_triplex5_1ld_	gi|790632| macrophage inflammatory protein-1alpha	E-22
TigCluster_2146_Contig1	gi|951175| limbic system-associated membrane prot.	E-11
MexSingletonClusters_v11_p54_o4_t3_	gi|971274| neurodegeneration associated protein 1	E-09
TigSingletonClusters_Salamander_2_J12_ab1		E-11

Table 2.10: *Ambystoma* limb regeneration contigs that show sequence similarity to zebrafish fin regeneration ESTs

A. mexicanum Contigs	Human ID	E-value	Zfish ID	E-value
Contig94	gi|10835079|	E-63	gnl|UG|Dr#S12319632	E-58
nm_30_a11_t3_	gi|32306539|	E-58	gnl|UG|Dr#S12312602	E-35
Contig615	gi|4502693|	E-70	gnl|UG|Dr#S12313407	E-34
nm_23_l13_t3_	No Human Hit	-	gnl|UG|Dr#S12320916	E-31
nm_9_e22_t3_	gi|4758788|	E-98	gnl|UG|Dr#S12309914	E-29
nm_8_l17_t3_	gi|21361310|	E-16	gnl|UG|Dr#S12313396	E-27
Contig531	gi|13775198|	E-27	gnl|UG|Dr#S12309680	E-26
Contig152	gi|5453712|	E-32	gnl|UG|Dr#S12239884	E-26
nm_32h_j20_t3_	gi|39777601|	E-79	gnl|UG|Dr#S12136499	E-25
Contig1011	gi|39752675|	E-65	gnl|UG|Dr#S12136499	E-24
v11_p50_b24_t3_	gi|41208832|	E-36	gnl|UG|Dr#S12319219	E-23
Contig589	gi|4506505|	E-56	gnl|UG|Dr#S12312662	E-22
Contig785	gi|33695095|	E-61	gnl|UG|Dr#S12264765	E-22
Contig157	gi|21361122|	E-138	gnl|UG|Dr#S12313094	E-21
v11_p42_j20_t3_049_ab1	gi|47591841|	E-100	gnl|UG|Dr#S12137806	E-21
Contig610	gi|10801345|	E-114	gnl|UG|Dr#S12310326	E-20
nm_27_o1_t3_	gi|7706429|	E-72	gnl|UG|Dr#S12310422	E-19
Contig439	gi|4504799|	E-25	gnl|UG|Dr#S12309233	E-19
nm_31_d5_t3_	gi|8923956|	E-50	gnl|UG|Dr#S12264745	E-17
v11_p41_h12_t3_026_ab1	No Human Hit	-	gnl|UG|Dr#S12320916	E-17
Contig129	gi|34932414|	E-103	gnl|UG|Dr#S12313534	E-17
nm_14_j21_t3_	gi|4505325|	E-42	gnl|UG|Dr#S12136571	E-17
Contig1321	gi|4501857|	E-80	gnl|UG|Dr#S12309233	E-17
nm_19_k3_t3_	gi|26051212|	E-106	gnl|UG|Dr#S12137637	E-17
Contig488	gi|4557525|	E-105	gnl|UG|Dr#S12311975	E-15
nm_35h_k19_t3_	gi|16950607|	E-43	gnl|UG|Dr#S12196214	E-15
Contig195	gi|4557231|	E-99	gnl|UG|Dr#S12309233	E-14
nm_14_h19_t3_	gi|4503787|	E-86	gnl|UG|Dr#S12310912	E-13
v11_p51_d20_t3_	gi|30520322|	E-19	gnl|UG|Dr#S12321150	E-13
g3-n14	gi|13654278|	E-23	gnl|UG|Dr#S12318856	E-13
nm_29_f2_t3_	gi|4506517|	E-65	gnl|UG|Dr#S12312662	E-13
g4-h23	gi|24111250|	E-33	gnl|UG|Dr#S12312651	E-13
Math_p2_A2_T3_	No human Hit	-	gnl|UG|Dr#S12078998	E-13
nm_35h_f4_t3_	gi|41148476|	E-67	gnl|UG|Dr#S12319663	E-13
Contig952	gi|21264558|	E-61	gnl|UG|Dr#S12318843	E-12
g4-g21	gi|11995474|	E-65	gnl|UG|Dr#S12192716	E-12
Contig854	gi|8922789|	E-117	gnl|UG|Dr#S12313534	E-11
Contig1105	gi|6912638||	E-83	gnl|UG|Dr#S12079967	E-11

Table 2.10 (continued)

A. mexicanum Contigs	Human ID	E-value	Zfish ID	E-value
nm_26_f7_t3_	gi|30181238|	E-83	gnl|UG|Dr#S12319880	E-11
Contig949	gi|21284385|	E-68	gnl|UG|Dr#S12290856	E-11
g3-n3	gi|18490991|	E-64	gnl|UG|Dr#S12320832	E-10
v11_p41_m16_t3_007_ab1	gi|4885661|	E-33	gnl|UG|Dr#S12310912	E-10
Contig653	gi|4505047|	E-124	gnl|UG|Dr#S12239868	E-9
Contig1349	gi|9665259|	E-46	gnl|UG|Dr#S12320840	E-9
6h12	gi|31317231|	E-43	gnl|UG|Dr#S12321311	E-9
v11_p43h_i14_t3_070_ab1	No Human Hit	-	gnl|UG|Dr#S12320916	E-9
nm_35h_d11_t3_	gi|7661790|	E-35	gnl|UG|Dr#S12196146	E-9
nm_35h_k22_t3_	gi|5031977|	E-124	gnl|UG|Dr#S12242267	E-9
v11_p48_g2_t3_087_ab1	gi|11496277|	E-60	gnl|UG|Dr#S12312396	E-9
nm_30_e11_t3_	gi|32483357|	E-56	gnl|UG|Dr#S12309103	E-8
nm_28_f23_t3_	gi|42544191|	E-25	gnl|UG|Dr#S12239884	E-8
nm_12_p16_t3_	gi|21361553|	E-21	gnl|UG|Dr#S12310912	E-8
nm_32h_a8_t3_	gi|11386179|	E-22	gnl|UG|Dr#S12312152	E-8

Figure 2.1: Results of BLASTX and TBLASTX searches to identify best BLAST hits for *Ambystoma* contigs searched against NCBI human RefSeq, nr, and *Xenopus* Unigene databases

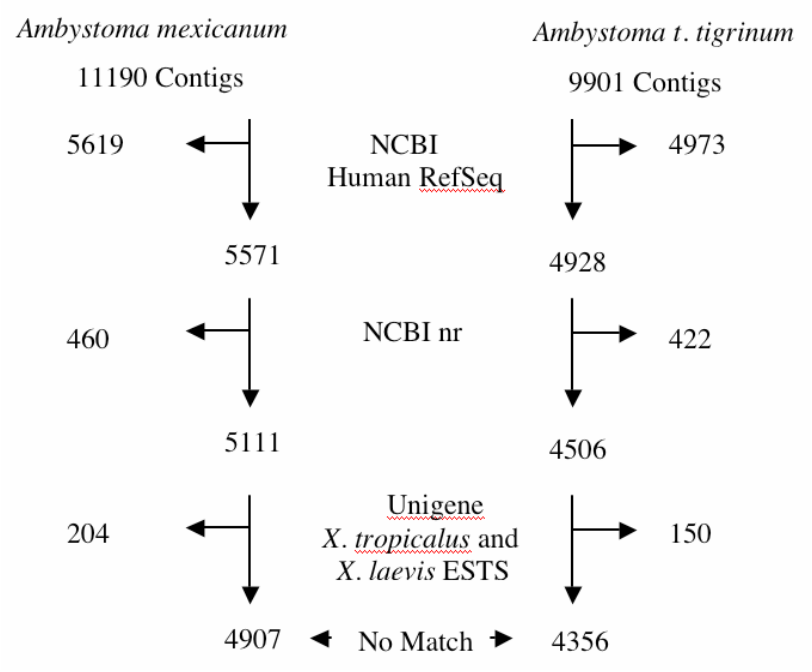


Figure 2.2: Venn diagram of BLAST comparisons among amphibian EST projects. Values provided are numbers of reciprocal best BLAST hits (E<10-20) among quality masked *A. mexicanum* and *A. t. tigrinum* assemblies and a publicly available *X. tropicalis* EST assembly (http://www.sanger.ac.uk/Projects/X_tropicalis).

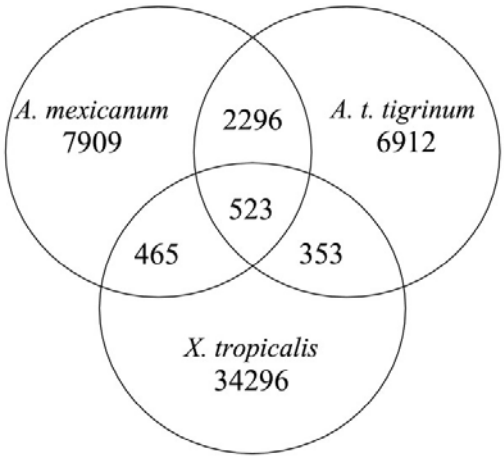


Figure 2.3: Results of BLASTN and TBLASTX searches to identify best BLAST hits for *A. mexicanum* regeneration ESTs searched against zebrafish EST databases. A total of 14,961 A. mexicanum limb regeneration ESTs were assembled into 4485 contigs for this search.

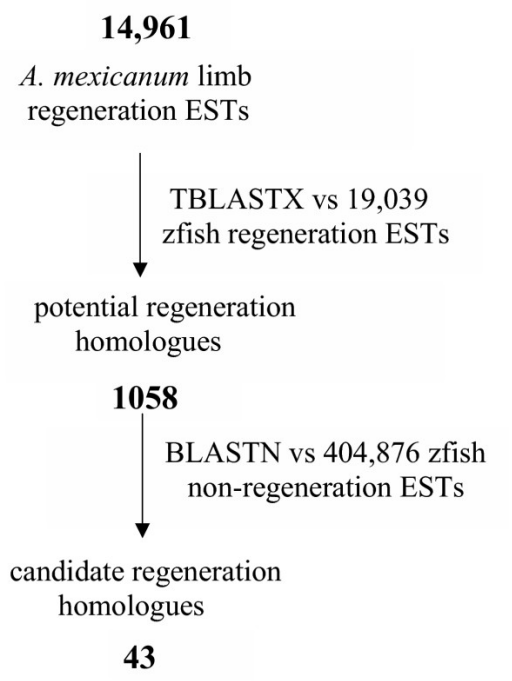
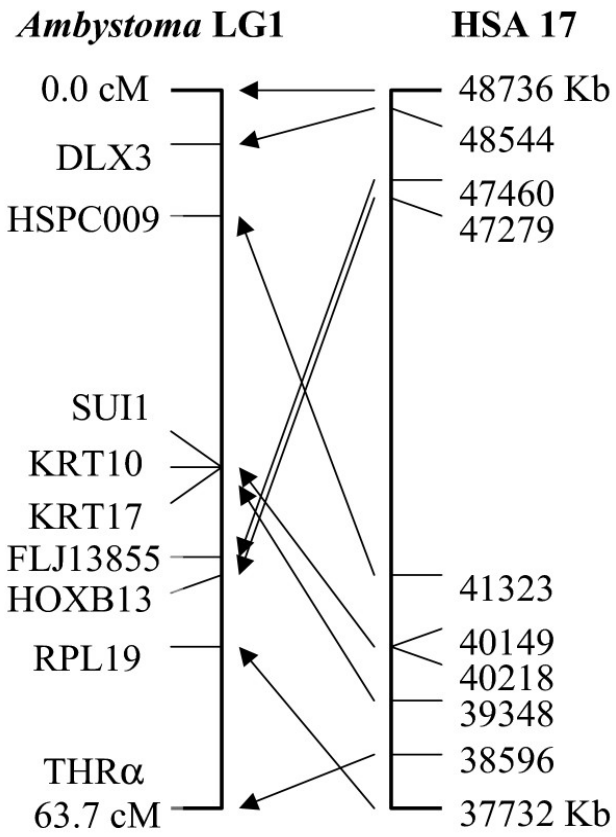


Figure 2.4: Comparison of gene order between *Ambystoma* linkage group 1 and an 11 Mb region of Hsa17 (37.7 Mb to 48.7 Mb). Lines connect the positions of putatively orthologous genes.



CHAPTER 3: A comprehensive EST Linkage Map for Tiger Salamander and Mexican Axolotl: Enabling Gene Mapping and Comparative Genomics in *Ambystoma*

Introduction

The tiger salamander (*Ambystoma tigrinum*) species complex consists of several closely related and phenotypically diverse taxa that range from central Mexico to southern Canada (Shaffer and McKnight, 1996). The complex as a whole is an important, naturalistic model system because taxa are characterized by extensive interspecific and intraspecific variation for a number of ecologically important traits including: expression of metamorphic *vs.* non-metamorphic (paedomorphic) life histories (Gould, 1977; Shaffer and Voss, 1996), timing of metamorphosis (Rose and Armentrout, 1976; Voss and Smith, 2005), cannibal *vs* normal larval morphologies (Powers, 1907; Hoffman and Pfennig, 1999), infectious disease (Collins *et al*, 2004), variation in adult coloration and pigment patterning (Reese 1969; Parichy, 1996; Parichy, 1998), and variation in general morphology (Shaffer, 1984; Irschick and Shaffer, 1997). In addition, these salamanders are important laboratory models for olfaction (Marchand *et al*, 2004; Park *et al*, 2004), vision (Chichilinski and Reike, 2005; Thoreson *et al*, 2004), cardiogenesis (Denz *et al*, 2004; Zhang *et al*, 2004), embryogenesis (Bachvarova, 2004; Ericsson *et al*, 2004), and post-embryonic development (Parichy, 1998; Voss and Smith, 2005), including organ and tissue regeneration (Christensen *et al*, 2002; Schnapp and Tanaka, 2005). Both natural- and laboratory-based research areas are in need of a comprehensive genome map that can be used to identify the position and effect of loci that contribute to phenotypic variation, and that can be used to compare features of the salamander genome to other vertebrates. Moreover, molecular markers that are used to develop linkage maps provide the material to generate nucleotide probes and sequences that can be used in a variety of ways, including *in situ* hybridization, population genetics, systematics, molecular evolution, and functional genomics.

Several lines of evidence suggest that meiotic or recombinational map size is large in tiger salamanders. The earliest mapping studies in *A. mexicanum* employed tetrad analysis and gynogenic salamanders to estimate genetic distance between a few phenotypic markers and their centromeres (*e.g.* 3 in Lindsley *et al*, 1955; 7 in Armstrong,

1984). Most of these phenotypes were mapped to positions distant from centromeres (unlinked); this suggests a large recombinational map size for *Ambystoma* relative to other vertebrates where the majority of loci typically show linkage with the centromere (Brown *et al*, 1998; Steen, 1999; Groenen *et al*, 2000; Kong *et al*, 2002; Kelly *et al*, 2000). Using a more direct method, Callan (1966) reported chiasmata counts from *A. mexicanum* oocytes nuclei ranging from 101 to 126, averaging 113 chiasmata per nucleus. Assuming that one crossover is equivalent to a 50 cM map distance (Sturtevant, 1913), this chiasmata frequency converts to an estimated map distance of 5,650 cM, which is approximately 4.2x larger than the mouse genome and 1.6x larger than the human genome (Table 3.1). Finally, Voss *et al* (2001) estimated map size from a linkage analysis in which 347 molecular genotypes were obtained for offspring from backcrosses between *A. mexicanum* and *A. mexicanum/A. t. tigrinum* hybrids. That analysis yielded one of the largest partial genetic linkage maps ever constructed (~3400 cM) and an overall map size estimate of 7291 cM, based on the method of Hulbert *et al* (1988). Thus, all studies to date suggest that recombinational map size is relatively larger in *Ambystoma* than it is in other vertebrates. Although recombinational map size in *Ambystoma* may simply be a consequence of large physical genome size (Table 3.1), the contribution of additional factors that are known to strongly influence map length in other vertebrates (*e.g.* chromosome number and morphology) have not been considered.

Recently completed expressed sequence tag (EST) sequencing projects for two tiger salamander species, *A. mexicanum* (Haberman *et al*, 2004; Putta *et al* 2004) and *A. tigrinum tigrinum* (Putta *et al*, 2004), permitted the development of a sufficient number of EST based markers to generate a comprehensive genetic map for *Ambystoma*, and the first such map for any amphibian. The following chapter describes new resource and shows its utility in two areas: (1) mapping phenotypic mutants and (2) investigating the relative effects of genome size and chromosome number on recombination-based estimates of genome size. Additionally, the importance of these EST-based PCR markers as a general resource for tiger salamander research is reviewed.

Materials and Methods

Study species

***Ambystoma tigrinum tigrinum*:** The Eastern tiger salamander is the archetype of the tiger salamander species complex. This relatively large salamander species breeds in ponds and temporary pools in the eastern and southern parts of the United States; the individual that was used in this study was purchased from Charles D. Sullivan Co. Inc., which collects and distributes amphibians from breeding ponds in the vicinity of Nashville, TN. These salamanders undergo a typical amphibian life cycle, wherein they hatch from the egg in an aquatic-larval form and eventually undergo a metamorphosis through which they achieve a terrestrial adult form.

***Ambystoma mexicanum*:** The Mexican axolotl is native to a single lake in central Mexico (Lake Xochimilco, Mexico D.F., Mexico). It is commonly used in biological research and strains have been maintained in laboratory culture for nearly a century and a half (Dumeril, 1870; see also review by Smith, 1989). The most notable feature of this salamander is that it exhibits a phenotype, known as paedomorphosis, wherein it forgoes metamorphic changes typical of most ambystomatid salamanders and retains a largely larval morphology throughout adult life. Several developmental mutants have been described for *A. mexicanum* (Malacinski, 1989) and are readily available (www.ambystoma.org). The genomic position of two recessive pigment mutants: *white* (Häcker, 1907) and *melanoid* (Humphrey and Bagnara, 1967) were estimated in relation to the new genetic map. In comparison to wildtype, white is characterized by a general lack of epidermal pigment cells, which is caused by failure of pigment cell precursors to migrate or survive in epidermal tissues (reviewed by Parichy *et al*, 1999). In comparison to wildtype, melanoid is characterized by a lack of iridophores in the skin and iris, relatively fewer xanthophores and greater numbers of melanophores (Humphrey and Bagnara, 1967).

Molecular markers

A total of 1095 molecular markers were genotyped using a previously described AxTg mapping family that was generated by backcrossing *A. mexicanum*/*A. t. tigrinum* F1 hybrids to a single *A. mexicanum* (Voss, 1995). In a previous study using AxTg, Voss *et*

al (2001) reported linkage analyses of 323 "anonymous" molecular markers (315 AFLPs and 8 RAPDs) and 24 markers for nuclear protein-coding sequences. This chapter reports an additional 202 AFLP markers (scored for 44 backcross offspring) and 546 EST based markers (scored for 91 backcross offspring). Marker fragments were amplified using standard PCR conditions (150 ng DNA, 50 ng each primer, 1.2 mM $MgCl_2$, 0.3 U Taq polymerase, 1x PCR buffer, 200 mM each of dATP, dCTP, dGTP, dTTP; thermal cycling at 94° C for 4 minutes; 33 cycles of 94° C for 45 s, 60° C for 45 s, 72° C for 30 s; and 72° C for 7 minutes). Polymorphisms were diagnosed by: primer extension (Perkin Elmer, AcycloPrime -FP chemistry and Wallac, Victor3 plate reader) (*e.g.* Chen *et al*, 1999; Hsu *et al*, 2001; Gardiner and Jack, 2002), thermal gradient capillary gel electrophoresis (SpectruMedix, Reveal), allele-specific amplification, PCR fragment size polymorphism, and restriction digestion. Primer sequences, diagnostic polymorphisms, polymorphism detection assays, numbers of individuals genotyped, and NCBI GI numbers for source sequences for all 570 protein-coding loci are summarized in Supplementary_Table_3_1. The design of AFLP markers, EST based markers, and primer extension probes is described in detail below.

Generation of AFLP and RAPD markers: The methods used to generate AFLPs and RAPDs were reported previously (Voss and Shaffer, 1997; Voss *et al*, 2001). An additional 202 AFLPs were generated for this study using the same methods and backcross progeny. Primer sequences for all AFLPs are provided in Supplementary_Table_3_2.

EST based marker design: The majority of EST based markers were designed to target 3′ unconserved regions of assembled EST sequences that were derived from *A. mexicanum* and *A. t. tigrinum* (Putta *et al*, 2004). These sequences were identified by first aligning (BLASTx) assembled sequences to a custom database of Human protein sequences (Ref Seq Release 3, NCBI). The 3′ most 70 base pairs of aligning sequence and all remaining non-aligning 3′ sequences were targeted for primer design. These regions presumably correspond to sequence encoding carboxy-terminal regions and 3′ untranslated regions. Primers were designed within these regions using Primer3 software (www.mit.wi.edu/Primer3.html). When presumably orthologous sequences (>90% identity) were available for both *A. mexicanum* and *A. t. tigrinum*, marker design was

additionally constrained to flank at least one polymorphism between the two sequences.

Primer extension probe design: Primer extension probes were designed to target polymorphisms that were observed between presumptively orthologous EST sequences or between EST sequences and sequenced marker fragments that were amplified from *A. mexicanum* and *A. t. tigrinum*. Design of primer extension probes was constrained such that probes were complementary to sequence immediately 3′ of targeted SNP positions. The software package Array Designer 2.1 (PREMEIR Biosoft International) was used to identify probes with Tm 60 ± 8° C and high Δg for primer hairpin and primer dimer formation.

Construction of the Linkage Map

A linkage map was constructed for the tiger salamander using genotypes for 570 protein-coding markers (EST and gene based) and 525 AFLP markers. Linkage analyses were performed using Mapmaker/EXP 3.0b and linkage groups were visualized using MapManagerQXTb19 (Meer *et al*, 2004) and MapChart 2.1 (Voorrips, 2002). This linkage map was constructed by first identifying sets of markers that formed linear linkage groups supported by log of odds (LOD) of 5.0 or greater, and then recursively adding markers to these initial groups at lower LOD thresholds (4.0, 3.5, and 3.0). At each step, linkage groups were visually inspected to verify the linear arrangement among markers. At each LOD threshold, the addition of new markers to an existing linkage group or merger of two existing linkage groups was constrained such that addition or merger did not disrupt arrangements that were previously set at a higher LOD threshold. A few markers could not be assigned to precise positions within a linkage group, but could be assigned to a single linkage group at LOD > 4.0 (Supplementary_Table_3_1). Linkage distances among markers in the final map were estimated using Kosambi (1944) mapping function in MapManagerQXTb19 (Meer *et al*, 2004).

QTL analysis

Interval analysis was used to identify the positions of two mutant phenotypes that segregated within the AxTg mapping family: *white*, which is specified by the recessive locus *d*; and *melanoid*, which is specified by the recessive locus *m*. Interval analysis was

also used to identify the position of a major effect QTL that contributes to segregation of metamorphosis *vs*. paedomorphosis and developmental timing. Segregation patterns of *white*, *melanoid*, and metamorphic phenotypes within the AxTg family have been previously described (Voss, 1995). Likelihood ratio statistics (LRS) for association of phenotypic variation with genotypic inheritance were estimated using the interval mapping (*e.g.* Haley and Knott, 1992; Martinez and Curnow, 1992; Zeng, 1993; Zeng, 1994) function of MapManagerQTXb19 (Meer *et al*, 2004). Significance thresholds for interval mapping were obtained through 10,000 permutations of trait values among backcross progeny. Confidence intervals for QTL positions were estimated by the method of Darvasi and Soller (1997).

Results

Segregation of Molecular Markers

The segregation of alternate genotypes for a majority of the molecular markers was consistent with the Mendelian expectation of 1:1. Only 7% of protein-coding markers had genotypic ratios exceeding the 95% tail of the χ^2 distribution (Figure 3.1a). On average, ratios of anonymous markers were more deviant than protein-coding markers with 17% of markers exceeding the 95% tail of the χ^2 distribution (Figure 3.1b). Plotting χ^2 values for both marker classes against genomic position showed that the few markers with significantly deviant ratios were not strongly clustered at any position within the genome (Figure 3.1c).

Linkage Analysis

Linkage analysis revealed 14 large linkage groups that ranged in size from 125.5 cM to 836.7 cM and consisted of between 14 and 123 markers (Figure 3.2, Table 3.2). A total of 790 markers were assigned to these 14 linkage groups, with 486 protein-coding loci and 169 anonymous loci placed at precise positions. The total linkage distance spanned by these markers was 5251.3 cM and the average inter-marker distance was 7.96 cM. Thirty-six smaller linkage groups were identified that ranged in size from 0 to 79 cM and consisted of between 2 and 10 markers. Thirteen of these linkage groups contained at least one EST or protein coding marker and spanned a total of 479.7 cM, 23 additional

linkage groups that consisted of only AFLP markers spanned a total of 630.5 cM. One hundred and sixty-five markers (134 AFLP and 31 protein coding) did not show linkage to any other marker and were not assigned to any of the linkage groups.

QTL Analyses

***White*:** Interval analysis identified two likliehood ratio statistic (LRS) peaks for *white* on LG1 (*D1*, LRS 37.2, 95% CI = 17 cM; *D2*, LRS 12.4, 95% CI = 44 cM) (Figure 3.3). Although this suggests the possibility that *white* may depend upon more than a single locus, the lesser LRS peak is only suggestive of a QTL and an equally supportive LRS is expected to be observed once at random within the genome. In the mapping cross, color phenotypes segregated as F2 type markers because F1 hybrids and the recurrent *A. mexicanum* carried a single recessive allele (d) at *white* (D^{tig}/d^{mex} and D^{mex}/d^{mex} genotypes respectively). In contrast, molecular genotypes segregated as backcross type markers. As a result, only the *A. t. tigrinum* marker alleles were fully informative for association testing; that is, it was expected that at the position of *white,* all individuals inheriting *A. t. tigrinum* molecular marker alleles would exhibit wildtype coloration, but never *white* coloration. Segregation ratios for two protein-coding markers (*E16A12* and *E23C5*) near the maximum LRS peak on LG1 are consistent with this expectation (Table 3.3). In contrast, incomplete association was observed for makers from the region with the suggestive LRS. These results indicate that *d* is most likely located within the 14.2 cM region that includes markers *E16A12* and *E23C5.*

***Melanoid*:** Interval analysis revealed a single region of LG14 that was strongly associated with segregation of the *melanoid* mutant phenotype (*M1*, LRS 17.5, 95% CI = 32 cM) (Figure 3.3). As was the case with *white*, the founding and recurrent *A. mexicanum* parents that were used to generate the AxTg family were heterozygous for the *melanoid* mutant (M^{mex}/m^{mex}). The expressed sequence marker *E17A7* is closest to the maximum inflection point of the LRS profile for *melanoid*. Genotypic ratios for this locus are consistent with tight linkage to *melanoid*, given the same expectations described above for segregation of molecular makers and the *white* phenotype (Table 3.4).

Metamorphosis *vs.* Paedomorphosis: Interval analysis revealed a single region of LG3 that was strongly associated with segregation of metamorphosis *vs.*

paedomorphosis in the AxTg family (LRS 101.7, 95% CI = 9 cM) (Figure 3.4). The marker *ctg325* lies closest to the maximum inflection point of the LRS profile. This QTL is located near the middle of LG3 and corresponds to the previously described *met* QTL (Voss and Shaffer, 1997, Voss and Smith, 2005). A region of LG4 also showed a suggestive association with segregation of metamorphosis *vs.* paedomorphosis, although it is much weaker relative to *met*.

BLAST Alignments to the Human RefSeq Database

The BLASTx analysis of 570 EST/gene based markers yielded alignments with multiple human protein-coding genes. In total, 390 presumptive protein-coding markers had best BLASTx alignments with 350 non-redundant human RefSeq proteins. Best BLASTx matches and corresponding bitscore and *E*-vales are provided in Supplementary_Table_3_3. Analyses of conserved synteny and segmental homology between *Ambystoma* and other vertebrate species will be presented elsewhere.

Discussion

This chapter described the first comprehensive linkage map for *Ambystoma*, which was generated by mapping over 1000 anonymous markers and EST loci in an interspecific mapping panel. Fourteen large (>125 – 836 cM) and 35 small (0 – 75 cM) linkage groups were identified. Because the number of large linkage groups equaled the haploid chromosome number in *A. t. tigrinum* and *A. mexicanum* (Paramenter, 1919; Fankhauser and Humphrey, 1942), and because these linkage groups yielded a map size estimate that agrees with a theoretical and empirical (see below) estimate of total map size, this set is proposed as a framework for the 14 chromosomes in the tiger salamander genome. It is unlikely that the collection of smaller linkage groups represent any additional chromosomes within the *Ambystoma* genome because the haploid chromosome number is indisputably 14 and microchromosomes are not known in this group.

Map coverage

The combined map distance of the 14 largest linkage groups (5251 cM) is consistent with previous studies that indicated a large genetic map for *Ambystoma*. By comparison, the

combined map length of LG1 and LG2 is greater than the total map length of the mouse genome! Marker based estimates of genome size vary greatly for *Ambystoma* (*e.g.* 2600 – 6276 cM in Armstrong, 1984; 7291 cM in Voss *et al*, 2001). This variation may be attributed to the large genome size of *Ambystoma* and non-robust nature of estimators; genotyping errors and missing data cause upward bias in size estimates and nonrandom distribution of markers with respect to recombinational distances causes downward bias (Chakravarti *et al*, 1991). Genome size was estimated using linkage data from this study and the method of Hulbert *et al* (1988). Analyses were performed separately for protein coding and anonymous marker classes, and for the combined dataset using several linkage thresholds (10, 15, 20, 25, and 30 cM; data not shown). Estimates of map size varied greatly depending on dataset and linkage threshold: estimates based on protein-coding markers ranged from 3741 to 4320 cM (15 and 30 cM thresholds respectively), estimates from anonymous markers ranged from 6925 to 10782 cM (25 and 10 cM thresholds respectively), and estimates based on the entire dataset ranged from 7624 cM to 8824 cM, (25 and 10 cM thresholds respectively). Presumably, higher intrinsic error rates of AFLP markers (*e.g.* Isidore *et al*, 2003) (Figures 3.1a and 3.1b) yield inflated estimates of map size in the anonymous and combined datasets.

Fortunately, a more direct and independent estimate of recombination frequency is available for *A. mexicanum*. Callan's (1966) counts of chiasmata in *A. mexicanum* oocyte nuclei convert to an estimated map size of 5650 cM. This estimate is only slightly higher than the sum of the 14 largest linkage groups (5251 cM). Several factors likely contribute to the relatively small difference between these two estimates including: incomplete coverage of the linkage map (especially near the telomeres), the degree to which the Kosambi (1944) mapping function models recombination in *Ambystoma*, variation in average recombination rate between the male F1s used in this study and the female *A. mexicanum* used by Callan (1966), and possibly small-scale differences in gene order between *A. t. tigrinum* and *A. mexicanum* (Sybengia, 1992; True *et al*, 1996). Close agreement between the Callan (1966) derived estimate and recombination-based estimates of genome size that were generated in this study suggest that the 14 large linkage groups frame the majority of the *Ambystoma* genome.

Some of the smaller linkage groups may represent distal portions of chromosomes

that are represented by larger linkage groups. Recombination rates are known to increase toward the ends of chromosomal arms in a wide array of organisms including mammals (*e.g.* Jensen-Semen *et al*, 2004), chicken (ICGSC, 2005), and grasses (King *et al*, 2002; Anderson *et al*, 2004). In agreement with this pattern, Callan (1966) reported a higher than average recombination rate between the nucleolar organizing region (NOR) and adjacent telomere in *A. mexicanum*. Alternately, some of these linkage groups may consist completely or partly of markers with genotyping error rates that prevent assignment to larger linkage groups at the LOD 4.0 threshold. Presumably as more loci are mapped, linkage groups that represent distal regions will be integrated into the framework map and spurious linkage groups will be excluded from the map.

Size of the tiger salamander genetic map

The salamander genome is one of the largest genomes and the linkage map is the largest yet reported for any vertebrate. The tiger salamander genome is estimated to be 10 – 25 times larger than other sequenced vertebrate genomes, yet the genetic map is only 1.5 – 5 times larger (Table 3.1). Thus, in terms of genome size, the salamander linkage map appears at first glance to be approximately 5 times smaller than expected. However, because physical and linkage distances do not scale linearly among organisms, variables in addition to physical size must be considered. Chromosome number contributes directly to the size of the genetic map because at least one chiasma must be formed per bivalent in order to ensure normal segregation of chromatids during meiosis in most species (Egel, 1995; Roeder, 1997; Moore and Orr-Weaver, 1998), although achiasmate meiosis has evolved in several plant and invertebrate species (reviewed by John, 1990). Therefore, the minimum map size of any eukaryotic genome is 50 cM times the haploid number of chromosomes. Furthermore, there is a strong correlation between chromosome arm number and genetic map size among mammals, suggesting that formation of at least one chiasma per chromosomal arm may be necessary in mammals (Durtilleaux, 1986; Pardo-Manuel de Villena and Sapienza, 2001). Map sizes for human, mouse, rat, chicken, and zebrafish are much closer to their theoretical obligatory minima, based upon chromosome arm number, than is the *Ambystoma* map (Table 3.1). Analyses that include multiple vertebrate taxa, especially those with large and intermediate genetic map lengths, are

needed to elucidate fully the effect of chromosome arm number and other variables on genetic map size. However, it is apparent from the comparisons made here that linkage map size in *Ambystoma* is influenced by factors in addition to chromosome arm number, the most likely factor being the large physical genome size of *Ambystoma*.

Refining the positional estimate of *white* and tentative assignment of LG1 to the chromosome harboring the NOR

The *white* phenotype has been a subject of several linkage analyses. Early tetrad mapping studies located *white* to the distal region of an unknown chromosome (Lindsley *et al*, 1955; Armstrong, 1984) and several studies have observed an association between segregation of *white* and segregation of NOR variants (reviewed by Sinclair *et al*, 1978). The NOR is located subterminally on one of the four largest chromosomes (Paramenter, 1919; Dearing, 1934; Haschuka and Brunst, 1965; Callan, 1966; Cuny and Malacinski, 1985). In this study a single region of the genome was identified that was significantly associated with *white*. The maximum inflection point of the LRS profile for association with *white* is located 95 cM from the end of the largest linkage group (LG1). Callan's (1966) chiasmata counts place the NOR at ~44 cM from the end of a chromosomal arm. Localization of *white* to the distal portion of a large linkage group is consistent with previous studies that indicated linkage between *white* and the NOR. It is therefore concluded that LG1 likely represents the NOR containing chromosome, with marker *E20A12* located near the distal portion of the arm harboring the NOR.

Identification of the genomic interval containing *melanoid*

A single region has been identified within the smallest of the 14 linkage groups that shows strong association with segregation of the *melanoid* phenotype. Armstrong's (1983) tetrad analyses yielded an estimated distance of 59.1 cM between *melanoid* and its centromere. Assuming that LG14 corresponds to one of the smaller chromosomes, the maximum inflection point of the LRS profile for *melanoid* is consistent with the earlier gene-centromere mapping study. The size of LG14 relative to the rest of the genetic map appears to support the assignment of LG14 to the smallest chromosome. Precise chromosomal measurements have been made for *A. tigrinum* (Paramenter, 1919) and *A.*

*mexicanum (*Callan, 1966); the proportion of the genome occupied by the smallest chromosome was estimated to be 2.65% for *A. tigrinum* and 2.4% for *A. mexicanum*. Linkage group 14 represents 2.4% of the current genetic map (LG1-14). Although no linear relationship is expected between genetic and physical distance, these results are consistent with *melanoid* being located on one of the smaller tiger salamander chromosomes.

The genetic basis of metamorphic failure in lab *A. mexicanum*

It has been known for some time that expression of paedomorphosis in laboratory *A. mexicanum* is largely determined by a single genetic factor (Tompkins, 1978; Voss, 1995). A previous QTL analysis of the AxTg mapping panel using 234 AFLP markers identified a major effect QTL (*met*) that is associated with expression of paedomorphosis *vs.* metamorphosis (Voss and Shaffer, 1997); *met* was later shown to affect metamorphic timing in crosses using wild-caught *A. mexicanum* (Voss and Smith, 2005). A QTL analysis using all of the newly mapped markers did not identify statistically significant factors in addition to *met,* however a second region of the genome yielded a statistically suggestive result (Figure 3.4). The relatively small number of individuals in the AxTg panel does not provide sufficient power to identify small effect QTL. This second genomic region will be tested using the much larger WILD2 mapping panel (Voss and Smith, 2005) below.

EST markers and *Ambystoma* research

The collection of EST makers that comprise the linkage map are significant new resources for *Ambystoma* research. Because these markers correspond to polymorphic expressed sequences with known linkage relationships, they are especially informative as probes for developmental, populational, systematic, molecular, comparative, and functional studies (Putta *et al*, 2004). Using the PCR primer information provided in Supplementary_Table_3_1, it will now be possible to develop probes for many protein coding genes for *A. mexicanum* and *A. t. tigrinum,* as well as other ambystomatid species (Riley *et al*., 2003; Fitzpatrick and Shaffer, 2004; Putta *et al*., 2004). Primer sequences for protein-coding markers, their corresponding assembled and curated EST sequences,

and BLAST alignments can also be obtained at the Salamander Genome Project Website.

Table 3.1: Relationships between genome size and map size in vertebrates.

organism	genome size (Gbp)[a]	N[b]	chromosomal arms[b]	map_size (cM)[c]	obligatory_map (N * 50cM)	proportion obligatory[d]	cM/arm
Mouse	2.7	20	20	1361	1000	0.73 (0.73)	68.1
Rat	2.6	21	33	1749	1050	0.60 (0.94)	53
Human	3	23	46	3615	1150	0.32 (0.64)	78.6
Chicken	1.2	38	44	3800	1900	0.50 (0.58)	86.4
Zebrafish	1.8	25	50	3011	1250	0.42 (0.83)	60.2
Ambystoma(map)[e]	30	14	28	5251	700	0.14 (0.27)	184
Ambystoma(χ)[f]	30	14	28	5650	700	0.12 (0.25)	201.8

a – References: (mouse - MGSC, 2002; rat - RGSPC, 2004; human - IHGSC, 2000; chicken - ICGSC, 2005; zebrafish – Kelly *et al*, 2000; *Ambystoma* - Licht and Lowcock, 1991). b – N = The haploid number of chromosomes. Smaller chicken microchromosomes (10-38 – Masabana *et al*, 2004) are tabulated as one-armed chromosomes. References: (mouse - MGSC, 2002; rat - RGSPC, 2004; human - IHGSC, 2000; chicken - ICGSC, 2005; zebrafish – Gornung *et al*, 2000; *Ambystoma* – Callan, 1966). c – References: (mouse – Steen *et al*, 1999; rat – Brown *et al*, 1998; human – Kong *et al*, 2002; chicken – Groenen *et al*, 2000; zebrafish – Kelly *et al*, 2000). d – Numbers in parentheses represent the obligatory proportion of the map assuming one obligate chiasma per arm. e – Map size estimated as the sum of LG1 – 14. f – Map size estimated as 50 cM times the average number of chiasmata (Callan, 1966).

Table 3.2: Distribution of markers on *Ambystoma* LG1–14.

LG	EST Based[a]	Anonymous Markers[a]	All Markers	cM	% Map	Intermarker Distance[b]
1	73 (70)	50 (37)	123 (107)	836.7	15.9	7.8
2	56 (51)	47 (24)	130 (75)	767.6	14.6	10.2
3	80 (78)	34 (21)	114 (99)	671.3	12.8	6.8
4	55 (54)	21 (14)	76 (68)	538.9	10.3	7.9
5	32 (30)	18 (9)	50 (39)	323.1	6.2	8.3
6	34 (34)	15 (5)	49 (39)	308.6	5.9	7.9
7	27 (25)	24 (17)	51 (42)	294.9	5.6	7
8	34 (34)	17 (8)	51 (42)	290.1	5.5	6.9
9	32 (31)	9 (5)	41 (36)	244.4	4.7	6.8
10	21 (21)	13 (3)	34 (24)	241	4.6	10
11	39 (38)	10 (6)	49 (44)	234.4	4.5	5.3
12	9 (9)	12 (7)	21 (16)	211.4	4	13.2
13	7 (7)	12 (8)	19 (15)	163.3	3.1	10.9
14	9 (9)	5 (5)	14 (14)	125.5	2.4	9
Total	508 (491)	287 (169)	795 (660)	5251.3	100	8
% mapped	87.4	54.7	71.9			

a – The number of markers that were placed at precise positions within the map are provided in parentheses. b – numbers provided are the average intermarker distance.

Table 3.3: Segregation of markers associated with *white*.

