Advancing the lizard, *Anolis carolinensis*, as a model system for genomic studies of evolution, development and regeneration.

by

Walter Eckalbar

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2012 by the
Graduate Supervisory Committee:

Kenro Kusumi, Chair
Matthew Huentelman
Jeffery Rawls
Norma Wilson-Rawls

ARIZONA STATE UNIVERSITY

December 2012

ABSTRACT

Well-established model systems exist in four out of the seven major classes of vertebrates. These include the mouse, chicken, frog and zebrafish. Noticeably missing from this list is a reptilian model organism for comparative studies between the vertebrates and for studies of biological processes unique to reptiles. To help fill in this gap the green anole lizard, *Anolis carolinensis*, is being adapted as a model organism. Despite the recent release of the complete genomic sequence of the *A. carolinensis*, the lizard lacks some resources to aid researchers in their studies. Particularly, the lack of transcriptomic resources for lizard has made it difficult to identify genes complete with alternative splice forms and untranslated regions (UTRs). As part of this work the genome annotation for *A. carolinensis* was improved through next generation sequencing and assembly of the transcriptomes from 14 different adult and embryonic tissues. This revised annotation of the lizard will improve comparative studies between vertebrates, as well as studies within *A. carolinensis* itself, by providing more accurate gene models, which provide the bases for molecular studies. To demonstrate the utility of the improved annotations and reptilian model organism, the developmental process of somitogenesis in the lizard was analyzed and compared with other vertebrates. This study identified several key features both divergent and convergent between the vertebrates, which was not previously known before analysis of a reptilian model organism. The

improved genome annotations have also allowed for molecular studies of tail regeneration in the lizard. With the annotation of 3' UTR sequences and next generation sequencing, it is now possible to do expressional studies of miRNA and predict their mRNA target transcripts at genomic scale. Through next generation small RNA sequencing and subsequent analysis, several differentially expressed miRNAs were identified in the regenerating tail, suggesting miRNA may play a key role in regulating this process in lizards. Through miRNA target prediction several key biological pathways were identified as potentially under the regulation of miRNAs during tail regeneration. In total, this work has both helped advance *A. carolinensis* as model system and displayed the utility of a reptilian model system.

DEDICATION

I dedicate this work to my loving and supportive family; it would not have been possible without them. To my parents who have fostered my education and interests my whole life, to my two children, Caleb and Kylie, who provide my inspiration, and especially to my wife Kelly who bravely moved to Arizona to support me through graduate school.

ACKNOWLEDGMENTS

I would first like to acknowledge my advisor, Dr. Kenro Kusumi. Without his support and guidance I could never have completed this dissertation. I would also like to acknowledge my committee members Dr. Alan Rawls, Dr. Jeanne Wilson-Rawls and Dr. Matthew Huentelman. They have provided wonderful support and advisement. I would also like to acknowledge Dr. Michael Caplan, who helped me greatly early in my graduate career, before my interests lead me away from the mathematical modeling aspects of biology.

TABLE OF CONTENTS

Page

LIST OF TABLES vii

LIST OF FIGURES ix

CHAPTER

1 INTRODUCTION 1

Annotation of the *Anolis carolinensis* Genome 3

Evolution of the Developmental Process of Somitogenesis ... 6

MicroRNA regulation of tail regeneration 12

2 GENOME REANNOTATION OF THE LIZARD ANOLIS CAROLINENSIS BASED ON 14 ADULT AND EMBRYONIC DEEP TRANSCRIPTOMES 18

Abstract 18

Background 19

Materials and methods 22

Results and discussion 27

Conclusions 43

3 SOMITOGENESIS IN THE ANOLE LIZARD AND ALLIGATOR REVEALS A TRANSITION IN EVOLUTION OF THE SEGMENTATION CLOCK 45

Abstract 45

Background 46

CHAPTER Page

Materials and methods 49

Results 56

Discussion....78

Conclusions 86

4 DEEP SEQUENCING IDENTIFIES NOVEL MICRORNAS LINKED TO STEM CELL PROLIFERATION AND DIFFERENTIATION DURING TAIL REGENERATION IN THE LIZARD ANOLIS CAROLINENSIS.... 87

Abstract....87

Background....88

Materials and methods 92

Results 97

Discussion....120

Conclusions 126

5 CONCLUSION 128

REFERENCES 135

LIST OF TABLES

Table Page

1. Overview of de novo transcript assembly for A. carolinensis based on RNA-Seq data from 14 adult and embryonic tissues and deposited EST sequence data 28
2. Comparison of ASU, NCBI and Ensembl gene annotations of the A. carolinensis genome 34
3. A. carolinensis genes unique to the ASU annotation with vertebrate orthologues 41
4. RNA expression levels of Anolis carolinensis orthologues of key somitogenesis genes 60
5. Overview of small RNA sequencing for annotation and quantification of miRNAs in the A. carolinensis in skeletal muscle, brain, as well as the tip and base of the regenerating tail 101
6. Summary of the top scoring novel miRNAs predicted by miRDeep2 103
7. Summary of the top differentially expressed miRNAs between the tip and base of the regenerating tail 110
8. Functional annotation clusters of GO biological processes predicted to be the targets of miRNAs up regulated in the tip of the regenerating tail 116

Table Page

9. Functional annotation clusters of GO biological processes predicted to be the targets of miRNAs down regulated in the tip of the regenerating tail 117

10. Functional annotation clusters of GO biological processes predicted to be the targets of miRNAs up regulated and transcriptionally down regulated in the tip of the regenerating tail 118

11. Functional annotation cluster of GO biological processes predicted to be the targets of miRNAs down regulated and transcriptionall up regulated in the tip 119

LIST OF FIGURES

Figure Page

1. Our methods and data sources for creating ASU annotations of the Anocarol2.0 genome 30
2. Increased N50 transcript length and number of predicted transcripts in the ASU annotation 35
3. Comparison of gene expression levels in the 14 adult and embryonic tissue transcriptomes using the *A. carolinensis* revised annotations 38
4. Next-gen transcriptome sequencing identifies segmentation clock genes in the green anole lizard, *Anolis carolinensis* 58
5. PSM gradient and determination front gene expression is conserved in the green anole, but *hes6a* PSM gradient expression is unique among amniotes 62
6. Lunatic fringe is not expressed in the PSM of green anole embryos 66
7. *Anolis carolinensis dll1* displays cycling expression in the presomitic mesoderm 71
8. *Anolis carolinensis dll3* is expressed in the presomitic mesoderm, and has a domain organization conserved with anamniotes 73

Figure Page

9. *Alligator mississippiensis* embryos do not display *LFNG* expression in the PSM, in a comparable pattern with *A. carolinensis* and unlike the mouse and the chicken, and expression of *DLL1* in the alligator PSM 76

10. Diagram comparing how the phylogenetic relationships of the zebrafish, frog *Xenopus laevis*, lizard *Anolis carolinensis*, chicken and mouse (Hedges & Kumar, 2009) with conserved and divergent features of the segmentation clock 83

11. An adult Male *Anolis carolinensis* lizard with a regenerating tail and the experimental design for small RNA sequencing 98

12. The annotation and orthology assignment of two novel miRNAs by miRDeep2: GL344214.1_1034 and 1_16347 105

13. MicroRNA expression in the tip and base of the regenerating tail at 25 dpa 107

14. Overview of the analysis of the predicted mRNA targets of A. carolinensis miRNAs 112

Chapter 1

INTRODUCTION

The course of evolution has led to seven distinct groups of vertebrates. Included in the vertebrates are the tetrapod groups of the mammals (Mammalia), the non-avian reptiles (*Reptilia*), the birds (*Aves*), and the amphibians (*Amphibia*). The vertebrates also include the fish, which are comprised of three classes, the bony fish (*Osteichthyes*), cartilaginous fish (*Chondrichthyes*) and the jawless fish (*Agnatha).* Well-established model systems with complete genomic sequence exist in four of these groups: 1) the mouse, *Mus musculus,* 2) the chicken, *Gallus gallus*, 3) the frogs *Xenopus tropicalis* and *laevis*, and 4) the zebrafish, *Danio rerio*. Noticeably absent from this list of model organisms is a representative from the non-avian reptiles. The lack of a reptilian model organism has limited comparative studies between the vertebrates, and particularly between the amniotes, which consist of mammals, birds and reptiles. The lack of reptilian model organism has also hindered the progress of studies relating to the unique attributes of reptiles.

In an attempt to address this limitation, recent efforts have been made by the reptilian research community to establish the green anole lizard, *Anolis carolinensis*, as the first reptilian model organism with genomic sequence. As part of this effort a genomic sequencing project, which included limited transcriptomic sequencing and an initial genome

annotation, for the green anole was recently completed (Alföldi et al., 2011). As a major out group of mammals, these sequencing efforts have helped fill in a key gap for comparative genomic and evolutionary studies between the amniotes. These genomic and transcriptomic sequences from a reptilian species will aid in identification of conserved genes and gene regulatory elements in amniotes, as well help researchers in understanding processes of gene duplications, genome rearrangements and the evolution of repetitive sequences.

Among the estimated over 17,000 extant reptilian species, there were a number of reasons *A. carolinensis* was chosen for complete genomic sequencing to aid in its adoption as a model organism. The first among these reasons was the green anole's demonstrated ability to be adapted as a model organism in a laboratory setting. Anoles can be housed in cages similar to mouse or rat cages and require less husbandry cost due to their low waste production and inexpensive diet. In addition, anoles lay a single egg at regular periods, every 2-3 weeks, and do not display parental care. These combined factors make the anole lizard a robust animal model in a laboratory setting (Losos, Braun, Brown, & Clifton, 2005).

The green anole lizard has been the established itself as a key model system in several important fields relating to human health. Due to the homology between the anatomy of the green anole's brain and that of a human's, the lizard has been used as a model to study neurological and

behavioral disorders. For example, lizard territorial and courtship displays utilize a similar mechanism that is involved in obsessive-compulsive disorder and Tourette's syndrome in humans, making them an excellent model system for these diseases (Baxter, 2001; Clark & Baxter, 2000a; 2000b). *Anolis carolinensis* has also been used study the role of steroid hormones in reproductive behaviors (Lovern, Holmes, & Wade, 2004; Wade, 1999), as well as the effect of neurochemicals and endocrine responses on aggression and depression (Summers, Forster, et al., 2005a; Summers, Watt, et al., 2005b). Another factor that created a case for a compete genomic sequencing effort of *Anolis carolinensis* is the fact that it belongs to one of the most diverse genera among all animals, with nearly 400 species. During this radiation of the *Anolis* genus, species have repeatedly evolved to occupy several distinct environmental niches (Losos, Jackman, Larson, Queiroz, & Rodriguez-Schettino, 1998; E. E. Williams, 1983). The creation of a robust model system with complete genomic sequences of *Anolis carolinensis* then gives researchers the opportunity to do molecular genetic experiments following up on hypothesis that arise from comparative genomic evolution studies between the nearly 400 species of *Anolis*.

Annotation of the *Anolis carolinensis* Genome

With the complete genome sequence for a reptilian model species, the power of comparative studies of evolution in amniotes, as well as

molecular studies in the gene anole, have been greatly improved. However, the current genome annotation for the lizard is largely based on predictive algorithms, which incorporate sequence conservation with genes from related species, open reading frame predictions for protein coding regions, the consensus motifs for splicing, and the limited transcriptomic sequencing that was done in combination with the genomic sequencing (Alföldi et al., 2011). While these predictions were highly informative of the gene content in the current genomic assembly for the lizard, many important features were likely missing and could be substantially improved through next generation RNA sequencing (RNA-Seq; Haas et al., 2008; Rhind et al., 2011). In particular, the initial genome annotations poorly represented the untranslated regions (UTR) of genes, since these sequences lack the high degree of conservation observed in protein coding sequences. Since, the current annotation relies largely on alignments from other species, genes novel to the reptilian lineage and not found in bird or mammalian genomes, are also not likely to be well represented in the current genome annotation. Furthermore, because these initial genome annotation tools are primarily predictive, alternative splice forms of genes are also not well represented. Instead, most genes are currently only represented as the single longest isoform. These issues limit the utility of the lizard genome for researchers seeking to use the anole for molecular, genomic and evolutionary studies and highlight the

need for more transcriptomic sequencing to improve the genome annotation.

Recent advances in next-generation sequencings technologies have greatly reduced the time, cost and difficulty of generating novel genomes and transcriptomes. These advances have led to efforts to sequence at least one species from each major vertebrate taxa (Genome 10K Community of Scientists, 2009), as well as a similar project to sequence 5,000 insect species (Robinson et al., 2011). Recently sequenced species with a draft genome available include the non-avian reptiles of the American alligator (*Alligator mississippiensis*), the saltwater crocodile (*Crocodylus porosus*), the gharial (*Gavialis gangeticus;*(John, Braun, & Isberg, 2012), the painted turtle (*Chrysemys picta bellii*) and the Burmesse python (*Python molurus bivittatus;* Castoe et al., 2011). The success of these projects will be impeded without the generation of high quality genome annotations. While RNA-Seq greatly reduces the cost of generating transcriptomes to aid genome annotations, a large number of tissue types and depth of sequencing are required to encompass all the transcripts an organism may express (Yandell & Ence, 2012). When it is not possible to generate multiple deep transcriptomes for a newly sequenced species the predictive annotations increasingly rely on the alignment high confidence gene annotations from closely related species. This highlights the need for the generation of high confidence genome annotation based on transcriptomic sequencing of multiple tissues from at

least one species in each major taxonomic group. These few key model species will serve as the anchor for the genome annotations of many newly sequenced species. As the first reptilian species with genomic and transcriptomic sequencing available, and as a species conducive to being used in a laboratory setting, *A carolinensis* is likely to become the species that anchors the genome annotations for the squamates, and potentially all of reptiles (Kusumi et al., 2011).

Evolution of the Developmental Process of Somitogenesis

The defining feature of vertebrates is the segmented spine, consisting of the vertebrae, spinal musculature, and associated tendons. In aquatic vertebrates such as teleosts, the vertebral column is key to locomotion through lateral flexion. However, the tetrapod vertebral column must bear the weight of the body for terrestrial life, while also allowing a flexible axis for movements. Tetrapod vertebrae have evolved to bear this weight through interlocking adjacent vertebrae at the zygapophysial joints. Among the amniotes there has been significant divergence of other aspects of the spine. In mammals, the tail is primarily tendinous and lacks muscle segments between the vertebrae. Also in mammals, the spinal cord ends with the first lumbar vertebrae (L1). While in birds most of the dorsal, sacral and caudal vertebrae have fused. Reptiles then may more closely represent the ancestral state of the amniote axial structure.

The vertebrae originate during early embryogenesis from paired columns of mesoderm tissue, termed somites, which lie on either side of a central neural tube. The somites are formed from the unsegmented presomitic mesoderm (PSM) during the process of somitogenesis. Somitogenesis is a continuous process, by which new paraxial mesodermal tissue is formed from the tailbud posterior region of the PSM, and new somites are formed from the anterior end of the PSM. The process of somitogenesis can be broken up into three key components: 1) the initial generation and maintenance of the PSM through cell proliferation, 2) an oscillatory mechanism that regulates the regular formation of somites, 3) the generation of boundaries between segments associated with cellular mesenchymal to epithelial transition. During somitogenesis a new pair of somites iteratively buds from the PSM at a species-specific period until the last somite pair is made and the PSM is exhausted.

Although the process of somitogenesis is highly conserved through vertebrates, several key changes to the genetic regulator mechanism controlling somitogenesis have been observed. Most of these changes have occurred in the oscillatory mechanism, termed the segmentation clock that controls the timing of the budding of each somite pair. The major changes in this process appear to have occurred concurrent with the divergence of the amniote vertebrates, including the birds, non-avian reptiles and mammals, from the anamniote vertebrates, including the

amphibians and fish. Thus, it may be possible to uncover key aspects of the evolution of the vertebrate segmentation process through investigating somitogenesis in the non-avian reptiles lineages. To fit this need the recently released complete genomic sequence for *Anolis carolinensis*, along with improved gene annotations, now allows for a through examination of the segmentation clock in a squamate reptile.

The core of the segmentation clock in all vertebrates is cyclical activation of Notch signaling, which is established through a negative feedback loop consisting of the hairy-enhancer of split (HES) bHLH genes (Ozbudak & Pourquié, 2008). In mammals these cycling *HES* genes include *Hes1*, *Hes5*, *Hes7*, *Heyl* and *Hey2* (Dunwoodie et al., 2002; Jouve et al., 2000; Leimeister, Schumacher, Steidl, & Gessler, 2000). The HES negative feedback loop is created by the ability of HES proteins, in particular HES7, to repress their own transcription (Bessho, Miyoshi, Sakata, & Kageyama, 2001). The HES proteins function by forming homo- or hetero-dimers, which act as transcription factors. HES dimers can then inhibit their own transcription through binding of N-box (CACNAG) or through forming heterodimers they can inhibit transcription factors that bind to E-boxes (CANNTG) establishing the negative feedback loop (Bessho et al., 2001).