		Phenotype	
QTL	Genotype	*White*	+
W1	$E16A12^{Am}$/ $E16A12^{Am}$	18	20
W1	$E16A12^{Am}$/ $E16A12^{Att}$	0	46
W1	$E23C5^{Am}$/ $E23C5^{Am}$	18	18
W1	$E23C5^{Am}$/ $E23C5^{Att}$	0	43
W2	$A32.20^{Am}$/ $A32.20^{Am}$	16	31
W2	$A32.20^{Am}$/ $A32.20^{Att}$	2	38

Table 3.4: Segregation of markers associated with *melanoid.*

		Phenotype	
QTL	Genotype	*Melanoid*	+
M1	$E24E9^{Am}$/ $E24E9^{Am}$	10	30
M1	$E24E9^{Am}$/ $E24E9^{Att}$	6	41
M1	$E17A7^{Am}$/ $E17A7^{Am}$	14	23
M1	$E17A7^{Am}$/ $E17A7^{Att}$	2	44

Figure 3.1: Statistical tests for Mendelian segregation. Panel A: Plot of the distribution of χ^2 statistic for 1:1 segregation of marker genotypes *vs.* standard χ^2 distributions for gene and EST based markers. Panel B: Plot of the distribution of χ^2 statistic for 1:1 segregation of marker genotypes *vs.* standard χ^2 distributions for anonymous markers (AFLPs and RAPDs). Panel C: Plot of χ^2 statistics for 1:1 segregation of marker genotypes *vs.* map position.

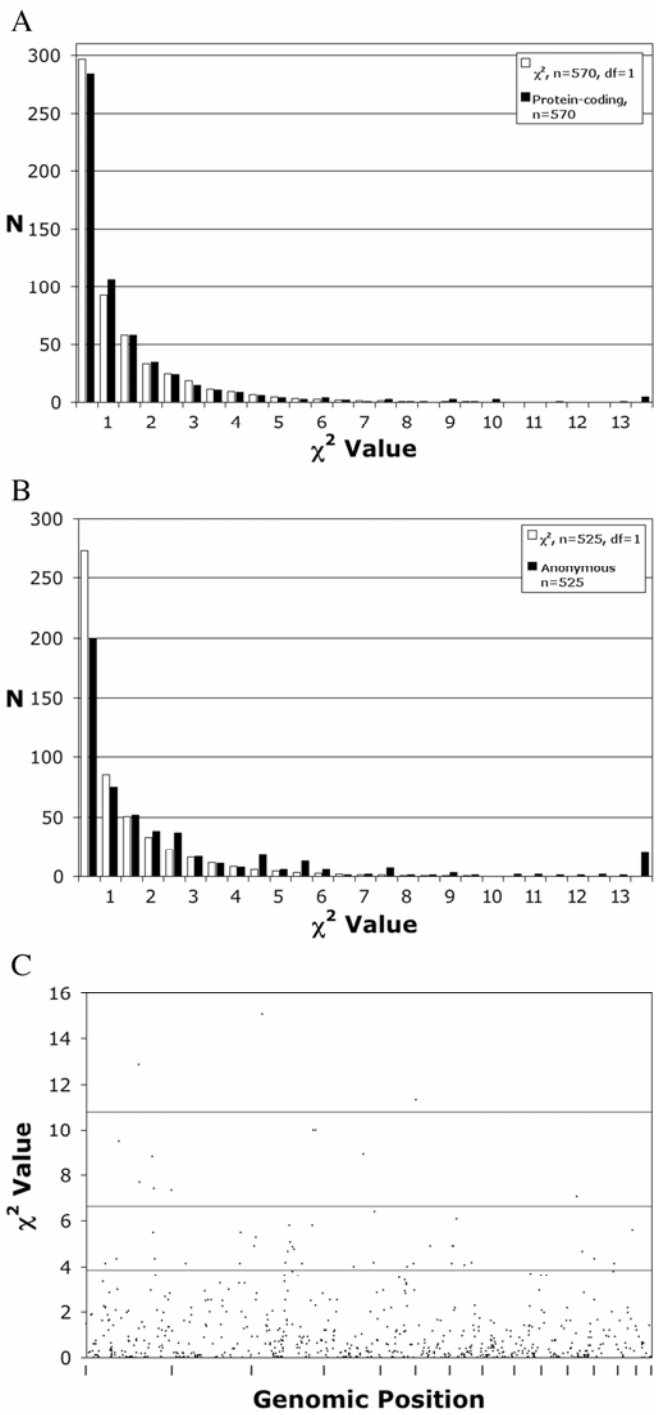


Figure 3.2: The tiger salamander linkage map (LG1-14).

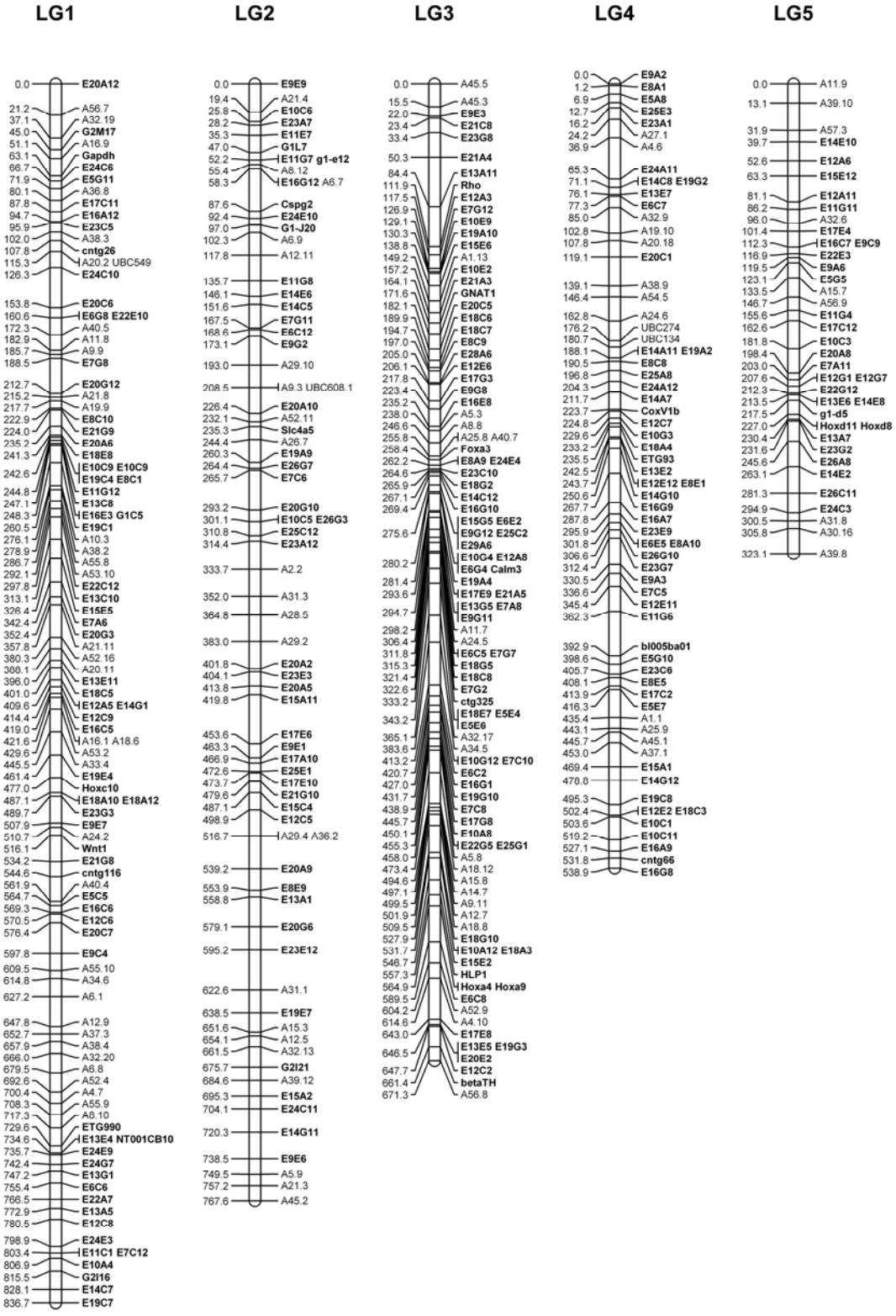


Figure 3.2 (continued)

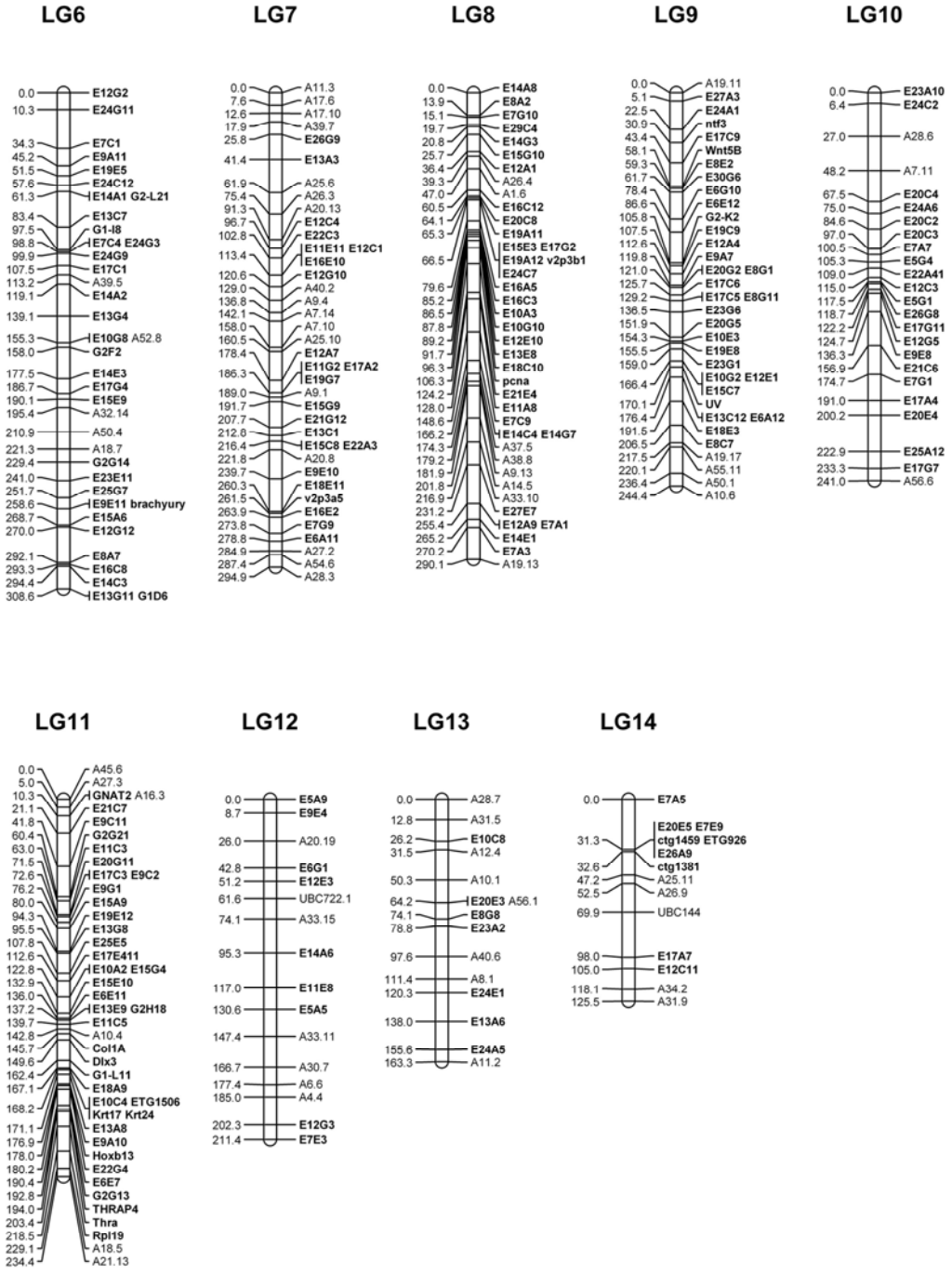


Figure 3.3: Likelihood ratio statistic (LRS) plot for association between segregation of pigment phenotypes and marker genotypes in LG1 and LG14. Horizontal lines represent LRS thresholds for suggestive (37th percentile), significant (95th percentile), and highly significant (99.9th percentile) associations (Lander and Krugiak, 1995) estimated using MapMaker QTXb19.

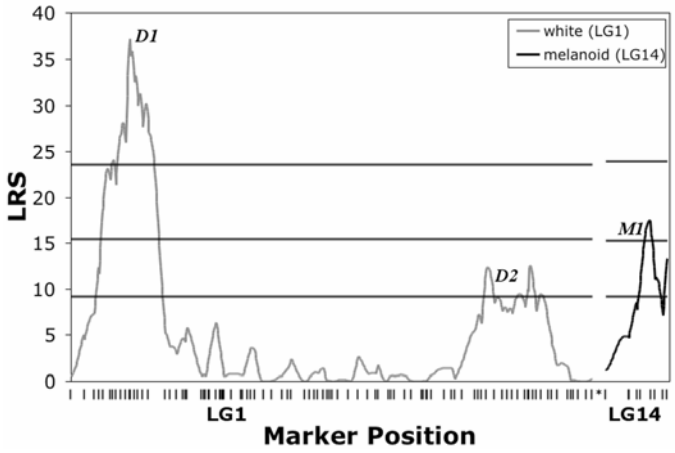
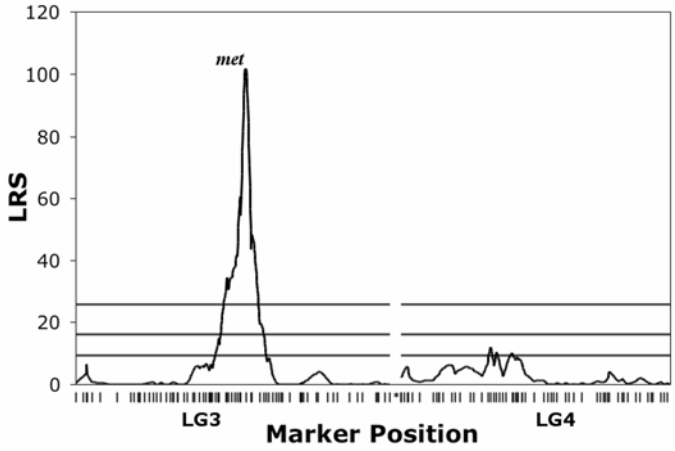


FIGURE 3.4: Likelihood ratio statistic (LRS) plot for association between segregation of metamorphosis *vs.* paedomorphosis and marker genotypes in LG3 and LG4. Horizontal lines represent LRS thresholds for suggestive (37th percentile), significant (95th percentile), and highly significant (99.9th percentile) associations (Lander and Krugiak, 1995) estimated using MapMaker QTXb19.



CHAPTER 4: Sal-Site: Integrating New and Existing Ambystomatid Salamander Research and Informational Resources

Introduction

Salamanders of the genus *Ambystoma* are important model organisms in biological research. Their seminal role in experimental embryology and broad utility in laboratory-based science is well known (Beetschen, 1996). Ambystomatid salamanders are currently used in multiple areas including olfaction, vision, cardiogenesis, embryogenesis, sensory system development, genomics, and post-embryonic development, including organ and tissue regeneration (Bachvarova *et al*, 2004; Chichilnisky and Rieke, 2005; Denz *et al*, 2004; Ericsson *et al*, 2004; Marchand *et al*, 2004; Park, *et al*, 2004; Schnapp and Tanaka, 2005; Thoreson *et al*, 2004; Zhang *et al*, 2004). Moreover, *Ambystoma* is very different from typical laboratory models because much is also known about their ecology, evolution, and natural history. The group is a model in studies of life history and natural phenotypic variation, infectious disease, evolutionary developmental biology, and conservation biology (Fitzpatrick and Shaffer, 2004; Hoffman and Phennig, 1999; Jancovich *et al*, 2005; Parichy, 1996; Riley *et al*, 2003; Rubbo and Kiesecker, 2005; Voss and Smith, 2005; Voss *et al*, 2003). In these respects, *Ambystoma* is a complete model organism system that offers integrative research opportunities spanning the continuum of biological organization.

Recent and on-going molecular resource development is providing new tools for research using ambystomatid salamanders. The Salamander Genome Project (SGP) has recently developed and annotated thousands of expressed sequence tags (ESTs) for *A. mexicanum* and *A. t. tigrinum* (Haberman *et al*, 2004; Putta *et al*, 2004), generated complete mtDNA sequence for 5 different ambystomatid species (Samuels *et al*, 2005), and completed the first comprehensive genetic linkage map (Smith *et al*, 2005). Markers that have been developed from these ESTs are providing new probes for molecular studies as well as markers for population and quantitative genetics, and phylogenetics (Fitzpatrick and Shaffer, 2004; Riley *et al*, 2003; Weisrock *et al*, 2006). This recent flurry of resource development stands to greatly increase the utility of ambystomatid salamanders, however there is a need to refine and integrate new resources with existing databases and information. To meet this need, Sal-Site (http://www.ambystoma.org) was

created. This web-portal functions to integrate new and existing ambystomatid resources. The collation of resources through Sal-Site will enhance communication across the *Ambystoma* community and provide a translational mechanism for researchers working in other model organism systems. Below describe six resources are described that are accessible through Sal-Site**.**

Salamander Genome Project (SGP)

The SGP (http://salamander.uky.edu), supported by the National Center for Research Resources at the National Institute of Health, continues to develop Expressed Sequence Tags (ESTs) and a genetic linkage map (see below). Expressed sequence tags are multifunctional resources because they can be developed for a number of uses, including population and quantitative genetics, comparative genomics, *in situ* hybridization, and functional genomics (Samuels *et al*, 2005; Smith *et al*, 2005; Weisrock *et al*, 2006). ESTs are especially useful in the ambystomatid system because sequence information from *A. mexicanum* and *A. t. tigrinum* can be easily extended to enable research in other species (Fitzpatrick and Shaffer, 2004; Riley *et al*, 2003; Samuels *et al*, 2005), as well as in distantly related vertebrates. Sequences deriving from assembled ESTs are also providing the majority of markers for the *Ambystoma* genetic linkage map (Smith *et al*, 2005). The SGP website was originally developed as a web-interface to allow registered members access to EST and gene mapping data as it was collected. These separate functions are now accomplished through separate but integrated databases that are described below, and there is no longer a requirement for users to register to access these databases. The SGP website now primarily functions to provide information about the project and update progress made in developing genome resources.

***Ambystoma* EST Database (AESTdb)**

Although all of the EST sequences developed under the SGP (Haberman *et al*, 2004; Putta *et al*, 2004) are available to the community through NCBI, these ESTs represent an immense collection of unedited data to sift through and sequencing errors are common. The AESTdb [http://salamander.uky.edu/ESTdb] was developed in order to organize *Ambystoma* ESTs into edited model RNA sequences and integrate these sequences with

related databases. To create the AESTdb, separate automated assemblies were first performed for all available ESTs that have been generated for the species *A. mexicanum* and *A. t. tigrinum* (see also Putta et al, 2004), including a subset of *A. mexicanum* ESTs that are available as unedited contigs from the Axolotl EST Database (Haberman *et al*, 2004; https://intradb.mpi-cbg.de/axolotl). Automated assembly methods may efficiently identify sequencing errors when large numbers of ESTs are analyzed because it is possible to efficiently identify sequencing and assembly errors against the backdrop of multiple overlapping sequences. However, the SGP has thus far generated an intermediate number of ESTs (~55,000) and many of the assembled contigs contained one or few EST members. As a result, automated methods for error detection were less efficient at detecting sequencing errors and many incorrect base calls were detected upon visual inspection of assembled trace data.

The prevalence of sequence errors among existing *Ambystoma* assemblies necessitated development of a quality-controlled manual editing methodology to minimize error rates within AESTdb contig sequences (Figure 4.1). All contigs were manually edited in two rounds by visually inspecting aligned trace files, removing low quality sequence ends, and correcting base miscalls. After manual editing, contigs were searched against the human RefSeq and *Xenopus* UniGene databases using the BLASTx and tBLASTx algorithms respectively. If a contig exhibited significant sequence similarity (E value > 1e-7) to a human or *Xenopus* protein-coding sequence, the presumptive ortholog, the reading frame, the translated sequence, and alignment statistics were recorded in individual files that also provide hyperlinks to associate contig sequences with gene and marker files in the *Ambystoma* Gene Collection (AGC) and *Ambystoma* Genetic Map and Marker Collection (AMAP) (Figure 4.2). For example, if the sequence of a contig was used to develop a PCR amplifiable molecular marker, a hyperlink was included to link the contig file to the molecular marker file in the AMAP database (see below). The AESTdb has a user-friendly web-interface that allows BLAST or textual searching of contigs and access to contig and associated raw EST sequence data.

Ambystoma Gene Collection (AGC)

Analyses of gene function are greatly facilitated by knowledge of full-length RNA and amino acid sequences. The AGC (http://salamander.uky.edu/AGC) acts as a repository for presumptive full-length sequences from multiple sources, including: AESTdb, sequences parsed from existing databases (*e.g.* NCBI), and sequences derived from the community at large. Several putative full-length sequences were identified for *A. mexicanum* (n = 940) and *A. t. tigrinum* (n = 717) contigs using the program MuSeqBox (Xing and Brendel, 2000). Comparisons against curated full-length protein-coding sequences from the human RefSeq database were performed using three length thresholds for variable amino-terminal and carboxy-terminal regions (20, 50, and 100 amino acids). The majority of putative full-length sequences that were identified showed greater than 50% amino acid sequence identity and greater than 80% sequence coverage when compared to their best human BLASTx hit (Tables 4.1 and 4.2). Individual AGC files contain source RNA sequences and predicted amino acid sequences, as well as supplementary fields that provide the best human RefSeq BLAST hit used for analysis of sequence coverage, percent sequence coverage of the best BLAST hit, and percent sequence identity between human and *Ambystoma* amino acid sequences. Files provided in the AGC are hyperlinked to corresponding ESTs and marker sequences in the AESTdb and AMAP as well as NCBI sequence records for corresponding human RefSeq proteins.

Ambystoma Genetic Map & Marker Collection (AMAP)

Although the genomes of ambystomatid salamanders are approximately 10 times larger than the human genome (Licht and Lowcock, 1991), the first complete genetic linkage map for any amphibian (including *Xenopus*) was recently assembled using an interspecific mapping cross between *A. mexicanum* and *A. t. tigrinum*. This resource is allowing the mapping of QTL and mutant phenotypes, and the identification of conserved vertebrate syntenies (Smith *et al*, 2005; Voss *et al*, 2001). The AMAP website (http://salamander.uky.edu/MAP) provides images showing the location of mapped markers on linkage groups that correspond to the 14 chromosome pairs in *Ambystoma*. Individual linkage groups can also be accessed as separate datasets in tabular format. These datasets provide precise map distances for all EST and gene-based markers, as well

as hyperlinks to separate marker files that provide additional marker specific information including assembly source sequence (hyperlinked to AESTdb records), primer, and polymorphism detection information.

System implementation

Sal-Site is implemented using a number of open-source software packages including Apache web server, Perl, CGI, BioPerl, PHP and MySQL. Sal-Site is hosted on a SMP (symmetric multi-processor) PC equipped with two processors, 4GB of RAM and running Linux 2.4.x. The program MySQL 4.0 was used as the backend relational database management system to store and manage all the information in a robust and efficient way.

Table 4.1: Full-length sequences identified among *A. mexicanum* assemblies.

	% Identity				
% Coverage	0-25	25-50	50-75	75-100	Total
90-100	1	44	228	241	514
80-90	3	36	101	80	220
70-80	0	35	78	93	206
Total	4	115	407	414	940

Distribution of percent amino acid sequence identity between *A. mexicanum* contigs and human RefSeq sequences (% Identity) and percent sequence coverage of corresponding human RefSeq sequences by aligning *A. mexicanum* sequence (% Coverage) for all putative full-length protein coding genes that were identified among *A. mexicanum* contigs.

Table 4.2: Full-length sequences identified among *A. t. tigrinum* assemblies.

	% Identity				
% Coverage	0-25	25-50	50-75	75-100	Total
90-100	0	37	122	180	339
80-90	0	35	72	64	171
70-80	2	55	77	73	207
Total	2	127	271	317	717

Distribution of percent amino acid sequence identity between *A. t. tigrinum* contigs and human RefSeq sequences (% Identity) and percent sequence coverage of corresponding human RefSeq sequences by aligning *A. t. tigrinum* sequence (% Coverage) for all putative full-length protein coding genes that were identified among *A. t. tigrinum* contigs.

Figure 4.1 – Schematic showing the curation methodology used to assemble and edit *Ambystoma* contigs. The numbers of ESTs and curated assemblies exclude mitochondrial transcripts.

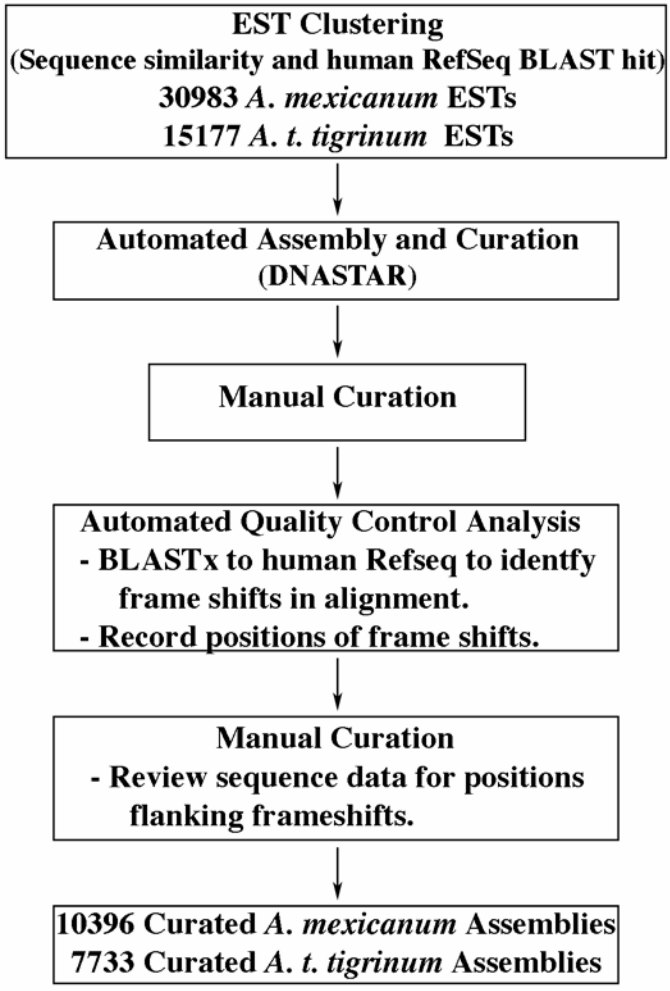
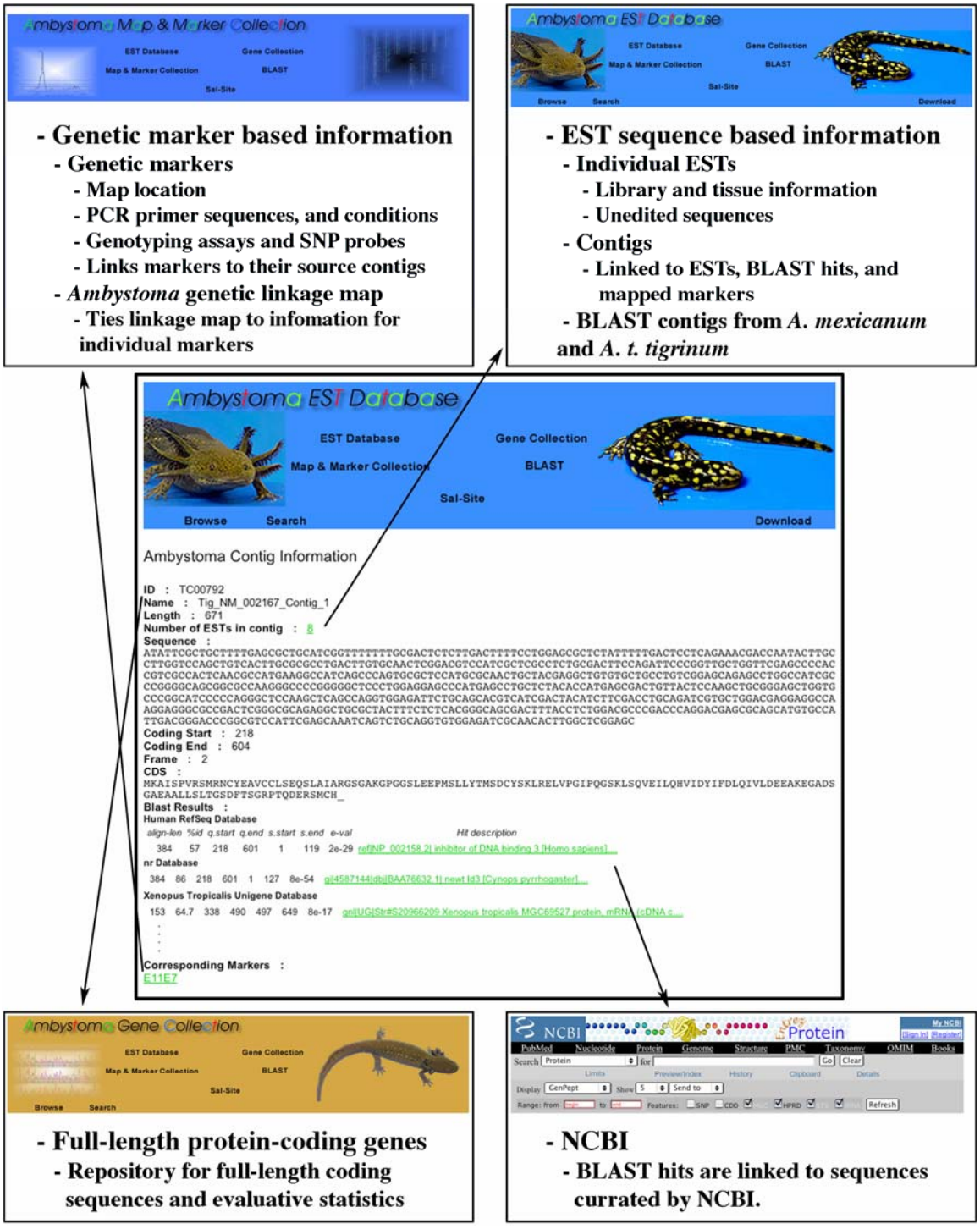


Figure 4.2: Overview of sequence-based information available through AESTdb and linked resources.



CHAPTER 5: Evolution of Salamander Life Cycles: A Major-Effect Quantitative Trait Locus Contributes to Discrete and Continuous Variation for Metamorphic Timing

Introduction

Alternate modes of development are often observed within and between closely related species. A number of classic examples have been studied, including flight and flightless forms of insects, feeding and nonfeeding echinoderm larvae, and metamorphic and nonmetamorphic salamanders (Roff, 1986; Matsuda, 1987; Raff, 1996; West-Eberhard, 2004). These examples indicate a taxonomically widespread potential for the evolution of discrete, morphological characteristics that allow for novel and alternative life histories. In each of these cases, the phenotypic transition is explainable using "heterochronic" terms that describe how development evolves through changes in the timing in which characters are expressed during ontogeny (Gould, 1977). For example, the evolution of salamander paedomorphosis presumably required a genetic change that blocked the initiation of metamorphosis, and this resulted in larval-form adults. Although such description is useful for explaining evolutionary shifts in development, it provides little insight about the structure of genetic architectures that underlie heterochronic, evolutionary transitions.

Like the majority of frogs and toads, many salamanders undergo an obligate metamorphosis that allows for the exploitation of both aquatic and terrestrial habitats during ontogeny. However, some salamander species express an alternate developmental mode in which they forego metamorphosis and remain in the aquatic habitat throughout their lifetimes (Figure 5.1). Nonmetamorphic forms are termed paedomorphic because they maintain juvenile features of the ancestral condition as they mature reproductively into large, larval forms (Gould, 1977). The exemplar of salamander paedomorphosis is the Mexican axolotl (*Ambystoma mexicanum*). *Ambystoma mexicanum* belongs to a group of several closely related species collectively known as the tiger salamander species complex (Shaffer and Mcknight, 1996). Salamanders of this complex occupy a variety of North American breeding habitats ranging from temporary vernal pools to large permanent lakes. Among these habitats, populations are highly variable for metamorphic timing and expression of paedomorphosis. Some populations express

metamorphosis (*e.g.*, *A. tigrinum tigrinum*) or paedomorphosis like *A. mexicanum*, while in other populations both phenotypes are observed at varying frequencies. Presumably, the expression of paedomorphosis is an opportunistic strategy that allows individuals to more successfully colonize relatively permanent aquatic niches (Wilbur and Collins, 1973; Sprules, 1974). Paedomorphic tiger salamanders are found in newly-created habitats like cattle watering troughs and wastewater treatment ponds (Rose and Armentrout, 1975; Collins, 1981), as well as in stable, large lake systems (Shaffer, 1984). This chapter further explores the genetic contribution of a major-effect QTL (*met*) that is strongly associated with the discrete expression of metamorphosis *vs.* paedomorphosis in interspecific crosses using *A. tigrinum tigrinum* and a laboratory strain of *A. mexicanum* (Voss and Shaffer, 1997). Previous studies have shown that a dominant allele from *A. tigrinum tigrinum* (met^{Att}) and a recessive allele from lab *A. mexicanum* (met^{lab}) results in metamorphosis in *A. tigrinum tigrinum* / *A. mexicanum* hybrids and that met^{Att}/met^{lab} and met^{lab}/met^{lab} backcross genotypes are strongly associated with metamorphosis and paedomorphosis, respectively (Voss, 1995; Voss and Shaffer, 1997, 2000) (Table 5.1). These studies use a newly identified and highly informative expressed sequence tag marker for *met* called *contig325*. This marker is informative in the majority of *Ambystoma* species (data not shown) and thus represents an important new candidate for studies of developmental timing variation in natural populations. This chapter also describes a large, newly created backcross population (called WILD2) that was generated using wild-caught *A. mexicanum* individuals. WILD2 and the smaller WILD1 backcrosses (Voss and Shaffer, 2000) may differ from lab *A. mexicanum* (LAB) backcrosses as a result of differences in the effects of *met* alleles and/or genetic background effects. To test this idea, *contig325* was examined within the context of all available backcross populations (LAB, WILD1, and WILD2) to infer genetic changes that have modified the paedomorphic response of the natural *A. mexicanum* population during domestication of the laboratory strain at Indiana University (The Axolotl Colony, now the *Ambystoma* Genetic Stock Center at the University of Kentucky). Because WILD2 is the largest *A. t. tigrinum* / *A. mexicanum* backcross developed thus far (N = 457), it was also possible to assess, in an accurate and reliable manner, the genetic contribution of *met* to a second form of phenotypic variation: continuous variation in

metamorphic timing. These results show that *met* contributes genetically to both discrete and continuous forms of metamorphic timing variation and suggest a linkage between the evolutionary maintenance of biphasic life cycles and the evolution of alternate developmental modes.

Materials and Methods

Genetic crosses

Phenotypic and genotypic segregation ratios were compared among three different backcross resources. Three strains were used to make the backcrosses: wild-caught *A. t. tigrinum*, wild-caught *A. mexicanum*, and laboratory *A. mexicanum* (Figure 5.1; Table 5.1). The LAB and WILD1 backcrosses have been described previously (Voss, 1995; Voss and Shaffer, 2000). For LAB, *A. mexicanum* were obtained from the Indiana University Axolotl Colony strain; for WILD1, *A. mexicanum* were collected from their natural habitat at Lake Xochimilco, Mexico D.F., Mexico. In these, and the WILD2 crosses described below, *A. t. tigrinum* were obtained from the same source population in Tennessee (Charles Sullivan).

The WILD2 backcrosses were created to obtain the largest-ever segregating population for genetic analysis of *Ambystoma* (Smith, 2002). WILD2 was created using *A. mexicanum* individuals collected from Lake Xochimilco to make F1 hybrids and first generation descendants of wild-caught *A. mexicanum* to make backcrosses. The F1 hybrids were generated from a single cross and backcross offspring were generated using three male *A. mexicanum* and four female *A. t. tigrinum* / *A. mexicanum* hybrids. A total of nine backcross families compose WILD2 (Table 5.2). Artificial insemination was used in all crosses (Armstrong and Duhon 1989; Voss 1995).

Rearing conditions

The rearing methods for offspring from LAB and WILD1 were described previously (Voss, 1995; Voss and Shaffer, 2000). Here, the rearing conditions for WILD2 offspring are described. At 21 days postfertilization, larvae were released from their eggs and placed individually in 5-oz paper cups of artificial pond water. Throughout the course of this experiment all individuals were maintained in a single room within which the

temperature fluctuated from 19°–22°. Individuals were reared in separate containers and rotated within the room after water changes to reduce effects of spatial temperature variation. Larvae were fed freshly-hatched *Artemia* twice daily for their first 30 days posthatching. After day 20 posthatching, diets were supplemented with small (<1 cm) California black worms (*Lumbriculus*). During this time, individuals were provided with fresh water and cups after every third feeding. On day 30, larvae were transferred to 16-oz plastic bowls, after which they were fed California black worms exclusively and water was changed every third day. Finally, at 80 days posthatching all individuals were transferred to 4-liter plastic containers and were otherwise maintained under the same regimen as the previous 50 days, until completion of metamorphosis or the end of the experiment (day 350). The majority of backcross offspring were euthanized, as described above, upon completion of metamorphosis or at day 350. At this time, individuals were dissected and tissue samples (liver and/or blood) were collected for DNA isolation. A few individuals were not euthanized and are currently being maintained for use in future studies. For these individuals, tissue samples were collected as tail clips.

Phenotypic scores

Individuals were scored as metamorphs upon complete resorption of all external gills (gills <1.0 mm in length). Age at metamorphosis was recorded as the number of days from fertilization to completion of metamorphosis. For WILD2, the experiment was terminated on day 350, at which point no individuals had completed metamorphosis within the previous 3 weeks. All remaining individuals showed no sign of having initiated metamorphosis (no apparent regression of the tail fin or external gills) and were scored as paedomorphs.

Genotyping

A total of 98, 112, and 457 individuals from LAB, WILD1, and WILD2, respectively, were genotyped for *contig325*, a molecular marker that was isolated as a result of ongoing EST and genetic linkage mapping projects that generate genome resources for *Ambystoma* research (http://salamander.uky.edu). This marker was isolated from an *A. mexicanum* tail regeneration blastema cDNA library (Putta *et al*, 2004). Additional

coding sequence for this EST was obtained by 5'-RACE and assembled with existing EST sequences. The resulting 985-bp DNA sequence shows strong similarity to a human nerve growth factor receptor precursor (sequence data not shown; NP_002498, bit score = 164; BLASTX).

A 221-bp DNA fragment corresponding to *contig325* was amplified from all individuals under standard PCR conditions (150 ng DNA, 50 ng each primer, 1.2 mM $MgCl_2$, 0.3 units Taq polymerase, 1x PCR buffer, 200 mM each of dATP, dCTP, dGTP, and dTTP; thermal cycling at 94° for 4 min; 33 cycles of 94° for 45 sec, 60° for 45 sec, 72° for 30 sec; and 72° for 7 min). Genomic DNA was isolated from all individuals using a previously described phenol extraction method (Voss, 1993). Primer sequences for amplifying *contig325* were forward, 5'-GTGAAGTCAGTGATGAAAGTCCATGT-3', and reverse, 5'-CTAGGATACCAGTGGGAGAGTGTAAT-3'. Genotypes were assayed by restriction digestion of PCR products with a diagnostic *Alu*I restriction enzyme (New England Biolabs, Beverly, MA) followed by agarose gel electrophoresis.

Linkage analysis

Linkage and QTL mapping studies were performed using the software package MapMakerQTXb19 (http://www.mapmanager.org/mmQTX.html; Meer *et al*, 2004). Linkage distance and arrangement among *contig325* and previously described amplified fragment length polymorphisms (AFLP) (Voss and Shaffer, 1997) was estimated using the Kosambi mapping function at a linkage threshold of P = 0.001. The maximum-likelihood position of the *met* QTL was estimated using the interval mapping function. Significance thresholds for interval mapping were obtained through 10,000 permutations of trait values among backcross progeny. Associations between *contig325* genotypes and phenotypic variation were measured using the marker regression function.

Results

Identification of a highly informative EST marker for *met*

The *met* QTL was originally identified in LAB using AFLPs and interval mapping (Voss and Shaffer, 1997). In LAB, met^{Att}/met^{lab} and met^{lab}/met^{lab} segregate as highly penetrant genotypes for metamorphosis and paedomorphosis, respectively; in fact, cosegregation of

associated AFLPs and morph phenotypes was statistically consistent with simple Mendelian inheritance. However, given the anonymous nature of AFLPs and the nonspecific way in which these markers are generated, a more informative and user-friendly marker was developed for *met*: an expressed sequence tag that we refer to as *contig325*. To show linkage between *contig325* and *met*, individuals from LAB were genotyped. A stronger association between *contig325* and *met* was observed than with the most closely linked AFLP marker (estimated proportion recombinants: AFLP32.17 = 0.15, N = 70; *contig325* = 0.07, N = 91). Genotypes at *contig325* explain 71% of discrete variation for segregation of metamorphosis *vs.* paedomorphosis in LAB. Interval mapping shows that *contig325* is located near the maximum inflexion point of the previously determined AFLP LOD profile for *met* (Figure 5.2).

Survival of offspring in the newly created WILD2 mapping panel

Overall, a high proportion (91%) of WILD2 backcross offspring survived from hatching to completion of metamorphosis or failed to metamorphose by day 350. There were no significant differences in survival probability among crosses. In total, the nine crosses generated 497 backcross offspring that survived through completion of metamorphosis or to the end of the experiment as paedomorphs.

Segregation of discrete developmental modes and *contig325* in WILD2 and WILD1

Segregation of metamorphs and paedomorphs was observed in all nine WILD2 crosses (Table 5.2). Segregation ratios were not significantly heterogeneous (G = 13.32, d.f. = 8, P = 0.10) among crosses; therefore segregation ratios were pooled for hypothesis testing. The majority of offspring generated (453 of 497) metamorphosed before day 350. In total, only 44 (9%) of the offspring exhibited paedomorphosis and ratios were significantly different from the simple Mendelian expectation of 1:1 (G = 392, d.f. = 1, P = 4×10^{-87}, N = 497). Significantly lower-than-expected numbers of paedomorphs (19%) were also observed in WILD1 (Voss and Shaffer, 2000). Thus, results from WILD1 and WILD2 indicate that the proportion of paedomorphs is significantly lower in backcrosses using wild-caught *A. mexicanum*, relative to laboratory *A. mexicanum*.

To determine if *met* contributed to the segregation of discrete developmental

modes in WILD2, all individuals were genotyped for *contig325* (*325*) (Table 5.3). Inheritance of $325^{Att}/325^{Am}$, and thus presumably of met^{Att}/met^{wild2} (Table 5.1), yielded the expected metamorphic phenotype in >99% of the cases. The $325^{Att}/325^{Am}$ genotype (presumably marking met^{wild2}/met^{wild2}) was not as penetrant for the paedomorphic phenotype as only 17% of individuals in this genotypic class were paedomorphs. However, inheritance of met^{wild2}/met^{wild2} is apparently necessary for expression of paedomorphosis as only one paedomorph inherited a met^{Att}/met^{wild2} genotype. To investigate linkage results between WILD2 and WILD1, which had previously been examined using only the informative AFLP makers (Voss and Shaffer 2000), individuals from WILD1 were genotyped for *contig325*. The pattern of segregation for *contig325* in WILD1 was the same as that observed for WILD2. All but one individual that inherited $325^{Att}/325^{Am}$ was metamorphic and $325^{Am}/325^{Am}$ yielded incomplete penetrance for paedomorphosis (only 16 of 51 $325^{Am}/325^{Am}$ genotypes were paedomorphic). Observation of the same pattern of segregation between WILD1 and WILD2 suggests no sex linkage or maternal effect on the segregation of genotypes and phenotypes because the crossing designs were reversed to create WILD1 and WILD2 backcrosses (*i.e.*, F1 hybrids were male in creating WILD1 but female in creating WILD2). Overall, these results show that $325^{Att}/325^{Am}$ is strongly associated with the metamorphic phenotype; this association did not vary across LAB, WILD1, or WILD2. However, the proportion of $325^{Am}/325^{Am}$ genotypes that were associated with paedomorphosis was 4.5 times higher in LAB than in WILD1 and WILD2. This indicates a genetic difference in the basis of paedomorphosis between the natural and domestic strains of *A. mexicanum*.

Continuous variation for age at metamorphosis and association with *met*

The large number of metamorphic individuals in both genotypic classes from the WILD2 panel provided the opportunity to test for association between *contig325* and a second form of metamorphic timing variation: age at metamorphosis. Age at metamorphosis was scored only for those individuals that did undergo a metamorphosis. Age at metamorphosis varied continuously from 115 to 300 days in WILD2. Plotting metamorphic ages separately for $325^{Att}/325^{Am}$ and $325^{Am}/325^{Am}$ revealed two overlapping yet distinct distributions (Figure 5.3; Table 5.2). The means of these two distributions,

171 days and 207 days, differ significantly (t = 14.48, d.f. = 413, P < 0.0001), with $325^{Am}/325^{Am}$ individuals metamorphosing on average 36 days later than $325^{Att}/325^{Am}$ individuals. A similar difference in age at metamorphosis between genotypic classes was also observed within the WILD1 panel (t = 6.99, d.f. = 93, P < 0.0001) with $325^{Am}/325^{Am}$ individuals metamorphosing on average 25 days later than $325^{Att}/325^{Am}$ individuals (Figure 5.4). This replicated result indicates that *met*, which is strongly associated with discrete variation for metamorphosis/paedomorphosis, is also strongly associated with continuous variation for metamorphic age.

Discussion

Novel developmental modes may evolve as a result of genetic changes in developmental timing or heterochrony (Gould, 1977; Roff, 1986; Matsuda, 1987; Ambros, 1988; Raff, 1996; West-Eberhard, 2004). The paedomorphic developmental mode of the *A. mexicanum* is a classic example of heterochrony. Paedomorphosis in *A. mexicanum* presumably evolved as a result of a genetic change that blocked the initiation of metamorphosis in a biphasic ancestor, and this resulted in larval-form adults. In support of this idea, it was observed that within interspecific crosses using *A. mexicanum* and metamorphic *A. t. tigrinum*, the segregation of genotypes at a major-effect QTL (*met*) was associated with the expression of metamorphosis *vs.* paedomorphosis. This result supports the long-held idea that paedomorphosis in *A. mexicanum* evolved via saltation (Goldschmidt, 1940; Gould, 1977, 1981; Tompkins, 1978; Ambros, 1988; McKinney and McNamara, 1991; Voss, 1995; Voss and Shaffer, 1997; Futuyma, 1998). However, differences in gene effect were also observed that have evolved rapidly between the laboratory and wild strains of *A. mexicanum* (Voss and Shaffer, 2000), and it was found that *met* contributed to a second form of phenotypic variation: continuous variation in age at metamorphosis. This later result indicates that expression of paedomorphosis is associated with genetic changes that alter developmental timing (*contra* Raff and Wray, 1989; Raff, 1996). The following paragraphs review the primary results of this study and explain how a genetic architecture that contributes to both continuous and discrete phenotypic variation supports a more gradual selection model for the evolution of paedomorphosis.

Genetic basis of discrete variation: expression of metamorphosis *vs.* paedomorphosis

A conceptual framework for understanding how polygenes give rise to discrete phenotypic variation is the threshold model (Falconer, 1989). Under this model, the expression of alternate phenotypes depends upon an individual's liability value relative to a threshold value, with liability values above and below the threshold yielding alternate phenotypes. Under the threshold model, *met* makes a major contribution to the liability or threshold underlying the expression of metamorphosis *vs.* paedomorphosis. Within LAB, both $325^{Att}/325^{Am}$ and $325^{Am}/325^{Am}$ were highly predictive of their expected phenotypes, indicating highly significant linkage to a single locus (*met*) ($\chi^2 = 84.97$, d.f. = 1, N = 98, $P < 0.001$; Table 5.3). Thus, in the LAB genetic background, the threshold for expressing metamorphosis *vs.* paedomorphosis is traversed by the segregation of alternate met genotypes at a single locus. Apparently, $325^{Att}/325^{Am}$ is not sensitive to genetic background because this genotype was also highly predictive of metamorphosis in WILD1 and WILD2. Thus, in both LAB and WILD genetic backgrounds, substitution of a single *A. mexicanum met* allele with a dominant *A. t. tigrinum* m*et* allele rescued the metamorphic phenotype in essentially all cases.

In contrast to $325^{Att}/325^{Am}$, the penetrance of $325^{Am}/325^{Am}$ for paedomorphosis varied between LAB and the WILD backcrosses. This suggests that met^{lab} and $met^{wild1,2}$ contribute differently to the underlying genetic architecture or that LAB and WILD genetic backgrounds influence the probability of paedomorphosis differently. Although it is currently not possible differentiate between these two possibilities, the genetic basis of paedomorphosis clearly differs between the natural population and a recently derived laboratory strain of *A. mexicanum*, thus indicating the potential for rapid evolution of genetic architecture. This supports the idea that the simple Mendelian basis of paedomorphosis in LAB evolved recently during the domestication of *A. mexicanum* (Voss and Shaffer, 2000; see also Malacinski, 1978). Although paedomorphosis is expressed by both the wild strain and the laboratory strain, these results indicate that selection has canalized expression of paedomorphosis to a greater degree in the laboratory strain, as assayed using an interspecific crossing design. Thus, although paedomorphosis has been cited as a classic example of heterochrony by a major gene

effect, this study shows that factors beyond a single major gene are important in discrete trait expression in *A. mexicanum*.

Genetic basis of continuous variation: variation in metamorphic age

Because the WILD backcrosses yielded a large number of metamorphosing offspring reared under identical conditions, it was possible to estimate the contribution of *met* to variation in metamorphic age. It wass observed that metamorphic age varied significantly between $325^{Am}/325^{Am}$ and $325^{Att}/325^{Am}$ genotypic classes. This indicates that met^{wild}/met^{wild} delays timing of metamorphosis relative to met^{Att}/met^{wild}. Because met^{wild}/met^{wild} was associated with paedomorphosis in WILD1,2 (all but two paedomorphs were met^{wild}/met^{wild}), these results show that both delayed metamorphosis and expression of paedomorphosis are associated with this genotype; it is important to note that these associations were observed in the same genetic background. Conversely, an earlier metamorphosis was associated with the alternate met^{Att}/met^{wild} genotype, again within the same WILD genetic backgrounds. This indicates that *met* alleles deriving from paedomorphic *A. mexicanum* delay metamorphosis while *met* alleles from the metamorphic *A. t. tigrinum* decrease the time to metamorphosis. These results imply that metamorphic age is a continuous variable that is closely associated with the underlying liability or threshold that determines the expression of alternate developmental modes. It is possible that *met* influences metamorphic timing via changes in the timing of the sensitive period for hormonal initiation of metamorphosis, as has been suggested for dung beetles (*Onthophagus taurus*) that express alternate male morphs (Moczek and Nijhout, 2002). A comparative mapping project is underway to identify likely candidate genes in the vicinity of *contig325* (http://salamander.uky.edu).

Evolutionary maintenance of the biphasic life cycle and evolution of paedomorphosis

These results suggest that two distinct evolutionary processes—(1) adaptation of biphasic life cycles through selection of metamorphic timing (Voss *et al*, 2003) and (2) evolution of novel paedomorphic developmental modes that isolate lineages and promote speciation (Shaffer, 1984)—are apparently linked by a common genetic architecture. Selection for

met alleles that increase or decrease age at metamorphosis is expected to allow the evolution of a continuum of metamorphic timing phenotypes. Because *met* did not account for all of the variation in metamorphic timing in WILD2, it is likely that other loci make a contribution to continuous variation (Voss *et al*, 2003). The average difference in metamorphic age that was observed between *met* genotypic classes was 36 days. This amount of variation may significantly affect larval survivorship in natural populations that use unpredictable, ephemeral ponds (Wilbur and Collins, 1973). In more predictable ephemeral ponds, selection is expected to favor alleles that delay metamorphic timing because larvae that attain larger body sizes have increased survival probabilities after metamorphosis (Semlitsch *et al*, 1988). In this study, inheritance of the same *met* genotype was associated with delayed metamorphosis and expression of paedomorphosis. Because both of these life history strategies would be favored in a stable aquatic habitat, it seems likely that the evolution of paedomorphosis in *A. mexicanum* occurred gradually via selection for delayed metamorphic timing. Overall, these results provide a framework for understanding how metamorphic timing and paedomorphic phenotypes can evolve to be fixed or variable within and between species and, thus, how microevolutionary processes lead to macroevolutionary patterns.

Table 5.1: Nomenclature for backcrosses, expected *met* genotypes, expected *contig325* genotypes, and expected morph phenotypes.

Cross ID	Backcross hybrid X Am	*met* Genotype	*contig325* Genotype	Morph type
LAB	(Att/Am-lab) X Am-lab	met^{Att}/met^{lab}	$325^{Att}/325^{Am}$	Met
		met^{lab}/met^{lab}	$325^{Am}/325^{Am}$	Paed
WILD1	(Att/Am-wild1) X Am-wild1	met^{Att}/met^{wild1}	$325^{Att}/325^{Am}$	Met
		met^{wild1}/met^{wild1}	$325^{Am}/325^{Am}$	Paed
WILD2	(Att/Am-wild2) X Am-wild2	met^{Att}/met^{wild2}	$325^{Att}/325^{Am}$	Met
		met^{wild2}/met^{wild2}	$325^{Am}/325^{Am}$	Paed

Dominant met^{Att} alleles derive from the same Att strain. Recessive met^{lab} and met^{wild} alleles derive from different Am strains. *Contig325* is a species-specific marker locus linked to *met*. Morph types are based upon a single gene model. Att, *A. tigrinum tigrinum*; Am, *A. mexicanum*. Met, metamorph; Paed, paedomorph.

Table 5.2: WILD2 backcrosses showing parentage, morph segregation, and mean age at metamorphosis for *contig325* genotypes.

	Parents		Offspring			Mean age at metamorphosis	
Cross ID	F1	P2	Met	Paed	% Paed	$325^{Att}/325^{Am}$	$325^{Am}/325^{Am}$
1	F1.1	P2.1	69	5	7	175.4	217.4
2	F1.2	P2.1	50	7	12	171.3	199.1
3	F1.3	P2.1	38	3	7	183.2	206.7
4	F1.4	P2.1	65	11	15	177.1	204.8
5	F1.1	P2.2	56	6	10	173.2	210.1
6	F1.2	P2.2	80	3	4	164.5	203.8
7	F1.3	P2.2	44	1	2	165.1	214.6
8	F1.4	P2.2	35	4	10	159.9	200.2
9	F1.2	P2.3	16	4	2	156.7	209.2
Total			453	44	9	170.9	207.1

Figure 5.1: Larval and adult phases of *Ambystoma*. (A) Larval *A. mexicanum*. (B) Adult *A. t. tigrinum* (metamorphic). (C) Adult *A. mexicanum* (paedomorphic).

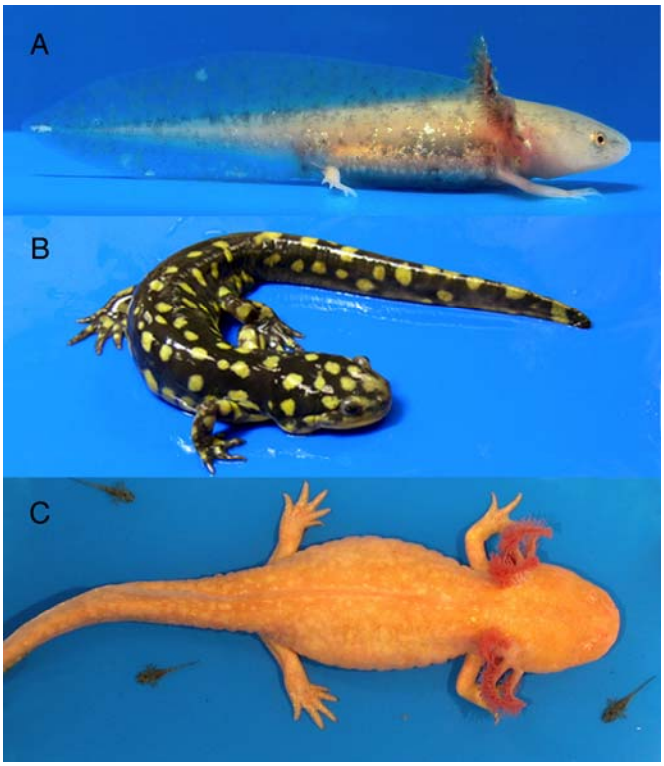


Figure 5.2: Likelihood-ratio statistic (LRS) plot for association of paedomorphosis with genetic factors in the *met* QTL region. The LRS for the contig325 marker is shown. Horizontal shaded lines represent LRS thresholds for suggestive (1.3), significant (6.6), and highly significant associations (15.6) as estimated using Map Manager QTXb19.

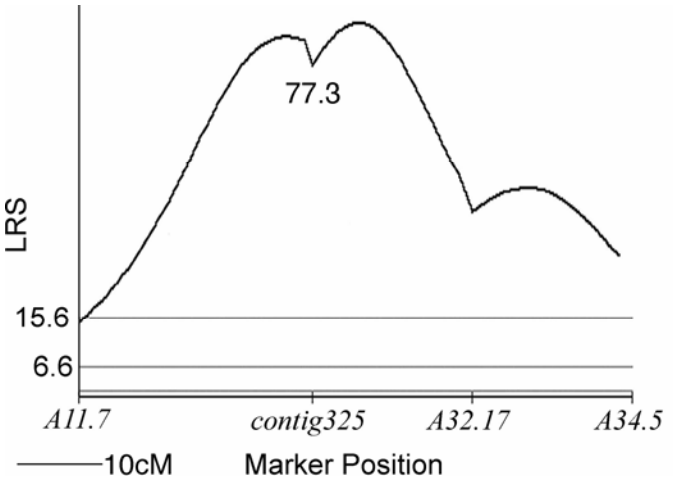


Figure 5.3: Distribution of ages at metamorphosis for WILD2 plotted separately for *contig325* genotypes.

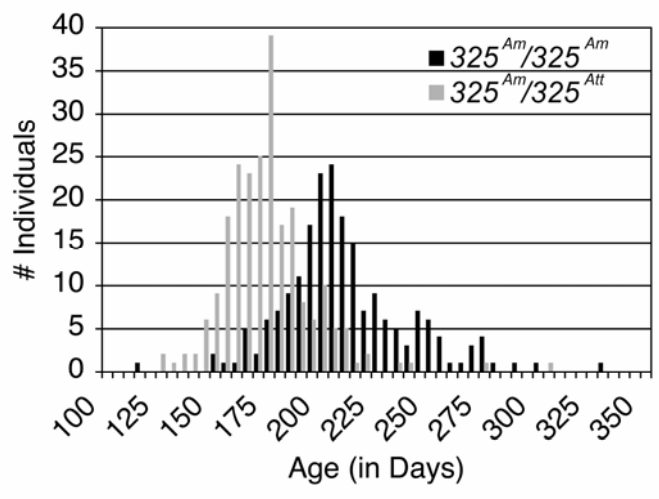
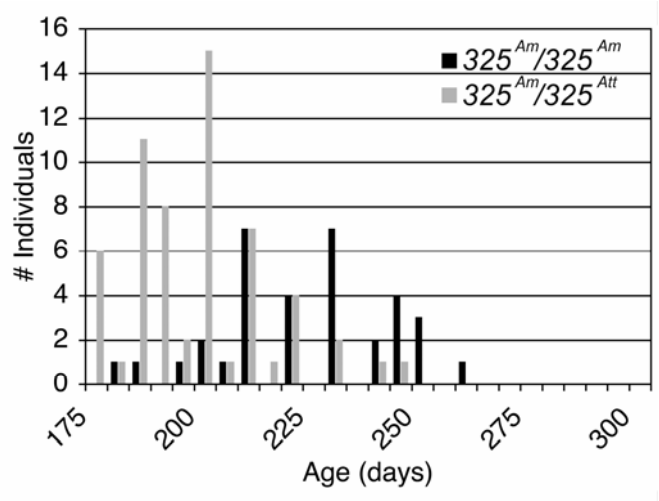


Figure 5.4: Distribution of ages at metamorphosis forWILD1 plotted separately for *contig325* genotypes.



CHAPTER 6: Gene Order Data from a Model Amphibian (*Ambystoma*): New Perspectives on Vertebrate Genome Structure and Evolution.

Introduction

Amphibians (salamanders, frogs, and cecilians) arise from a branch of the vertebrate evolutionary tree that is juxtaposed between aquatic fishes and more terrestrial amniotes (Figure 6.1). This phylogenetic location therefore positions amphibians to provide important comparative perspective for reconstructing character changes that have occurred during vertebrate evolution. For example, the amphibian perspective is essential for understanding molecular, developmental, and morphological changes of appendages that are associated with the transformation of obligatorily aquatic fish to terrestrial tetrapods (Gardiner *et al*, 1998; Simon, 1999; Hinchliffe, 2002; Shubin, 2002). In addition, because amphibians are the most basal tetrapod lineage, the amphibian perspective is essential for understanding the evolution of amniote characteristics among the "higher vertebrate" groups. Although many studies have exploited the phylogenetic position of amphibians for comparative perspective, very few comparisons have been made at the genome level. This chapter presents results from the first broad-scale comparison of genome structure between an amphibian and other representative vertebrate taxa.

One of the most fundamental structural characteristics of genomes is the order in which protein-coding genes are arranged on chromosomes. Gene order is determined using one of several approaches, including physical mapping, linkage mapping, and whole genome sequencing. The most powerful approach is whole genome sequencing (IHGSC, 2001; MGSC, 2002; Jaillon *et al*, 2004; RGSPC, 2004; ICGSC, 2005), but only if the final product is a complete (or nearly complete) genome assembly. Physical mapping refers to the direct localization of a gene to a whole or partial chromosome, for example by the method of somatic cell hybridization (Goss and Harris, 1975; Cox *et al*, 1990; Kwok *et al*, 1998; Wardrop *et al*, 2002) or chromosome in-situ hybridization (Gall and Pardue, 1969; John *et al*, 1969; Trask *et al*, 1989; Heng *et al*, 1992). In comparison to these physical genome approaches, genetic linkage mapping refers to the approach of estimating recombination frequencies among loci (genes) in a segregating cross for the purpose of ordering genes into linkage groups (*e.g.* Ott, 1991). Ultimately, the genomic

approach taken to order genes in a particular species is determined by genome characteristics and the availability of resources. For example, the extremely large genome size of some amphibians makes it difficult to justify a whole genome sequencing effort at this time (Morescalchi, 1975; Licht and Lowcock, 1991). However, genetic linkage mapping is an efficient strategy for amphibians because large numbers of offspring can be obtained from segregating crosses, thus allowing accurate estimates of map position (Smith *et al*, 2005).

Until recently, there were few amphibian gene order data available for comparative analyses of vertebrate genome structure (Voss *et al*, 2001; Ohta *et al*, 2006). Much physical genome sequence has been collected recently for an anuran amphibian (*Xenopus tropicalis*), but this sequence has not yielded a complete genome assembly and there are no large-scale genetic maps for *Xenopus* that can be used in comparative studies (http://tropmap.biology.uh.edu). The recently developed genetic linkage map for the salamander genus *Ambystoma,* however, now provides an amphibian resource that can provide structural and evolutionary perspective at the genomic level (Smith *et al*, 2005). This chapter presents the largest gene order dataset ever obtained for an amphibian. This dataset is used to describe the extent to which gene orders have been conserved between *Ambystoma* and other representative vertebrate species with assembled physical genome maps. Several other examples are also provided that illustrate the importance of the amphibian genome perspective for reconstructing gene orders of the ancestral tetrapod and amniote genomes, and for understanding the importance of gene order rearrangement in vertebrate evolution.

Materials and Methods

Sequence comparisons

Ambystoma orthologs were identified from assembled contigs of the Salamander Genome Project (Putta *et al*, 2004; Smith *et al*, 2005; http://www.ambystoma.org) and other sequences published in GenBank (see Supplementary_Document_6_1 and Supplementary_Table_6_1). These sequences ranged in length from 126 to 6167 bp and presumably correspond to partial and full-length RNA transcripts. A FASTA file of these sequences is included in Supplementary_Document_6_1. Similarity searches and

sequence alignments between translated *Ambystoma* sequences and translated genome sequences were performed using the program BLAT (Kent, 2002). Alignments were generated between the source sequences for 491 *Ambystoma* genetic markers (Smith *et al*, 2005) and genome assemblies for human, mouse, rat, dog, chicken, zebrafish, and *T. nigroviridis*. Source sequences for human, mouse, rat, dog, chicken, zebrafish, and *T. nigroviridis* (respectively: hg17 build 35, mm6 build 34, rn3, canFam1, galGal2, danRer2, tetNig1 V7) were downloaded from the UCSC Genome Browser Gateway (http://genome.ucsc.edu/cgi-bin/hgGateway). Cumulative bitscores were calculated for alignments between *Ambystoma* sequence and full genome sequences by summing across presumptive exons. This was accomplished by summing bitscores for otherwise continuous alignments that were interrupted by gaps of 10,000 or fewer bases.

Statistical analysis of conserved synteny

Houseworth and Postlethwait (2002) proposed two measures of synteny conservation: ρ and λ. Both of these statistics measure the degree of association between chromosomes (or other segments) from two genomes. The statistic ρ is equivalent to the square of Cramer's V statistic for frequencies of orthologs in a two-way table of chromosomes (Housworth and Postlethwait, 2002; Cramer, 1946). Cramer's V and ρ are scaled χ^2 statistics and as such may not be fully appropriate for measures of association when the average cell frequency within a two-way contingency table is less than 6 (Roscoe and Byars, 1971). In other words accurate estimation of the χ^2 statistic for comparisons between two genomes with 1N = 20 would require at minimum identification of 2400 (20*20*6) orthologies. Furthermore, χ^2 based measures of association are not directly comparable between analyses, nor interpretable in a probabilistic sense (*e.g.* Fisher, 1938; Blalock, 1958; Goodman and Kruskal, 1954; Kendall and Stuart, 1967).

In terms of pairwise comparisons between genomes, λ provides a measure of the proportional increase in ability to predict the chromosomal assignment of an ortholog in either of two species (or in probabilistic terms, "the relative decrease in probability of erroneous guessing" (Kendall and Stuart, 1967), when the ortholog's position is known in the other species, *vs.* when it is unknown) (Kendall and Stuart, 1967). The value of λ ranges from 0 to 1, with a value of $\lambda = 0$ representing the case where knowledge of the

positions of orthologous loci in either species is completely uninformative in predicting the location of orthologs in the other, and a value of $\lambda = 1$ representing the case where knowledge of the positions of orthologous loci in either species can be used to exactly predict the location of all orthologs in the other. Values of ρ and λ were highly similar among analyses. For simplicity and ease of interpretation, and because the λ statistic is seemingly more appropriate for the question at hand, only values for λ are reported with approximate 95% confidence estimated using the methods of Goodman and Kruskal (1963).

Statistical analysis of segmental homology

Segmental homologies were identified by comparing the positions of orthologs between the *Ambystoma* genetic map and the reference genomes for human, mouse, rat, dog, chicken, zebrafish, and *T. nigroviridis*. The *Ambystoma* map and reference genomes were formatted as concatenated (across linkage groups or chromosomes) series of orthologs and input into the program FISH (Calabrese *et al*, 2003). In effect, FISH identifies segmental homologies by comparing the distribution of points on an oxford plot to the expected null distribution for an equal number of randomly scattered points. Concatenating chromosomes of multichromosomal genomes permits correct calculation of the null distribution of orthologies by FISH. However, one potential caveat of using concatenated genomes is that the analysis does not take into account the position of chromosomal boundaries. The possibility therefore exists that clusters or orthologies that cross the boundaries of chromosomes or linkage groups will be identified as segmental homologies. Because these putative clusters involve artificially generated segments, they likely represent spurious segmental homologies. To check for this possibility, the locations of all identified segmental homologies were examined manually. A single segmental homology in the *Ambystoma*-mouse comparison was observed that crossed a boundary. This homology was removed from subsequent analyses. It is noted, however, that boundary-crossing clusters might alternately represent fission breakpoints that were placed (by chance) adjacent to one another in the concatenated genome. This possibility will be explored in future work.

Results

Identification of putative orthologs

In total, 491 protein-coding marker sequences from the *Ambystoma* genetic map were searched against the genome sequences of human (*Homo sapiens*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), dog (*Canis familiaris*), chicken (*Gallus gallus*), zebrafish (*Danio rerio*), and freshwater pufferfish (*Tetraodon nigroviridis*) to identify presumptive orthologs. For each search, orthologs were defined as the BLAT hit with the highest bitscore, plus all other hits within 1% of the highest bitscore. Using this definition, orthologs for 344 (70%) *Ambystoma* sequences were identified within the genome of at least one species in the reference set. Alignment summaries for all presumptive *Ambystoma*-vertebrate orthologs are provided in Supplementary_Table_6_2. The number of *Ambystoma* orthologs varied among species, ranging from 237 to 322. A low proportion of *Ambystoma* orthologs aligned to more than one presumptive ortholog in all comparisons (human, 4.0%; dog, 6.0%; mouse, 8.4%; rat, 4.4%; chicken, 1.8%; *T. nigroviridis*, 4.9%; zebrafish, 3.8%). In general, a greater number of *Ambystoma* orthologs were identified among amniote taxa *vs.* fish taxa (Table 6.1). This suggests greater conservation of orthologs among tetrapod taxa.

To gain insight into variation in genome coverage of *Ambystoma*-vertebrate orthologies, the distribution of *Ambystoma*-human orthologies was compared to the expected distribution under random sampling of human loci (Table 6.2). The human genome assembly for this comparison (http://www.ncbi.nih.gov/mapview/stats/BuildStats.cgi?taxid= 9606&build=36&ver=1) was selected because the assembly is relatively complete and contains a large number of gene annotations. The observed number of orthologs on three human chromosomes (HSA1, 12, and 17) deviated significantly from the expected number ($p>0.01$). Notably, all three of these chromosomes contained an excess of orthologies, rather than a deficiency. A single human chromosome (HSA4) contained a marginally significant deficiency of *Ambystoma* orthologies ($p = 0.05$, however, given the large number of comparisons, a similar deviation would be expected to occur by chance. Comparisons with the human genome assembly suggest that *Ambystoma*-amniote orthologies will provide coverage of most regions of ancestral vertebrate genomes.

Orthologs for a majority of *Ambystoma* marker sequences were identified in more than one reference genome (Figure 6.2). Of the 343 *Ambystoma* orthologs identified from all searches, 292 (85%) yielded hits to five or more genomes. A high proportion of *Ambystoma*-amniote orthologs were identified from four or more of the amniote taxa (88%). A lower proportion of *Ambystoma*-fish orthologs were identified from both fish taxa (68%). The relatively lower proportion of *Ambystoma*-fish orthologs may reflect lineage-specific gene losses and divergence that has occurred between these fish species, or differences in completeness of their genome assemblies. Below, *Ambystoma* orthologies are used as characters to identify conserved syntenies and gene orders and reconstruct key events in the evolution of vertebrate genomes.

Conservation of synteny

The association index λ describes the extent to which chromosomal assignments of loci (genes) in one species are predictive of chromosome assignments in another species (see Methods). High λ values indicate high predictability; such values are expected when few inter-chromosomal rearrangements of genes occur between two species after divergence from a common ancestor. Thus, λ provides a measure of the combined effects of phylogenetic distance and lineage specific rearrangement rates on the inter-chromosomal distribution of genes. The statistic λ□□□□□□□□□□□□□ for pairwise comparisons between *Ambystoma* and each of the seven reference vertebrate genomes. Significant (non-zero) association indices were observed for all comparisons and there was considerable variation in λ values (0.18 for *Ambystoma vs.* zebrafish or mouse to 0.33 for *Ambystoma vs.* chicken; see Table 6.1). Variable λ values for *Ambystoma*-amniote comparisons illustrate the importance of lineage specific effects, because all amniotes share the same divergence time. In this case of λvariability among amniotes, lower λ values for *Ambystoma*-murid rodents indicate an increased rate of genome rearrangement in the murid rodent lineage.

To obtain a more complete picture of genome similarity pairwise λ values were calculated for all possible species comparisons using two different datasets: 1) the set of all genes that showed 1:1 orthology in comparisons between *Ambystoma* each of the seven reference vertebrate genomes (ranging from N=170 for chicken-zebrafish to

N=309 for *Ambystoma*-human), and 2) a smaller set of genes (N=110) in which 1:1 orthology was established among all species. The cumulative gene set was expected to identify a greater number of associations while the smaller set controlled for comparison-wise differences among the gene sets used to estimate λ. The cumulative gene set yielded slightly lower values of λ than the smaller set. For both gene sets, λ was inversely correlated with phylogenetic distance (Figure 6.3) (see Supplementary_Table_6_3). As before, it was also observed that λ varied substantially among species with identical divergence times, consistent with lineage specific variation in rearrangement rates. Notably, λ values for the *Ambystoma*-chicken comparison are higher than or similar to λ values calculated between chicken and mammals, and similar to values calculated between murid rodents and non-rodent mammals, despite differences in divergence time among these comparisons on the order of 60 to 280 million years (MY) (Figure 6.1) (see Supplementary_Table_6_3). To better understand the effect of divergence time on variation in λ, the average rate of decrease in λ ((1-λ)/divergence time)) was estimated for all pairwise comparisons. Figure 6.4 shows that interchromosomal rearrangement rates are strikingly higher in murids and more variable among mammals in comparison to all other vertebrate groups. In contrast, genome rearrangements in non-mammalian vertebrate species appear to accumulate at a similar, lower rate.

Conservation of segmental homology

The map position of *Ambystoma* genes were compared to the physical positions of their presumptive orthologs in each of the seven reference genomes. These pair wise comparisons were visualized using Oxford plots to show intra-chromosomal positions of orthologies between *Ambystoma* and each of the reference genomes (Figures 6.5-6.8) (see Supplementary_Figures_6_1-6_3). In Oxford plots, conserved segmental homologies can be identified as diagonally oriented clusters of points. This visual approach was compared with a statistical approach using the program FISH (Calabrese *et al*, 2003). The algorithm underlying FISH appears to be somewhat conservative for *Ambystoma*-amniote comparisons because several clusters that are visually indicative of conserved segments were not identified as such, and several of the significant clusters did not always include orthologies that were very close to cluster margins. At any rate, the majority (57%) of the

334 *Ambystoma*-amniote orthologs were found within statistically significant, segmental homologies. The proportion of *Ambystoma* orthologs that were assigned to homologous segments varied greatly among comparative maps (Table 6.1) (see Supplementary_Table_6_2). A much higher percentage of *Ambystoma*-amniote orthologs were found in significant segment homologies than *Ambystoma*-fish orthologs. For example, the *Ambystoma*-chicken Oxford plot reveals a striking pattern of conservation of gene order (Figure 6.7). Overall, fewer segmental homologies were identified between *Ambystoma* and murid rodents *vs.* nonrodent amniotes. However, the number of segmental homology differences among amniotes was small in comparison to the nearly 2-fold difference in λ values observed for *Ambystoma*-murids *vs. Ambystoma*-nonrodents (Table 6.1). Thus, although there has been greater reordering of loci among murid chromosomes during evolution, orders of loci within murid chromosomes are conserved and identifiable in comparisons to *Ambystoma*.

Discussion

Amphibians occupy an important, intermediate position in the vertebrate evolutionary tree. This study is the first to include amphibian gene order data in a taxonomically broad comparison of vertebrate genome structure. Comparisons of genome structure between *Ambystoma* and representative fish, reptilian, and mammalian species revealed extensive conservation of gene location at the intra- and inter-chromosomal levels. Overall, conserved syntenies and segmental homologies were identified for hundreds of *Ambystoma* protein-coding sequences (see Supplementary_Table_6_1). These data provide evidence, beyond nucleotide identity, that *Ambystoma* genes are annotated with the correct vertebrate orthology. Information about gene orthology, conserved synteny, and segmental homology will extend *Ambystoma* as a research model because it will enable development of orthologous probes for comparative molecular studies, and the identification of candidate genes for *Ambystoma* mutants and QTL.

This study shows that the *Ambystoma* Genetic Map can identify conserved syntenies and segmental homologies when compared to any of the primary vertebrate model organism genome assemblies. Overall, greater conservation of genome structure was observed between *Ambystoma* and amniotes, however, many conserved syntenies are

identifiable between *Ambystoma* and fish (*T. nigroviridis*, zebrafish). It was also observed that genome rearrangement rates are not simply a function of phylogenetic distance; there are clear differences in inter-chromosomal rearrangement rates, especially within mammals, as well as between mammals and "lower vertebrates". These points are elaborated upon below, as well as and several new insights that amphibians provide concerning vertebrate genome evolution.

Genome conservation between *Ambystoma* and fish

Fewer presumptive orthologs, conserved syntenies, and segmental homologies were identified between *Ambystoma* and fish (*T. nigroviridis*, zebrafish) than between *Ambystoma* and amniotes. This result is expected because of the deeper divergence time of *Ambystoma* and fish; in other words, there has been more time for nucleotide substitutions (that make it difficult to identify orthologs) and synteny disruptions to accumulate since the divergence of *Ambystoma* and fish from a common ancestor. Nevertheless, 57% of *Ambystoma* orthologs were observed in conserved syntenies with four or more orthologs in at least one fish species, and with the exception of *Ambystoma* linkage group (LG)13 (which shows strong synteny with GGA3), all *Ambystoma* linkage groups show discreet regions of synteny with chromosomes of *T. nigroviridis* and zebrafish. Assuming conservation of gene order during evolution, several regions of conserved synteny between *Ambystoma* and fish were likely present in the ancestral euteleostean genome. These include: the right hand portion of *Ambystoma* LG6, which shows extensive synteny with TNI21 and segmental homology with DRE19; and *Ambystoma* LG10, which shows extensive synteny with TNI15 and DRE20 (Figure 6.8) (see Supplementary_Figure_6_3). Observation of extensive synteny between *Ambystoma* and fish is interesting because recent evidence suggests a whole genome duplication predating the common ancestor of *T. nigroviridis* and zebrafish, followed by differential losses of paralogous loci (*e.g.* Jaillon *et al*, 2004; Bernot and Weissenbach, 2004; Naruse *et al*, 2004; Woods *et al*, 2005). Under such a model of genome evolution, the positions of syntenic *Ambystoma* genes are expected to map to overlapping positions on different fish chromosomes. This pattern is not observed for *Ambystoma- T. nigroviridis* orthologs on a few of the smaller *Ambystoma* linkage groups (*e.g. Ambystoma* LG9 *vs.* TNI13 and

19), however this pattern is not as obvious in larger *Ambystoma* linkage groups, or in comparisons between *Ambystoma* and zebrafish. The observed patterns appear to be consistent with chromosomal duplications in some instances, but may alternately reflect ancient large-scale rearrangements that have since been shuffled to yield interleaving sets of conserved syntenies. Better reconstruction of the pre-duplicated, ancestral teleost genome is needed to differentiate between these possibilities.

Genome conservation between *Ambystoma* and amniotes

Results from this study indicate extensive conservation of gene orders between *Ambystoma* and amniotes, and especially between *Ambystoma* and chicken. Many of the orthologs identified on the smaller chicken chromosomes exist in nearly exclusive synteny or segmental homology with discreet regions of the *Ambystoma* genome (Figure 6.7). This is interesting in view of the large difference in chromosome number and genome size between these species. *Ambystoma* has a much larger genome and haploid complement of 14 chromosomes (Licht and Lowcock, 1991), whereas chicken has a haploid complement of 39 chromosomes (ICGSC, 2005). Given that an ancestral chromosomal number of 12-14 chromosomes seems most likely for euteleost (Jaillon *et al*, 2004; Bernot and Weissenbach, 2004; Naruse *et al*, 2004; Woods *et al*, 2005; Postlethwait *et al*, 2000; Kohn *et al*, 2006) tetrapod (Smith, unpublished data), and reptilian ancestors (Norris *et al*, 2004), differences between *Ambystoma* and chicken genomes are largely explained by lineage specific fissions (mostly giving rise to individual chicken microchromosomes) and a moderate number of large rearrangements. The very high number of segmental homologies observed between *Ambystoma*-chicken suggests they share a large portion of the ancestral tetrapod genome structure. When considering additional segmental homologies identified between *Ambystoma* and mammals, more than half of the *Ambystoma*-amniote orthologs that are currently located on the *Ambystoma* Genetic Map identify segmental homologies within at least one amniote genome, and by extension, the ancestral amniote and tetrapod genomes.

Variation in interchromosomal rearrangement rates

This study corroborates the idea that mammalian genomes are characterized by higher

and more variable rates of genome rearrangement in comparison to other vertebrate groups (*e.g.* Kohn *et al*, 2006; Bourque *et al*, 2004; Bourque *et al*, 2005). In comparison to mammals, lower estimates of genome rearrangement rates were obtained for *Ambystoma*, chicken, zebrafish, and *T. nigroviridis*. These estimates are consistent with cytogenetic data that indicate extensive conservation of the avian karyotype over approximately 80 to100 MY of evolution (Ohno, 1967; de Boer, 1980; Shetty *et al*, 1999), with estimates of genome rearrangement rates between chicken and mammals (Bourque *et al*, 2005; Burt *et al*, 1999), and with comparisons between chicken and reptiles (Matsuda *et al*, 2005). It is curious to find similar rearrangement rates among non-mammalian vertebrates that differ so greatly in life history and genome structure, and whose genomes have been shaped differently by lineage-specific processes during evolution. Birds, amphibians, and fish have very different generation times, chromosome numbers, and genome sizes. However, these results suggest relatively constant rates of genome rearrangement from the euteleost ancestor to the origin of mammals.