In mammals, the HES and HEY proteins are inhibit the transcription of the Notch receptor modulator and glycosyltransferase, lunatic fringe (Lfng; Cole, Levorse, Tilghman, & Vogt, 2002; Morales, Yasuda, & Ish-

Horowicz, 2002), because HES and HEY expression oscillates, Lfng expression consequently also oscillates (Aulehla & Johnson, 1999; Forsberg, Crozet, & Brown, 1998; McGrew, Dale, Fraboulet, & Pourquie, 1998). Lunatic fringe is a fucose-specific beta 1,3 *N*-acetylglucosaminyltransferase that modifies glycosyl groups added to the extracellular domain of NOTCH1 (Moloney et al., 2000). By modifying the glycosylation of the NOTCH1, LFNG acts to modulate the ability NOTCH1 to be later activated by its ligands, such as DLL1, on the cell surface (Dale et al., 2003; Hicks et al., 2000). The activation of the NOTCH1 receptor leads to cleavage of the Notch intracellular domain (Notch-ICD) which then translocates to the nucleus and acts as a transcription factor activating its downstream targets, including genes such as *Hes7* and *Lfng* (Bessho, Hirata, Masamizu, & Kageyama, 2003; Cole et al., 2002; Morales et al., 2002). Additionally, some evidence has suggested that cyclical NOTCH activation is controlled through oscillations in expression of its ligand, DLL1 (Krol et al., 2011; Maruhashi, Van De Putte, Huylebroeck, Kondoh, & Higashi, 2005). It is through this mechanism that the segmentation clock causes cycles of Notch activation and oscillatory expression of *Hes, Lfng,* and *Dll1* genes.

In other vertebrate model organisms the core negative feedback loop established by the hes and hey genes is conserved, however, the secondary oscillator driving cyclical activation of NOTCH is divergent from that found in mammals. In the anamniote model systems of the zebrafish,

Danio rerio (a teleost), and the African clawed frog, *Xenopus laevis* (an anuran amphibian), expression of the lunatic fringe orthologues have not been found to oscillate within the PSM, unlike in the mouse. Instead, it has been found that the orthologue of the mammalian delta-like ligand 3 (*Dll3*) deltaC and X-Delta-2 in zebrafish and the frog, respectively, display oscillating expression and likely drive cyclical activation of NOTCH. In the chicken, *Gallus gallus* (an avian reptile), the *LFNG* gene has been observed to cycle, as in mammals (Dale et al., 2003). However, the *DLL1* orthologue does not display cycling expression in the PSM, as it does in mammals (Caprioli, Goitsuka, Pouget, Dunon, & Jaffredo, 2002). Additionally, the *Dll3* orthologue has not been identified in the current genomic assembly of the chicken, and may have been lost in avian lineages. Together, these data suggest that during the evolution of amniotes from anamniotes, the secondary oscillatory used by the segmentation clock switches from the *Dll3* orthologue to the *Lfng* orthologue, and that mammals may have independently evolved to also incorporate the oscillation of the *Dll1* in the segmentation clock.

Two other pathways have been shown to be part of the segmentation clock in mammals, the WNT and FGF pathways. In canonical WNT signaling, binding of the frizzled receptor to WNT proteins causes the downstream stabilization of beta-catenin. Beta-catenin then is able to enter the nucleus and acts as a transcriptional activator for WNT inhibitors *Dkk1* and *Dact1*, which sets up a negative feedback loop.

Oscillation in expression of the WNT inhibitors *Axin2*, *Dkk1* or *Dact1* would then be expected to cause oscillations in WNT signalling (Aulehla et al., 2003; Dequéant, 2008; Suriben, Fisher, & Cheyette, 2006). In mice, *Axin2* was found to display cycling expression in the PSM (Aulehla et al., 2003). This oscillating expression of *Axin2* in mice then drives the expression of other downstream targets in the WNT pathway, including *Dkk1*, *Dact1*, *Sp5*, and *Myc* (Dequéant, 2006; Glinka et al., 1998; He et al., 1998; Krol et al., 2011; Weidinger, Thorpe, Wuennenberg-Stapleton, Ngai, & Moon, 2005; William et al., 2007). Members of the FGF pathway have been found to display oscillatory expression in phase with Notch pathway cyclic genes (Dequéant & Pourquié, 2008). In mice, these genes include the FGF inhibitors sprouty homolog 2 (*Spry2*) and dual specificity phosphatase 4 and 6 (*Dusp4* and *Dusp6*). Downstream targets of the FGF pathway have also been found to oscillate, including snail homolog 1 and 2 (*Snai1* and *Snai2;* Dale et al., 2006; Niwa et al., 2007).

Several key changes in the role of the WNT and FGF signalling pathways in the somitogenesis clock mechanism have been observed between the vertebrate model systems. Recent microarray screens have demonstrated that the zebrafish and chicken segmentation clocks generally involve fewer genes displaying oscillatory expression in the WNT and FGF signalling pathways than compared to mammals (Krol et al., 2011). Specifically, the key WNT inhibitor, *Axin2* displays cycling expression in the chicken, but not zebrafish. Similarly, the key FGF

pathway members of mouse *Dusp6/2* and *Snail1/2* orthologues display cycling expression in the chicken, but not the zebrafish. Together, these findings have shown that the number of Notch, WNT and FGF pathway members participating in the segmentation clock has increased in the mammalian lineage. However, these studies have not included a representative on non-avian reptiles, such as *Anolis carolinensis*.

MicroRNA regulation of tail regeneration

Squamate lizards, including *Anolis carolinensis*, have evolved the ability to regenerate their tail after self-amputation, or autotomy (Alibardi, 2010; R. E. Fisher et al., 2012; Goss, 1992; Maginnis, 2006; Reichman, 1984; Ritzman, Stroik, & Julik, 2012; Vitt, 1981). Through this regenerative process lizards are able to regrow *de novo* muscle groups, hyaline cartilage and skin, as well as a spinal cord and the peripheral nervous system. This represents the most significant regenerative capacity among the amniotes, which are in general poor regenerators compared to amphibians. Given the more significant regeneration found in the amphibians, tail regeneration in lizards may also more closely correspond to the basal regenerative capacity of amniotes.

The fully regenerated tail is organized in radial symmetry, with the spinal cord surrounded by a tube of hyaline cartilage, which it self is surrounded by blocks of muscle fibers and finally the skin (R. E. Fisher et al., 2012; Ritzman et al., 2012). The fully regenerated tail is created in

through four general steps. First, following autotomy is wound healing, which last up to 10 days post-autotomy (dpa). In the second stage a cone of mesenchymal and ependymal cells form at the location of autotomy during 10-15 dpa. In the third stage begins to grow outward beginning at 15 dpa. During this stage the regenerated muscle fibers, cartilage and ependymal cells are observed in the base, and a loose mesenchyme of less differentiated cells is observed in the tip (Hutchins, et al., in preparation). The fourth stage is maturation of the regenerated tissues, that takes place after approximately 60 dpa (R. E. Fisher et al., 2012; Ritzman et al., 2012).

The tail regeneration process that is observed in lizards is substantially different than that of what is found in other vertebrate species. The amphibian species, axolotl (*Ambystoma mexicanum*) displays the ability to regenerate its limbs and tail as nearly perfect copies to their original structures. This feature has been the subject of much research, and it has been shown that the regenerative process found in the axolotl involves the dedifferentiation of adult tissues around the site of injury and the forms a blastema. This blastema is made up of multipotent stem cells that can be transplanted to create ectopic limbs on the adult axolotl (McCusker & Gardiner, 2011; Nye, Cameron, Chernoff, & Stocum, 2003). It is also purposed that this dedifferentiation and blastema guided mechanism controls regeneration in the zebrafish fin (Poss, Keating, & Nechiporuk, 2003). However, this process has not been observed in the

tail regeneration in lizard and is likely regulated through a currently unknown mechanism (Hutchins, et al., in preparation). Mammals are generally thought to be poor regenerators and several key findings have outlined the limited extent of the mammalian regenerate capacity (Muneoka, Allan, Yang, Lee, & Han, 2008). The most striking example of mammalian regeneration is in response to the amputation of digit tips, where a similar, but not perfect, recapitulation of the original tissue results. Similar to what has been shown in lizards, a blastema as found in axolotl regeneration is not observed during the regeneration of digit tips in mammals.

These initial findings then support the lizard as a preferred model organism of regeneration over the anamniotes, with the ultimate goal of translating findings towards improving the regenerative capacity in mammals. Additionally, a lizard model species for regeneration will share a more recent common ancestor with mammals, likely meaning that the regulation of molecular genetic pathways controlling regeneration are more likely to be conserved between the two taxa. With the release of the complete genomic sequence and a corresponding genome annotation of *A. carolinensis*, it is now possible to study the molecular basis for this regeneration in a reptile. Furthermore, recent advances in next generation sequencing have leveled the playing field between traditional model organisms, such the mouse or zebrafish, and emerging model systems, such as the lizard. These developments have particularly improved the

capability to study the transcriptional regulation of a biological process through RNA-Seq. However, RNA-Seq can also be applied to investigate post-transcriptional regulation through sequencing of MicroRNAs.

MicroRNAs (miRNAs) are short 18-22 bp, which inhibit gene expression through blocking transcription and/or causing the degradation of their mRNA targets (Bartel, 2009; R. C. Lee & Ambros, 2001; Sun, Julie Li, Huang, Shyy, & Chien, 2010). miRNAs are first transcribed as pri-miRNAs 60-80bp in length that form a hairpin loop structure. While still inside the nucleus pri-miRNAs are cleaved by a complex made up of DGCR8 and Drosha to create a pre-miRNA (Denli, Tops, Plasterk, Ketting, & Hannon, 2004; Gregory et al., 2004). Once the pre-miRNA is exported to the cytoplasm it is cleaved by Dicer to form the mature miRNA of just 18-22 bp (Grishok et al., 2001; Hutvágner et al., 2001). The mature miRNA can then bind to the 3' UTR of target mRNA transcripts through imperfect complementary base pairing and complexes with members of the Argonaute family of proteins to form the miRNA-induced silencing complex (miRISC) complex. This miRISC complex then causes the silencing of transcription of or degradation of the mRNA target (Doench & Sharp, 2004) and it has been estimated that each miRNA could inhibit more than 100 mRNA transcripts (L. P. Lim et al., 2005).

Through this mechanism of inhibition, miRNAs have been shown to play critical regulatory roles in a wide range of biological processes, particularly in stem cell proliferation and differentiation during development

(Stefani & Slack, 2008). However, only a limited number of studies have investigated the role of miRNAs in regeneration and none of these studies have involved tail regeneration in a lizard. Additionally, until next-generation sequencing technologies emerged, miRNA expressional studies were largely limited to model organisms through the use of microarray technology, which require known probe sequences for each miRNA. In non-model and emerging model organisms, such as *A. carolinensis*, annotations of what miRNAs are present in the organism are primarily based on sequence conservation with miRNAs from closely related model organisms (Griffiths-Jones, 2006; Griffiths-Jones, Saini, & Van Dongen, 2008; Jones, 2004; Kozomara & Griffiths-Jones, 2011). However, next generation small RNA sequencing now allows for the improved annotation, as well as expressional studies, of miRNAs in non-model organisms (Friedlander, Mackowiak, Li, Chen, & Rajewsky, 2011). As part of this work we utilize next generation sequencing to annotate novel miRNAs in the *A. carolinensis* lizard, as well as to investigate the regulatory role of miRNAs during the tail regeneration process.

Throughout this thesis I aim to improve *Anolis carolinensis* as a model organism, as well utilize the lizard for two studies uniquely suited to a reptilian species. The first research chapter of this dissertation describes my efforts to improve the genome annotation of *A. carolinensis* through next generation sequencing and *de novo* assembly of the transcriptomes from 14 adult and embryonic tissues. This improved

genomic annotation can then aid molecular studies in the lizard, as well comparative studies between vertebrates, which both rely on accurate gene models. Adoption of a reptilian model organism with accurate gene models then allows for improved comparative studies of development and evolution. Specifically, through the second research chapter of this dissertation, I display the utility of the lizard model organism by investigating the developmental process of somitogenesis. Additionally, the addition of *A. carolinensis* as a model organism allows for molecular studies of biological processes unique to reptiles. In the third research chapter of this dissertation I investigate role of miRNAs in regulating the tail regeneration process in lizards. It is through these three chapters that I aimed to not only advance A. carolinensis as a new model organism, but to utilize it to uncover novel findings only possible through studies utilizing a reptilian model species.

Chapter 2

GENOME REANNOTATION OF THE LIZARD ANOLIS CAROLINENSIS BASED ON 14 ADULT AND EMBRYONIC DEEP TRANSCRIPTOMES

Abstract

The green anole lizard, *Anolis carolinensis*, is a key species for both laboratory and field-based studies of evolutionary genetics, development, neurobiology, physiology, behavior, and ecology. As the first non-avian reptilian genome sequenced, *A. carolinesis* is also a prime reptilian model for comparison with other vertebrate genomes. The public databases of Ensembl and NCBI have provided a first generation gene annotation of the anole genome that relies primarily on sequence conservation with related species. A second-generation annotation based on tissue-specific transcriptomes would provide a valuable resource for molecular studies. Here we provide an annotation of the *A. carolinensis* genome based on *de novo* assembly of deep transcriptomes of 14 adult and embryonic tissues. This revised annotation describes 59,373 transcripts, compared to 16,533 and 18,939 currently for Ensembl and NCBI, and 22,962 predicted protein-coding genes. A key improvement in this revised annotation is coverage of untranslated region (UTR) sequences, with 79% and 59% of transcripts containing 5' and 3' UTRs, respectively. Gaps in genome sequence from the current *A. carolinensis* build (Anocar2.0) are highlighted by our identification of 16,542 unmapped transcripts, representing 6,695 orthologues, with less than 70% genomic

coverage. Incorporation of tissue-specific transcriptome sequence into the *A. carolinensis* genome annotation has markedly improved its utility for comparative and functional studies. Increased UTR coverage allows for more accurate predicted protein sequence and regulatory analysis. This revised annotation also provides an atlas of gene expression specific to adult and embryonic tissues.

Background

Recent advances in sequencing technologies and *de novo* genome assembly algorithms have greatly reduced the time, cost, and difficulty of generating novel genomes (Mardis, 2011). This has led to organized efforts to sequence a representative species from all major vertebrate taxa, referred to as the Genome 10K Project (Genome 10K Community of Scientists, 2009), as well as a similar project to sequence five thousand insect genomes, the i5K project (Robinson et al., 2011). While these efforts have the potential to transform comparative studies, many applications including studies of biological function will be limited without quality genome annotations. Genome annotations of newly sequenced species initially rely primarily on *ab initio* gene predictions and alignment of reference transcripts of related species; however, the quality of gene models is greatly improved when incorporating same species transcriptomic sequencing (Haas et al., 2008). In particular, information from high-density next-generation RNA sequencing, i.e., deep

transcriptomes, greatly improves even well-annotated genomes (Rhind et al., 2011).

While 39 mammalian genomes and 3 avian genomes have been published (Flicek et al., 2011), whole genome sequences have only recently been available for non-avian reptiles. The first published non-avian reptilian genome was that of a squamate, the lizard *Anolis carolinensis* (Anocar2.0 assembly; Alföldi et al., 2011). Subsequently, releases of draft genomes from another squamate, the Burmese python, *Python molurus bivittatus*, (Castoe et al., 2011) and three crocodilian species: the American alligator, *Alligator mississippiensis*, the gharial *Gavialis gangeticus*, and the saltwater crocodile *Crocodylus porosus* (John et al., 2012) were published. As an emerging model system with its genome sequence available, the green anole has already proved useful in a variety of fields including comparative genomics (Fujita, Edwards, & Ponting, 2011; Janes et al., 2011; Novick, Smith, Floumanhaft, Ray, & Boissinot, 2011; Tollis & Boissinot, 2011), functional genomics (Grassa & Kulathinal, 2011; C. H. Lim, Hamazaki, Braun, Wade, & Terada, 2011), behavior (M. A. Johnson & Wade, 2010; M. A. Johnson, Cohen, Vandecar, & Wade, 2011), evolutionary genetics (Kolbe, Revell, Szekely, Brodie, & Losos, 2011; Sanger, Revell, Gibson-Brown, & Losos, 2012), and development and evolution (Eckalbar et al., 2012; Koshiba-Takeuchi et al., 2009). In all of these areas of research, the green anole genome, in

combination with avian and mammalian data, provides a key perspective on conserved and divergent features among amniotes.

Currently, the public databases of the National Center for Biotechnology Information (NCBI), Ensembl, and University of California, Santa Cruz (UCSC) have devoted anole genome portals. NCBI and Ensembl provide first generation genome annotations, which are based primarily on conservation with other species (Alföldi et al., 2011). These first generation annotations rely heavily on conservation of protein-coding sequences, and as such, predicted green anole genes generally lack untranslated regions (UTRs) and often do not contain start and/or stop codons. Furthermore, alternative splice forms and evolutionarily divergent orthologues are not represented in the first generation annotations. These issues have limited the ability of researchers to carry out comparative and functional genomic studies based on the *A. carolinensis* genome sequence.