Evolution of human chromosomes

In the remainder of the discussion a few examples are provided to show how *Ambystoma* provides perspective on the evolution of gene orders within the human genome. In general, *Ambystoma* comparative mapping data are useful because they help establish ancestral amniote and tetrapod gene orders. The *Ambystoma* ancestral perspective is needed to identify conserved syntenies and disruptions, and to corroborate evolutionary inferences based only on comparisons between chicken and mammals (ICGSC, 2005; Bourque *et al*, 2004; Schmid *et al*, 2000; Nanda *et al*, 2002; Schmid *et al*, 2005) or only mammals (Bourque *et al*, 2004; Murphy *et al*, 2003 ; Richard *et al*, 2003; Yang *et al*, 2003).

Synteny of HSA1 and HSA19 loci in the ancestral amniote and tetrapod genomes.

A region of segmental homology between *Ambystoma* LG4 and GGA28 overlaps regions of segmental homology between *Ambystoma* LG4 and two human chromosomes, HSA1 and HSA19 (Figure 6.9). This arrangement suggests that portions of HSA1 and 19 were joined in the ancestral tetrapod and amniote genomes. Fission of this ancestral gene order

presumably occurred before the diversification of eutherian mammals (87 MY ago) because *Ambystoma* LG4 orthologies are distributed similarly among the chromosomes of human, mouse, rat, and dog. The overall distribution of conserved syntenies among *Ambystoma* and amniotes indicates that many *Ambystoma* LG4 genes were syntenic in the ancestral tetrapod genome.

Synteny of HSA7 and HSA12 loci in the ancestral amniote, tetrapod, and euteleost genomes

Regions of synteny and segmental homology between *Ambystoma* LG9 and GGA1 overlap the positions of syntenic markers located on HSA7 and 12 (Figure 6.9). This arrangement suggests that loci of HSA7 and 12 were syntenic in the ancestral tetrapod and amniote genomes. As was observed above for *Ambystoma* LG4, fission of this ancestral gene order presumably occurred before the diversification of eutherian mammals because *Ambystoma* LG9 orthologies are distributed similarly among the chromosomes of human, mouse, rat, and dog. Because *Ambystoma* LG9 also shows conserved synteny and segmental homology with much of DRE4, many *Ambystoma* LG9 genes were apparently syntenic in the euteleost ancestral genome.

Value of multiple species in comparative genomics

Ambystoma LG12 and 13 show extensive conserved synteny and segmental homology with portions of GGA1 and 3, respectively. Apparently, these homologous chromosomal segments have changed little since diversification of the tetrapod lineage, approximately 370 MY ago. However, neither *Ambystoma* LG12 nor *Ambystoma* LG13 show substantial conserved synteny or segmental homology with any human chromosome. This suggests the possibility of lineage-specific synteny disruptions in the primate lineage, because *Ambystoma* LG12 does show conserved synteny with portions of the X-added region of rat and dog (Graves, 1995, Ross *et al*, 2005). This example shows that conserved chromosomal segments may not always be identifiable in the human genome or other mammalian genomes; a multi-species perspective is essential to identify lineage specific effects in comparative vertebrate genomics.

Fissions derived within the mammalian lineage

Several pairs of human chromosomes may have been fused in the ancestral mammalian genome: HSA3/21, 4/8, 10/12, 7/16, 14/15, 16/19, and two regions of 12/22, (Bourque *et al*, 2004; Murphy *et al*, 2003 ; Richard *et al*, 2003; Yang *et al*, 2003). Although the current dataset is insufficient to test all of these hypotheses, the *Ambystoma*-human Oxford plot (Figure 6.5) suggests that two of these chromosome pairs were fused in the ancestral tetrapod and amniote genomes (Figure 6.10). Conserved syntenic regions of HSA16 and 19 adjoin each other in the middle of *Ambystoma* LG3 and at the left end of *Ambystoma* LG4. Additionally, conserved syntenic regions of HSA7 and HSA16 adjoin each other on the right end of *Ambystoma* LG3. These data suggest that some of the chromosomal arrangements that have been proposed for the ancestral mammalian genome may trace back to the ancestral tetrapod genome.

Conclusions

These studies demonstrate the importance of amphibians in revealing key events and trends in vertebrate genome evolution. Measurements of conserved synteny using *Ambystoma* orthologies suggest relatively constant rates of genome rearrangement from the euteleost ancestor to the origin of mammals. *Ambystoma* comparative mapping data are also useful in establishing ancestral amniote and tetrapod gene orders and identifying synteny disruptions that have occurred in amniote lineages. More than half of the *Ambystoma*-amniote orthologs that are currently located on the *Ambystoma* Genetic Map identify segmental homologies within at least one amniote genome, and by extension, the ancestral amniote and tetrapod genomes. Comparisons between *Ambystoma* and amniotes also reveal some of the key events that have structured the human genome since diversification of the ancestral amniote lineage.

Table 6.1: **Summary of sequence alignments and analyses of synteny and segmental homology using the full set of mapped *Ambystoma* sequences.**

Species	Divergence Time[a]	Orthologs	BLAT Alignments	λ	% in SH[b]
Human	370 MY	322	341	0.25±0.12	32
Dog	370 MY	317	340	0.24±0.12	25
Mouse	370 MY	308	350	0.18±0.12	27
Rat	370 MY	316	340	0.19±0.13	23
Chicken	370 MY	284	301	0.33±0.13	34
T.nigroviridis	450 MY	243	255	0.23±0.14	2
Zebrafish	450 MY	237	248	0.18±0.15	3
Total Numbers		343	2175		57

a – The approximate divergence time in millions of years (MY) between *Ambystoma* and the vertebrate species used in this study. See Figure 1 for references. b – SH = Segmental Homologies.

Table 6.2: Distribution of human/*Ambystoma* orthologies across human chromosomes.

Human	Annotated	Human/Ambystoma Orthologies				
Chromosome	Genes[a]	Observed	Expected[b]	Deviation[c]	χ^2	p
HSA1	2,580	47	33.3	+	5.67	0.017
HSA2	1,742	16	22.5	-	1.86	0.173
HSA3	1,378	21	17.8	+	0.59	0.443
HSA4	1,017	6	13.1	-	3.86	0.050
HSA5	1,188	10	15.3	-	1.85	0.174
HSA6	1,385	11	17.9	-	2.63	0.105
HSA7	1,367	21	17.6	+	0.65	0.421
HSA8	927	12	12	+	0.00	0.989
HSA9	1,069	12	13.8	-	0.23	0.631
HSA10	983	8	12.7	-	1.72	0.189
HSA11	1,692	15	21.8	-	2.13	0.145
HSA12	1,261	28	16.3	+	8.48	0.004
HSA13	491	3	6.3	-	1.75	0.186
HSA14	1,173	10	15.1	-	1.74	0.188
HSA15	884	11	11.4	-	0.01	0.906
HSA16	1,031	16	13.3	+	0.55	0.458
HSA17	1,354	29	17.5	+	7.63	0.006
HSA18	400	4	5.2	-	0.26	0.610
HSA19	1,584	28	20.4	+	2.81	0.094
HSA20	710	11	9.2	+	0.37	0.542
HSA21	337	4	4.3	-	0.03	0.869
HSA22	693	4	8.9	-	2.73	0.099
HSAX	1,125	13	14.5	-	0.16	0.693

a – The number of genes that were assigned to each human chromosome (NCBI Build 35.1). b – The expected number of human salamander orthologies is based on random sampling from the human genome. This number is calculated as the proportion of genes that occur on a given human chromosome multiplied by the entire number of human/*Ambystoma* orthologies that were identified (N = 340). A single HSAY orthology is excluded from this table. c – "+" designates a positive deviation from the expected number,"-" designates a negative deviation from the expected number.

Figure 6.1: An abridged phylogeny of the vertebrates showing the species used and divergence times. Letters mark ancestral nodes: A – the euteleost (bony vertebrate) ancestor (Kumar and Hedges, 1998), B – the tetrapod ancestor (Ruta *et al*, 2003), C – the amniote ancestor (Hedges and Kumar, 2004; Riesz and Muller, 2004), D – the (eutherian) mammalian ancestor (Springer *et al*, 2003), E – the (murid) rodent ancestor (Springer *et al*, 2003), and the teleost fish (euteleostei) ancestor (Nelson, 1994; Wittbrodt *et al*, 2002).

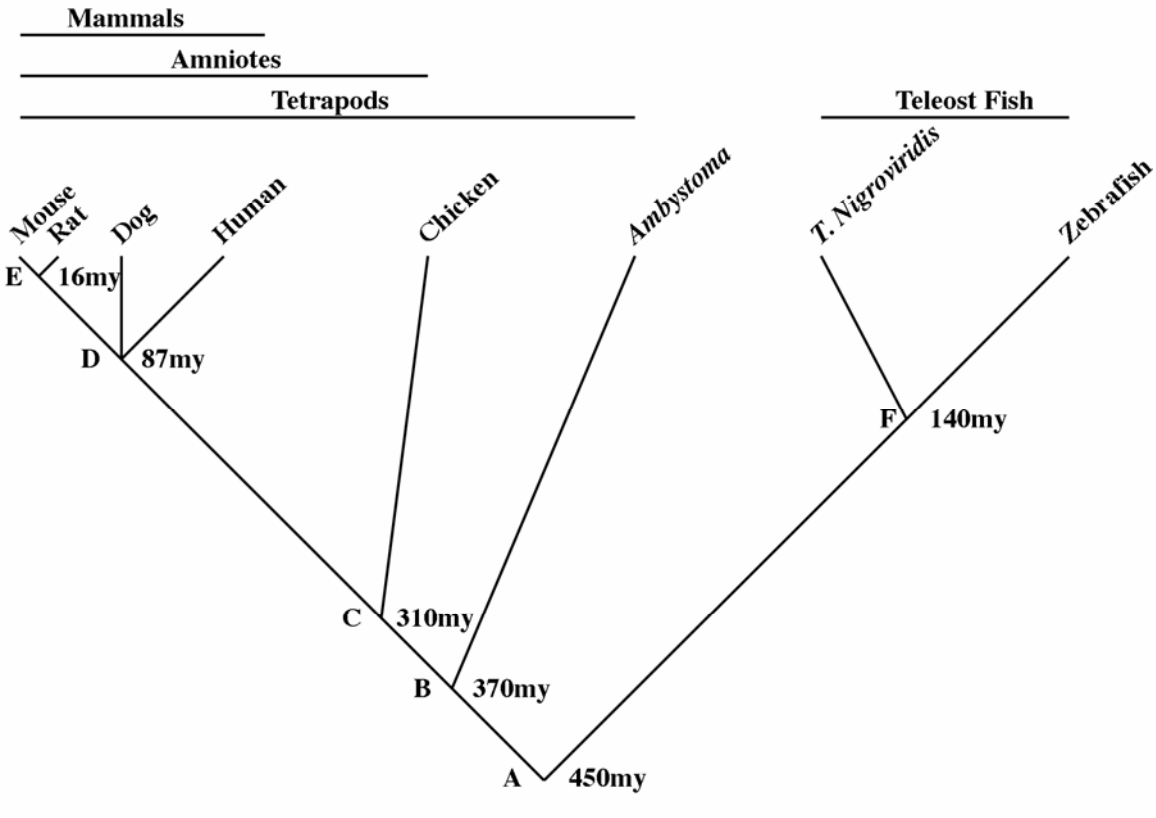


Figure 6.2: Frequency distributions of the numbers of presumptive *Ambystoma* orthologs that were identified among major vertebrate groups. Categories on the X-axes represent the total number of species within each major vertebrate group that yielded a presumptive *Ambystoma* ortholog.

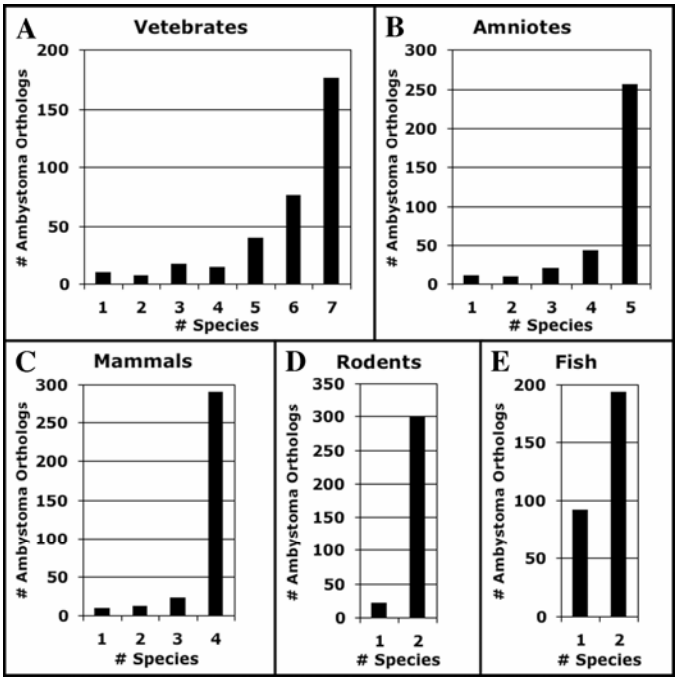


Figure 6.3: Plot of the λ association index *vs.* the divergence time of vertebrate species. Labels A-F denote comparisons that hinge on the common ancestral nodes shown in Figure 1.

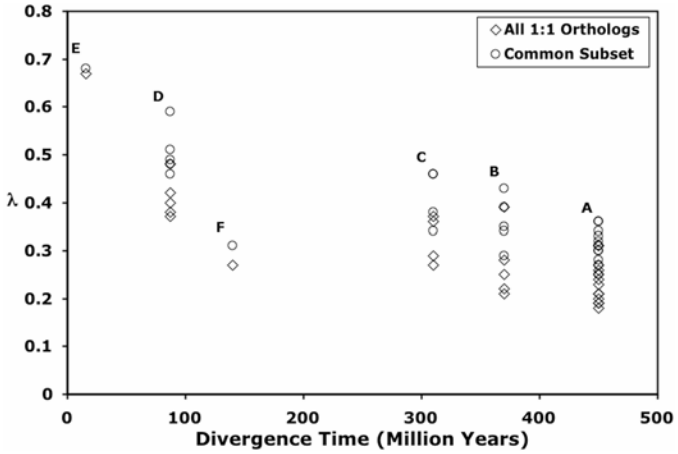


Figure 6.4: The average rate of decrease in λ for multiple vertebrate species, plotted relative to the *Ambystoma* rate. Rates are based on the set of all genes that showed 1:1 orthology in comparisons between *Ambystoma* each of the seven reference vertebrate genomes (All), and a smaller set of genes in which 1:1 orthology was established among all species (Common).

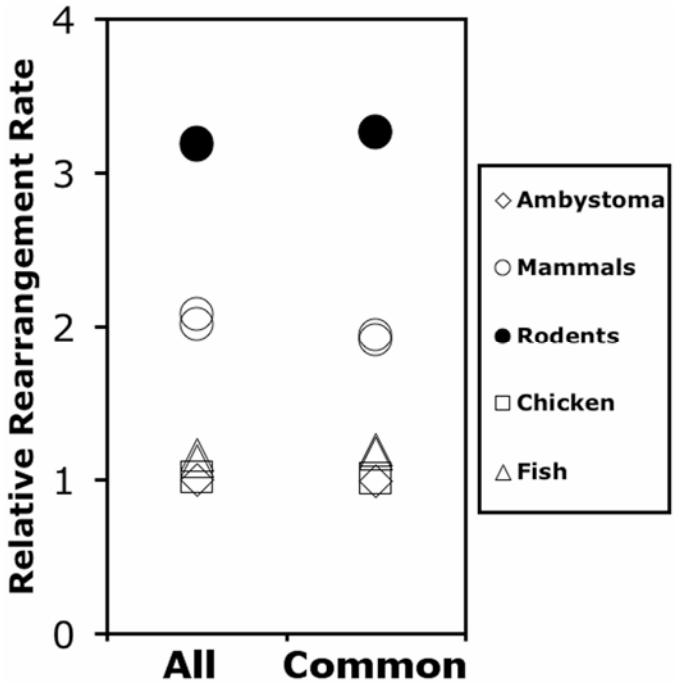


Figure 6.5: Oxford plot of the positions of presumptive orthologies between *Ambystoma* linkage groups and human chromosomes. Lines represent the boundaries of chromosomes (horizontal) and linkage groups (vertical). Dots represent the relative position of orthologs within the *Ambystoma* map. Orthologs that are members of synteny groups of three or more are shown in black. Orthologs that are members of significant (FISH) segmental homologies are shown in red. Some smaller chromosomes are not labeled. All chromosomes are presented ordinally from bottom to top.

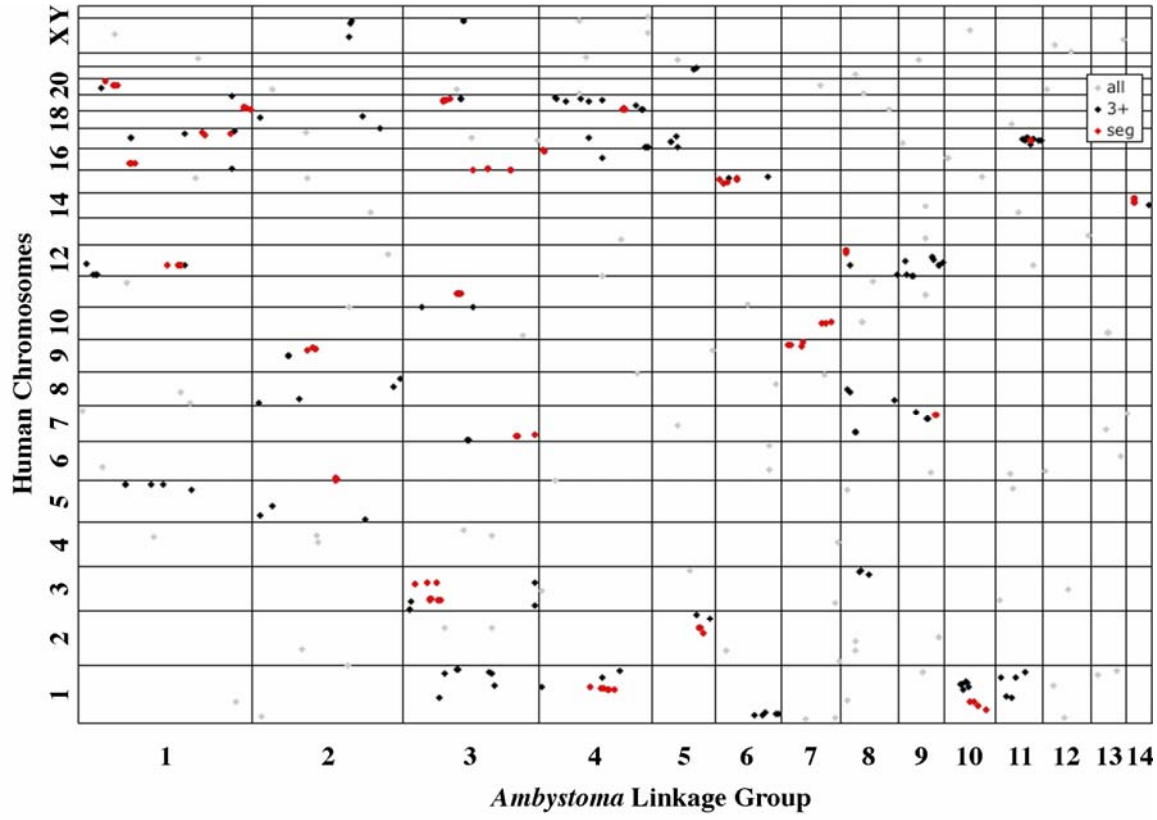


Figure 6.6: Oxford plot of the positions of presumptive orthologies between *Ambystoma* linkage groups and mouse chromosomes. See Figure 5 for further details.

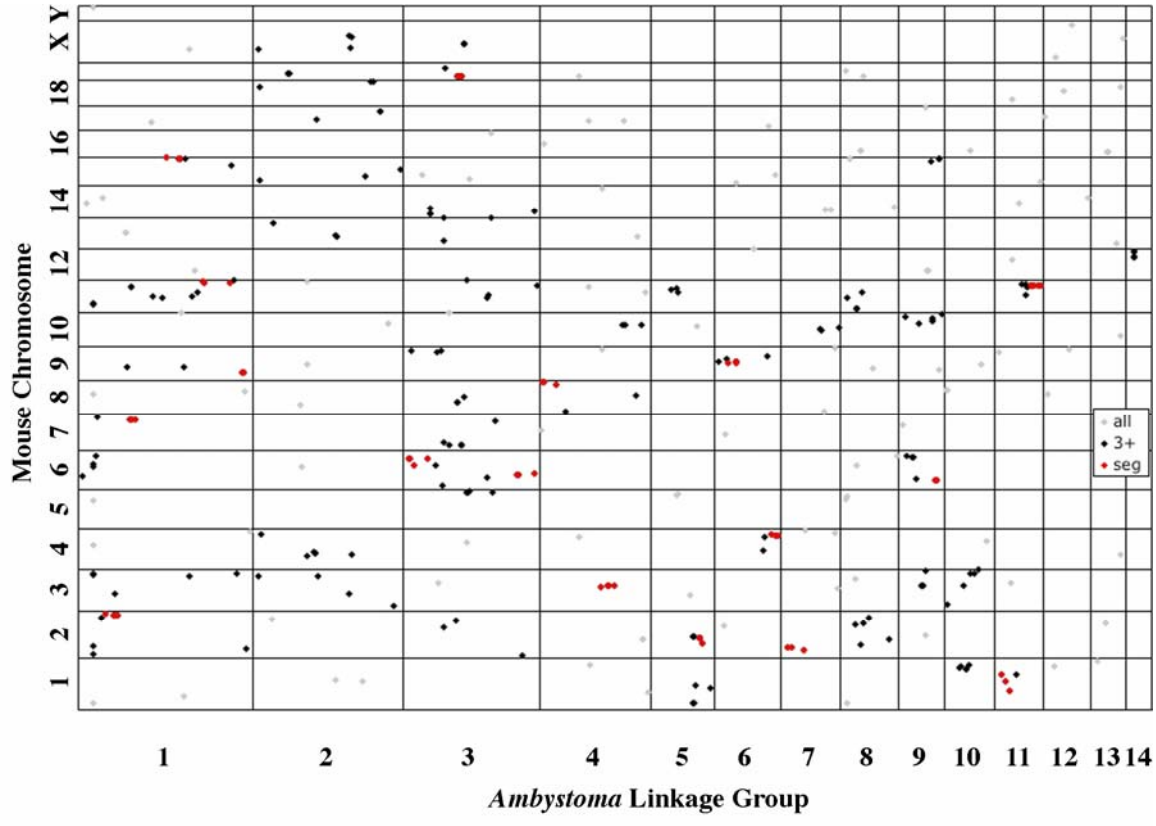


Figure 6.7: Oxford plot of the positions of presumptive orthologies between *Ambystoma* linkage groups and chicken chromosomes. The minichromosomes GGA16, 25, 29-31 and 33-38 are not shown; GGA16 contains no mapped *Ambystoma* orthologs and others are not represented in the genome assembly. See Figure 5 for further details.

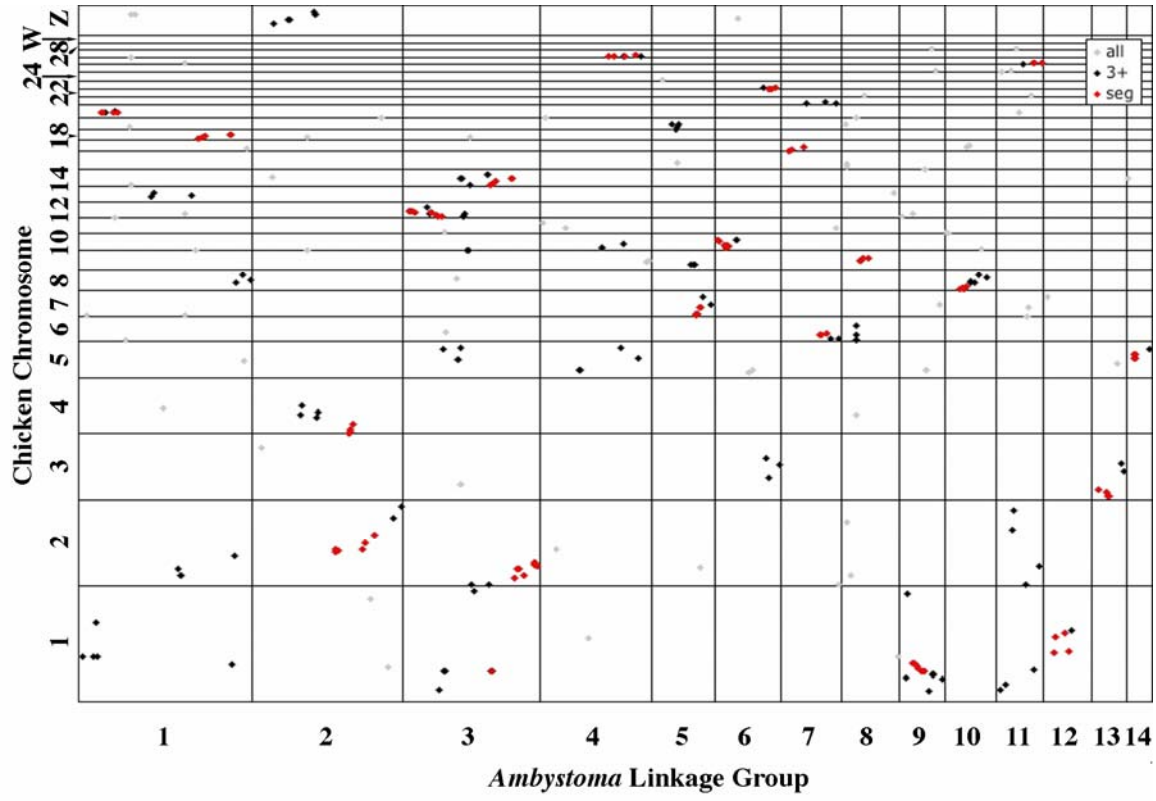


Figure 6.8: Oxford plot of the positions of presumptive orthologies between *Ambystoma* linkage groups and *Tetraodon nigroviridis* chromosomes. See Figure 5.5 for further details.

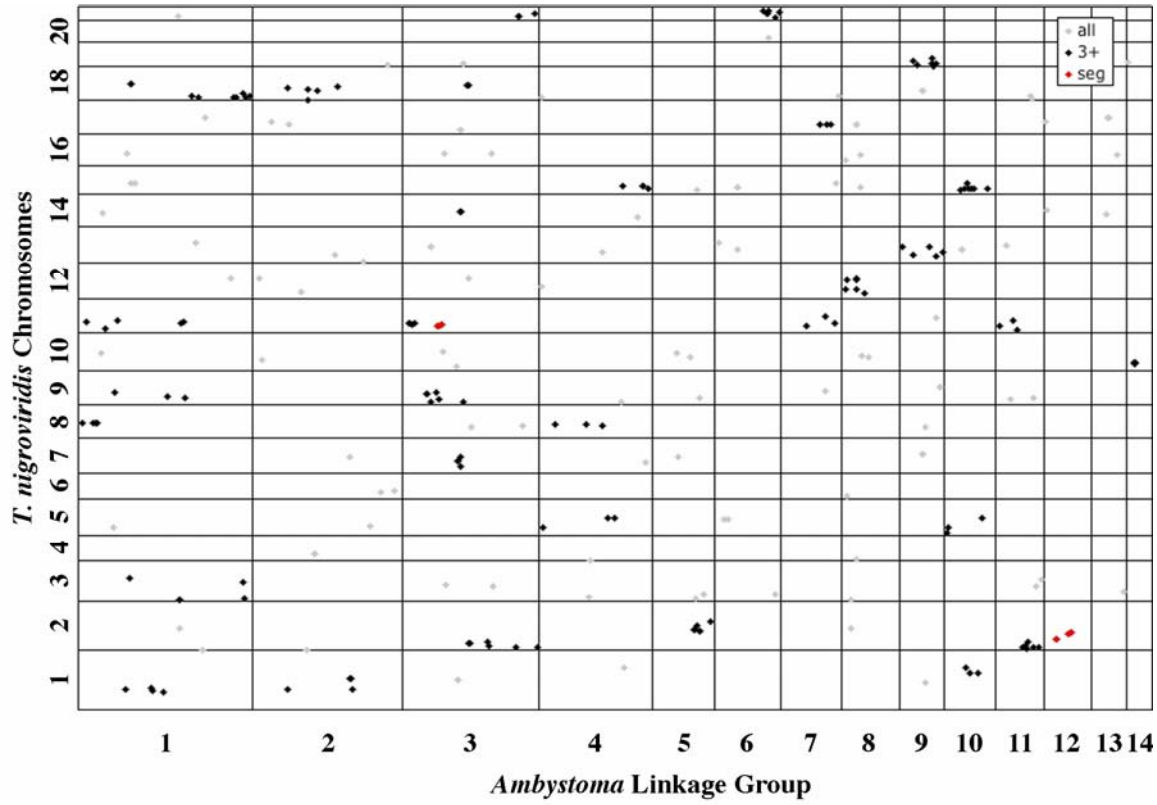


Figure 6.9: Oxford plot of the positions of presumptive orthologies between *Ambystoma* LG4 and 9, GGA1 and 28, and HSA1, 7, 12, and 19. The distribution of *Ambystoma*-chicken and *Ambystoma*-human conserved syntenies suggests that much of the content of *Ambystoma* LG4 and 9 was present in the ancestral tetrapod genome.

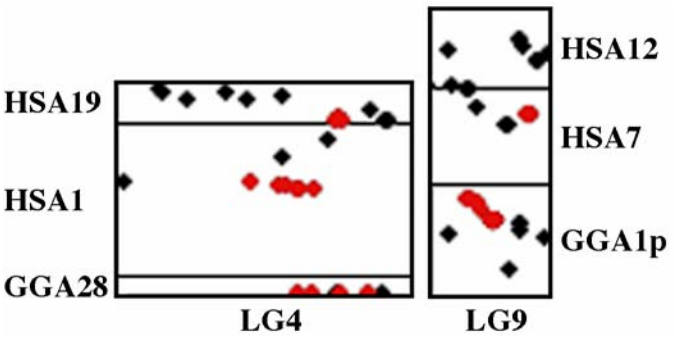
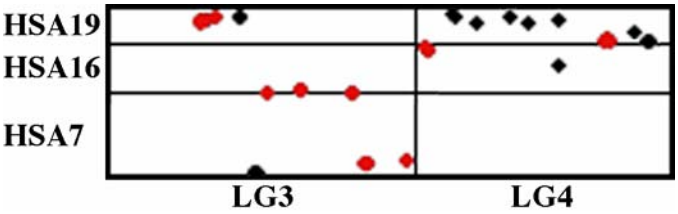


Figure 6.10: Oxford plot of the positions of presumptive orthologies between *Ambystoma* LG3 and 4 and HSA7, 16, and 19. The arrangement of *Ambystoma*-human orthologies suggests that some of the proposed ancestral chromosome arrangements of mammals can be traced back to the ancestral tetrapod genome.



CHAPTER 7: MapToGenome: A Comparative Genomic Tool that Aligns Transcript Maps to Sequenced Genomes.

Introduction

Many evolutionary analyses require the identification of orthologous gene loci among the genomes of unrelated organisms. This task is not easily accomplished because genomes are rarely characterized in the same way or to the same extent. Partial and whole-genome sequences are only available for a few primary vertebrate model organisms (*e.g.* pufferfish, chicken, mouse, zebrafish) and gene annotations vary considerably. For many more eukaryotic species, genomes are characterized by a subset of markers or transcripts that collectively constitute a physical or genetic map. The recent and continuing explosion of expressed sequence tag (EST) projects will likely increase the number of transcript maps in the next few years and these data stand to complement and enhance comparative studies. However, to exploit these resources, tools are needed to better facilitate comparisons of orthologous genes between transcript maps and whole genome sequences.

Transcript maps can be compared directly to genome assemblies by first identifying orthologous alignments between mapped transcripts and a genomic sequence, then cross-referencing the positions of orthologies. Several programs have been developed to align transcribed sequences to the genome from which they originated (*e.g.* Mott, 1997; Florea *et al*, 1998; Wheelan *et al*, 2001; Usuka *et al*, 2000; Schlueter *et al*, 2003; Lee *et al*, 2003; Ranganathan *et al*, 2003; Kruger *et al*, 2004; Wu and Watanabe, 2005; van Nimwegen *et al*, 2006), or to generate transcript-anchored alignments of genomic sequences among divergent taxa (*e.g.* Bray *et al*, 2003; Bray and Patcher, 2004; Yap and Patcher, 2004; Flannick and Batzoglou, 2005; Ye and Huang, 2005; Hsieh *et al*, 2006; Huang *et al*, 2006). Currently, no programs exist for directly aligning transcript maps to divergent genome assemblies. As in the case of within-species transcript mapping, divergent transcript-genome comparisons necessitate gapped alignments because introns often account for a significant fraction of the primary transcript - roughly 95% of the average human primary transcript (Duret *et al*, 1995; Venter *et al*, 2001). The process of aligning transcripts to divergent genomic sequence is additionally complicated by nucleotide divergence and genome-level variation (*e.g.* duplication and genome size

evolution) that has evolved since the two species last shared a common ancestor. Precise delimitation of intron-exon boundaries is challenging for within-species mapping and nucleotide mismatches can dramatically affect alignments in the vicinity of these boundaries (Nimwegen *et al*, 2006). Because transcripts maps are often derived from incompletely sequenced transcripts with relatively high intrinsic error rates (*e.g.* ESTs), rigorous annotation of intron-exon boundaries may be impossible for many comparisons. Clearly, tools for comparing divergent transcript and genome maps must contain features that can accommodate variation in nucleotide and genome diversity parameters to join gapped alignments optimally.

A further challenge toward identifying divergent orthologs for mapped transcripts stems from the fact that they often represent only a small fraction of the transcriptome. Often, sequences are defined as orthologs if they identify each other as their highest scoring alignment, otherwise known as reciprocal best alignment orthology or monotopoorthology (reviewed by Dewey and Patcher, 2006). This strict definition of orthology is convenient for comparisons of homologous chromosome segments or complete genomes (*e.g.* Bray and Patcher, 2004; Borque *et al*, 2004; Borque *et al*, 2005; Woods *et al*, 2005; Kohn *et al*, 2006) because nearly all transcripts are known and comparisons can be described in terms of 1:1 relationships. However, reciprocal best orthologies are only a subset of the true orthologous relationships that are possible between two genomes. Most notably, reciprocal best orthologies may not distinguish homologs arising from duplication events, which are known to have played a major role in shaping the genomes of extant vertebrate species (*e.g.* Jaillon *et al*, 2004; Meyer and Yan de Peer, 2005; Panopoulou and Poustka, 2005; Blome *et al*, 2006). Due to the above factors and because substantial portions of the transcriptome are commonly missing from transcript maps, it is prudent to use a more flexible definition of orthology for aligning maps and genomes that are derived from distantly-related organisms.

MapToGenome was developed as a flexible tool for aligning transcript maps and genome assemblies. MapToGenome processes sequence alignments between mapped transcripts and whole genome sequence while accounting for the presence of intronic sequences, and permits user-specifiable thresholds for maximum intron length and splice site fidelity. MapToGenome defines orthology based on two user definable thresholds:

the proportion of best alignments and minimum bitscore. It also cross-references mapping and genome positional information and generates Oxford plots of transcript map/genome alignments. The speed and flexibility of MapToGenome permits optimization of the parameters that are used to produce gapped alignments and assign orthologies. This chapter describes the operation and implementation of MapToGenome. Datasets from several vertebrate species are used to illustrate the utility of MapToGenome in identifying orthologous loci among distantly related vertebrate species, including whole genome sequence from rat, chicken, zebrafish, and *Tetraodon nigroviridis,* and genetic maps for rat, *Ambystoma*, zebrafish, and medaka. These analyses demonstrate that map-genome alignments can be improved by optimizing maximum intron length thresholds and by ignoring alignments for SSLP marker sequences. Specific examples are also provided that illustrate how MapToGenome allows visualization of the correspondence of mapped genes to their presumptive genome localizations, conservation of synteny, and disruption of genomes by chromosomal rearrangements and duplications.

Methods

Mapping and Sequence Data

Sequences for whole-genome assemblies were downloaded from the University of Santa Cruz Genome Browser Gateway http://genome.ucsc.edu/cgi-bin/hgGateway. These include genome assemblies for: rat (rn4), zebrafish (danRer4), chicken (galGal2), and *T. nigroviridis* (tetNig1). Linkage maps were obtained from the literature. The rat radiation hybrid (RH) map v3.4 (Kwitek *et al*, 2004) was downloaded from ftp://rgd.mcw.edu/pub/rhmap/3.4. Many of the markers on this map correspond to only a single EST sequence, therefore, a modified version was generated for use in this study by first aligning all ESTs to the rat RefSeq database, and replacing all EST sequences with their corresponding RefSeq sequence (Supplementary_Table_7_1). The Zebrafish RH (Geisler *et al*, 1999) and linkage (Gates *et al*, 1999; Kelly *et al*, 2000) maps were downloaded from the Zfin website http://zfin.org/zf_info/downloads.html#map. The salamander linkage map was from Smith *et al* (2005). The medaka linkage map was generated using genotypes from Naruse *et al* (2004) and the Kosambi mapping function of MapMaker QTXb20 (Supplementary_Table_7_2).

Sequence Alignment

Similarity searches and sequence alignments were accomplished using the program BLAT (Kent, 2002). All BLAT analyses were performed using default alignment criteria and were output in NCBI blast tabular format (*e.g.* -out=blast8). Intraspecific comparisons of transcript and genome assembly sequence were accomplished using nucleotide/nucleotide alignments. Interspecific comparisons were accomplished using alignments between transcript sequences that were translated in three forward frames and genome sequence that was translated in six frames.

MapToGenome

Algorithm: MapToGenome processes tabular alignment output from BLAT (Kent, 2002) (*e.g.* -out=blast8) or similarly formatted tabular alignment output, such as BLASTn (Altschul *et al* 1990) (*e.g.* -m 8). It examines all HSPs (high-scoring segment pairs) for a given marker and generates cumulative alignments and summary statistics by summing across presumptive exons for every query-subject pair. This is achieved by consolidating otherwise continuous alignments that are interrupted by gaps (presumptive introns). Consolidation is defined as: updating start and end positions for hit and query; summing identities, mismatches, gaps, and bitscores; and updating the % identity between query and hit (identities divided by the sum of identities, mismatches, and gaps). The maximum allowable length of presumptive introns (IL) and splice site fidelity (SF) are user definable. Splice site fidelity is defined as the number of base pairs that are permitted to overlap between alignments of adjacent regions of the query sequence, that have been consolidated across a presumptive intron. The algorithm works by first ordering all of the sequence alignments that have been generated for each query by their position within each chromosome of the subject genome. Next, the alignment orientation, and summary statistics are recorded for the first alignment. The orientation and alignment ends (query end and hit end for forward oriented alignments or query start and hit end for forward oriented alignments) of the first alignment are then compared to the orientation and the adjacent ends of the next alignment. If the two sequences are in the same orientation and the distance between adjacent query ends is within the SF and IL thresholds, the

alignments are consolidated and compared to the next alignment. If any of these conditions is not met, consolidation is terminated and a new consolidation round is initiated.

MapToGenome also permits assignment of two user-definable thresholds for selecting the alignments that will be presented in the final alignment summary and oxford plot files. These are 1) the minimum cumulative bitscore that is necessary for an alignment to be reported, and 2) the minimum ratio of an alignment's cumulative bitscore relative to the maximum cumulative bitscore for the query-genome alignment (*i.e.* proportion of best-in-genome). The program also generates an output file of all cumulative alignments prior to enforcing user-defined thresholds.

After generating a final alignment summary, MapToGenome appends information from files (provided by the user) to add the marker information and update marker/subject base positions. MapToGenome also provides a file of marker and subject positions that are concatenated across linkage groups and chromosomes, which is used for generating an oxford plot of map by genome alignments. Oxford plots are generated as postscript files.