In order to help resolve many of these issues, here we present a second-generation revised annotation based on a foundation of 14 *de novo* deep transcriptomes and published cDNA sequences. We used a customized pipeline based on the Program to Assemble Spliced Alignments (PASA; Haas, 2003; Loke, 2005; Rhind et al., 2011; Shen et al., 2008), EVidenceModeler (EVM; Haas et al., 2008) and MAKER2 (Cantarel et al., 2007; Holt & Yandell, 2011) to combine the following data: i) *de novo* and reference based assemblies of 14 RNA-Seq

transcriptomes, ii) 7 publicly available EST libraries, iii) RefSeq alignments of the available vertebrate transcripts, iv) RefSeq alignments of zebrafish, *Xenopus* frog, chicken, mouse, and human protein sequences, v) NCBI and Ensembl current annotations, and vi) *ab initio* gene predictions based on analysis by SNAP and Augustus (Korf, 2004; Stanke & Waack, 2003; Stanke, Schöffmann, Morgenstern, & Waack, 2006).

Materials and methods

Animals. All animals were maintained according to Institutional Animal Care and Use Committee guidelines of Arizona State University. *Anolis carolinensis* lizards were purchased from approved vendors (Charles D. Sullivan Co., Inc., Nashville, TN; Marcus Cantos Reptiles, Fort Myers, FL) and were housed at 70% humidity. Lighting and temperature were maintained for 14 hours at 28°C daylight and 10 hours at 22°C night. Adult tissues were collected immediately after euthanasia. Eggs were collected within one day of laying, typically at the 25-30 somite pair stage.

RNA-Seq. Samples for RNA-Seq, including embryos, regenerating tail, original tail, dewlap skin, brain, heart, lung, liver, adrenal, ovary, and skeletal muscle were collected for extraction using the total RNA protocol of the miRVana kit (Ambion). For the regenerating tail, original tail, 28 and 38 somite-pair staged embryos, total RNA samples were prepared using the Ovation RNA-Seq kit (NuGEN) to generate double stranded cDNA. Illumina manufacturer protocols were followed to generate paired-end

sequencing libraries. Sequencing was carried out on the Illumina HiSeq 2000 using paired-end chemistry with read lengths of 104 base pairs. Strand-specific RNA sequencing libraries were prepared for adrenal, brain, dewlap, pooled 28 and 38 somite staged embryos, heart, liver, lung, skeletal muscle, and ovary RNA samples using the dUTP protocol (Parkhomchuk et al., 2009). The dUTP strand-specific libraries were sequenced on the Illumina HiSeq using paired-end chemistry with read lengths of either 76 or 101 bp.

***De novo* assembly**. Non-directional RNA-Seq data was assembled using the ABySS and Trans-ABySS pipeline (Birol et al., 2009; Robertson et al., 2010; Simpson et al., 2009). Each sample was assembled in ABySS using every 5^{th} kmer ranging from 26bp to 96bp. These assemblies were then combined using trans-ABySS to create a merged assembly with reduced redundancy. This merged assembly was then mapped to the genome using BLAT inside trans-ABySS. *De novo* assembled contigs were then filtered to require at least 90% coverage of the contig to the genome and to require at least one 25 bp gap. Because of its ability to utilize stranded information, Trinity was used to assemble the strand-specific RNA sequencing data using default parameters (Grabherr et al., 2011).

PASA alignment assembly. The *de novo* assembled transcripts from ABySS/Trans-ABySS and Trinity (Grabherr et al., 2011), as well as the contigs from the EST data sets, were then assembled using the PASA

reference genome guided assembly. Seqclean was first used to remove Illumina adapters and any contaminants from the UniVec databases from the *de novo* assembled transcripts and the EST libraries. PASA alignment and assembly was then executed using default parameters and utilizing the strand-specific data when possible (Haas, 2003; Loke, 2005; Rhind et al., 2011; Shen et al., 2008).

***Ab initio* training (PASA/CD-HIT) and prediction.** In order to train *ab initio* gene prediction algorithms, a set of high confidence transcripts were extracted from the PASA assemblies from each RNA sequencing data set. These transcripts were then combined and redundancy removed using CD-HIT (Y. Huang, Niu, Gao, Fu, & Li, 2010; W. Li & Godzik, 2006). This set of transcripts was then used to train gene identification parameters for Augustus (Stanke et al., 2006; Stanke & Waack, 2003), as well as SNAP (Korf, 2004) inside MAKER2 (Cantarel et al., 2007; Holt & Yandell, 2011). Each gene finder was then run to produce a set of *ab initio* predictions for the *A. carolinensis* genome.

EVM annotation combiner/PASA updates. EVidenceModeler (Haas et al., 2008) was utilized to combine *ab initio* gene predictions from Augustus and SNAP, the Ensembl Build 65, and the NCBI ref_Anocar2.0 gene predictions, in combination with UCSC reference protein alignments and *A. carolinensis* transcriptomic data from the PASA assembles. This initial annotation ignored alternate splicing and UTRs. Cufflinks assemblies derived from TopHat alignments of the raw reads, as well as

human and chicken RefSeq protein alignments carried out using Exonerate, were used to guide a MAKER2 annotation update to include novel genes, UTR sequences and alternative splicing isoforms (Langmead & Salzberg, 2012; Langmead, Trapnell, & Pop, 2009; Roberts, Pimentel, Trapnell, & Pachter, 2011a; Roberts, Trapnell, Donaghey, Rinn, & Pachter, 2011b; Slater & Birney, 2005; Trapnell, Pachter, & Salzberg, 2009; Trapnell et al., 2012; 2010). PASA was then again utilized to update this initial genome annotation from EVM and MAKER2 to add alternate splicing and UTR based on transcript data (Haas, 2003; Haas et al., 2008; Holt & Yandell, 2011; Loke, 2005; Rhind et al., 2011; Shen et al., 2008). Orthologues were then assigned to these annotations through finding overlapping gene annotations from NCBI or Ensembl gene models. ASU gene predictions that did not have an overlap with NCBI or Ensembl genes were then assigned orthology by identifying the most similar vertebrate RefSeq protein using blastx inside Blast2GO (Conesa & Götz, 2008; Conesa et al., 2005; Gotz et al., 2008).

TopHat/Cufflinks. In order to estimate gene expression levels in each tissue type, reads from each sample were first trimmed based on quality and mapped to the *A. carolinensis* genome using Bowtie and TopHat as described previously (Langmead et al., 2009; Langmead & Salzberg, 2012; Trapnell et al., 2009). Cufflinks version 2.0.0 was then used to generate an estimation of transcript abundance as Fragments per Kilobase of exon per Million fragments mapped (FPKM), based on the

number of reads aligned to each transcript. The Cufflinks results where then analyzed for differential expression using cuffdiff on the non-strand specific and stand specific RNA-seq data sets separately. CummeRbund was then used to generate a heat map and dendogram for each Cuffdiff result (Roberts, Pimentel, Trapnell, & Pachter, 2011a; Trapnell et al., 2010; 2012).

Analysis of transcript contigs that failed to align to the Anocar2.0 genome assembly. All tissue-specific contigs that failed to align to the Anocar2.0 genome assembly were processed using CAP3 (X. Huang & Madan, 1999) for a second assembly to reduce redundancy and extend any partial transcripts contained in each sample. Transcripts were filtered to remove >33% percentage of repetitive sequences using RepeatMasker (Tarailo-Graovac & Chen, 2002) and remaining transcripts that contained open reading frames longer than 100 amino acids were then extracted from the CAP3 assembly for further analysis. Because CAP3 tended to reconstruct more complete transcripts, these transcripts containing ORFs longer than 100 amino acids were then realigned to the genome using BLAT and alignments covering greater than 70% of the transcript at 90% identity were removed (Kent, 2002). The filtered transcript contigs were then assigned orthology based on a best blastx match to vertebrate RefSeq proteins using Blast2GO (Conesa et al., 2005; Conesa & Götz, 2008; Gotz et al., 2008).

Revised A. carolinensis annotation files. The ASU_Acar version 2.1 annotation files have been deposited to NCBI and Ensembl for release through their organism-specific portals. Assemblies and the meta-assembly of unmapped transcripts have also been distributed to the *A. carolinensis* community research portals, *lizardbase* (http://www.lizardbase.org) and *AnolisGenome* (http://www.anolisgenome.org).