Implementation: MapToGenome is written in C++ for UNIX platforms. It has been tested on Linux (2.4.x) and OS X (10.3, 10.4). It uses gnuplot (http://www.gnuplot.info) to generate oxford plots. Based on analyses of data sets ranging from several hundred to several thousand alignments, implementation of MapToGenome is limited only by memory size and processor speed. MapToGenome processed 100,438 BLAT alignments from the Rat/Chicken dataset into 4949 presumptive orthologies (at a 10 kb intron length threshold) in <30 seconds, using a standard desktop computer [1.25Ghz PowerPC G4 (3.2) CPU with 1GB DDR SDRAM running Mac OS X 10.3.9]. Using MapToGenome to consolidate alignments is much faster than rerunning BLAT at various gap thresholds. The initial BLAT alignment of the same Rat/Chicken dataset took 10.8 hours using a substantially more powerful computer [PowerPC G5 (3.0) CPU with 4 GB RAM running Mac OS X Server 10.3.9].

Availability: MapToGenome is written in C++ and is freely available to non-commercial users via web download at http://www.ambystoma.org/Tools/mapToGenome/.

Software Requirements: MapToGenome requires GNU make and a C++ compiler like g++ to compile the source. It uses UNIX sort command, gnuplot (http://www.gnuplot.info/) for plotting oxford grid in postscript. It also requires POSIX pipe support. This program was tested with g++ version 3.2, 3.6, gnuplot versions - 3.7, 3.8, 4.0 running on Linux (2.4.x, 2.6.x), Mac OS X versions - 10.2, 10.3, 10.4.

Results and Discussion

Within Species Comparisons

MapToGenome was used to process sequence alignments between mapped transcripts and genome assemblies of rat and zebrafish. Initially, the complete datasets were aligned using several maximum IL thresholds (ranging from 0 to 100 kb), and a SF threshold of 10 bp. In all cases, summing bitscores across presumptive introns yielded a higher proportion of coordinately aligned sequences (*i.e.* alignments that localize to the same chromosome between conspecific maps and genomes) relative to comparisons that do not account for intron structure (Figure 7.1). Furthermore, a majority of these localized to similar relative positions within chromosomes (Figure 7.2).

The initial comparison of the rat RH map to the rat genome yielded a high, yet lower-than-expected, proportion of coordinately-aligned sequences (88.2%). In light of this, the rat alignment data were reexamined, revealing that a relatively high proportion of misalignments involved marker sequences that were developed as simple sequence length polymorphisms (SSLPs). To examine the effect of these markers on the rat alignment, all SSLP markers were removed from the rat map and used MapToGenome to reprocess alignments. Exclusion of SSLP markers increased the proportion of coordinately aligning sequences from 88.2% to 93.6% (Figure 7.1). In some cases, SSLP "orthologies" formed vertical or horizontal lines in the Oxford plot (Figure 7.2). Vertical lines represent cases where a single marker identifies a large number of orthologies that are scattered across many chromosomes, and horizontal lines represent cases where particular genomic regions tend to attract presumptively non-orthologous alignments. The patterns exhibited by SSLP markers are not particularly surprising because their genomic distribution is often strongly biased toward non-coding and repetitive regions (Arcot *et al*, 1995; Nadir *et al*, 1996; Ramsay *et al*, 1999; Metzgar *et al*, 2000; Toth *et al*, 2000; but

see Morgante *et al*, 2002 for plant genomes). In general, including SSLP sequence alignments inhibited optimal alignment of the map and genome.

To examine further the nature of coordinately *vs.* non-coordinately aligned transcripts, the distributions of bitscore values were compared within both classes of alignments using the 20 kb intron length threshold (Figure 7.3). Alignment bitscores <2,000 were overrepresented among non-coordinate alignments relative to the coordinate class, the excess corresponding to ~48 of the 239 non-coordinate alignments. Imposing an alignment bitscore cutoff of 2,000 results in a modest increase in the proportion of coordinately aligning sequences from 93.6% to 95.0%. However, this increase is achieved at the expense of 657 coordinate alignments (17.7% of the total). Thus, even at a strict bitscore threshold, ~5% of the alignments were assigned to the wrong chromosome within the rat RH map or genome assembly. Several factors likely contribute to these "errors": 1) biological factors, such as recent gene duplicates and processed pseudogene insertions, 2) sequence and assembly errors within whole genome and transcript datasets, 3) mapping errors, and 4) failure of MapToGenome to identify "true" alignments between rat transcripts and their corresponding genome sequences. In practice, alignment "finishing" algorithms such as SPA (Nimwegen *et al*, 2006) could be used to improve intraspecific transcript/genome alignments that are generated by BLAT. However, SPA does not currently permit integration of linkage or RH mapping data, or tabular output formatting.

By comparison to the rat dataset, genetic maps for zebrafish show much less correspondence to the zebrafish genome assembly. At intron length thresholds of 10 and 20 kb, only 63.5% of sequence alignments mapped to coordinate chromosomes within the zebrafish linkage map and genome assembly. The proportion of coordinately aligned sequences was 57.6% when SSLP markers were included in the dataset and 64.3% when SSLP markers were excluded (Figure 7.1). As was observed for the rat dataset, the lower proportion of coordinately aligning sequences within the complete RH dataset suggests that inclusion of SSLP mapping data into genome alignments may inhibit optimal alignment of maps and assemblies. Notably, low proportions of coordinately aligning transcripts (~64%) were observed for comparisons between the zebrafish genome assembly and both zebrafish genetic maps. This replicated observation suggests that a

substantial proportion of zebrafish genes are not currently localized to the correct chromosome within the zebrafish (Zv6) genome assembly.

Inspection of the zebrafish Oxford plots reveals another conspicuous difference between the RH map and linkage group based comparisons. Specifically, the slopes of the lines that are created by coordinate alignments are less linear within the zebrafish linkage map/genome plot. This non-linear relationship between linkage distance and physical distance is expected because linkage distances are based on recombinational frequencies, rather than physical distance. Regions of the Oxford plot with steeper slopes presumably represent regions of the genome wherein recombinational distance changes at a relatively slower rate (*e.g.* suppression of recombination near the centromeres) and regions of the Oxford plot with gentler slopes presumably represent regions of the genome wherein recombinational distance changes at a relatively higher rate (*e.g.* near the telomeres) (Jensen-Seaman *et al*, 2004). It is, therefore, important to consider not only the relative position of orthologies within chromosomes, but also the relationship between map units and physical distance, when interpreting the Oxford plots that are output by MapToGenome.

Comparing Divergent Genomes

Although self/self comparisons are useful for evaluating genome maps and assemblies, comparisons between divergent genomes are more interesting in an evolutionary context. Furthermore, these comparisons provide a measure of the cross predictability of gene location information between species and a means of cross-referencing gene location information between highly developed whole genome assemblies and less developed genetic maps.

In order to evaluate between-species alignments, the association index λ was used as a measure of the conservation of broad sense synteny (*i.e.* the distribution of genes among chromosomes) between species. The λ index measures the extent to which knowing the location of genes in either of two species is predictive of their location in the other (Goodman and Kruskal, 1954; Housworth and Postlethwait, 2002; Smith and Voss, 2006). In probabilistic terms, the λ index measures "the relative decrease in probability of erroneous guessing" (Goodman and Kruskal, 1954) that is gained by knowing the

position of an ortholog in either of two species. For example, a λ value of 0.5 means that knowing the location of an ortholog in either of two species results in a 50% decrease in the probability of incorrectly guessing its location in the other species, or doubles the probability of correctly guessing its location. For this study, the λ index was primarily used to evaluate orthology calls based on different IL thresholds. Errors in assigning orthology should tend to decrease λ because they will cause an ortholog to be assigned to the wrong chromosome, and hence obscure associations that have been conserved through time. When the λ index is estimated using accurate orthology assignments it provides a measure of the extent to which the inter-chromosomal distribution of genes has been conserved between two species, or the conservation of broad-sense synteny. Because this measurement is interpretable in a probabilistic sense, λ values can be informatively compared between independent studies. The same is not true for similar chi-square based statistics (Fisher, 1938; Blalock, 1958; Goodman and Kruskal, 1954; Kendall and Stuart, 1967).

Vertebrates *vs.* chicken: Genetic maps for one representative mammal (rat), one representative amphibian (*Ambystoma*), and two fish (zebrafish and medaka) were aligned to a representative reptile (chicken). These comparisons span divergence times of ~310 MY (rat/chicken), ~370 MY (*Ambystoma*/chicken), and ~450 MY (fish/chicken) (Kumar and Hedges, 1998). Maps were aligned to the chicken genome using the same initial IL and SF thresholds that were used for self/self alignments. In general, gapped alignments yielded higher values of λ, although the maximum intron length threshold that maximized λ varied among alignments (Figures 7.4 and 7.5).

Estimation of λ can provide a measure of the cross-predictability of chromosomal assignments. However, smaller chromosomal segments, rather than entire chromosomes, are often the functional units of conservation. Conserved chromosomal segments can be identified as discreet clusters of points within oxford plots. Statistical methods exist for identifying chromosomal segments with highly conserved gene orders (Calabrese *et al*, 2003), although these methods have not been fully adapted for use in multi-chromosomal genomes (Smith and Voss, 2006). The degree to which the linear order of segments has been conserved can also be assessed by directly inspecting oxford plots (Figure 7.6). Highly conserved segments can be visualized as semi-linear clusters of points, whereas

segments within which linear order has been less conserved appear as unordered clusters. In comparison to the fish/chicken plots, orthologies appear to be more tightly clustered and somewhat more linear in rat/chicken and *Ambystoma*/chicken plots. This presumably reflects the greater divergence time spanned by the fish/chicken comparison, during which intrachromosomal rearrangements have accumulated.

Fish *vs.* fish: The whole genome sequence of *Tetraodon* was also compared to three fish maps: the medaka linkage map, the zebrafish linkage map, and the zebrafish RH map. Comparisons between fish linkage maps and the *Tetraodon* genome can permit better prediction of the chromosomal location of large numbers of *Tetraodon* gene orthologs within the genomes of less developed fish species. These comparisons can also provide insight into the structure of the ancestral teleost genome, which will lend a critical outgroup perspective for interpreting patterns of conservation and changes that are observed among amphibian and amniote vertebrates (***e.g.*** Smith and Voss, 2006). The zebrafish/*T. nigroviridis* comparisons span a large fraction of the history of the euteleostei (~110 to160 MY) whereas the medaka/*T. nigroviridis* comparison spans ~60 to 80 MY (Nelson, 1994; Wittbrodt *et al*, 2002). As was observed for the above comparisons, gapped alignments yielded larger λ values for comparisons between linkage maps and the *T. nigroviridis* genome (Figure 7.7). Values of λ that were estimated for the medaka/*T. nigroviridis* were substantially higher than those estimated for the zebrafish/*T. nigroviridis* comparison. Presumably this difference is due in part to the additional ~60 MY of shared ancestry between medaka and *T. nigroviridis*.

In a pattern similar to comparisons between the zebrafish RH map and genome, comparisons between the zebrafish RH map and the *T. nigroviridis* genome yielded substantially lower values of λ relative to comparisons that used the zebrafish linkage map (Figure 7.7). Removing SSLP markers from the RH dataset resulted in an overall increase in estimated λ values, yielding values that were similar to, though slightly higher than, those based on the zebrafish linkage map. Notably, several of the existing fish genetic maps consist largely of SSLP based markers, and these SSLP markers have served as the basis for numerous comparative genetic studies (***e.g.*** Woram *et al*, 2004; Danzmann *et al*, 2005; Lee *et al*, 2005; Gharbi *et al*, 2006; Stemshorn *et al*, 2006). It was generally observed that excluding SSLP markers results in higher cross-predictability (%

coordinate alignments and λ) indicates that caution should be exercised when interpreting comparisons that are based on SSLP marker data.

Oxford plots of fish/fish comparisons provide further insight into the evolution of fish genomes. The Oxford plot for medaka/*T. nigroviridis* reveals that the structure of these genomes has been highly conserved since these two species last shared a common ancestor (Figure 7.8). Many medaka/*T. nigroviridis* chromosomes show strong evidence of conserved synteny. Furthermore, many of these orthologs lie in nearly diagonal clusters, indicating higher order conservation of chromosome segments. Oxford plots for zebrafish/*T. nigroviridis* reveal several dense clusters of syntenies that are consistent with widespread conservation of synteny, although there appears to be very little conservation of the large-scale linear order of orthologs. Thus, the zebrafish/*T. nigroviridis* reveals a greater extent of interchromosomal and intrachromosomal rearrangement relative to the medaka/*T. nigroviridis* comparison, although the general chromosomal location of orthologs is apparently highly predictable on both species when compared to the *T. nigroviridis* assembly.

Summary

MapToGenome was developed as a flexible tool for aligning transcript maps and genome assemblies. The speed and flexibility of MapToGenome permits optimization of the parameters that are used to produce gapped alignments and assign orthologies. The above-described studies demonstrate that, in general, map/genome alignments can be improved by optimizing maximum intron length thresholds, and by ignoring alignments for SSLP marker sequences. Moreover, these studies illustrate the utility of MapToGenome in visualizing: 1) the correspondence of mapped genes to their presumptive genomic localizations, 2) differential scaling of recombinational and physical map distances, 3) conservation of synteny, and 4) disruption of genomes by chromosomal rearrangements and duplications.

An Additional Note

A new version of the zebrafish genome assembly has been released since this paper was submitted. This new assembly (ZF6.41, http://www.ensembl.org/Danio_rerio/index.html)

shows much stronger correspondence to the zebrafish genetic maps. When MapToGenome is used to compare zebrafish maps to the ZF6.41 assembly, the percentages of coordinately aligning transcripts increased from 63.5% to 88.9% for the linkage map and from 64.3% to 86.1% for the radiation hybrid map.

Figure 7.1: Plot of the percentage of coordinate alignments for within species comparisons of maps and whole genome assemblies. Values were calculated using several maximum intron length thresholds.

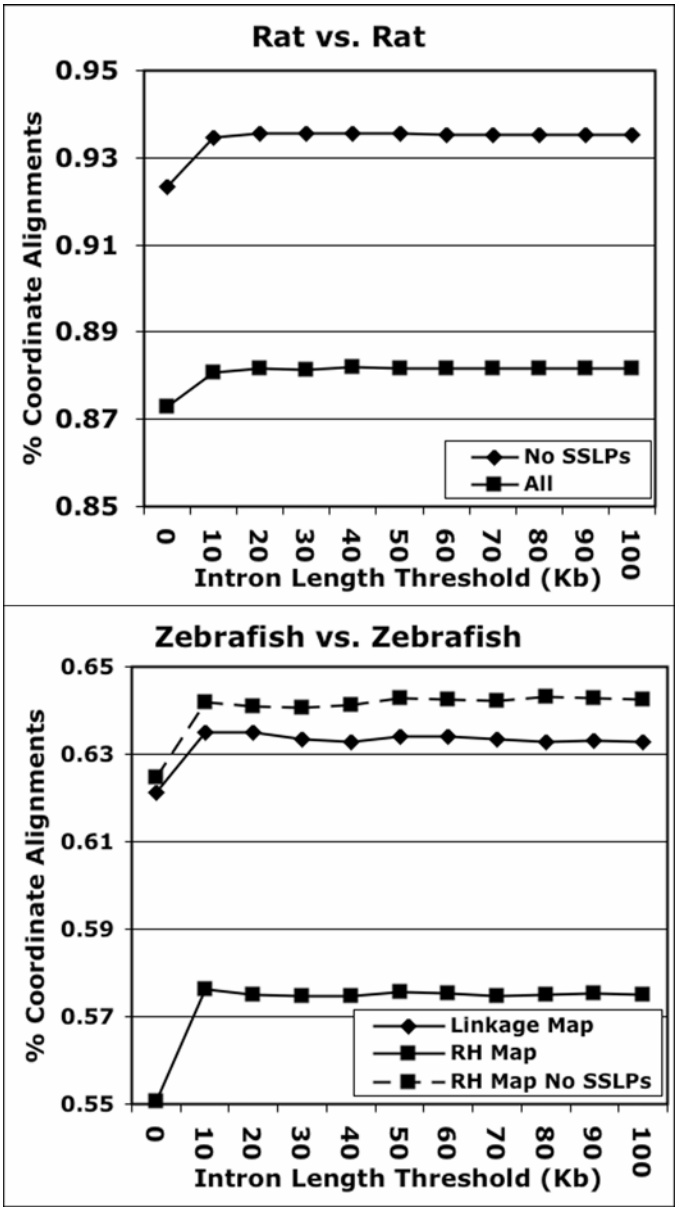


Figure 7.2: Oxford plots of alignments for within species comparisons of maps and whole genome assemblies. The rat plot was generated using a maximum intron length threshold of 20Kb. Zebrafish plots were generated using a maximum intron length threshold of 10Kb. Chromosomes are presented in order from left to right and bottom to top (for zebrafish: 1-25, for rat: 1-20 followed by X). Markers that are based on SSLPs are shown in red, all other markers are shown in black.

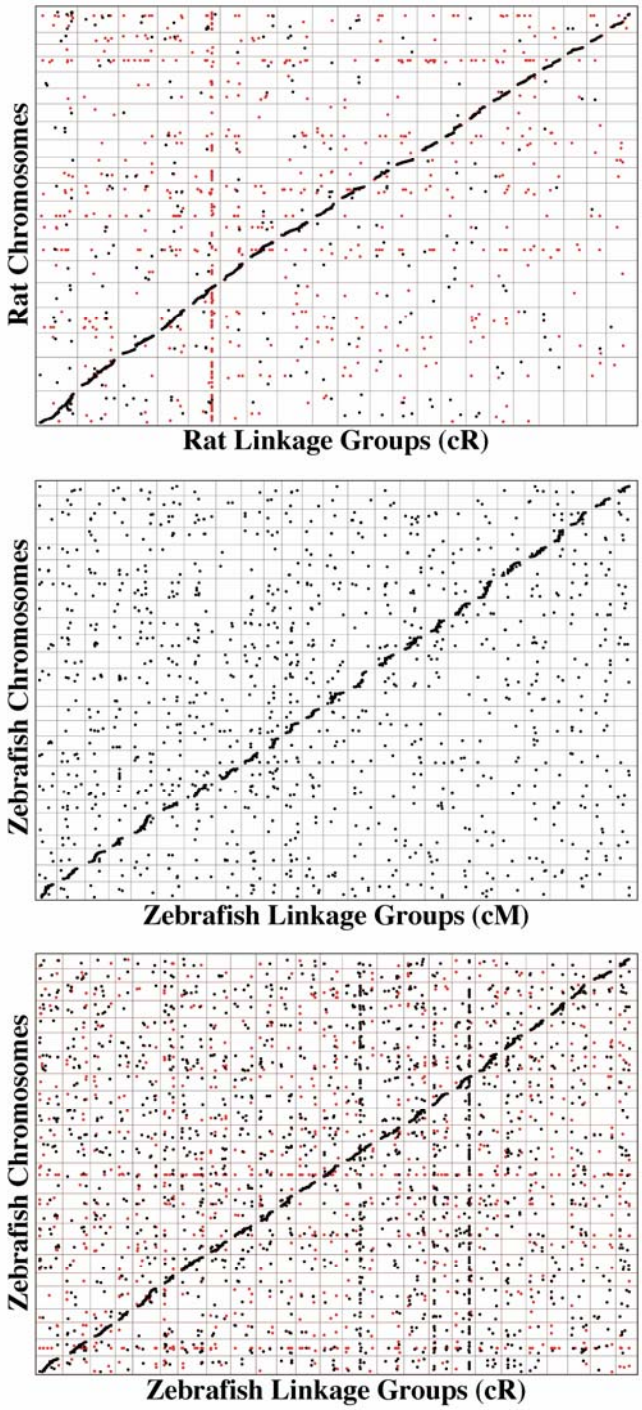


Figure 7.3: Distribution of bitscore values for coordinate and non-coordinate alignments. Values are based on alignment of the rat RH map and genome assembly, using a maximum intron length threshold of 20Kb.

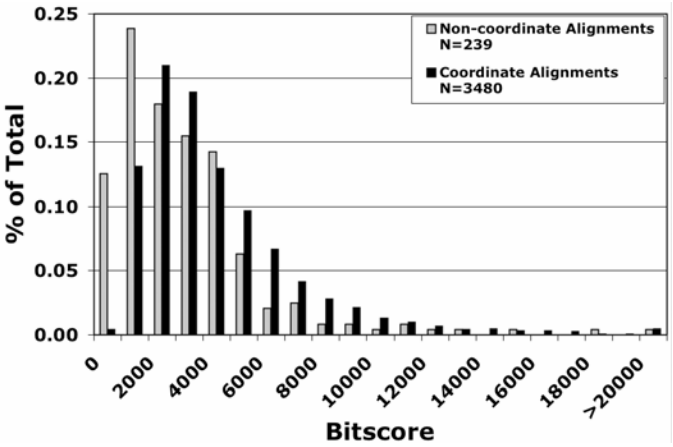


Figure 7.4: Plot of the λ values for comparisons between tetrapod (*Ambystoma*, and rat) genetic maps and the chicken genome assembly. Values were calculated using several maximum intron length thresholds.

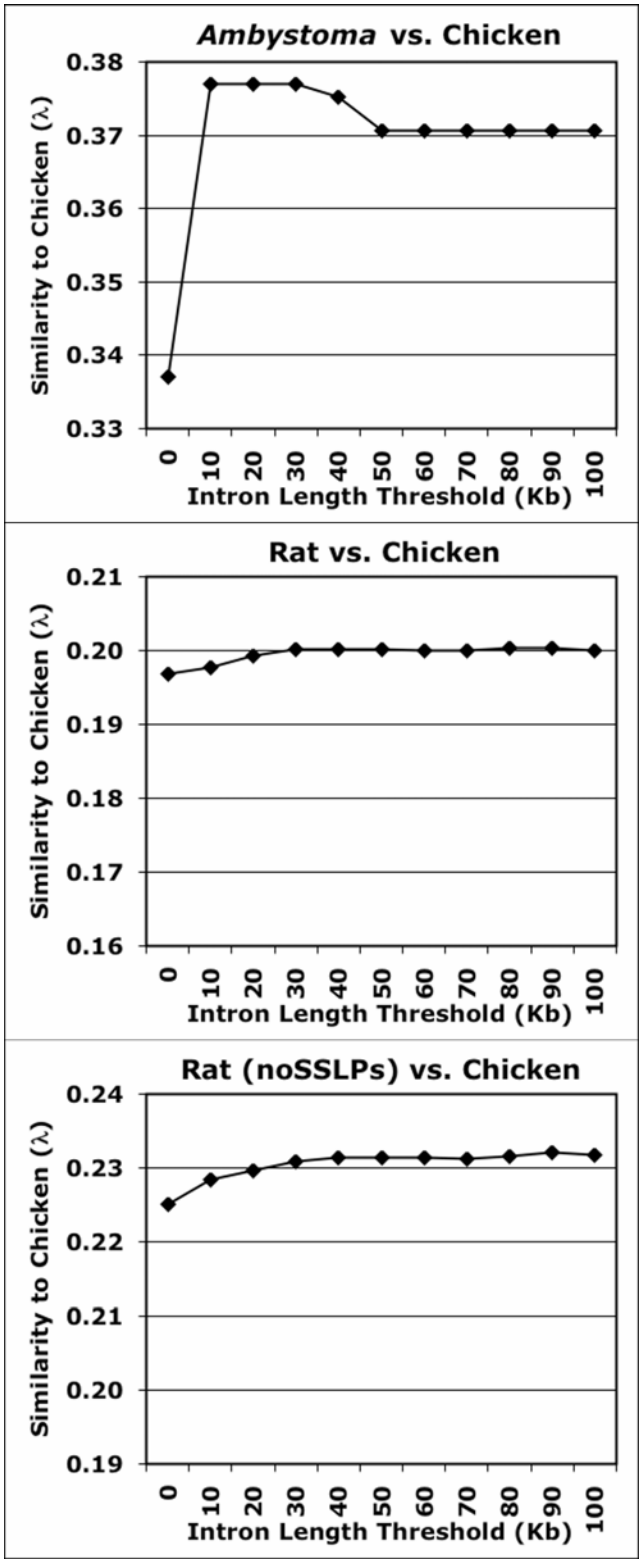


Figure 7.5: Plot of the λ values for comparisons between fish (medaka, and zebrafish) genetic maps and the chicken genome assembly. Values were calculated using several maximum intron length thresholds.

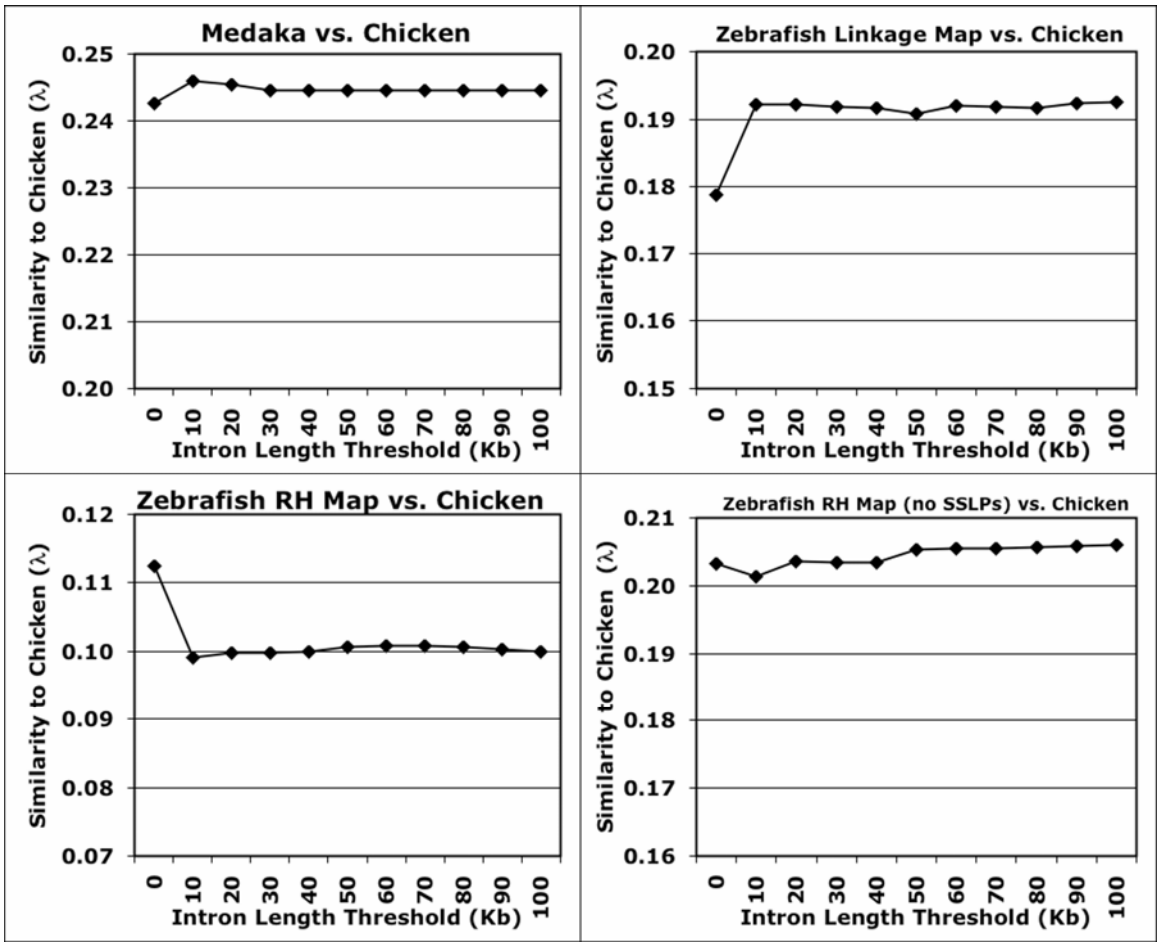


Figure 7.6: Oxford plots of vertebrate/chicken alignments. Maximum intron length thresholds of 10, 30, 90, and 100 Kb were used for the medaka, *Ambystoma*, rat, and zebrafish alignments, respectively. Chromosomes and linkage groups are presented in order from left to right (for zebrafish: 1-25; for rat: 1-20 followed by X; for *Ambystoma*: 1-14; for medaka: 1-24). Chicken chromosomes are presented in order from bottom to top, 1-32, followed by W and Z. Chicken chromosomes 25, and 29-31 are omitted, these chromosomes are not represented in the current chicken genome assembly (v2.1).

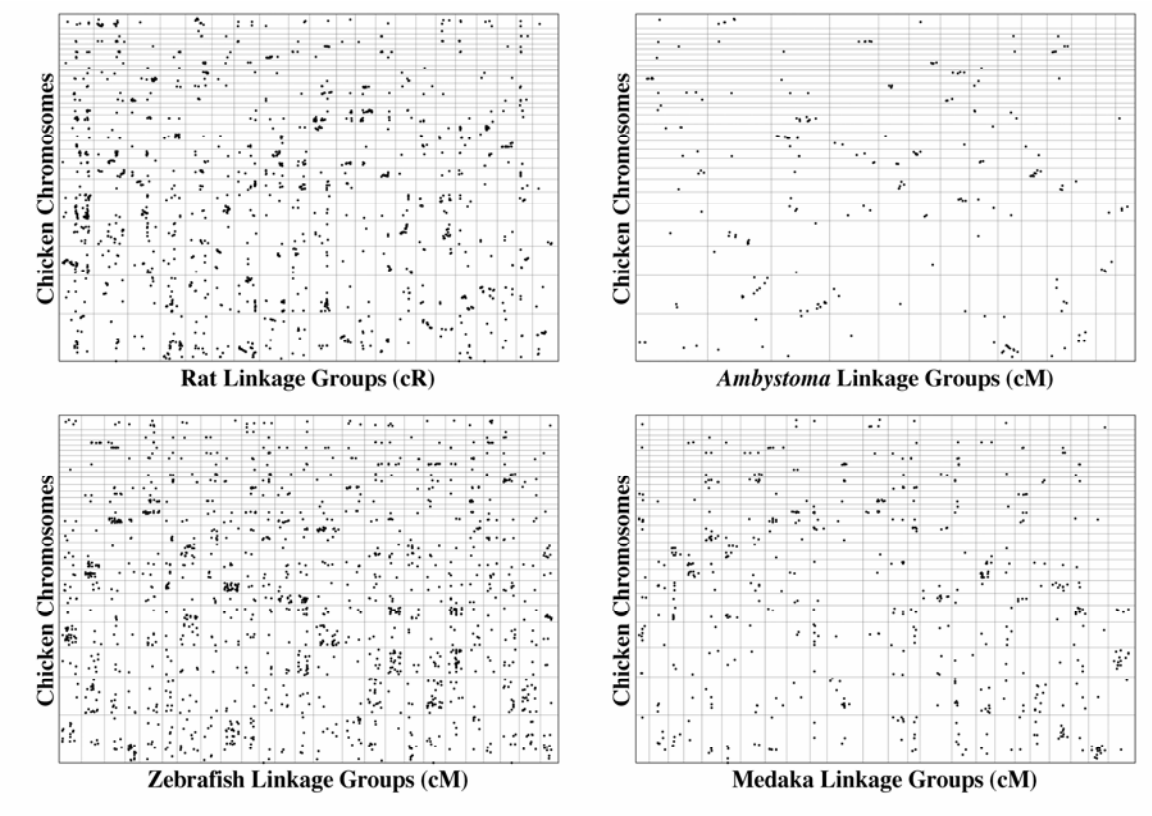


Figure 7.7: Plot of the λ values for comparisons between fish (medaka, and zebrafish) genetic maps and the *T. nigroviridis* genome assembly. Values were calculated using several maximum intron length thresholds.

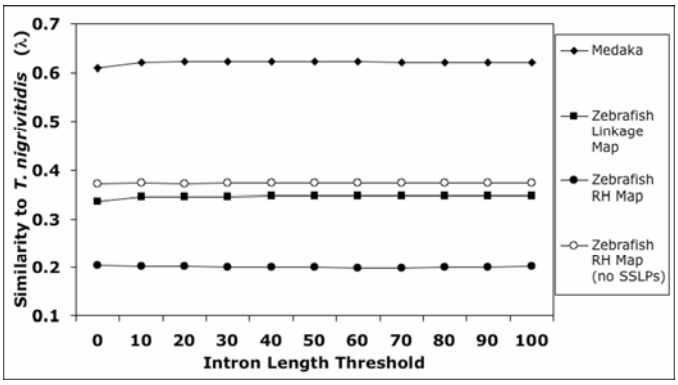
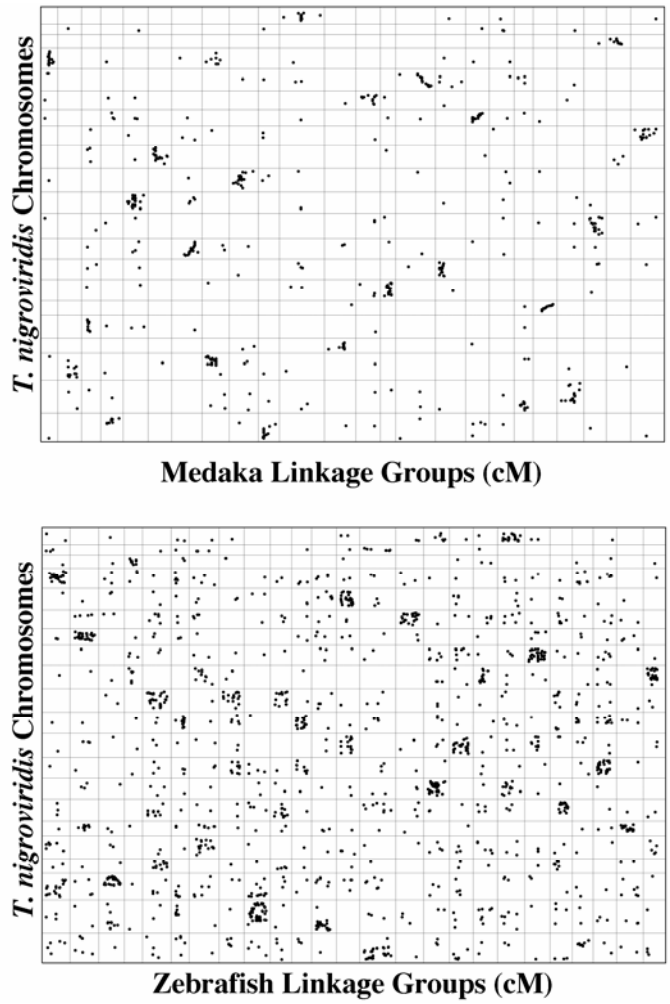


Figure 7.8: Oxford plots of Fish/ *T. nigroviridis* alignments. A maximum intron length thresholds of 60 Kb was used for both alignments. Chromosomes and linkage groups are presented in order from left to right (for zebrafish: 1-25; for medaka: 1-24). *T. nigroviridis* chromosomes are presented in order from bottom to top, 1-21.



CHAPTER 8: Bird and Mammal Sex Chromosome Orthologs Map to the Same Autosomal Region in a Salamander (*Ambystoma*)

Introduction

A classic problem in evolution concerns the origin of sex chromosomes among amniote vertebrates (Ohno, 1967). In mammals, females have two identical (XX) sex chromosomes while males have an X and Y (XY). In contrast, birds have a ZZ-ZW determination system wherein females are the heterogametic sex (ZW). The mammalian Y and chicken W chromosomes are conspicuously smaller than their X and Z counterparts and they contain fewer genes. Presumably, these sex chromosome homologs have undergone extreme, divergent evolution since their recruitment as sex-determining factors, a pattern observed broadly among animals and plants (Ohno, 1967; Bull, 1983; Lahn *et al*, 2001; Ayling and Griffin, 2002; Charlesworth and Charlesworth, 2005; Charlesworth *et al*, 2005; Khil and Camerini-Otero, 2005). Ohno (1967) proposed nearly four decades ago that bird and mammalian sex chromosomes are homologous. Recent comparative genomic analyses have observed that human chromosome (HSA) X contains many orthologs of chicken chromosome (GGA) 4 genes (Ross *et al*, 2005) and that GGAZ contains many orthologs of HSA9 genes, and fewer orthologs on HSA5 and 8 (Burt *et al*, 1999; Nanda *et al*, 2002; Fridolfsson *et al*, 1998). Because HSAX and GGAZ share few if any orthologs, these comparative data have been interpreted as strong evidence that the sex chromosomes of birds and mammals evolved independently, through separate recruitments of bird and mammalian sex chromosomes from independent ancestral autosomes (*e.g.* Fridolffson *et al*, 1998; Nanda *et al*, 1999; Ellegren, 2000; Nanda *et al*, 2000; Graves *et al*, 2002; Nanda *et al*, 2002; Kohn *et al*, 2004; Handley *et al*, 2004; Khil and Camerini-Otero, 2005; Kohn *et al*, 2006).

Comparisons between chicken and human are powerful for identifying features of the ancestral amniote genome that have been conserved in both lineages, but they provide no evolutionary insight about features that have changed within amniote lineages. To determine whether the precursors of GGAZ and HSAX were or were not linked ancestrally, it is necessary to consider the condition of these ancestral regions within an appropriate outgroup species (Stevens, 1980; Watrous and Wheeler, 1981; Maddison *et al*, 1994; Futuyma, 1986; Martin, 2001; Bourque *et al*, 2005). In general, the most

appropriate outgroup is the taxon that is most closely related to the last common ancestor of the clade but not included within the clade (the most proximate outgroup). In the case of amniote/amniote comparisons, amphibians represent the most proximate living outgroup (Figure 1). Until recently, there were few amphibian gene order data available for comparative analyses of vertebrate genome structure (Voss *et al*, 2001; Smith *et al*, 2004; Ohta *et al*, 2006). However, the recently developed genetic linkage map for the salamander genus *Ambystoma* provides a new outgroup perspective for reconstructing amniote genome evolution (Smith *et al*, 2005; Smith and Voss, 2006). The ambystomatid genome contains relatively few, large chromosomes that show extensive synteny conservation with chromosomes from fish and amniote genomes (Smith *et al*, 2005; Smith and Voss, 2006).

A few studies of amniote sex chromosome evolution have used teleost (ray-finned) fish to provide an outgroup perspective (Kohn *et al*, 2004; Kohn *et al*, 2006). The results of these studies have been interpreted as supporting the hypothesis of separate autosomal recruitments because amniote sex chromosome orthologues are observed to be distributed among several fish chromosomes. However, these studies have not explicitly tested for the presence or absence of the ancestral association of amniote sex chromosomes. Indeed, analyses across deep phylogenetic distances have rarely used statistical approaches to investigate the possibility of conserved syntenies (but see Danchin and Pontarotti, 2004). Moreover, it is generally accepted that the ancestor of most teleosts experienced a whole genome duplication, which was followed by massive losses of paralogous duplicates (Amores *et al*, 1998, Postlethwait *et al*, 1998; Jaillon *et al*, 2004; Woods *et al*, 2005). Such events, especially in combination with several hundred million years of independent evolution, would be expected to distribute ancestral syntenies among chromosomes.

Interestingly, a recent study of the sex-determining chromosomes of a monotreme (egg-laying) mammal seemingly lends support to the idea that Z and W chromosome loci may have been linked in the ancestral amniote genome. The deepest split within the mammalian lineage is between monotremes (platypus and echidna) and therians (all other mammals, *i.e.* marsupials and placental mammals) (van Rheede *et al*, 2006). The platypus X1 chromosome contains many genes from the mammalian X-conserved region

(XCR) (Graves, 1995; Ross *et al*, 2005) and is linked, via a meiotic translocation chain of five X and five Y chromosomes, to a chromosome that harbors the DMRT1 gene (Grutzner *et al*, 2004; Rens *et al*, 2004). The gene DMRT1 is located within the sex-determining region of the avian Z chromosome and is a primary candidate for the avian sex-determining gene, along with two W linked genes: ASW and FET1 (Smith and Sinclair, 2004). Currently, it is unclear whether the localization of Z and X orthologs to the platypus sex-determining chromosomes is representative of the condition in the ancestral amniote genome, or of rearrangements that were derived after the monotreme/therian divergence (Grutzner *et al*, 2004; Rens *et al*, 2004; Charlesworth and Charlesworth, 2005; Ezaz *et al*, 2006).

Here, the *Ambystoma* genetic map is used to provide an outgroup perspective on the origin of bird and mammalian sex chromosomes. Genes from the XCR and GGAZ map to adjacent regions of ALG2. Furthermore, the proportion of sex-chromosome orthologies observed on ALG2 is dramatically different from the proportion that would be expected by chance. Further comparisons between chicken and human genomes, and with the draft genome of the pufferfish *Tetraodon nigroviridis*, support the *Ambystoma* outgroup perspective and reveal further traces of this common ancestry.

Materials and Methods

Linkage Mapping and QTL Analysis

Linkage analyses were performed using previously described mapping panels AxTg (Voss, 1995) and WILD2 (Voss and Smith, 2005). Primers and probes for all genetic markers have been reported previously (Smith *et al*, 2005), except for 13 markers on ALG2. Primer sequences, diagnostic polymorphisms, and polymorphism detection assays for these 13 markers are summarized in Supplementary_Table_8_1. Linkage mapping and association analyses were performed using MapManagerQTXb21 (Meer *et al*, 2004).

Identification of orthologs

Presumptive orthologies were identified by aligning salamander, human RefSeq, chicken RefSeq, and *T. nigroviridis* transcripts to human, chicken, and *T. nigroviridis* genome assemblies. Similarity searches and sequence alignments were accomplished using the

program BLAT (Kent, 2002). Source sequences for human (IHGSC, 2001), chicken (IGSC, 2004), and *T. nigroviridis* (Jaillon *et al*, 2004) (hg17 build 35, galGal3, and tetNig1) genomes were downloaded from the UCSC Genome Browser Gateway (http://genome.ucsc.edu/cgi-bin/hgGateway). Cumulative bitscores were calculated for alignments between transcripts and full genome sequences by summing across presumptive exons. This was accomplished by summing bitscores for otherwise continuous alignments that were interrupted by gaps of 10,000 or fewer bases using the program MapToGenome (Putta *et al*, 2007). Positions of transcripts within conspecific genomes required an alignment with ≥98% nucleotide sequence identity. A genomic region was considered to be orthologous to a transcript if translated sequences yielded an alignment bitscore of ≥99% of the highest alignment bitscore for that transcript. Positions of orthologous loci were plotted using MapChart 2.1 (Voorrips, 2002).

Distribution of orthologies

Adjusted G statistics (Sokal and Rohlf, 1995) were used to test for non-random distribution of sex chromosome orthologs and orthologs from human and chicken chromosomes that share ancestry with sex chromosome from the other species (reciprocal amniote sex chromosome orthologs) on ALG2. First, the following question was asked: are the frequencies of X or Z orthologies on ALG2 significantly higher than expected by chance? The ALG2 loci (the ones that have orthologs) fall into two classes: sex chromosome orthologs and non-sex chromosome orthologs (Supplementary_Table_8_2). The observed numbers of orthologs on ALG2 from the human alignment were 5 from the XCR and 38 from other chromosomes. If ALG2 orthologs were randomly drawn from the human genome, the proportion of XCR orthologs on ALG2 would be expected to approximate the proportion within the human genome. This expected proportion (0.030) was based on the number of genes that are located between 60 Mb and 155 Mb of the X chromosome (N=810) relative to the total number of genes for the Human Genome Assembly build 36.2 (http://www.ncbi.nih.gov/mapview/stats/BuildStats.cgi?taxid=9606&build=36&ver=2) (N=28,617), excluding mitochondrial and HSAY genes. A similar statistic was also calculated based on the expected proportion (0.053) of GGAZ genes (N=840) relative to the total number of genes for the Chicken Genome Assembly

build 2.1 (http://www.ncbi.nih.gov/mapview/stats/BuildStats.cgi?taxid=9031&build=2&ver=1) (N=15,928), excluding mitochondrial and GGAW genes.

Based on extensive comparisons between chicken and human genomes (Fridolffson *et al*, 1998; Nanda *et al*, 1999; Ellegren, 2000; Nanda *et al*, 2000; Graves *et al*, 2002; Nanda *et al*, 2002; Kohn *et al*, 2004; Handley *et al*, 2004; Khil and Camerini-Otero, 2005; Kohn *et al*, 2006), a second question can be asked: Is the frequency of reciprocal amniote sex-chromosome orthologies on ALG2 significantly higher than expected by chance? Here, ALG2 loci fall into two classes: reciprocal sex chromosome orthologs and non-sex chromosome orthologs. The observed numbers based on the human set (HSAX, 9, 5, 8) are 26 ALG2/sex chromosome orthologs and 17 ALG2/non-sex chromosome orthologs. If ALG2 orthologs were randomly drawn from the human genome, the proportion of HSAX, 9, 5, 8 orthologs on ALG2 would be expected to approximate the proportion within the human genome. This expected proportion (0.165) was based on the number of genes on HSA X, 5, 8, and 9 (N=4,736) relative to the total number of genes for the Human Genome Assembly build 36.1 (N=28,617). Again, a similar statistic was also calculated based on the expected proportion (0.128) of GGAZ/4 genes (N=2,043) relative to the total number of genes for the Chicken Genome Assembly build 2.1 (N=15,928).

Adjusted G statistics (Sokal and Rohlf, 1995) were also used to test for non-random distribution of *T. nigroviridis* orthologs among human and chicken chromosomes (Supplementary_Tables_8_3-8_6). The frequency of orthologies that were identified on all amniote/*T. nigroviridis* chromosome pairs were tested for goodness of fit to the frequencies of all orthologs on each of the two chromosomes, relative to the grand total of orthologies that were identified. A similar adjusted G statistic was also calculated to test for non-random distribution of orthologs from the human RefSeq/Chicken genome comparison (925 reciprocal amniote sex-chromosome orthologies and 11,116 non-sex-chromosome orthologs; Supplementary_Table_8_7) on three ancestrally duplicated pairs of *T. nigroviridis* chromosomes (H, A, B; *sensu* Jaillon *et al*, 2004), relative to all other *T. nigroviridis* chromosomes.

In practice, some of the statistical analyses that are outlined above can be implemented as a G test or a permutation test (*e.g.* Fisher's Exact Test). Both tests are

appropriate for the question at hand and will generally give similar results (Sokal and Rohlf, 1995). G tests were selected for the analysis of ALG2 orthologies because they permitted incorporation of external data on gene frequencies from the human and chicken genome databases to better approximate sampling probabilities. For consistency, G statistics are reported throughout this chapter. It should be noted, however, that comparing G statistics, or Fisher's Exact Tests, among different contrasts does not provide insight into the relative strength of association among contrasts. That is, the statistics provide an estimation of the probability of obtaining a given distribution of orthologies under a random sampling scheme, but the values of these statistics are not directly interpretable in a probabilistic sense (Fisher, 1938; Goodman and Kruskal, 1954; Kendall and Stuart, 1967).

Results and Discussion

Twenty amniote sex chromosome orthologs were mapped to *Ambystoma* linkage groups. The majority of Z orthologs (85%) mapped to a single *Ambystoma* linkage group (ALG2) (Figure 2, Supplementary_Table_8_2). The frequency of Z orthologs on ALG2 was greater than would be expected by chance ($G_{ajdusted\ (adj)}$=6.2, P=0.013). Many orthologs from the XCR (42%) mapped to neighboring, but non-overlapping regions of ALG2 and no more than two X chromosome orthologs were identified on any other *Ambystoma* linkage group. The frequency of XCR orthologs on ALG2 was also greater than would be expected by chance (G_{ajd}=6.3, P=0.009). When all ALG2 genes were searched against the full genome assemblies for human and chicken, they were found to be distributed non-randomly among human chromosomes (G_{ajd}=42.6, P=6.9e^{-11}) and chicken chromosomes (G_{ajd}=32.9, P=9.7e^{-9}) that harbor reciprocal amniote sex chromosome loci. Although some chromosomal associations might be expected by chance given that the salamander genome is composed of relatively few, large chromosomes, statistical analysis shows that the distribution of sex-chromosome orthologies on ALG2 is not likely to have occurred by chance. This pattern of orthologies on ALG2 is consistent with the idea that the X-Y and Z-W chromosomal regions were linked on an ancestral chromosome prior to the divergence of the amphibian and amniote lineages. The pattern of orthologies on ALG2 also provides support for all other conserved syntenies that have been previously

identified between amniote sex chromosomes and autosomes (Fridolfsson *et al*, 1998; Burt *et al*, 1999; Nanda *et al*, 2002; ICGSC, 2004; Bourque *et al*, 2005; Ross *et al*, 2005). The ALG2 region defined by Z orthologs also includes orthologs from HSA 4, 5, 8, and 9 and the ALG2 region defined by the XCR included loci from GGA4. These comparisons show that many of the gene orders conserved between the sex chromosomes and autosomes of chickens and humans are interspersed but conserved on the same salamander chromosome. Thus, ALG2 apparently retains some of the gene content of an ancestral chromosome that gave rise to the X and Z sex chromosomes.

To test further the idea that the sex chromosomes of birds and mammals were derived from the same ancestral chromosome, the location of (N=14,922) human RefSeq orthologs were identified in the chicken genome and (N=3,120) chicken RefSeq orthologs were identified in the human genome (Supplementary_Tables_8_7 and 8_8). As has been reported previously (Fridolfsson, 1998; Burt *et al*, 1999; Nanda *et al*, 2002), we identified a large region of conserved synteny between GGAZ and HSA9 (N=211 loci)(Figure 3) and smaller, but very confined regions of conserved synteny between GGAZ and HSA8 (N=19 loci), and between GGAZ and HSA5 (N=227 loci). In the reciprocal comparison of human to chicken, a large region of synteny was identified between XCR and GGA4 (N=272). Thus, when only amniote sex chromosome loci are considered, there is seemingly no support for the idea of a common autosomal origin for mammalian and bird sex chromosomes, because conserved syntenies are not observed between X an Z loci. However, regions of synteny are observed between GGA4 and HSAX and Y, and between GGA4 and human autosomes (HSA 5 and 8) that show synteny with GGA Z (Figure 3). Thus, syntenies that are observed in comparisons of ALG2 and amniote genomes are also observed in comparisons between human and chicken genomes. This pattern supports the idea that amniote chromosomal regions from GGA Z, 4 and HSA 9, 4, X, 5, 8 were linked ancestrally.

To investigate a deeper outgroup perspective on the evolution of amniote sex chromosomes, the following question was posed: Do comparisons with *T. nigroviridis* provide evidence for deep conservation of HSAX/GGAZ/ALG2 syntenies? Although these comparisons span an additional 670 MY of independent evolution (Kumar and Hedges, 1998), and a teleost whole-genome duplication (Amores *et al*, 1998; Postlethwait

et al, 1998; Jaillon *et al*, 2004; Woods *et al*, 2005), statistically significant, non-random distributions of reciprocal amniote sex chromosome orthologies can be observed among three pairs of *T. nigroviridis* chromosomes (Figure 4, Supplementary_Tables_8_3 and 8_4). These chromosome pairs correspond to the proposed duplicates of ancestral teleost chromosomes A, B, and H (Jaillon *et al*, 2004). To address more directly the common ancestry of sex-chromosome orthologs among these three duplicate chromosome pairs, sex-chromosome orthologs that were identified in comparisons between human RefSeq and the chicken genome were more closely examined (Supplementary_Table-8_7). These sex-chromosome orthologs correspond to 1194 human RefSeq genes that fall on HSA 5, 8, 9, or X, and identified an ortholog on GGA 4 or Z. Of these genes, 925 identify orthologs in the *T. nigroviridis* genome, and the overwhelming majority of these (n=724) fall on the three chromosome pairs (H, A, B) that contain a significant excess of reciprocal amniote sex chromosome loci (Figure 4). Of 11,116 non-sex chromosome/*T. nigroviridis* orthologies, 2,906 fall on H, A, B chromosome pairs and 8,210 fall on other *T. nigroviridis* chromosomes. The distribution of sex-chromosome *vs.* non-sex chromosome orthologs on *T. nigroviridis* H, A, B, chromosome pairs is very unlikely to have occurred by chance (G_{ajd} = 959.8, P = $9.8e^{-211}$). Thus, a large fraction the orthologous loci that are associated in the *T. nigroviridis* duplicates of A, B, and H are from the same amniote chromosomal regions that support ancestral XCR/GGAZ linkage in the salamander genome. Of particular importance is the observation that GGAZ and GGA4 orthologs are linked in both *T. nigroviridis* and *Ambystoma* genomes (Figure 2, 4). Because conserved syntenies also reveal common ancestry of GGA4 and XCR chromosomal regions, the most parsimonious interpretation of *T. nigroviridis* comparative mapping data is ancestral linkage of Z and X chromosomal regions in the tetrapod and amniote lineages. The distribution of ancestral amniote sex chromosome regions among different *T. nigroviridis* chromosome also reveals the confounding effects of genome duplications and rearrangements that have occurred within the lineage that gave rise to *T. nigroviridis* subsequent to the diversification of the bony vertebrate (euteleost) ancestor.

These comparative analyses provide the first amphibian outgroup perspective on the evolution of amniote sex chromosomes and are most parsimoniously interpreted as

evidence for ancestral linkage of XCR and GGAZ regions. However, it should be recognized that the most parsimonious evolutionary scenario may not always be correct. It is possible (though presumably less probable) that these associations are derived from rearrangements that occurred independently within the amniote, amphibian, and fish lineages. Resolution of this issue will necessitate additional comparative data from amphibian outgroups and, to some extent, perspective from the preduplicated fish genome. Ongoing progress toward improving linkage (Smith *et al*, 2005) and physical (http://genome.jgi-psf.org/Xentr4/Xentr4.home.html) maps for representative amphibian species will likely result in more accurate reconstructions of the ancient events that have structured vertebrate genomes. It is noted, however, that even with the modest number of *Ambystoma*/amniote orthologies that are currently mapped, statistical analyses strongly support conserved synteny of ALG2/amniote sex-chromosome orthologs. It is therefore anticipated that additional mapping studies will tend to support the primary findings of this study.

This study does not resolve the question of what type of sex-determining system was present in the ancestral amniote lineage and how modifications on this ancestral system gave raise to the mammalian XY and avian ZW sex determining systems. The genetic changes that can initiate the evolution of chromosomal sex determining mechanisms are best understood in cases where a genetic sex determining mechanism evolves from an ancestor with genetically identical sexes (***e.g.*** temperature dependant sex determination) (Bull, 1993; Janzen and Krenz, 2004; Charlesworth *et al*, 2005). However, the distribution of ZW and XY systems within fish and amphibian phylogenies suggests that direct transitions between ZW and XY systems may be a relatively common occurrence (reviewed by Ezaz *et al*, 2006). The unusual sex chromosomes of platypus have been interpreted as evidence supporting the idea that the mammalian XY system evolved from something very similar to the avian ZW (Ezaz *et al*, 2006, Waters *et al*, 2007). Here again, information from appropriately positioned outgroups (amphibians) and a whole-genome perspective will be critical for testing alternate hypotheses: that platypus sex chromosomes represent an ancestral state ***vs.*** a state that was derived independently within the monotreme lineage (Grutzner *et al*, 2004; Rens *et al*, 2004; Charlesworth and Charlesworth, 2005; Ezaz *et al*, 2006).

In conclusion, gene-mapping data show that amphibian orthologs for loci on chicken and human sex chromosomes are linked in the *Ambystoma* genome. This pattern of linkage, which is revealed by including an amphibian outgroup perspective, is interpreted as a signature of shared ancestry between avian and mammalian sex chromosomes. This signature is likely to have been retained as a vestige for two reasons: (1) In comparison to amniotes with chromosomal sex determination, our mapping results show that sex in *Ambystoma* is determined by a single locus on a chromosome with autosomal characteristics (Smith and Voss, unpublished data). Gene orders on ALG2 have not been disrupted by amniote specific rearrangements or mechanisms that are associated with the divergence of dimorphic sex chromosomes (Ohno, 1967; Bull, 1983; Lahn *et al*, 2001; Ayling and Griffin, 2002; Charlesworth and Charlesworth, 2005; Charlesworth *et al*, 2005; Khil and Camerini-Otero, 2005). (2) In comparison to amniote genome evolution, *Ambystoma* has experienced relatively lower rates of genome rearrangement and fission (Smith and Voss, 2006). The signature of shared sex chromosome ancestry is difficult to see when only comparing bird and mammalian genomes because mutational processes have fractured and re-arranged gene orders within these groups, especially in the mammalian lineage (Burt *et al*, 1999; Bourque *et al*, 2005; Smith and Voss, 2006). This study shows that clarity in comparative vertebrate genomics can be greatly increased by including relevant and phylogenetically well positioned outgroups like *Ambystoma*.

Figure 8.1: An abridged phylogeny of the major groups of bony vertebrates. Divergence times were obtained from the literature (Kumar and Hedges, 1998; Rutta *et al*, 2003). Birds represent an ancient reptile lineage that diverged from other reptilian groups approximately 220 million years ago (MYA) (Kumar and Hedges, 1998)

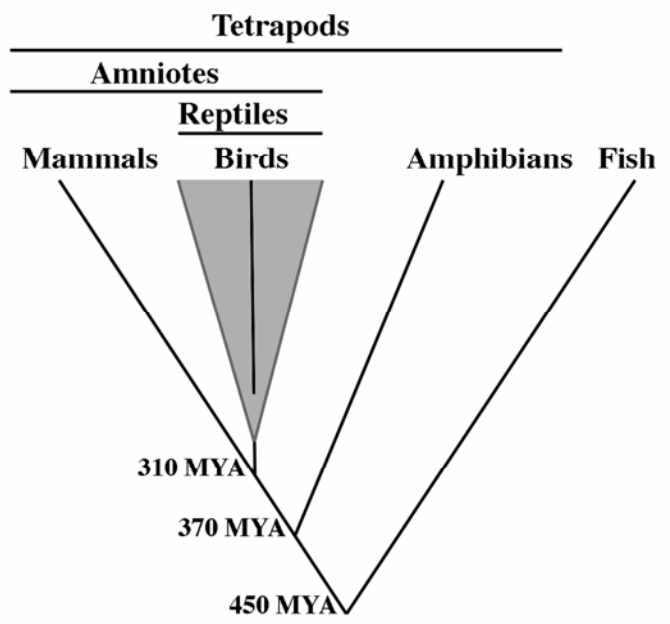


Figure 8.2 : A comparative map of *Ambystoma* LG2 and syntenic chromosomes from human and chicken. Vertical bars within ALG2 represent the position of *Ambystoma* transcripts that yielded an alignment in either the human (HS) or chicken (GG) genome. ALG2 loci that do not correspond to orthologs on GGA Z, 4 or HSA 4, 5, 8, 9, X are highlighted in green. Red lines connect human/*Ambystoma* orthologies and blue lines connect chicken/*Ambystoma* orthologies. The sex chromosomes of human (X) and chicken (Z) are both syntenic with *Ambystoma* LG2.

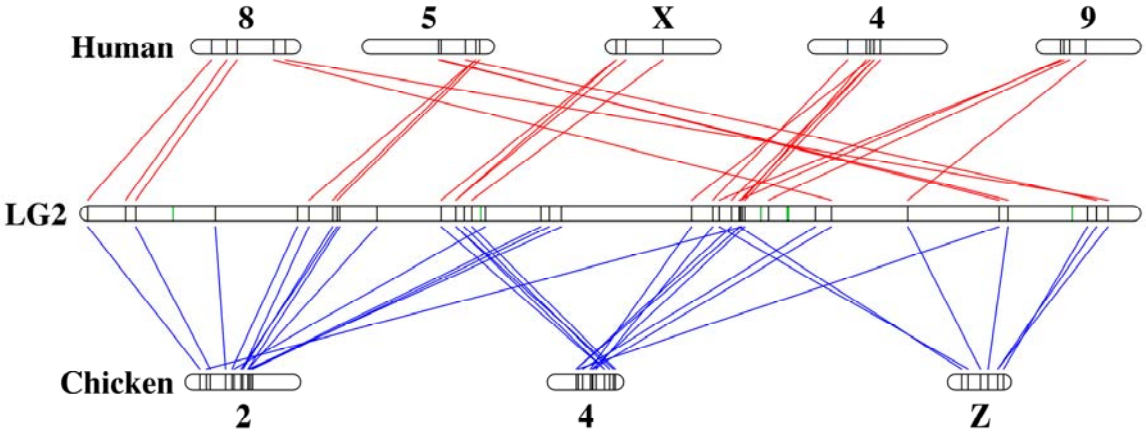


Figure 8.3: A comparative map of the human and chicken chromosomes that are syntenic with *Ambystoma* LG2. Vertical bars within chromosomes represent the position of mapped human RefSeq transcripts or chicken orthologs. Lines connect the positions of human/chicken orthologies. Bars above human chromosomes show the proportions of chicken chromosome orthologs that were identified within a given segment, and bars below chicken chromosomes show the proportions of human chromosome orthologs that were identified within a given segment. Note that the chicken chromosomes Z and 4 map to adjacent and overlapping regions of HSA8 and HSA5.

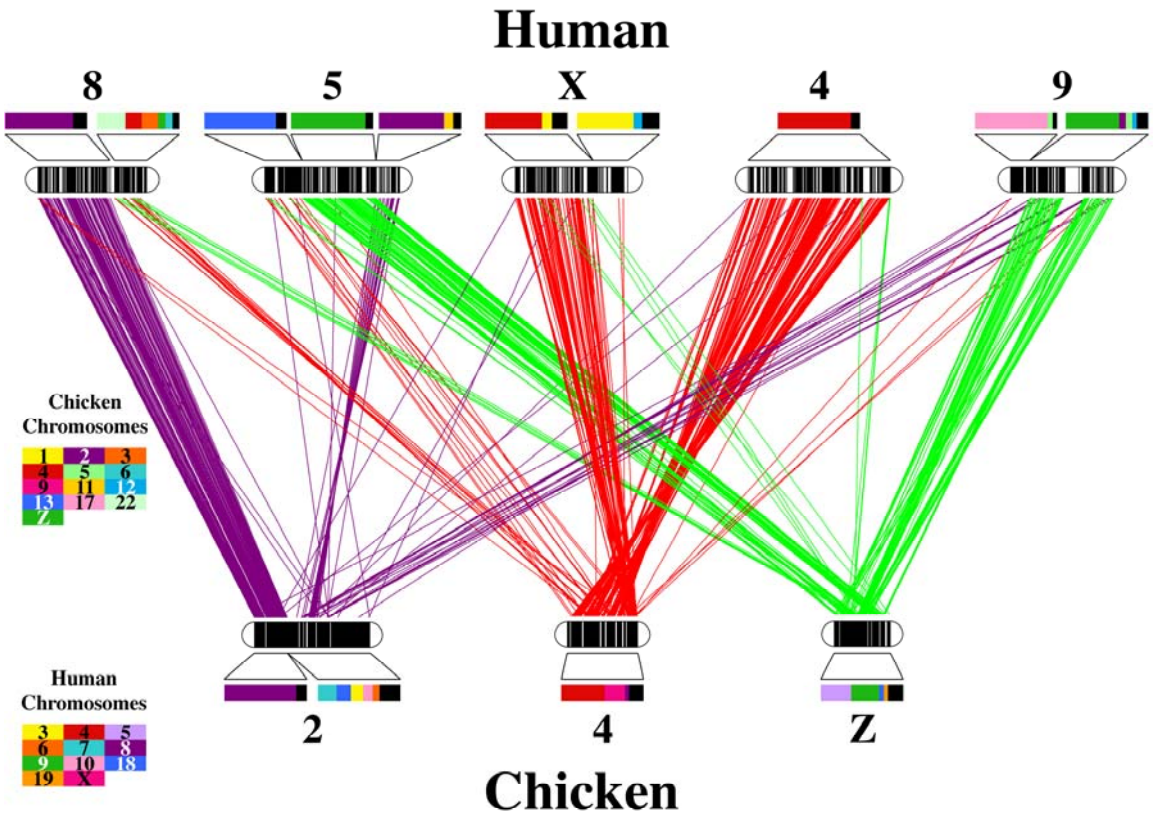
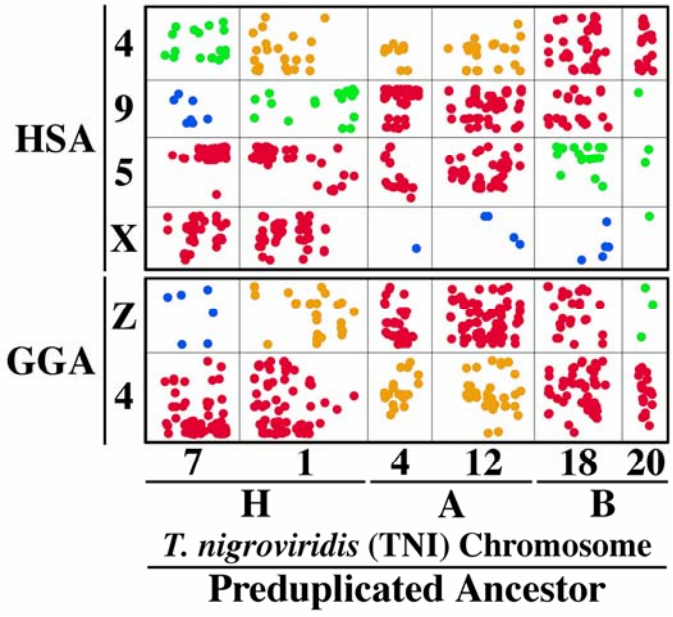


Figure 8.4: Oxford plot of the position of amniote sex chromosome loci in the *T. nigroviridis* genome. The Y-axis represents the relative position of orthologs on human (HSA) and chicken (GGA) chromosomes. The X-axis represents the relative position of orthologs on *T. nigroviridis* chromosomes. The labels "H", "A", and "B" correspond to preduplicated ancestral chromosomes (Jaillon *et al*, 2005). Cells containing an excess of orthologies are highlighted in red ($p < 0.005$) and yellow. Cells containing a deficiency of orthologies are highlighted in blue ($p < 0.005$) and green.



CHAPTER 9: Amphibian Sex Determination: Segregation and Linkage Analysis Using Members of the Tiger Salamander Species Complex (*Ambystoma mexicanum* and *A. t. tigrinum*).

Introduction

Distinct sexual phenotypes (*e.g.* male and female) are one of the most conserved forms of segregating phenotypic variation among vertebrates, and possibly all eukaryotes. Although mechanisms underlying vertebrate sex determination remain largely unknown, the available evidence suggests incredible diversity among and within each of the major groups. Sex determination in fishes ranges from Mendelian to polygenic, but in some cases, sex is entirely determined by environmental factors (Baroiller *et al*, 1999; Delvin and Nagahama, 2002; Mank *et al*, 2006). Environmental sex determination and sex chromosomes are distributed throughout reptilian phylogeny, suggesting independent evolutionary losses and gains, but conserved mechanisms are known for crocodilians (temperature) and birds (chromosomal sex determination) (Olmo, 1990; Ewert and Nelson, 1991; Pieau *et al*, 1999; Shine, 1999; Nada *et al*, 2000; Nada and Schmid, 2002). Of all the major vertebrate groups, sex determination is perhaps the most conserved within the order Mammalia, however even within this group there are a few interesting exceptions to the SRY gene rule (Sinclair *et al*, 1990; Just *et al*, 1995; Soullier *et al*, 1998; Pask and Graves, 1999). Overall, amazing diversity in sex determination is observed among vertebrates. It is expected that studies of additional groups will not only reveal new diversity, but will also elucidate conservative aspects of sex-determining and sex-differentiating pathways that presumably characterized the ancestral vertebrate condition.

The vast majority of, if not all, amphibian species are thought to exhibit genetic modes of sex determination, however several evolutionary transitions between ZW and XY type sex-determining mechanisms (heterogametic females or males respectively) may have occurred (Hillis and Green, 1990; Schmid *et al*, 1991; Hayes, 1998; Wallace *et al*, 1999). Under chromosomal sex determination, sexual differentiation depends upon the inheritance of a homologous (ZZ or XX) or heterologous (ZW or XY) pair of sex chromosomes. Presumably alternate modes of sexual differentiation are determined by a 'genetic switch' that is manifested either by presence or absence of one or more loci

between sex chromosomes, or by gene dosage. In either case, alternate sexual phenotypes (male *vs.* female) are expected to segregate as simple Mendelian traits, with a resulting 1:1 sex ratio among siblings of a large cross. While results consistent with a simple Mendelian mode of inheritance have been observed in the few amphibian species examined to date, non-Mendelian segregation of sexual phenotypes has also been observed (*Pleurodeles waltl*, Collenot *et al*, 1994; *Rana* and *Hyla spp,* Richards and Nance, 1978). Moreover, several studies (involving at least 4 urodele and 5 anuran species) have demonstrated that exposure to extreme temperatures during larval development can skew the sex ratio of amphibian clutches, yielding extremely female or male-biased sex ratios (reviewed by Wallace *et al*, 1999; Wallace and Wallace, 2000; Sakata et al, 2005; Kuntz et al, 2006). These studies suggest that sex determination in some amphibians may depend upon more than a single, segregating genetic factor.

Salamanders from the genus *Ambystoma* have served as important amphibian models for studying development, ecology, and evolution, and with the recent development of a complete genetic map for *Ambystoma* the prospects of understanding the genetic basis of biologically important trait variation has become a reality. A number of developmental and cytogenetic experiments have established that sex is specified by a ZW type mechanism of chromosomal sex determination in the species *Ambystoma mexicanum* (Humphrey, 1945; Lindsley *et al*, 1956; Humphrey, 1957; Brunst and Hacshuka, 1963; Armstrong, 1985; Cuny and Malacinski, 1985), and possibly other Ambystomatid species, *e.g. A. tigrinum tigrinum* (Cuny and Malacinski, 1985) and the *A. jeffersonianum* species complex (Sessions, 1982). In addition, sex ratios from laboratory crosses of (Humphrey, 1945; Lindsley *et al*., 1956; Humphrey, 1957; Armstrong, 1985) suggest a single gene basis for sex determination in the laboratory strain of *A. mexicanum*.

In this study I test the hypothesis that sex segregates in a 1:1 ratio in crosses derived from the tiger salamander species complex (*sensu* Shaffer 1984) and report the identification of a single genomic region that shows nearly complete association with sex-determination. Segregation ratios of males *vs.* females were largely consistent with the presence of a single sex-determining factor in the previously described WILD2 (Voss and Smith, 2005) backcross mapping family (wild collected *A. mexicanum*/*A. t. tigrinum*

X wild collected *A. mexicanum*) as well as a new "wild" *A. mexicanum* X "lab" *A. mexicanum* F2 intercross (MEX1). Interestingly though, a small but consistent female bias was observed among WILD2 crosses. To localize the major sex-determining factor to the *Ambystoma* genetic map, a subset of the WILD2 cross (Smith and Voss 2005) was genotyped 156 previously developed markers (Smith *et al*, 2005). Analysis of association between segregating genotypes and sex-phenotypes identified a single genomic region, linked to the marker *E24C3* (LG 5), which was strongly associated with segregation of sex. Two new probes were developed to track segregation of *E24C3* alleles within the MEX1 cross. As was observed for the WILD2 cross, segregating genotypes for *E24C3* were strongly associated with segregating sex phenotypes in MEX1. These studies validate the existence of a single quantitative trait locus (QTL) (*ambysex)* that acts as the primary sex-determining factor in *A. mexicanum*. Patterns of segregating sex phenotypes and marker genotypes within the WILD2 cross indicate that other environmental or genetic factors may have influenced segregation of sex phenotypes. Finally, the patterns of recombination that were observed among *Ambystoma* mapping panels is consistent with the idea that Z and W sex chromosomes of *A. mexicanum* have not differentiated to any great extent, implying a recent evolutionary origin for this sex determining system.

Materials and Methods

Genetic Strains and Crosses

Two crosses were used in this study: WILD2 and MEX1. The cross WILD2 was generated by backcrossing F1 hybrids between *A. mexicanum* and *A. t. tigrinum* to *A. mexicanum*. See Voss and Smith (2005) for a detailed description of the crossing design and rearing conditions that were used to generate WILD2. Two strains of *A. mexicanum* were used to generate the MEX1 cross. A female from the laboratory strain of *A. mexicanum* that is maintained by the *Ambystoma* Genetic Stock Center (http://www.ambystoma.org/AGSC/) was crossed to a strain that was more recently derived from the single natural population of *A. mexicanum* that occurs in Lake Xochimilco, Mexico D.F., Mexico. Two of the resulting F1 offspring were then mated to generate the MEX1 cross.

Rearing conditions

At approximately 20 days post-fertilization, larvae were released from their eggs and placed individually in 5 oz paper cups of artificial pond water. Throughout the course of these experiments all individuals from each of the mapping panels were maintained in a single room within which the temperature fluctuated from 19-22º. Individuals were reared in separate containers and rotated within the room after water changes to reduce effects of spatial temperature variation. Larvae were fed freshly hatched *Artemia* twice daily for their first 30 days post-hatching. After day 20 their diet was supplemented with small (<1cm long) California black worms (*Lumbriculus).* During this time, individuals were provided with fresh water and cups after every third feeding. On day 30 larvae were transferred to 16 oz plastic bowls, after which they were fed exclusively California black worms and water was changed every third day. Finally, at 80 days post-fertilization, all individuals were transferred to 4 L plastic containers and were otherwise maintained under the same regime as the previous 50 days**.**

Phenotypic Scores

WILD2: The majority of WILD2 offspring were euthanized upon completion of metamorphosis or at day 350. At this time, individuals were dissected, tissue samples (liver and/or blood) were harvested for DNA isolation, and gonads were examined in order to identify each individual's sex phenotype. Individuals with gonads consisting of a membrane surrounding translucent (immature ova), or opaque/pigmented (more mature ova) spheres were classified as females. Individuals with gonads appearing as opaque-ovoid (immature testes) or lobed (more mature testes) structures were classified as males. Individuals metamorphosing early in the experiment often could not be unequivocally assigned to either sex; these individuals were classified as immatures. Gonads of immatures appeared as a thin strip of tissue (undifferentiated gonadal primordia or early stages of differentiation) adjacent to the abdominal fat bodies. Individuals at early stages of gonadal differentiation cannot be unequivocally assigned to either sex (Humphrey, 1929; Gilbert, 1936) and are excluded from subsequent analyses. A few individuals were not euthanized and are currently being maintained for use in future studies. For these individuals, sex was scored after the development of secondary sexual characteristics. In

particular, the cloacal opening of male ambystomatids is much longer than that of females and male cloacal lips are several times larger than those of females. Dissected animals invariably corroborated cloacal sex scores. Sex was identified for a total of 360 WILD2 individuals within this study.

MEX1: Sex phenotypes were scored in the same manner as described for the WILD2 cross in the MEX1 cross, except that all animals from this cross remained paedomorphic and all individuals were scored at day 310 post fertilization. At this age, all animals had completed gonadal differentiation and sex phenotypes were unequivocally identified for all 91 individuals.

Genotyping

Based on previous studies, which suggested that the sex-determining gene of ambystomatids resides on one of the smaller chromosomes (*A. mexicanum* and *A. t. tigrinum* – Brunst and Haschuka, 1963; Cuny and Malacinski, 1985; *A. jeffersonianum* species complex – Sessions, 1982), an initial screen for the sex-determining factor was targeted to the five smallest linkage groups, LG10 – LG14 (Smith *et al*, 2005). In this screen, 54-161 individuals of known sex phenotype (>195 days old) were genotyped for 23 molecular markers that were broadly scattered across the five smallest linkage groups (Supplementary_Tables_9_1 and 9_2). A second screen targeted the remaining fraction of the map. This screen used an initial panel of 24 paedomorphs and late-metamorphosing individuals (12 males and 12 females) from WILD2.3 and targeted 124 previously described genetic markers (Smith *et al*, 2005) that were spaced at roughly 30 cM intervals (Supplementary_Tables_9_1 and 9_2). Based on analyses of genotypes from the initial screens, *E24C3* and *E26C11* were also genotyped for a larger subset of the WILD2 panel (N=117 for *E24C3*; N = 74 for *E26C11*).

Three novel genetic markers were developed for this study, using previously described methods (Smith *et al*, 2005). One of these probes (E26C11_329_3.1_A/G – GGGTTCTCTCAATGAACTGTATGTTGATTG) was designed to genotype the marker *E26C11* in WILD2. Two additional fluorescence polarization probes were developed to genotype the *E24C3* marker fragment in the MEX1 cross (E24C3_79_3.1_G/A – CTGTGGTGTATTCGAACATGTCGC, and E24C3_318_3.1_G/A –

AGGGCCTTCACATATTTTTCTGCAAAATAT). These markers were genotyped using standard primer extension and fluorescence polarization protocols (Perkin Elmer, AcycloPrime -FP chemistry and Wallac, Victor3 plate reader) (*e.g.* Chen *et al*, 1999; Hsu *et al*, 2001; Gardiner and Jack, 2002).

Statistical Analyses: Segregation of sex

Sex ratios within crosses were examined for fit to a 1:1 ratio expected under single locus sex determination, using G tests (Sokal and Rohlf, 1995). Because WILD2 consisted of several crosses, tests for homogeneity ($G_{\text{heterogeneity (h)}}$) among crosses and fit to a 1:1 sex ratio in the data pooled across crosses ($G_{\text{pooled (p)}}$), were performed in addition to tests for fit to a 1:1 sex ratio for individual crosses (G) (Sokal and Rohlf, 1995). Lack of fit to a 1:1 sex ratio would suggest that genetic factors, in addition to a single locus, might contribute to sex determination in *Ambystoma*.

Association analyses

Likelihood ratio statistics (LRS) for association of phenotypic variation with genotypic inheritance were estimated using the interval mapping and marker regression functions of MapManagerQTXb19 (Meer *et al*, 2004). Significance thresholds for interval mapping were obtained through 10,000 permutations of trait values among backcross progeny.

Results

Segregation of Sex: WILD2

The sex of 360 WILD2 individuals from was identified by gonadal and/or cloacal morphology. All remaining larvae (N = 137) were scored immature. The ratios of males to females from 8 of 9 crosses were consistent with a 1:1 ratio expected under the hypothesis that a single locus specifies sex in *Ambystoma* (Table 1). A significantly female-biased sex ratio was observed in Cross 1 (G = 8.26, DF = 1, P = 0.004) and non-significant female bias was observed in 5 other crosses. Sex ratios among crosses were not significantly heterogenous (G_h= 8.06, DF = 8, P = 0.427); therefore, data for sex were pooled and tested for fit to 1:1. Pooling of female-biased sex ratios among individual crosses revealed an overall female-biased sex ratio (203 females and 157 males) that

deviated significantly from 1:1 (G_p= 8.06, DF = 1, P = 0.015, N = 360).

MEX1: Sex phenotypes were obtained for a total of 93 individuals from the MEX1 cross: 46 males and 47 females. As was observed for the majority of WILD2 crosses, the sex ratio in MEX1 is consistent with a 1:1 ratio expected under the hypothesis that a single locus specifies sex in *Ambystoma* (G = 0.01, DF = 1, P = 0.917, N = 93).

Segregation of Molecular Markers in WILD2

A total of 159 markers were genotyped for the WILD 2 cross. The distribution of markers across the known linkage map (Smith *et al*, 2005) places, 92% of map within 30cM of at least one molecular marker and 73% of map within 15cM. Most of the regions that were not covered by this screen are represented only by anonymous (AFLP) markers. Segregation ratios of molecular markers deviated from 1:1 less often than expected based on the distribution of values for the χ^2 test for 1:1 segregation (Figure 1). A similar pattern of deviation from χ^2, towards unity, was observed in a previous study that mapped the same markers using independently generated crosses (Smith *et al*, 2005).

Association Analysis of Sex in WILD2

Genetic screens for sex-associated regions of the *Ambystoma* genome identified a single marker (*E24C3*) that was completely associated with segregation of sex phenotypes in the subpopulation of 24 WILD2 offsping that were used for the initial genome scan. This marker is the most terminal gene/EST based marker at the bottom of LG5. The WILD2 mapping family was generated using a backcross mating design, therefore individuals segregated two genotypes for the *E24C3* marker, $E24C3^{G/G}$ (homozygous for the *A. mexicanum* genotype) and $E24C3^{G/A}$ (*A. mexicanum*/*A. t. tigrinum* heterozygotes) The male phenotype was associated with inheritance of the heterozygous $E24C3^{G/A}$ genotypes, whereas the female phenotype was associated with inheritance of the homozygous $E24C3^{G/G}$ genotype. The pattern of segregation of $E24C3^{G/G}$ *vs.* $E24C3^{G/A}$ genotypes is consistent with the previously demonstrated ZW (heterogametic female) sex-determining mechanism of *A. mexicanum* (Humphrey, 1945; Lindsley, 1956; Humphrey, 1957; Armstrong, 1984), given that the P1 *A. mexicanum* and the F1 hybrid

were females (respectively, $Z^{A.\ mexicanum\ (A.mex)}/W^{A.mex}$ and $Z^{A.\ t.\ tigrinum\ (A.t.t)}/W^{A.mex}$) and the P1 *A. t. tigrinum* and P2 *A. mexicanum* were males (respectively, $Z^{A.t.t}$ / $Z^{A.t.t}$ and $Z^{A.mex}$ /$Z^{A.mex}$). In other words, females presumably inherited a $W^{A.mex}$ locus from their F1 parent.

To further resolve the position of the sex-associated region on LG5, larger panels of late-metamorphosing individuals were genotyped for the markers *E24C3* (N = 112) and E26C11 (N = 74). Again, segregating genotypes for these markers were strongly associated with segregation of the sexes (E24C3: LRS = 141.1, P < 1E-5; E26C11: LRS = 13.8, P = 2E-4) (Table 2 and 3). Based on these genotypes, the most likely position for a single sex determining is approximately 8.1 cM distal to the position of *E24C3* on LG5 (Figure 2).