Results and Discussion

De novo transcriptome generation and assembly. We carried out RNA-Seq to generate 11 adult tissue and 3 embryonic transcriptomes (Table 1). Strand-specific directional sequences were generated from adrenal gland, brain, dewlap skin, heart, liver, lung, ovary, and skeletal muscle. RNA-Seq generated by directional library construction can be used to distinguish between coding transcripts and antisense noncoding transcripts. The adrenal, lung, liver and skeletal muscle samples were derived from a single male individual (Supplemental Table 1). The brain, dewlap skin, heart, and ovary samples were pooled from several individuals (Supplemental Table 1). Standard non-directional RNA-Seq libraries were prepared from regenerating tail and embryonic tissues. Lizards including the green anole can regenerate their tail following autotomy, or self-amputation (Ritzman et al., 2012). Regenerating tissues from 3 tails at 15 days post-autotomy were divided into pools of the

Table 1.

Overview of de novo transcript assembly for A. carolinensis based on RNA-Seq data from 14 adult and embryonic tissues and deposited EST sequence data.

De novo RNA-Seq	# Reads	*De novo* assembled transcripts	Transcripts aligning to Anocar2.0 assembly	PASA assembled transcripts
Embryo-28 somite stage	52,548,024	83,627	81,032	22,670
Embryo-38 somite stage	55,048,179	99,578	95,753	24,595
Regenerating tail tip	122,099,352	92,275	88,150	22,278
Regenerating tail stump	31,721,054	78,005	73,516	24,897
Original tail	109,404,060	96,450	91,601	20,240
Adrenal	55,858,836	110,349	101,449	20,482
Brain	32,518,977	203,519	192,407	33,912
Dewlap skin	31,785,178	81,598	76,866	25,853
Embryos (pooled)	59,681,427	118,949	110,124	19,969
Heart	34,068,834	154,255	144,617	26,582
Liver	50,782,350	89,010	81,441	21,549
Lung	48,723,049	272,071	255,035	37,985
Ovary	35,139,647	80,306	75,807	26,827
Skeletal Muscle	42,707,477	75,006	69,250	18,857
Total	762,086,444	1,634,998	1,537,048	346,696

EST Library (NCBI)	# Sequences	NCBI defined UniGene groups	Transcripts aligning to Anocar2.0 assembly	PASA assembled transcripts
Brain	19,139	5,631	9,991	1,715
Dewlap skin	19,809	5,453	10,180	2,216
Embryo	38,923	8,714	9,991	4,158
Mixed Organ	19,863	5,657	9,327	2,053
Ovary	19,410	5,467	7,394	3,737
Regenerating Tail	19,851	6,751	11,064	6,757
Testis	19,807	4,261	8,677	2,594
Total	156,802	41,934	66,624	23,230

regenerating epithelial tip and the adjacent tail stump. RNA-Seq was also performed on the original autotomized tail from those same animals. Embryos between zero to one day after egg laying (28 and 38 somite-pair stages) were analyzed individually by standard RNA-Seq as well as pooled for directional library construction and sequencing. More than 762 million paired-end reads were generated from these adult and embryonic tissue samples (Table 1).

The pipeline for *de novo* assembly of RNA-Seq data involved two steps. First, strand-specific transcriptome sequence libraries were assembled using Trinity (Fig. 1A; Grabherr et al., 2011). Standard non-directional RNA-Seq libraries were assembled using ABySS and Trans-ABySS (Birol et al., 2009; Robertson et al., 2010; Simpson et al., 2009). In total, this generated more than 1.62 million *de novo* assembled transcript contigs. Second, these assembled contigs were then aligned to the *A. carolinensis* Anocar2.0 assembly (Alföldi et al., 2011) using the gmap tool within PASA, with the aim of i) eliminating sequences not aligning to the genome and ii) merging *de novo* assembled sequences to remove redundancy. We observed that over 94% of these sequences aligned to the green anole genome at a cutoff of 95% identity and 90% transcript coverage. This first step of the *de novo* assembly pipeline reduced the number of RNA-Seq based transcript contigs down to 669,584.

As part of the *A. carolinensis* genome sequencing effort, EST sequences were generated from five adult organs (brain, dewlap skin,

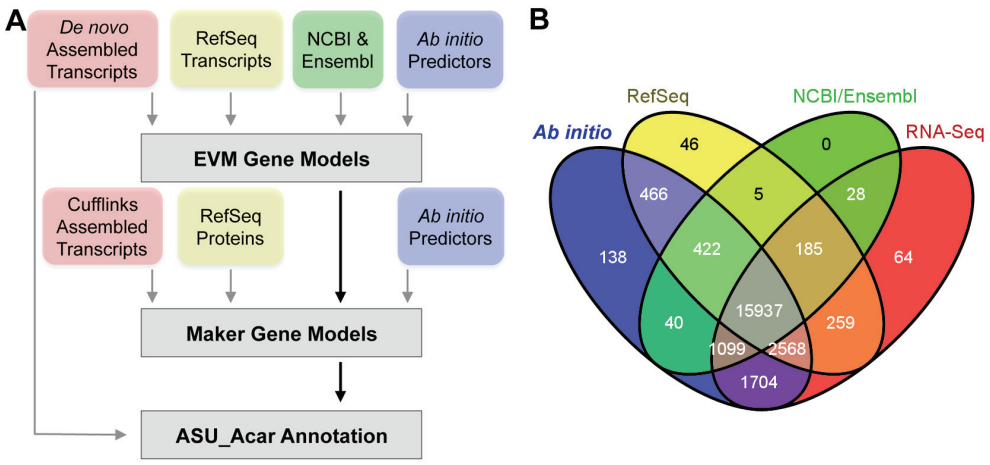


Figure 1. Our methods and data sources for creating ASU annotations of the Anocarol2.0 genome. **A.** Diagram of the bioinformatic pipeline for the *A. carolinensis* reannotation. **B.** Venn diagram illustrating the sources of data for the *A. carolinensis* reannotation. *Ab initio*, algorithm based gene predictions using Augustus and SNAP [26-28]. RefSeq, alignments of zebrafish, *Xenopus* frog, chicken, mouse, and human protein and available vertebrate transcripts to the Anocar2.0 genome assembly. NCBI/Ensembl, combined data of *A. carolinensis* genome annotations from NCBI ref_Anocar2.0 and Ensembl Build 65. RNA-Seq, transcriptomic data from analysis of 14 adult and embryonic tissues.

ovary, regenerated tail, and testis), embryo, and a seventh mixed organ library that included heart, kidney, liver, lung, and tongue (Alföldi et al., 2011). These EST sequences were introduced at the second step of this pipeline and aligned to the *A. carolinensis* Anocar2.0 assembly using gmap, identifying another 35,188 transcript contigs not present from the RNA-Seq deep transcriptomic data. This yielded a total of 704,772 transcript contigs that were then used as the basis of the second-generation *A. carolinensis* genome annotation (Table 1).

Generating a revised annotation of the A. carolinensis genome. The reannotation of the *A. carolinensis* genome incorporates four classes of evidence that were combined using EVM tool (Fig. 1A; Haas et al., 2008). First, the 704,772 *de novo* assembled transcript contigs were given the highest weight to generate the revised annotation. Second, two *ab initio* gene prediction tools, SNAP and Augustus (Korf, 2004; Stanke et al., 2006; Stanke & Waack, 2003), were trained using a subset of the PASA transcriptome assemblies after removing redundancy using CD-HIT (W. Li & Godzik, 2006). In brief, 9,064 *A. carolinensis* coding sequences were used to train SNAP, and 1,041 complete predicted protein sequences were used to train Augustus. Third, the first generation *A. carolinensis* gene annotations from NCBI ref_Anocar2.0 (abbreviated as NCBI) and Ensembl Build 65 (abbreviated as Ensembl) were used as an input to EVM. Finally, regions of alignment to RefSeq homologous transcript

sequences from the UCSC Genome Bioinformatics portal were also incorporated into the EVM predictions.

Since EVM currently generates only a single protein-coding sequence for each gene and the transcript evidence requires at least 90% alignment to the genome, further steps were necessary to improve the annotation. First, the RNA-Seq reads were aligned to the Anocar2.0 assembly using TopHat and reference guided assemblies were completed using Cufflinks. Second, the EVM predictions, the Cufflinks assemblies, as well as zebrafish, *Xenopus* frog, chicken, mouse, and human protein alignments, were used as input into MAKER2 to annotate novel genes, extend UTR sequence, and annotate alternative splicing (Fig. 1A). These models were updated to further incorporate UTR sequences and alternate splice forms present in the *de novo* assembled transcripts described above. We have named this second-generation annotation for *A. carolinensis* ASU_Acar version 2.1 (abbreviated as ASU).