Because sex-linked markers are now available for WILD2 it is possible to test whether the observed female bias was due to the fact that only paedomorphic and late metamorphosing individuals (individuals with clearly differentiated gonads) were used in the tabulation of sex ratios. If the sex ratio of early metamorphosing individuals was male biased, this sampling methodology would have yielded the observed female-biased results, assuming a 1:1 sex ratio overall. To address this possibility, the *E24C3* locus was used to sex 66 individuals that underwent metamorphosis prior to achieving gross morphological differentiation of their gonads; these animals completed metamorphosis between 116 and 162 days post hatching. Among these animals there were 31 $E24C3^{G/G}$ heterozygotes (presumptive males), and 35 $E24C3^{G/A}$ heterozygotes (presumptive females). This sex ratio is not different from 1:1 nor is it male biased, thus, it appears that sex-specific maturation rates did not play a large role in the observed pattern of female-biased sex ratios in WILD2.

Association Between E24C3 and Sex in MEX1

In order to test the previous mapping result, and to more specifically test the hypothesis that E24C3 is linked to the sex-determining factor in the species *A. mexicanum*, 85 individuals from the cross MEX1 were genotyped for two SNPs that segregated within the *E24C3* marker fragment. Genotypes for *E24C3* were strongly associated with segregation of male and female phenotypes in the MEX1 panel (Figure 3) (for $E24C3^{318}$

– LRS = 20.3, P = $1E^{-5}$, N = 85; for $E24C3^{79}$ – LRS = 25.0, P < $1E^{-5}$, N = 85). As was the case in the WILD2 mapping family, the pattern of association in the MEX1 cross is consistent with the previously demonstrated ZW sex-determining mechanism of *A. mexicanum* (Humphrey, 1945; Lindsley, 1956; Humphrey, 1957; Armstrong, 1984) (Figure 3). Using MEX1, the distance between *E24C3* and the sex-determining factor is estimated at approximately 5.9 cM. This is similar to the estimate obtained using the WILD2 panel.

Discussion

The studies outlined in this chapter describe genome wide segregation analyses that localize the major sex-determining factor in the WILD2 mapping panel and provide independent validation of the linkage between the marker *E24C3* and the sex-determining factor of the species *A. mexicanum*. From the standpoint of genetic linkage and gene content, *Ambystoma* LG 5 represents most completely characterized amphibian sex chromosome to date. Sex-associated allozyme (reviewed by Sumida and Nishioka, 2000) and nucleotide (Ogata *et al*, 2003) markers have been previously identified for a few anuran species and a sex-associated allozyme has been identified for one urodele species (*Pleurodeles waltl*: Ferrier *et al*, 1983). Localization of *ambysex* to the tip of LG5, sheds light on previous genetic and cytogenetic studies of sex-determination in ambystomatid salamanders, the evolution of sex chromosomes, and provides a powerful new set of molecular tools for studying the role of the genetic sex-determining mechanism in the evolution, ecology, and development of tiger salamanders, and amphibians in general. Results from WILD2 indicate that the genetic sex-determining mechanism of *A. mexicanum* is compatible with the genomes of F1 hybrids between *A. mexicanum* and *A. t. tigrinum* and their backcross offspring. This suggests that the genetic sex determining mechanism of *A. mexicanum* is likely to be conserved in many of the species that comprise the tiger salamander species complex. Although the evidence is strong for the presence of a major sex-determining locus in WILD2, the observation of slight, but consistent, female bias indicates that additional factors may have also influenced segregation of sex in this mapping family.

Localization of the major sex-determining factor of *A. mexicanum*

The localization and inheritance of the sex-determining factor in these studies is consistent with results from previous genetic studies of sex-determination in *A. mexicanum*. Genetic studies have firmly established a ZW type sex-determining mechanism for *A. mexicanum* (Humphrey, 1945; Lindsley *et al*, 1956; Humphrey, 1957; Armstrong, 1985), and linkage analyses have indicated that the sex-determining factor is not linked to a centromere (Lindsley *et al*, 1956; Armstrong, 1985). The patterns of association between E24C3 and sex are consistent with segregation of a female specific "W" locus and the terminal location of *ambysex* is inconsistent with linkage to its centromere.

Although linkage analyses for sex-determination are consistent with transmission genetic studies of sex-determination in *A. mexicanum*, these results seem to be at odds with a few cytogenetic studies that have reported association between sex and segregation of heteromorphisms on one of the two smallest chromosome pairs. Numerous descriptions have been made of tiger salamander (*A. mexicanum* and *A. t. tigrinum*) karyotypes (Parmenter 1919; Dearing, 1934; Haschuka and Brunst, 1965; Callan, 1966; Cuny and Malacinski, 1985). Two cytogenetic studies have reported subtle heteromorphisms of the smallest chromosome pair that are consistent with ZW segregation (Haschuka and Brunst, 1965; Cuny and Malacinski, 1985). However, other studies have reported no sex-specific heteromorphisms (Parmenter, 1919; Dearing, 1934) or have directly disputed the existence of ZW heteromorphisms (Callan, 1966). The mapping studies that are presented below are inconsistent with the idea that the smallest the presences of the SD locus on the smallest chromosome pair and, by extension, imply that the Z and W chromosomes of *A. mexicanum* are not morphologically differentiated to any appreciable degree.

Estimates of chiasmata frequencies in female *A. mexicanum* oocytes (Callan, 1966) place the results of this study at further odds with the cytogenetic studies that have reported ZW heteromorphisms in ambystomatid salamanders. Callan's chiasmata counts provide map length estimates for chromosomes 13 and 14 of 169.3 and 135.5 cM respectively. These estimated lengths are similar to recently published map lengths for

linkage groups 13 (163.3 cM) and 14 (125.5 cM) (Smith *et al*, 2005), but are substantially different from the estimated length for linkage group 5 of 323.1 cM (Smith *et al*, 2005) in the LAB cross (Voss, 1995) and >292 in WILD2 (Voss and Smith, 2005). It therefore seems unlikely that the sex-determining locus of *A. mexicanum* lies on either of the two smallest chromosomes.

Female bias in WILD2

A slight, but statistically significant, female bias was observed in WILD2 and the segregation of sex-linked (*E24C3*) genotypes among early metamorphosing individuals is not consistent with sex-specific differences in age at metamorphosis. It is interesting to note that the number of females with presumptively recombinant genotypes (N = 8 $E24C3^{G/A}$ females) is also larger than the number of presumptively recombinant males (N = 1 $E24C3^{G/G}$ male). The larger number of female might reflect the presence of additional genetic factors that cause ZZ individuals to develop as females, or an increased mortality of male recombinants during pre-hatching stages. Identification of markers that are more tightly linked to the sex-determining locus, or the sex-determining locus itself, should ultimately permit recombinant individuals to be differentiated from potentially sex-reversed individuals.

It is also possible, though perhaps less likely, that female-biased sex ratios were generated as a result of environmental conditions under which WILD2 offspring were reared. Although all WILD2 individulas experienced very similar environmental conditions, it is formally possible that these conditions increased the probability that individuals developed as females, irrespective of genotype. Temperature is known to affect sex ratio in some salamanders (reviewed by Wallace *et al*., 1999), but only at extreme levels that are not encountered during normal laboratory culture. Perhaps more importantly, the only study to test for environmental effects on sex ratios in *Ambystoma* found no deviation from a 1:1 sex ratio in *A. tigrinum* or *A. maculatum* reared at 13 or 22º (Gilbert, 1936), which encompasses the range of temperature variation within this study. Thus, it seems unlikely that the observed slight female bias was caused by rearing temperature in WILD2 crosses. With validated markers for sex, it will be possible to systematically test for environmental or effects on sex determination in *A. mexicanum*

and possibly other members of the tiger salamander species complex.

Recombination rates in the *ambysex*-linked region imply a recent origin for *ambysex*

Previously published linkage analyses were based on recombination that occurred in the testes of male (presumably ZZ) F1 hybrids between *A. mexicanum* and *A. t. tigrinum* (Voss, 1995; Smith *et al*, 2005). Estimates of linkage distance in WILD2 are based on female (ZW) recombination, and estimates of linkage distance in MEX1 are based on recombination in both sexes. Divergence of tiger salamander Z and W chromosomes should tend to reduce recombinational distances in female meiosis relative to male meiosis. The recombinational distance between *E24C3* and a Mendelian *ambysex* locus in WILD2 is similar to the frequency of recombination between *E24C3* and *ambysex* in the male and female F1s that founded MEX1 (Table 2, Figures 3 and 4). Moreover, the estimated ZW recombinational distance between *E24C3* and *E26C11* was 30.2 cM in WILD2, higher than the estimated rate of ZZ recombinational distance (13.6 cM) for these same loci in AxTg (Smith *et al*, 2005). In the few other cases where recombination frequencies have been characterized for young sex chromosomes (*e.g.* medaka: Kondo *et al*, 2004; stickleback: Peichel *et al*, 2005; and papaya: Lui *et al*, 2004; Ma *et al*, 2004) the frequency of recombination in the 30-100 cM adjacent to the sex-determining factor is substantially reduced within the heterogametic sex, relative to the homogametic sex. Given the lack of evidence for differences in recombination frequency, and the lack of convincing cytogenetic evidence for structurally differentiated sex chromosomes in the tiger salamander lineage, it stands to reason that the *ambysex* locus may have arisen quite recently, within the last 5-10 million years.

Sex and the biology of ambystomatid salamanders

Ambystomatid salamanders are important models for studying many aspects of development (*e.g.* Stopper and Wagner, 2007; Zhang *et al*, 2007; Sobkow *et al*, 2006), ecology (*e.g.* Trenham and Shaffer, 2005; Fitzpatrick and Shaffer, 2004; Brunner JL *et al*, 2005), and evolution (*e.g.* Voss and Smith, 2005; Robertson *et al*, 2006; Weisrock *et al*, 2006), and sex is one of most ubiquitous sources of developmental and ecological variation. Localization of the sex-determining factor to the *Ambystoma* linkage map

provides an important tool for understanding the role of sex in shaping the biology of tiger salamanders and amphibians in general.

Table 9.1: Segregation of sex among backcross progeny and corresponding G tests for goodness of fit to a 1:1 sex ratio.

Cross	*Male*	*Female*	*N*	*Proportion Sexed*	Proportion *Female*	*Test (G)*	*df*	*G*	*p*
1	19	41	60	0.81	0.68		1	8.26	0.004
2	17	23	40	0.70	0.58		1	0.90	0.342
3	18	13	31	0.76	0.42		1	0.81	0.368
4	21	24	45	0.59	0.53		1	0.20	0.655
5	22	23	45	0.73	0.51		1	0.02	0.881
6	26	36	62	0.75	0.58		1	1.62	0.203
7	16	20	36	0.80	0.56		1	0.45	0.505
8	13	13	26	0.67	0.50		1	0.00	1.000
9	5	10	15	0.75	0.67		1	1.70	0.192
Total	157	203	360	0.72	0.56	Total	9	13.96	0.124
						Pooled	1	5.89	0.015
						Heterogeneity	8	8.06	0.427

Table 9.2: Segregation of *E24C3* genotypes and sex phenotypes in a subset of the WILD2 cross.

Sex	$E24C3^{G/G}$	$E24C3^{G/A}$
Female	50	8
Male	1	53

Table 9.3: Segregation of *E26C11* genotypes and sex phenotypes in a subset of the WILD2 cross.

Sex	$E26C11^{A/A}$	$E26C11^{A/G}$
Female	24	14
Male	8	28

Figure 9.1: Plot of the distribution of the χ^2 statistic for 1:1 segregation of marker genotypes in WILD2 *vs.* χ^2 distribution.

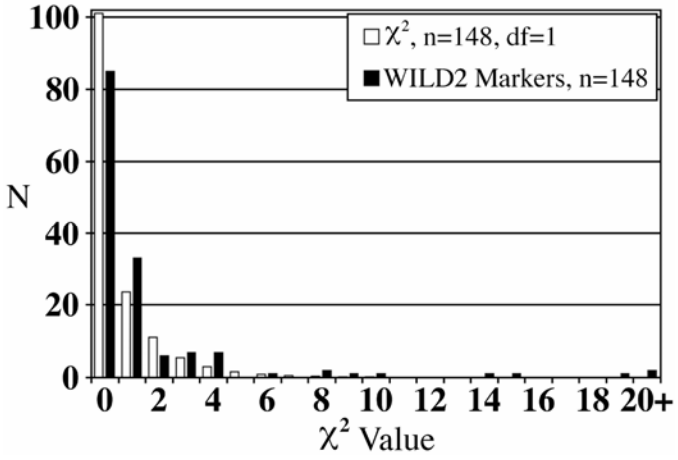


Figure 9.2: Plot of Likelihood Ratio Statistics (LRS) for association between segregation of sex phenotypes and marker genotypes on LG5 (WILD2). Horizontal lines represent (from bottom to top) linkage group-wide LRS thresholds for suggestive (37th percentile), significant (95th percentile), and highly significant (99.9th percentile) associations (Lander and Kruglyak, 1995) estimated using MapMaker QTXb19.

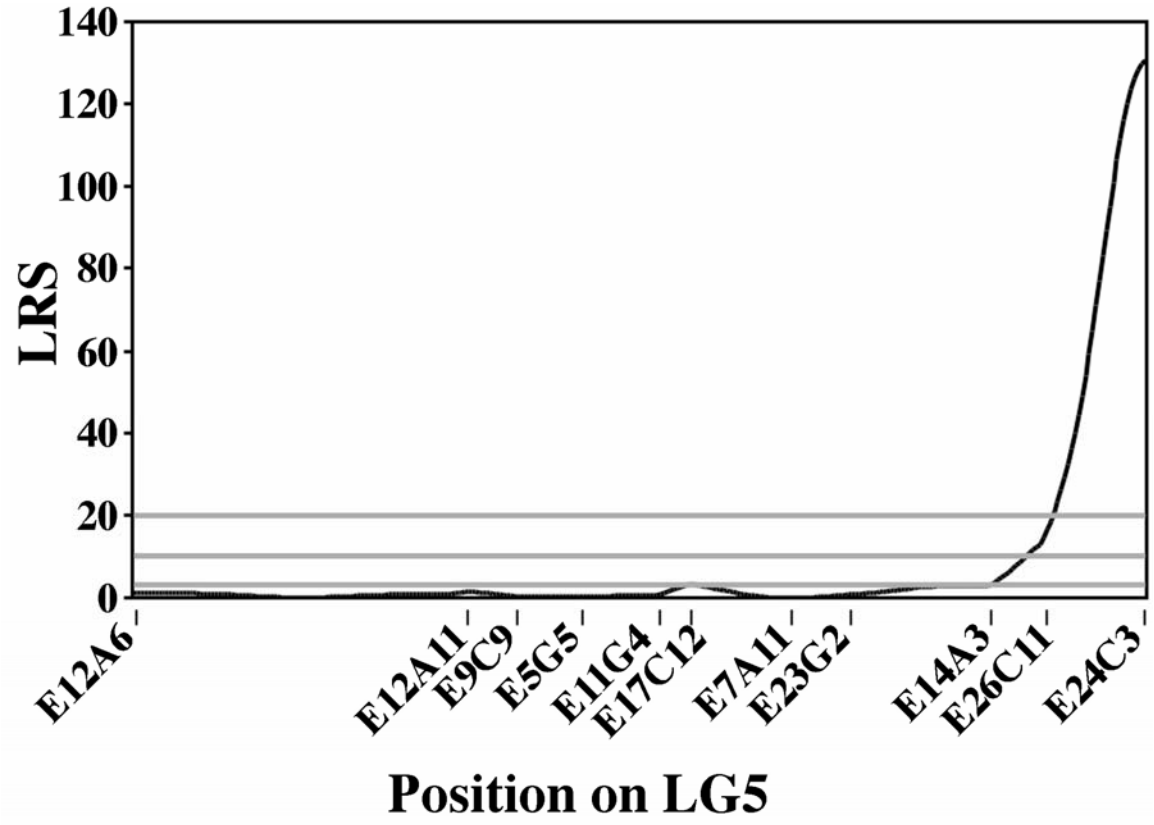
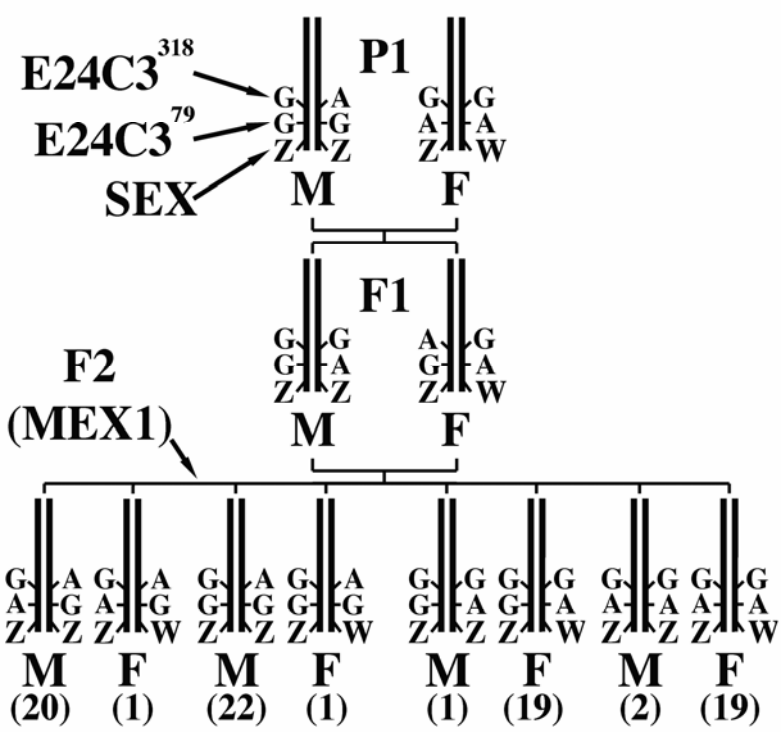


Figure 9.3: Diagram of the crossing design that was used to generate the MEX1 cross and segregation of *E24C3* marker genotypes and Z/W alleles of the *ambysex* locus in this crossing design. The numbers of individuals that inherited each multilocus genotype are in parentheses. M = male, F = female.



CHAPTER 10: Future Directions.

The studies that are described in the preceding chapters set the stage for future studies using ambystomatid salamanders and other amphibian species to gain insights about the genetic basis of development and the evolution of vertebrate genomes. Below, I outline a few future avenues of research that build directly upon the studies that are presented in this dissertation. Specifically, I briefly outline avenues that: 1) generate more precise and insightful evolutionary comparisons which will shed further light on the evolutionary history of vertebrate genomes; 2) more precisely characterize the *met* locus and identify other QTL that contribute to variation in age at metamorphosis or the expression of metamorphic *vs.* paedomorphic life histories; and 3) shed insight into the effects of sex on variation in development and gene expression.

Comparative Genomics

By improving the methods that are used to identify orthologous genes and genetic maps for other phylogenetically well-positioned amphibians, it should be possible generate more precise and insightful evolutionary comparisons to shed light on the evolutionary history of vertebrate genomes.

Improvements to existing methods

Although the alignments that are generated by MapToGenome (Chapter 7) are generally robust at identifying orthologous loci even at relatively deep levels of evolutionary divergence (*e.g.* alignments between *Ambystoma* and human), there is certainly room for improvement. For example, MapToGenome relies on the user to optimize the thresholds that are used to identify intron breaks. Other programs that have been developed to align transcripts to their source genomes have employed more sophisticated models for calling intron breaks, thereby generating more precise alignments (*e.g.* van Nimwegen, 2006). Incorporating similar models, should improve alignment quality and the interpretability of homology maps. Improving the algorithms that are used to generate genome wide alignments will also facilitate comparisons that extend well beyond questions of *Ambystoma*/vertebrate synteny.

Other amphibians

As genetic maps become available for additional amphibian species, they will provide important perspective on the genome wide comparisons that are described in preceding chapters of this dissertation. As is mentioned in Chapter 8, it is possible that some of the features that are shared between the *Ambystoma* genome and amniote genomes have arisen through convergent evolutionary events. When genetic maps are developed for other amphibian species, it will be possible to more completely address the possibility that changes have evolved independently. Features that are shared between *Ambystoma* and another amphibian species are more likely to have been shared by a common ancestor, and presumably less likely to have evolved in parallel. With the progress that has been made toward developing genome wide resources for the frog *Xenopus* the possibility of obtaining independent validation of known amphibian (*Ambystoma*)/ amniote syntenies may soon become reality.

It is not an impossibility that mapping data could be generated for many other amphibian species. It should be possible to apply essentially the same methods that we describe in Chapters 2 and 3 to generate large numbers of genetic markers for many amphibian species. In theory, any amphibian species that can be manipulated to produce meiotic products (*i.e.* gametes or offspring) could be used to generate an informative mapping panel.

met

In the near future it should be possible to more precisely characterize the *met* locus and identify other QTL that contribute to variation in age at metamorphosis or the expression of metamorphic *vs.* paedomorphic life histories. Identification of the gene(s) that underlies *met* will provide a critical next step toward understanding the molecular basis of this adaptive phenotype. Indeed, much of the time that I have spent at the University of Kentucky has been dedicated toward this goal. The comparative genomic studies that are described in Chapters 6 and 7 were largely motivated by the prospect of leveraging information from other vertebrate species (especially human and chicken) to identify candidates for the *met* gene: genes that participate in thyroid hormone signaling and are

likely to have orthologs that lie within the *met* QTL region.

It should also be possible to gain additional insights into the evolutionary/ developmental effects of *met* and other loci hat contribute to metamorphic variation by using the markers that were developed for mapping in the LAB cross (also known as AxTg) to perform genome-wide scans for metamorphosis QTLs in WILD2 and crosses that can and have been made between other species pairs (*e.g.* Voss and Shaffer, 1996; Voss *et al*, 2003). In essence, every tiger salamander population represents a genetic line that has experienced a unique history of selection for metamorphic timing and the expression of metamorphosis *vs.* paedomorphosis. Natural populations of *Ambystoma* therefore represent an almost limitless source of genetic information on the molecular processes that govern metamorphosis.

Sex

Much in the same way that identifying the *met* gene will shed critical insight into the molecular basis of adaptation and development, identifying the sex-determining gene(s) will shed new light on the causes and consequences of sex-determination. By identifying the sex locus, it will be possible to achieve immense insight into the developmental/ genetic basis of sex and the effects of sex on variation in physiology, development and gene expression in *Ambystoma*, and in vertebrates in general.

It is important to recognize, however, that many insights can also be gained before the sex-determining locus is identified. More in-depth studies of sex-linked regions of the genome can elucidate the evolutionary consequences of single locus (ZW) sex determination, and the sex-linked markers that were identified in Chapter 9 can be used to identify males and females prior to the completion of gonadal differentiation (with a minimal amount to error due to recombination). The ability to ascertain the genetic sex of individuals at all stages of development will also be invaluable in describing the transcriptional and developmental events that characterize the normal process of sex determination, as well as investigating the effects of natural variation (e.g. quantitative trait loci) or perturbations of gene function (*e.g.* endocrine disruptors, RNA interference, transgenes).

REFERENCES

Adrian, E. K. Jr. and B. E. Walker, 1962 Incorporation of thymidine-H3 by cells in normal and injured mouse spinal cord. J Neuropathol Exp Neurol. 21: 597–609.

Ambros, V., 1988 Genetic basis for heterochronic variation, pp. 269–286 in *Heterochrony in Evolution*, edited by M. L. McKinney. Plenum Press, New York.

Amores, A., A. Force, Y. L. Yan, L. Joly, C. Amemiya *et al*., 1998 Zebrafish hox clusters and vertebrate genome evolution. Science 282: 1711-1714.

Anderson, J. D. and Worthington R. D. 1971 The life history of the Mexican salamander *Ambystoma ordinarium* Taylor. Herpetologica 27: 165–176.

Anderson, L. K., N. Salameh, H. W. Bass, L. C. Harper, W. Z. Cande, *et al*., 2004 Integrating genetic linkage maps with pachytene chromosome structure in maize. Genetics 166: 1923-1933.

Arcot, S. S., Z. Wang, J. L. Weber, P. L. Deininger and M. A. Batzer, 1995 *Alu* repeats: a source for the genesis of primate microsatellites. Genomics 29:136-44.

Armstrong, J. B. and S. T. Duhon, 1989 Induced spawnings, artificial spawnings, and other genetic manipulations, pp. 228–236 in *Developmental Biology of the Axolotl*, edited by J. B. Armstrong and G. M. Malacinski. Oxford University Press, New York.

Armstrong, J. B., 1984 Genetic Mapping in the Mexican axolotl, *Ambystoma mexicanum*. Can. J. Genet. Cytol. 26: 1-6.

Ayling, L. J. and D. K. Griffin, 2002 The evolution of sex chromosomes. Cytogenet. Genome Res. 99: 125-140.

Bachvarova, R. F., T. Masi, M. Drum, N. Parker, K. Mason, *et al*., 2004 Gene expression in the Axolotl germ line: Axdazl, Axvh, Axoct-4, and Axkit. Dev. Dyn. 231: 871-880.

Baroiller, J. F., Y. Guigen and A. Fostier, 1999 Endocrine and environmental aspects of sex differentiation in fish. Cell. Mol. Life Sci. 55: 910-931.

Beetschen, J.-C., 1996 How did urodele embryos come into prominence as a model system? Int. J. Dev. Biol. 40: 629–636.

Bernot, A. E. and J. E. Weissenbach, 2004 Estimation of the extent of synteny between *Tetraodon nigroviridis* and *Homo sapiens* genomes. J. Mol. Evol. 59: 556-569.

Blalock, H. M. Jr. 1958 Probabilistic interpretations for the mean square contingency. J. Am. Stat. Assoc. 53: 102-105.

Blomme, T., Vandepoele K., De Bodt S., C. Simillion, S. Maere, Y. Van de Peer, 2006 The gain and loss of genes during 600 million years of vertebrate evolution. Genome Biol. 7: R43.

Borland, S., K. Crawford and V. Brand 2003 Setting the stage: Developmental biology in pre-college classrooms. Int. J. Dev. Biol. 47: 85–91.

Bourque, G., P. A. Pevzner and G. Tesler, 2004 Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes. Genome Res. 14: 507-516.

Bourque, G., E. M. Zdobnov, P. Bork, P A. Pevzner and G. Tesler, 2005 Comparative architectures of mammalian and chicken genomes reveal highly variable rates of genomic rearrangements across different lineages. Genome Res. 15: 98-110.

Bray, N., I. Dubchak and L. Pachter, 2003 AVID: A global alignment program. Genome Res. 13: 97-102.
Bray, N. and L. Pachter, 2004 MAVID: constrained ancestral alignment of multiple sequences. Genome Res. 14: 693-699.
Brown, D. M., T. C. Matise, G. Kioke, J. S. Simon, E. S. Winer, *et al*., 1998 An integrated linkage map of the laboratory rat. Mamm. Genome 9: 521-530.
Bruni, J. E. and W. A. Anderson, 1985 Ependyma of the rat fourth ventricle and central canal: Response to injury. Acta Anat. 128: 265–273.
Brunner, J. L., K. Richards and J. P. Collins, 2005 Dose and host characteristics influence virulence of ranavirus infections. Oecologia 144: 399-406.
Bull, J. J. 1983 *Evolution of Sex Determining Mechanisms*. Benjamin Cummings, Menlo Park.
Burt, D. W., C. Bruley, I. C. Dunn, C. T. Jones, A. Ramage *et al*., 1999 The dynamics of chromosome evolution in birds and mammals. Nature 402:411-413.
Calabrese, P. P., S Chakravarty and T. J. Vision, 2003 Fast identification and statistical evaluation of segmental homologies in comparative maps. Bioinformatics 19: i74-i80.
Callan, H. G., 1966 Chromosomes and nucleoli of the axolotl, *Ambystoma mexicanum*. J. Cell Sci. 1: 85-108.
Cannatella, D. C. and R. O. De Sa, 1993 *Xenopus laevis* as a model organism. Syst. Biol. 42: 476–507.
Carmel, J. B., A. Galante, P. Soteropoulos, P. Tolias and M. Recce, 2001 Gene expression profiling of acute spinal cord injury reveals spreading inflammatory signals and neuron loss. Physiol. Genomics 7: 201–213.
Chakravarti, A., L. K. Lasher, and J. E. Reefer, 1991 A maximum-likelihood method for estimating genome length using genetic-linkage data. Genetics 128: 175-182.
Charlesworth, D. and B. Charleswoth, 2005 Sex chromosomes: evolution of the weird and wonderful. Curr. Biol. 15: R129-131.
Charlesworth, D., B. Charleswoth and G. Marais, 2005 Steps in the evolution of heteromorphic sex chromosomes. Heredity 95: 118-128.
Chen, X., L. Levine and P-Y. Kwok, 1999 Fluorescence polarization in homogenous nucleic acid analysis. Genome Res. 9: 492-498.
Chernoff, E. A. G., D. L. Stocum, H. L. D. Nye and J. A. Cameron 2003 Urodele spinal cord regeneration and related processes. Dev. Dyn. 226: 295–307.
Chichilnisky, E. J., and F. Rieke, 2005 Detection sensitivity and temporal resolution of visual signals near absolute threshold in the salamander retina. J. Neurosci. 25: 310-330.
Christensen, R. N., M. Weinstein and R. A. Tassava 2002 Expression of fibroblast growth factors 4, 8, and 10 in limbs, flanks, and blastemas of *Ambystoma*. Dev. Dyn. 223: 193-203.
Collenot, A., D. Durand, M. Lauther, R. Dorazi, J.C. Lacroix, *et al*., 1994 Spontaneous sex reversal in *Pleurodeles waltl* urodele amphibia): Analysis of its inheritance. Genet. Res. 64: 43–50.
Collins, J. P., 1981 Distribution, habitats and life history variation in the tiger salamander, *Ambystoma tigrinum*, in east-central and southeast Arizona. Copeia 1981: 666–675.

Collins, J. P., J. L. Brunner, J. K. Jancovich and D. M. Schock, 2004 A model host-pathogen system for studying infectious disease dynamics in amphibians: Tiger salamanders (*Ambystoma tigrinum*) and *Ambystoma tigrinum* virus. Herpetol. J. 14 195-200.

Cox, D. R., M. Burmeister, E. R. Price, S. Kim and R. M. Myers, 1990 Radiation hybrid mapping: a somatic cell genetic method for constructing high resolution maps of mammalian chromosomes. Science 250: 245-250.

Cramer, H., 1946 *Mathematical Methods of Statistics*, Princeton: Princeton University Press.

Cuny, R., and G. M. Malacinski, 1985 Banding differences between tiger salamander and axolotl chromosomes. Can. J. Genet. Cytol. 27: 510-514.

Danchin, E. G. and P. Pontarotti, 2004 Statistical evidence for a more than 800-million-year-old evolutionarily conserved genomic region in our genome. J. Mol. Evol. 59: 587-597.

Danzmann, R. G., M. Cairney, W. S. Davidson, M. M. Ferguson, K. Gharbi, *et al.*, 2005 A comparative analysis of the rainbow trout genome with two other species of fish (Arctic char and Atlantic salmon) within the tetraploid derivative Salmonidae family (subfamily: Salmoninae). Genome 48: 1037–1051.

Davis, R. H. 2004 The age of model organisms. Nat. Rev. 5: 69–77.

de Boer, L. E. M., 1980 Do the chromosomes of the kiwi provide evidence for a monophyletic origin of the ratites? Nature 287: 84-85.

Dearing, W. H. Jr., 1934 The maternal continuity an individuality of the somatic chromosomes of *Ambystoma tigrinum*, with special reference to the nucleolus as a chromosomal component. J. Morph. 56: 157-179.

Del Rio-Tsonis, K., J. C.Jung, I. M.Chiu and P. A. Tsonis, 1997 Conservation of fibroblast growth factor function in lens regeneration. Proc. Natl. Acad. Sci. USA 94: 13701–13706.

Delvin, R. H. and Y. Nagahama, 2002 Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. Aquaculture 208: 191-364.

Denz, C. R., A. Narshi, R. W. Zajdel and D. K. Dube, 2004 Expression of a novel cardiac-specific tropomyosin isoform in humans. Biochem. Biophys. Res. Comm. 320: 1291-1297.

Dewey, C. N and Pachter L. 2006 Evolution at the nucleotide level: the problem of multiple whole-genome alignment. Hum. Mol. Genet. 15: 51-56.

Dumeril, A. H. A., 1870 Quatrième notice sur la menagerie des reptiles du Muséum d'Historie Naturelle. Noveau Archives du Muséum d'Historie Naturelle, Paris: 5: 47-60.

Duret, L., D. Mouchiroud and C. Gautier, 1995 Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochors. J. Mol. Evol. 40: 308-317.

Durtilleaux, B., 1986 Le role des chromosomes dans l'evoluiton: une nouvelle interpretation. Ann. Genet. 29: 69-75.

Echeverri, K. and E. M. Tanaka, 2002 Ectoderm to mesoderm lineage switching during axolotl tail regeneration. Science 298: 1933–1936.

Egel, E. R., 1995 The synaptonemal complex and the distribution of meiotic recombination events. Trends Genet. 11: 206-208.
Ellegren, H., 2000 Evolution of the avian sex chromosomes and their role in sex determination. Trends Ecol. Evol. 15: 188-192.
Ericsson, R., R. Cerny, P. Falck, and L. Olsson, 2004 Role of cranial neural crest cells in visceral arch muscle positioning and morphogenesis in the Mexican axolotl, *Ambystoma mexicanum*. Dev. Dyn. 231: 237-247.
Ewert, M. A. and C. E. Nelson, 1991 Sex determination in turtles: diverse patterns and some possible adaptive values. Copeia 1: 50–69.
Ewing, B. and P. Green, 2000 Analysis of expressed sequence tags indicates 35,000 human genes. Nat. Gen. 25: 232–234.
Ewing, B. and P. Green, 1998 Base-calling of automated sequencer traces using PHRED. II. Error probabilities. Genome Res. 8: 186–194.
Ezaz, T., R. Stiglec, F. Veyrunes and J. A. M. Graves, 2006 Relationships between vertebrate ZW and XY sex chromosome systems. Curr. Biol. 16: R736–R743.
Falck P., J. Hanken and L. Olsson, 2002 Cranial neural crest emergence and migration in the Mexican axolotl (*Ambystoma mexicanum*). Zoology-Jena 105: 195–202.
Falconer, D. S., 1989 *Quantitative Genetics*. Longman Scientific & Technical, Essex, UK.
Fankhauser, G. and R. R. Humphrey, 1942 Induction of triploidy and haploidy in axolotl eggs by cold treatment. Biol. Bull. 83: 367-374.
Ferretti, P., 1996. Re-examining jaw regeneration in urodeles: what have we learnt? Int. J. Dev. Biol. 40: 807–811.
Ferrier, V., F. Gasser, A. Jaylet and C. Cayrol, 1983 A genetic-study of various enzyme polymorphisms in *Pleurodeles waltlii* (urodele amphibian) .2. peptidases - demonstration of sex linkage. Biochem. Genet. 21: 535-549.
Fisher, R.A., 1938 Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd.
Fitzpatrick ,B. M. and H. B. Shaffer, 2004 Environment-dependent admixture dynamics in a tiger salamander hybrid zone. Evolution 58: 1282-1293.
Flannick, J. and S. Batzoglou, 2005.Using multiple alignments to improve seeded local alignment algorithms. Nucleic Acids Res. 33: 4563-77.
Florea, L., G. Hartzell, Z. Zhang, G. M. Rubin, W. Miller, 1998 A computer program for aligning a cDNA sequence with a genomic DNA sequence. Genome Res. 8: 967-74.
Fridolfsson, A. K., H. Cheng, N. G. Copeland, N. A. Jenkins, H.-C. Liu *et al*., 1998 Evolution of the avian sex chromosomes from an ancestral pair of autosomes. Proc. Natl. Acad. Sci. USA 95: 8147-8152.
Futuyma, D. J., 1998 *Evolutionary Biology*. Sinauer Associates, Sunderland, MA.
Gall, J. G. and M. L. Pardue, 1969 Formation and detection of RNA-DNA hybrid molecules in cytological preparations. Proc. Natl. Acad. Sci. USA 63: 378-383.
Gardiner, D. M., T. Endo and S. V. Bryant, 2002 The molecular basis of amphibian limb regeneration: integrating the old with the new. Semin. Cell Dev. Biol.13: 345–352.

Gardiner, D. M., M. A. Torok, L. M. Mullen and S. V.Bryant, 1998 Evolution of vertebrate limbs: Robust morphology and flexible development. Am. Zool. 38: 659-671.
Gardiner, A. F. and W. E. Jack, 2002 Acyclic and dideoxy terminator preferences denote divergent sugar by archaeon detection. BioTechniques 31: 560-570.
Gates, M. A., L. Kim, E. S. Egan, T. Cardozo, H. I. Sirotkin, *et al*., 1999 A genetic linkage map for zebrafish: comparative analysis and localization of genes and expressed sequences. Genome Res. 9: 334-347.
Geisler, R., G. J. Rauch, H. Baier, F. van Bebber, L. Bross, *et al.*, 1999 A radiation hybrid map of the zebrafish genome. Nat. Genet. 23: 86–89.
Gharbi, K., A. Gautier, R. G. Danzmann, S. Gharbi, T. Sakamoto, *et al*., 2006 A linkage map for brown trout (*Salmo trutta*): chromosome homeologies and comparative genome organization with other Salmonid fish. Genetics 172: 2405–2419.
Gilbert, W. M., 1936 *Amphisexuality and sex differentiation in Ambystoma*. Unpublished Thesis, State University of Iowa.
Goldschmidt, R. B., 1940 *The Material Basis of Evolution*. Yale University Press, New Haven, CT.
Goodman, L. A. and W. H. Kruskal, 1954 Measures of association for cross classifications. J. Am. Stat. Assoc. 49: 732-764.
Goodman, L. A. and W. H. Kruskal, 1963 Measures of association for cross classifications III: approximate sampling theory. J. Am. Stat. Assoc. 58: 310-364.
Gornung, E., I. Gabrielli, S. Cataudella and L. Sola, 1997 CMA(3)-banding pattern and fluorescence in situ hybridization with 18s rRNA genes in zebrafish chromosomes. Chromosome Res. 5: 40-46.
Goss, S. J and H. Harris, 1975 New method for mapping genes in human chromosomes. Nature 255: 680-684.
Gould, S. J., 1977 *Ontogeny and Phylogeny*. Belkap Press, Cambridge, MA.
Gould, S. J., 1981 Change in developmental timing as a mechanism of macroevolution, pp. 333–346 in *Evolution and Development*, edited by J. T. Bonner. Springer-Verlag, Berlin.
Graves, J. A. M. 1995 The origin and function of the mammalian Y-chromosome and Y-borne genes – an evolving understanding. Bioessays 17: 311-320.
Graves, J. A. M., J. Gecz and H. Hameister, 1995 Evolution of the human X-a smart and sexy chromosome that controls speciation and development. Cytogenet. Genome Res. 99: 141-145.
Groenen, M. A. M., H. H. Cheng, N. Bumstead, B. F. Benkel, W. E. Briles, *et al*., 2000 A consensus linkage map of the chicken genome. Genome Res. 10: 137-147.
Grützner, F., W. Rens, E. Tsend-Ayush, N. El-Mogharbel, P. C. O'Brien *et al*., 2004 In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. Nature 432: 913-917.
Haberman, E. T., A-G. Bebin, S. Herklotz, M. Volkmer, K. Eckelt, *et al*., 2004 An *Ambystoma mexicanum* EST sequencing project: Analysis of 17,352 expressed sequence tags from embryonic and regenerating blastema cDNA libraries. Genome Biol. 5: R67.
Häcker, V.,1907 Uber mendelschen vererbung bei axolotln. Zoologischer Anzeiger 31: 99-102.