Sources for genome reannotation. The improvements in the ASU annotation derive from multiple sources. The largest group of annotated genes, 69% or 15,937, were based on all sources of data, and an additional 30% (6,776) were based on two or three sources of data (Fig. 1B). In addition, RNA-Seq was a key source of data for this reannotation, contributing to 95% of all gene predictions. Only 1% of predicted genes were based solely on one source of data (transcriptome, NCBI or Ensembl annotation, RefSeq alignment, or *ab initio* predictions). The *ab initio* gene

predictions, which do not make use of any empirically derived data, contribute less than 1% (138) of the gene predictions for this reannotation. Since both the first and second-generation annotation pipelines rely on open reading frame and coding sequence predictions, noncoding transcripts are likely underrepresented. The generation of long noncoding and microRNA-Seq data and sampling of more tissues by the research community will contribute to improved *A. carolinensis* genome annotations in the future.

Improvements in gene annotation. To quantify the differences between the first and second-generation genome annotations, we compared ASU with NCBI, and Ensembl annotations. First, the ASU annotation has identified more genes than either NCBI or Ensembl (22,962 vs. 15,645 in NCBI and 17,792 in Ensembl; Table 2). Second, the ASU annotation greatly increases the number of annotated transcript isoforms (59,373 vs. 16,533 for NCBI and 18,939 for Ensembl; Table 2). Third, predicted transcripts in the ASU annotation appear to be more complete in a number of different parameters. Of the 59,373 annotated transcript isoforms, 90% (53,401) are predicted to be complete protein-coding sequences. Furthermore, 59% (34,926) are predicted to contain 3' UTR sequences, and 79% (46,782) to contain 5' UTR sequences (Table 2). In addition, the ASU annotation greatly improves transcript lengths (5,355 bp vs. 2,364 bp for NCBI and 2,037 bp for Ensembl; Fig. 2A & B; Table 2). An example of the improvements in gene annotation is evident

Table 2.

Comparison of ASU, NCBI and Ensembl gene annotations of the A. carolinensis genome.

Overview	**ASU**	**NCBI**	**Ensembl**
Annotated genes	22,962	15,645	17,792
Annotated transcript isoforms	59,373	16,533	18,939
Annotated isoforms/gene	2.59	1.06	1.06
Annotated Transcripts			
All transcript isoforms	59,373	16,533	18,939
Transcripts with start & stop codons	53,401	14,667	4,170
Transcripts missing start or stop codon	5,972	1,866	14,769
Single exon transcripts	2,070	983	364
Transcript N50 length	5,355	2,364	2,037
Average coding sequence length	1,964	1,701	1,531
Exons			
Total number of exons	229,204	156,742	174,545
Exons with start with codon	29,677	13,512	5,971
Exons without start or stop codon	168,367	128,486	158,935
Exons with stop codon	29,727	13,779	9,278
Exons/annotated transcript	12.05	10.11	9.62
Average exon length	170	170	160
Total exon length	38,902,806	26,658,387	27,910,718
3' UTR			
Total transcripts with 3'UTR	34,926	5,861	0
Average length of transcripts with 3'UTR	1,168	456	0
Total 3'UTR sequence length	40,798,794	2,674,388	0
5' UTR			
Total transcripts with 5'UTR	46,782	6,168	0
Average length of transcripts with 5'UTR	244	86	0
Total 5'UTR sequence length	11,422,626	527,454	0
Introns			
Total number of introns	192,418	141,362	155,949
Average intron length	4,525	4,463	2,553
Total intron sequence length	870,771,088	630,937,171	398,124,572

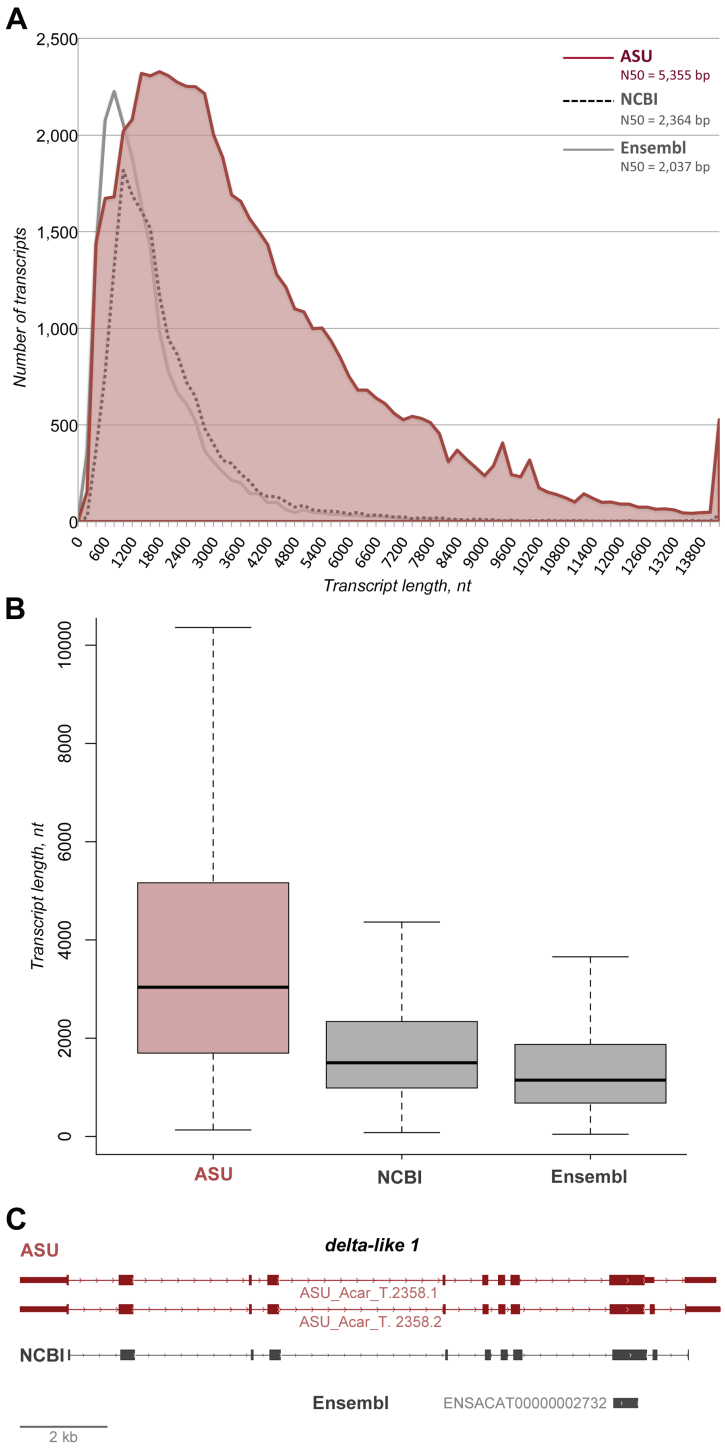

A
2,500
2,000
1,500
1,000
500
0
Number of transcripts
ASU
N50 = 5,355 bp
NCBI
N50 = 2,364 bp
Ensembl
N50 = 2,037 bp
0
600
1200
1800
2400
3000
3600
4200
4800
5400
6000
6600
7200
7800
8400
9000
9600
10200
10800
11400
12000
12600
13200
13800
Transcript length, nt
B
10000
8000
6000
4000
2000
0
Transcript length, nt
ASU
NCBI
Ensembl
C
ASU
delta-like 1
ASU_Acar_T.2358.1
ASU_Acar_T. 2358.2
NCBI
Ensembl
ENSACAT00000002732
2 kb

Figure 2. Increased N50 transcript length and number of predicted transcripts in the ASU annotation. **A.** The distribution of transcript lengths is shown for the ASU, NCBI and Ensembl genome annotations. The ASU reannotation transcript N50 length of 5,355 bp is greater than values for the first generation annotations from Ensembl (2,037 bp) and NCBI (2,364 bp). **B.** A boxal plot showing the median (horizontal line) and boundaries for the 25^{th} and 75^{th} percentiles (box) as well as the range for the ASU, NCBI, and Ensembl predicted transcripts. **C.** The Notch ligand *dll1* is an example of gene whose annotation has been markedly improved in the reannotation.

the Notch pathway ligand, delta-like 1 (Fig. 2C; gene symbol *dll1* following guidelines of the *Anolis* Gene Nomenclature Committee; (Kusumi et al., 2011).

With the increased transcript length and incorporation of UTR sequences, the improvement in the *A. carolinensis* annotations will increase the accuracy of RNA-Seq based studies of gene expression levels. We selected adult and developing tissues for deep transcriptomic analysis that displayed divergent gene expression profiles in other model organisms, and this divergence was demonstrated in RNA-Seq-derived gene expression levels by heat map (Fig. 3C, D) and similarity of expression profiles determined by dendograms generated by CummeRbund analysis (Fig. 3A, B; Trapnell et al., 2012).