Haley, C. S. and S. A. Knott, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69: 315-324.
Handley L. L.-J., H. Ceplitis and H. Ellegren, 2004 Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. Genetics 167: 367-376.
Hauschuka, T. S. and V. V. Brunst, 1965 Sexual dimorphism in the nucleolar autosome of the axolotl (*Sirenodon mexicanum*). Hereditas 52: 345-356.
Hayes, T. B., 1998 Sex determination and primary sex differentiation in amphibians: Genetic and developmental mechanisms. J. Exp. Zoo. 281: 373-399.
Hedges, S. B and S. Kumar, 2004 Precision of molecular time estimates. Trends Genet. 20: 242-247.
Heng, H. H. Q., J. Squire and L.-C. Tsui, 1992 High-resolution mapping of mammalian genes by in situ hybridization to free chromatin. Proc. Natl. Acad. Sci. USA 89: 9509-9513.
Hillis, D. M. and D. M. Green, 1990 Evolutionary changes of heterogametic sex in the phylogenetic history of amphibians. J. Evol. Biol. 3: 49-64.
Hinchliffe, J. R. 2002 Developmental basis of limb evolution. Int. J. Dev. Biol. 46: 835-845.
Hinegardner, R., 1968 Evolution of cellular DNA content in teleost fishes. Am. Nat. 102: 517-520.
Hoffman, E. A., and D. W. Phennig, 1999 Proximate causes of cannibalistic polyphenism in larval tiger salamanders. Ecology 80: 1076-1080.
Horner, P. J, A. E. Power, G. Kempermann, H. G. Kuhn, T. D. Plamer *et al*., 2002 Proliferation and differentiation of progenitor cells throughout the intact adult spinal cord. J. Neurosci. 20: 2218–2228.
Housworth, E. A and J. Postlethwait, 2002 Measures of synteny conservation between species pairs. Genetics 162: 441-448.
Hsieh S. J., C. Y. Lin, N. H. Liu, W. Y. Chow, C. Y. Tang, 2006 GeneAlign: a coding exon prediction tool based on phylogenetical comparisons. Nucleic Acids Res. 34: 280-84.
Hsu, T. M., X. Chen, S. Duan, R. D. Miller and P-Y. Kwok, 2001 Universal SNP genotyping assay with fluorescence polarization detection. Nucleic Acids Res. 30: 605-613.
Huang, W, D. M. Umbach and L. Li, 2006 Accurate anchoring alignment of divergent sequences. Bioinformatics 22: 29-34.
Huang, X. and A. Madan, 1999 CAP3: A DNA sequence assembly program. Genome Res. 9: 868–877.
Hulbert, S., T. W. Ilott, E. J. Legg, S. E. Linclon, E. S. Lander, *et al.*, 1988 Genetic analysis of the fungus, *Bremia lactucae*, using restriction fragment length polymorphisms. Genetics 120: 947-958.
Humphrey, R.R., 1929 Studies on sex reversal in *Ambystoma*: bisexuality and sex reversal in larval males uninfluenced by ovarian hormones. Anat. Rec. 42: 119-155.
Humphrey, R.R., 1945 Sex determination in the Ambystomatid salamanders: a study of the progeny of females experimentally converted into males. Am. J. Anat. 76: 33-66.

Humphrey, R.R., 1957 Male homogamety in the Mexican axolotl: a study of the progeny obtained when germ cells of a genetic male are incorporated into the developing ovary. J. Exp. Zoo. 134: 91-101.
Humphrey, R. R., 1967 Albino axolotls from an albino tiger salamander through hybridization. J Hered. 58: 95-101.
Humphrey, R. R. and J. T. Bagnara,1967 A color variant in the Mexican axolotl. J. Hered. 58: 251-256.
Ikegami Y., S. Mitsuda and M. Araki, 2002 Neural cell differentiation from retinal pigment epithelial cells of the newt: an organ culture model for the urodele retinal regeneration. J. Neurobiol. 50: 209–220.
International Chicken Genome Sequencing Consortium (ICGSC), 2005 Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432: 695-716.
International Human Genome Sequencing Consortium (IHGSC), 2001 Initial sequencing and analysis of the human genome. Nature 409: 860-921.
Irschick, D. J. and H. B. Shaffer, 1997 The polytypic species revisited: morphological differentiation among tiger salamanders (*Ambystoma tigrinum*) (Amphibia: Caudata). Herpetologica 53: 30-49.
Isidore, E., H. van Os, S. Andrzejewski, J. Bakker, I. Barrena, *et al*., 2003 Toward a marker dense linkage map of the potato genome: lesions from linkage group I. Genetics 165: 2107-2116.
Jaillon ,O., J. M. Aury, F. Brunet, J.-L. Petit, N. Stange-Thomann, E. Mauceli, *et al*., 2004 Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431: 946-957.
Jancovich, J. K., E. W. Davidson, N. Parameswaran, J. Mao, V. G. Chinchar, *et al*., 2005 Evidence for emergence of amphibian iridoviral disease because of human-enhanced spread. Mol. Ecol. 14: 213-224.
Jancovich, J. K., J. Mao, V. G. Chinchar, C. Wyatt, S. T. Case, *et al*., 2003 Genomic sequence of a ranavirus (family Iridoviridae) associated with salamander mortalities in North America. Virology. 316: 90–103.
Janzen, F. J. and J. G. Krenz, 2004 Phylogenetics: which was first, TSD or GSD?, pp. 121-130 in *Temperature Dependent Sex Determinaiton in Vertebrates*, edited by N. V. A. V. Lance. Smithsonian Books, Washington, DC.
Jensen-Semen, M. I., T. S. Furley, B. A. Payseur, Y. Lu, K. M. Roskin, *et al*., 2004 Comparative recombination rates in the rat, mouse, and human genomes. Genome Res. 14: 528-538.
John, H. A., M. L. Birnstiel and K. W. Jones, 1969 RNA-DNA hybrids at the cytological level. Nature 223: 582-587.
John, B., 1990 *Meiosis*. Cambridge University Pres, Cambridge.
Just, W., W. Rau, W. Vogel, M. Akhverdian, K. Fredga, *et al*., 1995 Absence of SRY in species of the vole *Ellobius*. Nature Genet. 11: 117-118.
Kalyanaraman, A., S. Aluru, S. Kothari and V. Brendel, 2003 Efficient clustering of large EST data sets on parallel computers. Nucleic Acids Res. 31: 2963–2974.
Kauer, J.S., 2002. On the scents of smell in the salamander. Nature 417: 336–342.
Kawai, J, A. Shinagawa, K. Shibata, M. Yoshino, M. Itoh, *et al*., 2001 Functional annotation of a full-length mouse cDNA collection. Nature 409: 685–690.

Kelly, P. D., F. Chu, I. G. Woods, P. Ngo-hazelett, T. Cardozo, *et al*., 2000. Genetic linkage mapping of zebrafish genes and ESTs. Genome Res. 10: 558-567.

Kendall, M. G. and A. Stuart 1967 *The Advanced Theory of Statistics: Volume 2 inference and relationship*. New York: Hafner Publishing Company.

Kent, W. J. 2002 BLAT - The BLAST-like alignment tool. Genome Res. 12: 656-664.

Kihl, P. P., and R. D. Camerini-Otero, 2005 Molecular features and functional constraints in the evolution of the mammalian X chromosome. Crit. Rev. Biochem. Mol. Biol. 40: 313-330.

King, J., L. A. Roberts, M. J. Kearsey, H. M. Thomas, R. N. Jones, *et al*., 2002 A demonstration of a 1:1 correspondence between chiasmata frequency and recombination using a *Lolium perenne*/*Festuca pratensis* substitution. Genetics 161: 307-314.

Kohn M., J. Hogel, W. Vogel, P. Minich, H. Kehrer-Sawatzki, *et al*., 2006 Reconstruction of a 450-MY-old ancestral vertebrate protokaryotype. Trends Genet. 22: 203-210.

Kohn, M., H. Kehrer-Sawatzki, W. Vogel, J. A. M. Graves and H. Hameister, 2004 Wide genome comparisons reveal the origins of the human X chromosome. Trends Genet. 20: 598-603.

Kondo M., I. Nanda, U. Hornung, M. Schmid and M. Schartl, 2004 Evolutionary origin of the medaka Y chromosome. Curr. Biol. 14: 1664-1669.

Kong, A., D. F. Gubdjartsson, J. Sainz, G. M. Jonsdottir, S. A. Gudjonsson, *et al*, 2002 A high resolution recombination map of the human genome. Nature Genet. 31: 241-247.

Kosambi, D.D., 1944 The estimation of map distances from recombination values. Ann. Eugen. 12: 172-175.

Kruger J, A. Sczyrba, S. Kurtz and R. Giegerich, 2004. e2g: an interactive web-based server for efficiently mapping large EST and cDNA sets to genomic sequences. Nucleic Acids Res. 32: W301-4.

Kumar, S., and B. Hedges, 1998 A molecular timescale for vertebrate evolution. Nature 392: 917-920.

Kuntz, S., D. Chardard, C. I. Ko, H. Dumond, M. Ducatez, *et al*., 2006 Female-enriched and thermosensitive expression of steroidogenic factor-1 during gonadal differentiation in *Pleurodeles waltl*. J. Mol. Endocrinol. 36: 175-186.

Kwitek A. E, J. Gullings-Handley, J. Yu, D. C. Carlos, K. Orlebeke, *et al*., 2004 High-density rat radiation hybrid maps containing over 24,000 SSLPs, genes, and ESTs provide a direct link to the rat genome sequence. Genome Res. 14: 750-757.

Kwok C, R. M. Korn, M. E. Davis, D. W. Burt, R. Critcher, L. McCarthy, *et al.*, 1998 Characterization of whole genome radiation hybrid mapping resources for non-mammalian vertebrates. Nucleic Acids Res. 26: 3562–3566.

Lahn, B. T., N. M. Pearson and K. Jegalian, 2001 The human Y chromosome, in the light of evolution. Nat. Rev. Genet. 2: 207–216.

Lander, E. and L. Kruglyak, 1995 Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genetics 11: 241-247.

Lee B.-Y., W.-J. Lee, J. T. Streelman, K. L. Carleton, A. E. Howe, *et al*., 2005 A second-generation genetic linkage map of tilapia (*Oreochromis spp*.). Genetics 170: 237-44.

Licht, L. E., L. A. Lowcock, 1991 Genome size and metabolic-rate in salamanders. Comp. Biochem. Physiol. B, Biochem. Mol. Biol. 100: 83-92.
Lindsley, D. L., G. Fankhauser and R. R. Humphrey, 1956 Mapping centromeres in the axolotl. Genetics 41: 5 8-64.
Liu Z.Y., Moore P.H., Ma H., Ackerman C.M., Ragiba M, *et al*., 2004 A primitive Y chromosome in papaya marks incipient sex chromosome evolution. Nature 427: 348-352.
Ma H., P.H. Moore, Z.Y. Liu, M.S. Kim, Q.Y. Yu, *et al*., 2004 High-density linkage mapping revealed suppression of recombination at the sex determination locus in papaya. Genetics 166: 419-436.
Maddison, W. P., M. J. Donoghue, and D. R. Maddison, 1984 Outgroup analysis and parsimony. Syst. Zool. 33: 83-103.
Malacinski, G. M., 1978 The Mexican axolotl, *Ambystoma mexicanum*: its biology and developmental genetics, and its autonomous cell-lethal genes. Am. Zool. 18: 195–206.
Malacinski, G. M., 1989 Developmental genetics pp. 102-109 in *Developmental Biology of the Axolotl*, edited by J.B. Armstrong and G.M. Malacinski. Oxford University Press, New York.
Mank J. E., D. E. L Promislow and J. C Avise, 2006 Evolution of alternative sex-determining mechanisms in teleost fishes. Biol. J. Linn. Soc. Lond. 87: 83-93.
Marchand, J. E., X. H. Yang, D. Chikaraishi, J. Krieger, H. Breer, *et al*., 2004 Olfactory receptor gene expression in tiger salamander olfactory epithelium. J. Comp. Neurol. 747: 453-467.
Martin A. P. and S. R. Palumbi, 1993 Rate of mitochondrial DNA evolution is slow in sharks compared to mammals. Proc. Natl. Acad. Sci. USA 90: 4087–4091.
Martin, A., 2001 The phylogenetic placement of Chondrichthyes: inferences from analysis of multiple genes and implications for comparative studies. Genetica 111: 349-357.
Martinez, O. and R. N. Curnow, 1992 Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. 85: 480-488.
Matsabana, J. S., D. W. Burt, C. M. O'Brien, A. Vignal, V. Fillon, *et al*., 2004 Molecular cytogenetic definition of the chicken genome: the first complete avian karyotype. Genetics 166: 1367-1373.
Matsuda Y., C. Nishida-Umehara, H. Tarui, A. Kuroiwa, K. Yamada, *et al*., 2005 Highly conserved linkage homology between birds and turtles: bird and turtle chromosomes are precise counterparts of each other. Chromosome Res. 13: 601-615.
Matsuda, R., 1987 *Animal Evolution in Changing Environments With Special Reference to Abnormal Metamorphosis*. John Wiley & Sons, New York.
McKinney, M. L. and K. J. McNamara, 1991 *Heterochrony: The Evolution of Ontogeny*. Plenum Press, New York.
Meer, J. M., R. H. Cudmore Jr. and K.F. Manly, 2004 MapManager QTX, http://ww.mapmanager.org/mmQTX.html.
Metzgar M., J. Bytof and C. Willis, 2000 Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res. 10: 72-80.

Meyer A, Van de Peer Y., 2005 From 2R to 3R: evidence for a fish-specific whole genome duplication (FSGD). BioEssays 27: 937-945.
Moczek, A. P., and H. F. Nijhout, 2002 Developmental mechanisms of threshold evolution in a polyphenic beetle. Evol. Dev. 4: 252–264.
Moore, D. P., and T. L. Orr-Weaver, 1998 Chromosome segregation during meiosis: building an unambivalent bivalent. Curr. Top. Dev. Biol. 37: 263-299.
Morescalchi A., 1975 Chromosome evolution in the caudate Amphibia. Evol. Biol. 8: 339-387.
Morgante M., M. Hanafey, W. Powell, 2002 Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Genet. 30: 194-200.
Mott R., 1997. EST_GENOME: a program to align spliced DNA sequences to unspliced genomic DNA. Comput. Appl. Biosci. 13: 477-478.
Mouse Genome Sequencing Consortium (MGSC), 2002 Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520-562.
Murphy W. J.,G. Bourque, G. Tesler, P. Pevzner, S. J. O'Brien, 2003 Reconstructing the genomic architecture of mammalian ancestors using multispecies comparative maps. human genome. Hum. Genomics 1: 30-40.
Nadir, E., H. Margalit, T. Gallily, S. A. Ben-Sasson, 1996 Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. Proc. Natl. Acad. Sci. USA 93: 6470-6475.
Namiki, J. and C. H. Tator, 1999 Cell proliferation and nestin expression in the ependyma of the adult rat spinal cord after injury. J. Neuropathol. Exp. Neurol. 58: 489-498.
Nanda, I., T. Haaf, M. Schartl, M. Schmid, D. W. Burt, 2002 Comparative mapping of Z-orthologous genes in vertebrates: implications for the evolution of avian sex chromosomes. Cytogenet. Genome Res. 99: 178-184.
Nanda, I., Z. Shan, M. Schartl, D. W. Burt, M. Koehler, *et al.*, 1999 300 million years of conserved synteny between chicken Z and human chromosome 9. Nat. Genet. 21: 258-259.
Nanda, I., E. Zend-Ajusch, Z. Shan, F. Grutzner, M. Schartl, *et al.*, 2000 Conserved synteny between the chicken Z sex chromosome and human chromosome 9 includes the male regulatory gene DMRT1: a comparative (re)view on avian sex determination. Cytogenet. Genome Res. 89: 67-78.
Nanda, I. and M. Schmid, 2002 Conservation of avian Z chromosomes as revealed by comparative mapping of the Z-linked aldolase B gene. Cytogenet. Genome Res. 96: 176-178.
Naruse, K., M. Tanaka, K. Mita, A. Shima, J. Postlethwait, and H. Mitani, 2004 A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. Genome Res. 14: 820-828.
Nelson, J. S., 1994 *Fishes of the World*. 3rd edition. New York: John Wiley.
Norris T. B, G. K. Rickards, and C. H. Daugherty, 2004 Chromosomes of the tuatara, *Sphenodon*, a chromosome heteromorphism and an archaic reptilian karyotype. Cytogenet. Genome Res. 105: 93-99.
Ogata, M., H. Ohtani, T. Igarashi, Y. Hasegawa, Y. Ichikawa and I. Miura, 2003 Change of the heterogametic sex from male to female in the frog. Genetics 164: 613-620.

Ohno, S., 1967 Sex chromosomes and sex linked genes, pp. 1-192 in *Monographs on Endocrinology, Vol. 1*, edited by A. Labhart, T. Mann, and L. T. Samuels. Springer, New York.

Ohta, Y., W. Goetz, M. Z. Hossain, M. Nonaka, M. F. Flajnik, 2006 Ancestral organization if the MHC revealed in the amphibian *Xenopus*. J. Immunol. 176: 3674-3685.

Olmo, E., 1990 *Cytogenetics of Reptiles and Amphibians*. Birkhauser Verlag: Basel, Switzerland.

Ott, J., 1991 *Analysis of Human Genetic Linkage. 3rd edition*. Baltimore: Johns Hopkins University Press.

Panopoulou G. and A. J. Poustka, 2005 Timing and mechanism of ancient vertebrate genome duplications -- the adventure of a hypothesis. Trends Genet. 21: 559-567.

Pardo-Manuel De Villena, F. and C. Sapienza, 2001 Recombination is proportional to the number of chromosome arms in mammals. Mamm. Genome 12: 318-322.

Parichy, D. M., 1996 Pigment patterns of larval salamanders (Ambystomatidae, Salamandridae): The role of the lateral line sensory system and the evolution of pattern-forming mechanisms. Dev. Biol. 175: 265-282.

Parichy, D. M., 1998 Experimental analysis of character coupling across a complex life cycle: Pigment pattern metamorphosis in the tiger salamander, *Ambystoma tigrinum tigrinum*. J. Morph. 237: 53-67.

Parichy, D. M., M. Stigson and S. R. Voss, 1999 Genetic analysis of steel and the PG-M/versican-encoding gene AxPG as candidates for the white (d) pigment mutant in the salamander *Ambystoma mexicanum*. Dev. Genes Evol. 209: 349-356.

Park, D., J. M. McGuire, A. L. Majchrzak, J. M. Ziobro, and H. L. Eisthen, 2004 Discrimination of conspecific sex and reproductive condition using chemical cues in axolotls (*Ambystoma mexicanum*). J. Comp. Physiol. A 190: 145-427.

Parmenter, C. L., 1919 Chromosome number and pairs in the somatic mitoses of *Ambystoma tigrinum*. J. Morph. 33: 169-242.

Pask, A. J. and J. A. Graves, 1999 Sex chromosomes and sex-determining genes: insights from marsupials and monotremes. Cell. Mol. Life Sci. 55: 864-875.

Pieau, C., M. Dorizzi and N. Richard-Mercier, 1999 Temperature dependant sex determination and gonadal differentiation in reptiles. Cell. Mol. Life Sci. 55: 887-900.

Postlethwait, J. H., I. G. Woods, P. Ngo-Hazelette, Y.-L. Yan, P. D. Kelly, *et al*., 2000 Zebrafish comparative genomics and the origins of vertebrate chromosomes. Genome Res. 10: 1890-1902.

Powers, J. H., 1907 *Morphological variation and its causes in Ambystoma tigrinum.* Student University of Nebraska 7: 197-274.

Putta, S., J. J. Smith, J. Walker, M. Rondet, D. W. Weisrock, *et al.*, 2004 From biomedicine to natural history research: EST resources for ambystomatid salamanders. BMC Genomics 5: 54.

Raff, R. A., 1996 *The Shape of Life: Genes, Development, and the Evolution of Animal Form*. University of Chicago Press, Chicago.

Raff, R. A. and G. A. Wray, 1989 Heterochrony: developmental mechanisms and evolutionary results. J. Evol. Biol. 2: 409–434.

Ramsay, L., M. Macaulay, L. Cardle, M. Morgante, S. delgi Ivanissevich, *et al*., 1999 Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant J. 17: 415-425.

Rat Genome Secuencing Project Consortium (RGSPC) 2004 Genome sequence of the Brown Norway Rat yields insights into mammalian evolution. Nature 428: 493-521.

Reese, R. W., 1969 *The taxonomy and ecology of the tiger salamander Ambystoma tigrinum of Colorado*. Unpublished Ph.D. Dissertation.

Reese, R. W., 1972 Taxonomy and Distribution of the tiger salamanders in Colorado. Transactions Kansas Acad. Sci. 75: 128-140.

Rens, W., F. Grützner, P.C. O'Brien, H. Fairclough, J.A.M. Graves, *et al*., 2004 Resolution and evolution of the duck-billed platypus karyotype with an X1Y1 X2Y2 X3Y3 X4Y4 X5Y5 male sex chromosome constitution. Proc. Natl. Acad. Sci. USA 101: 16257-16261.

Richard, F., M. Lombard and B. Dutrillaux, 2003 Reconstruction of the ancestral karyotype of eutherian mammals. Chromosome Res. 11: 605-618.

Richards, C.M. and G.W. Nace, 1978 Cytogenetic and hormonal sex reversal used in testes of the XX-XY hypothesis of sex determination in *Rana pipiens*. Growth 42: 319–331

Riesz, R. R. and J. Muller, 2004 Molecular timescales and the fossil record: a paleontological perspective. Trends Genet. 20: 237-241.

Riley, S. P. D., H. B. Shaffer, S. R. Voss and B. M. Fitzpatrick, 2003 Hybridization between a rare, native tiger salamander (*Ambystoma californiense*) and its introduced congener. Ecol. Appl. 13: 1263-1275.

Robertson A.V., C. Ramsden, J. Niedzwiecki, J.Z. Fu and J.P. Bogart, 2006 An unexpected recent ancestor of unisexual *Ambystoma*. Mol. Ecol. 15: 3339-3351.

Roeder, G. S., 1997 Meiotic chromosomes: it takes two to tango. Genes Dev. 11: 2600-2621.

Roff, D. A., 1986 The genetic basis of wing dimorphism in the sand cricket, *Gryllus firmus* and its relevance to the evolution of wing dimorphism in insects. Evolution 40: 1009–1020.

Roscoe, J. T. and J. A. Byars, 1971 An investigation of the restraints with respect to sample size commonly imposed on the use of the chi-square statistic. J. Am. Stat. Assoc. 66: 755-759.

Rose, F. L. and D. Armentrout, 1975 Adaptive strategies of *Ambystoma tigrinum* Green inhabiting the Llano Estacado of West Texas. J. Anim. Ecol. 45: 713-739.

Ross, M. T., D. V. Grafham, A. J. Coffey, S. Scherer, K. McLay, *et al*., 2005 The DNA sequence of the human X chromosome. Nature 434: 325-337.

Rozen, S. and H. J. Skaletsky, 1999 Primer 3. http://frodo.wi.mit.edu/primer3/primer3_code.html

Rubbo, M. J. and J. M. Kiesecker, 2005 Amphibian breeding distribution in an urbanized landscape. Conserv Biol. 19: 504-511.

Ruta, M., M. I. Coates and D. J. Quicke, 2003.Early tetrapod relationships revisited. Biol. Rev. Camb. Philos. Soc. 78: 251-345.

Sakata, N, Y. Tamori, M. Wakahara, 2005 P450 aromatase expression in the temperature-sensitive sexual differentiation of salamander (*Hynobius retardatus*) gonads. Int. J. Dev. Biol. 49: 417-425.

Samuels, A. K., D. Weisrock, J. J. Smith, S. Putta, J. Walker, K. France, and S. R. Voss, 2005 Transcriptional and phylogenetic analysis of 5 complete ambystomatid salamander mitochondrial genomes. Gene 349: 43-53.

Schmid, M, I. Nanda, M. Guttenbach, C. Steinlein, H. Hoehn, *et al*., 2000 First report on chicken genes and chromosomes. Cytogenet. Cell Genet. 90: 169–218.

Schmid, M., I. Nanda, C. Steinlein, K. Kausch, J. T. Epplen, *et al*., 1991 Sex determining mechanisms and sex chromosomes in amphibia, pp. 393-430 in *Amphibian Cytogenetics and Evolution*, edited by D. M. Green and S. K. Sessions. Academic Press: New York, pp. 393-430.

Schnapp, E. and E. M. Tanaka, 2005 Quantitative evaluation of morpholino-mediated protein knockdown of GFP, MSX1, and PAX7 during tail regeneration in *Ambystoma mexicanum*. Dev. Dyn. 232: 162-170.

Semlitsch, R. D., D. E. Scott and J. H. K. Pechmann, 1988 Time and size at metamorphosis related to adult fitness in *Ambystoma talpoideum*. Ecology 69: 184–192.

Semlitsch, R. D., R. N. Harris and H. M. Wilbur, 1990 Paedomorphosis in *Ambystoma talpoideum*: maintenance of population variation and alternative life history pathways. Evolution 44: 1604–1613.

Sessions, S.K., 1982 Cytogenetics of diploid and triploid salamanders of the *Ambystoma jeffersonianum* complex. Chromosoma 77: 599-621.

Shaffer, H. B., 1984a Evolution in a paedomorphic lineage. 1. An electrophoretic analysis of the Mexican ambystomatid salamanders. Evolution 38: 1194–1206.

Shaffer, H. B., 1984b Evolution in a paedomorphic lineage. II. allometry and form in the Mexican ambystomatid salamanders. Evolution 38: 1207-1218.

Shaffer, H. B., and M. L. McKnight, 1996 The polytypic species revisited: genetic differentiation and molecular phylogenetics of the tiger salamander species *Ambystoma tigrinum* (Amphibia: Caudata) species complex. Evolution 50: 417-433.

Shaffer, H. B., and S. R. Voss, 1996 Phylogenetic and mechanistic analysis of a developmentally integrated character complex: alternate life history modes in ambystomatid salamanders. Am. Zoo. 36: 24-35.

Shetty, S., D. K. Griffin and J. A. M. Graves, 1999 Comparative painting reveals strong chromosome homology over 80 million years of bird evolution. Chromosome Res. 7: 289-295.

Shine, R., 1999 Why is sex determined by nest temperature in so many reptiles? Trends Ecol. Evol. 14: 186-189.

Shubin, N. H., 2002 Origin of evolutionary novelty: Examples from limbs. J. Morphol. 252: 15-28.

Simon, H.-G., 1999 T-box genes and the formation of vertebrate forelimb- and hindlimb specific pattern. Cell Tissue Res. 296: 57-66.

Sinclair, A., H. P. Berta, M. S. Palmer, J. R. Hawknis, B. L. Griffiths, *et al*., 1990 A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346: 240-244.

Sinclair, J. H., C. R. Carrol and G. M. Malacinski, 1978 Composition and behavior of the nucleolar organizer of the nucleolar variant of axolotls; review and comparison to other systems. Am. Zool. 18: 225-236.

Smith, C. A. and A. H. Sinclair, 2004 Sex determination: Insights from chicken. Bioessays 26: 120-132.

Smith, H. M., 1989 Discovery of the axolotl and its early history in biological research, pp. 1-12 in *Developmental Biology of the Axolotl*, edited by J.B. Armstrong and G.M. Malacinski. Oxford University Press, New York.

Smith, J. J., D. K. Kump, J. A. Walker, D. M. Parichy and S. R. Voss, 2005 A comprehensive expressed sequence tag linkage map for tiger salamander and Mexican axolotl: enabling gene mapping and comparative genomics in *Ambystoma*. Genetics 171: 1161-1171.

Smith, J. J., S. Putta, J. A. Walker, D. K. Kump, A. K. Samuels, *et al*., 2005 Integrating new and existing ambystomatid salamander research and informational resources. BMC Genomics 2005, 6: 181.

Smith, J. J. and S. R. Voss, 2006 Gene order data from a model amphibian (*Ambystoma*): new perspectives on vertebrate genome structure and evolution. BMC Genomics 7: 219.

Smith, J. J., 2002 *Genetics in Ambystoma: metamorphic failure and sex determination*. MS Thesis, Colorado State University, Fort Collins, CO.

Sobkow, L., H. H. Epperlein, S. Herklotz, W. L. Straube and E. M. Tanaka, 2006 A germline GFP transgenic axolotl and its use to track cell fate: Dual origin of the fin mesenchyme during development and the fate of blood cells during regeneration. Dev. Biol. 290: 386-397.

Sokal, F. J., and R. R. Rohlf, 1995 *Biometry: the principles and practice of statistics in biological research 3rd ed.* W.H. Freeman and Company, New York.

Song G., C. Cechvala, D. K. Resnick, R. J. Dempsey and V. L. R. Rao, 2001 GeneChip analysis after acute spinal cord injury in rat. J Neurochem. 79: 804–815.

Springer, M. S., W. J. Murphy, E. Eizirik and S. J. O'Brien, 2003 Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc. Natl. Acad. Sci. USA 100: 1056-1061.

Sprules, W. G., 1974 The adaptive significance of paedogenesis in North American species of *Ambystoma* (Amphibia: Caudata): an hypothesis. Can. J. Zool. 52: 393–400.

Steen, R. G., A. E. Kitweck-Black, C. Glenn, J. Gullings-Handley, W. Van Etten, *et al*., 1999 A high density integrated genetic linkage and radiation hybrid map of the laboratory rat. Genome Res. 9: AP1-8.

Stemshorn, K. C., A. W. Nolte and D. A. Tautz, 2005 Genetic Map of *Cottus gobio* (Pisces, Teleostei) based on microsatellites can be linked to the Physical Map of *Tetraodon nigroviridis*. J. Evol. Biol. 18: 1619-1624.

Stevens, P. F., 1980 Evolutionary polarity of character states. Annu. Rev. Ecol. Syst. 11: 333-358.

Stopper, G. F. and G. P. Wagner, 2007 Inhibition of sonic hedgehog signaling leads to posterior digit loss in *Ambystoma mexicanum*: Parallels to natural digit reduction in urodeles. Dev. Dyn. 236: 321-331.

Straus, N. A., 1971 Comparative DNA renaturation kinetics in amphibians. Proc. Nat. Acad. Sci. USA 68: 799–802.
Sturtevant, A. H., 1913 The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J. Exp. Zool. 14: 43-59.
Sumida, M. and M. Nishioka, 2000 Sex-linked genes and linkage maps in amphibians. Comp. Biochem. Physiol. Biochem. Mol. Biol. 126: 257-270.
Sybengia, J., 1992 Cytogenetics in plant breeding. *Monographs on Theoretical and Applied Genetics 17.* Springer-Verlag, Berlin.
Takahashi M., A. Yasuhisa, H. Kurosawa, N. Sueyoshi and S. Shirai, 2003 Ependymal cell reactions in spinal cord segments after compression injury in adult rat. J Neuropath. Exp. Neurol. 62: 185–194.
Thoreson, W. B., K. Rabl, E. Townes-Anderson and R. Heilelberger, 2004 A highly Ca2+-sensitive pool of vesicles contributes to linearity at the rod photoreceptor ribbon synapse. Neuron 42: 595-605.
Tompkins, R., 1978 Genetic control of axolotl metamorphosis. Am. Zool. 18: 313-319.
Toth, G., Z. Gaspari and J. Jurka, 2000 Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res. 10: 967-981.
Trask, B., D. Pinkel and G. van den Engh, 1989 The proximity of DNA sequences in interphase cell nuclei is correlated to genomic distance and permits ordering of cosmids spanning 250 kilobase pairs. Genomics 5: 710-717.
Trenham, P.C. and H.B. Shaffer, 2005 Amphibian upland habitat use and its consequences for population viability. Ecol. Appl. 15: 1158-1168.
True, J. R., B. S. Weir and C. C. Laurie, 1996 A genome-wide survey of hybrid incompatibility factors by the introgression of marked segments of *Drosophila mauritiana* chromosomes into *Drosophila simulans*. Genetics 142: 819-837.
van Nimwegen E, N. Paul, R. Sheridan, and M. Zavolan, 2006 SPA: a probabilistic algorithm for spliced alignment. PLoS Genet. 2: e24.
van Rheede, T., T. Bastiaans, D. N. Boone, S. B. Hedges, W. W. de Jong, *et al*., 2006. The platypus Is in Its place: nuclear genes and indels confirm the sister group relation of monotremes and therians. Mol. Biol. Evol. 23: 587-597.
Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, *et al*., 2001. The sequence of the human genome. Science 291: 1304 –1351.
Voorrips, R. E., 2002 MapChart. http://www.plant.wageningenur.nl/products/mapping/MapChart.
Voss, S. R., 1993 Randomly amplified polymorphic DNA (RAPD) analysis of ambystomatid salamanders. Axolotl Newsl. 22: 28–32.
Voss, S. R., 1995 Genetic basis of paedomorphosis in the axolotl *Ambystoma mexicanum*: a test of the single gene hypothesis. J. Hered. 86: 441-447.
Voss, S. R. and H. B. Shaffer, 1996 What insights into the developmental traits of urodeles does the study of interspecific hybrids provide? Int. J. Dev. Biol. 40: 885-893.
Voss, S. R. and H. B. Shaffer, 1997 Adaptive evolution via a major gene effect: paedomorphosis in the Mexican axolotl. Proc. Natl. Acad. Sci. USA 94: 14185-14189.

Voss, S. R. and H. B. Shaffer, 2000 Evolutionary genetics of metamorphic failure using wild-caught *vs.* laboratory axolotls (*Ambystoma mexicanum*). Mol. Ecol. 12: 1217–1223.
Voss, S. R. and J. J. Smith, 2005 Evolution of salamander life cycles: A major effect QTL contributes to discreet and continuous variation for metamorphic timing. Genetics 170: 275-281.
Voss, S. R., J. J. Smith, D. M. Gardiner and D. M. Parichy, 2001 Conserved vertebrate chromosomal segments in the large salamander genome. Genetics 158: 735-746.
Voss, S. R., K. L. Prudic, J. C. Oliver and H. B. Shaffer, 2003 Candidate gene analysis of metamorphic timing in ambystomatid salamanders. Mol. Ecol. 12: 1217–1223.
Wallace, H. G., M. I. Badawy and B. M. Wallace, 1999 Amphibian sex determination and sex reversal. Cell. Mol. Life Sci. 55: 901-909.
Wallace, H. G. and B. M. Wallace, 2000 Sex reversal of the newt *Triturus cristatus* reared at extreme temperature. Int. J. Dev. Biol. 44: 807-810.
Wardrop, J., J. Snape, W. Powell and C. G. Machray, 2002 Constructing plant radiation hybrid panels. Plant J. 31: 223-228.
Waters, P. D., M. C. Wallis and J. A. M. Graves, 2007 Mammalian sex – origin and evolution of the Y chromosome and SRY. Semin. Cell Dev. Biol. doi:10.1016/j.semcdb.2007.02.007.
Watrous, L. E. and Q. E. Wheeler, 1981 The out-group comparison method of character analysis. Syst. Zool. 30: 1-11.
Weisrock, D. W., H. B. Shaffer, B. L. Storz, S. R. Storz and S. R. Voss, 2006 Multiple nuclear gene sequences identify phylogenetic species boundaries in the rapidly radiating clade of Mexican ambystomatid salamanders. Mol. Ecol. 15: 2489-2503.
West-Eberhard, M. J., 2004 *Developmental Plasticity and Evolution*. Oxford University Press, New York.
Wilbur, H. M. and J. P. Collins, 1973 Ecological aspects of amphibian metamorphosis. Science 182: 1305–1314.
Wittbrodt, J., A. Shima and M. Schartl, 2002 Medaka – a model organism from the far East. Nat. Rev. Genet. 3: 53–64.
Woods, I. G., C. Wilson, B. Friedlander, P. Chang, D. K. Reyes, *et al*., 2005 The zebrafish gene map defines ancestral vertebrate chromosomes. Genome Res. 15: 1307-1314.
Woram R. A., C. McGowan, J. A. Stout, K. Gharbi, M. M. Ferguson, *et al*., 2004 A genetic linkage map for Arctic char (*Salvelinus alpinus*): evidence for higher recombination rates and segregation distortion in hybrid versus pure strain mapping parents. Genome 47: 304–315.
Wu, T. D. and C. K. Watanabe, 2005 GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21: 1859-1875.
Xing, L. and V. Brendel, 2000 MuSeqBox: a program for multi-query sequence BLAST output examination. Bioinformatics 17: 744-745.
Yamamoto, S., M. Nagao, M. Sugimori, H. Kosako, H. Nakatomi, *et al*., 2001 Trancription factor expression and notch-dependent regulation of neural progenitors in the adult rat spinal cord. J. Neurosci. 21: 9814–9823.
Yang, F., E. Z. Alkalaeva, P. L. Perelman, A. T. Pardini, W. R. Harrison, *et al*., 2003 Reciprocal chromosome painting among human, aardvark, and elephant

(superorder Afrotheria) reveals the likely eutherian ancestral karyotype. Proc. Natl. Acad. Sci. USA 100: 1062-1066.
Yap, V. B. and L. Pachter, 2004 Identification of evolutionary hotspots in the rodent genomes. Genome Res. 14: 574-579.
Ye, L. and X. Huang, 2005 MAP2: multiple alignment of syntenic genomic sequences. Nucleic Acids Res. 33: 162-170.
Zeng, Z.-B., 1993 Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl. Acad. Sci. USA 90: 10972-10976.
Zeng, Z.-B., 1994 Precision mapping of quantitative trait loci. Genetics 136: 1457-1468.
Zhang, C., D. K. Dube, X. Huang, R. W. Zajdel, R. Bhatia, *et al*., 2003 A point mutation in bioactive RNA results in the failure of mutant heart correction in Mexican axolotls. Anat. Embryol. 206: 495–506.
Zhang, C., K.M. Pietras, G.F. Sferrazza, P. Jia, G. Athauda, *et al*., 2007 Molecular and immunohistochemical analyses of cardiac troponin T during cardiac development in the Mexican axolotl, *Ambystoma mexicanum*. J. Cell Biochem. 100: 1-15.
Zhang, J. and S. M. Wu, 2003 $G_O\alpha$ labels ON bipolar cells in the tiger salamander retina. J. Comp. Neurol. 461: 276–289.
Zhang, C., F. Meng, X. P. Huang, R. Zajdel, S. L. Lemanski, *et al*., 2004 Down regulation of N1 gene expression inhibits the initial heart beating and heart development in axolotls. Tissue and Cell 36: 71-81.

Vita

JERAMIAH J. SMITH

DATE AND PLACE OF BIRTH: July 4, 1976
Rapid City, South Dakota

EDUCATION: Black Hills State University, B.S. in Biology, *cum laude*, 1998
Colorado State University, M.S. in Biology, 2002.

PROFESSIONAL POSITIONS HELD:

Undergraduate Research Assistant, Black Hills State University, Fall 1996 – Fall 1999.
Teaching Assistant, Colorado State University, Spring 1999, Fall 2000, Fall 2001.
Research Assistant, Colorado State University, Summer 1999 – Summer 2000, Spring 2001 – Summer 2001, Spring 2002.
Research Assistant, University of Kentucky, Summer 2002 – Current.
Research Fellow, University of Kentucky, 2002-2003, 2006-Current.

SCHOLASTIC AND PROFESSIONAL HONORS:

Black Hills State University Biology Scholarship (1997)
Kentucky Opportunity Fellowship, University of Kentucky (2002)
Travel Support: G. Flora Ribble Enrichment Fund (2002 – 2005)
Graduate Student Support Grant, University of Kentucky (2003)
Presidential Graduate Fellowship, University of Kentucky (2006)

PROFESSIONAL PUBLICATIONS: (* denotes equal contribution for primary authorship)

***** Putta, S., J. J. Smith, C. Staben, S. R. Voss, 2007 MapToGenome: A Comparative Genomic Tool that Aligns Transcript Maps to Sequenced Genomes. Evolutionary Bioinformatics Online 2: 15-25.

Smith, J. J., S. R. Voss, 2006 Gene order data from a model amphibian (*Ambystoma*): new perspectives on vertebrate genome structure and evolution. BMC Genomics 7: 219.

Page, R. B., J. R. Monaghan, A. K. Samuels, J. J. Smith, C. K. Beachy, S. R. Voss, 2006 Microarray analysis identifies keratin loci as sensitive biomarkers for thyroid hormone disruption in the salamander *Ambystoma mexicanum*. Comparative Biochemistry and Physiology, Part *C*. 145: 15-27.

Smith, J. J., D. K. Kump, J. A. Walker, D. M. Parichy, S. R. Voss, 2005 A Comprehensive Expressed Sequence Tag Linkage Map for Tiger Salamander and Mexican Axolotl: Enabling Gene Mapping and Comparative Genomics in Ambystoma. Genetics 171: 1161-1171.

* Voss, SR, J. J. Smith, 2005 Evolution of salamander life cycles: A major-effect quantitative trait locus contributes to discrete and continuous variation for metamorphic timing. Genetics 170: 275-281.

Smith, J. J., S. Putta, J. A. Walker, D. K. Kump, A. K. Samuels *et al*., 2005 Sal-Site: Integrating new and existing ambystomatid salamander research and informational resources. BMC Genomics 6: 181.

Samuels, A. K., D. W. Weisrock, J. J. Smith, K. J. France KJ, J. A. Walker JA *et al*., 2005 Transcriptional and phylogenetic analysis of five complete ambystomatid salamander mitochondrial genomes. Gene 349: 43-53.

* Putta, S., J. J. Smith, J. A. Walker, M. Rondet, D. W. Weisrock *et al*., 2004 From biomedicine to natural history research: EST resources for ambystomatid salamanders. BMC Genomics 5: 54.

Voss, S. R., J. J. Smith, D. M. Gardiner, D. M. Parichy, 2001 Conserved vertebrate chromosome segments in the large salamander genome. Genetics 158: 735-746.