Assignment of gene orthology. Identification of orthologous relationships between genes in *A. carolinensis* and other vertebrate model systems is a key step in comparative studies. However, this is a complex task due to gene deletions and genome duplications and rearrangements during vertebrate evolution. For protein-coding genes, metrics have been proposed (Kusumi et al., 2011) that consider both protein sequence similarity and synteny conservation. For comparison of ASU annotation, we have used the current orthology assignments in the NCBI and Ensembl gene models. Given the longer transcript lengths in the ASU annotation we identified that 16,303 genes overlapped with Ensembl

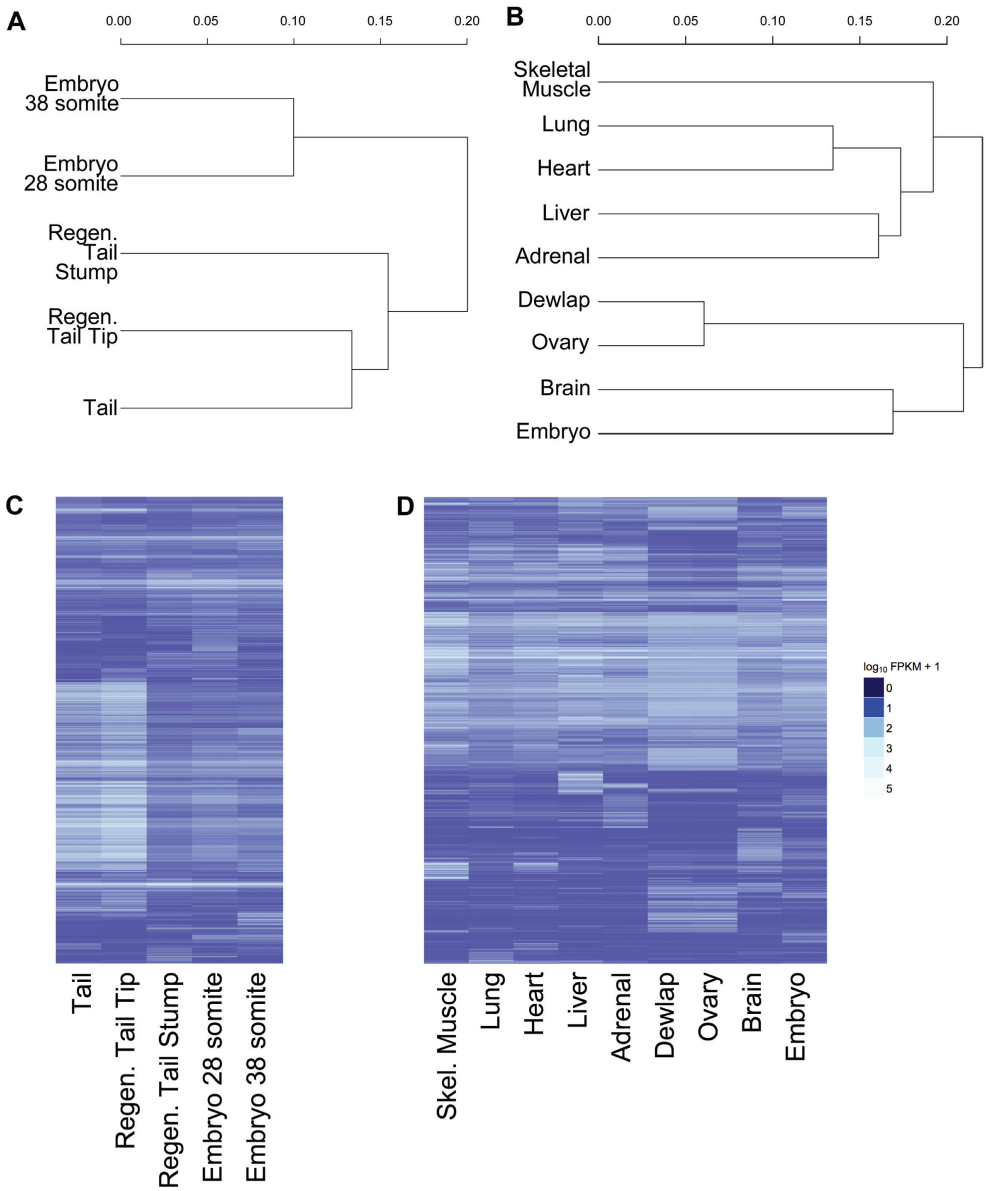

A
0.00
0.05
0.10
0.15
0.20
Embryo 38 somite
Embryo 28 somite
Regen. Tail Stump
Regen. Tail Tip
Tail
B
0.00
0.05
0.10
0.15
0.20
Skeletal Muscle
Lung
Heart
Liver
Adrenal
Dewlap
Ovary
Brain
Embryo
C
Tail
Regen. Tail Tip
Regen. Tail Stump
Embryo 28 somite
Embryo 38 somite
D
Skel. Muscle
Lung
Heart
Liver
Adrenal
Dewlap
Ovary
Brain
Embryo
log10 FPKM + 1
0
1
2
3
4
5

Figure 3. Comparison of gene expression levels in the 14 adult and embryonic tissue transcriptomes using the *A. carolinensis* revised annotations. Dendrograms of non-directional (**A**) and directional (**B**) RNA-Seq data sets demonstrate the degree of similarity between tissue samples. Distances were calculated using Jensen-Shannon divergence [36]. B. Relative gene expression levels for tissues analyzed by non-directional (**C**) and directional (**D**) RNA-Seq analysis is also shown by heat map.

predicted genes and 16,908 overlapped with NCBI predicted genes (Supplemental Table 2).

However, this comparison left 5,246 ASU predicted genes with no orthology assignment based on NCBI or Ensembl annotations. Gene orthology for these remaining predicted genes were next evaluated by Blast2GO against the vertebrate RefSeq database (Conesa et al., 2005; Conesa & Götz, 2008; Gotz et al., 2008). This analysis demonstrated that 56% of these predicted genes (2,928/5,246) had a Blast2GO Expect (E) value score of at least 10^{-3} with a vertebrate gene (Table 3; Supplemental Table 3). Of these predicted genes, 90% (2,627/2,928) contain multiple exons with an average of 6.4 exons/gene and a N50 value of 2,157 bp. These may reflect genes that have been newly identified in the ASU annotation but were missing in the NCBI and Ensembl annotations. The remaining 10% of the predicted genes (301/2,928) contain only a single annotated exon, which could result from gaps in the Anocar2.0 reference genome assembly. The remaining group of genes (2,318/5,246) aligned to the Anocar2.0 assembly but had poor vertebrate homology. This group may include novel lizard genes and rapidly diverging genes such as noncoding RNAs.

Transcripts with vertebrate homology not present in the Anocar2.0 genome assembly. Given a 7.1x genome coverage for the *A. carolinensis* Anocar2.0 assembly, only 81% of the 2.2 Gbp genome is

Table 3.

A. carolinensis genes unique to the ASU annotation with vertebrate orthologues.

Gene	**ASU**
Annotated genes	2,928
Annotated transcript isoforms	3,612
Annotated isoforms/gene	1.23
Transcript	
All transcript isoforms	3,612
Transcripts with start & stop codons	2,698
Transcripts missing start or stop codon	914
Single exon transcripts	301
Transcript N50 length	2,157
Average coding sequence length	1,182
Exon	
Total number of exons	18,921
Exons with start with codon	2,468
Exons without start or stop codon	13,901
Exons with stop codon	2,300
Exons/annotated transcript	6.35
Average exon length	188
Total exon length	3,569,265
3' UTR	
Total transcripts with 3'UTR	1,323
Average length of transcripts with 3'UTR	761.2
Total 3'UTR sequence length	1,007,040
5' UTR	
Total transcripts with 5'UTR	1,816
Average length of transcripts with 5'UTR	238.7
Total 5'UTR sequence length	433,533
Intron	
Total number of introns	15,835
Average intron length	5,304
Total intron sequence length	83,999,254

predicted to be included in the current contig sequences (Alföldi et al., 2011). In addition, approximately 30% of the *A. carolinensis* genome consists of repetitive mobile element sequences, which leads to a lower than typical N50 given the sequencing depth. Thus, some transcripts identified by RNA-Seq analysis would not align to the Anocar2.0 assembly, and these transcripts would not included in the ASU annotation. This category of genes missing from the Anocar2.0 assembly may include important developmental or regulatory genes.

We developed a pipeline to analyze the genes poorly represented in the Anocar2.0 assembly (see Materials and Methods). Starting with 638,802 *de novo* assembled contigs, the pipeline reduced this group down to 29,706 and increased the N50 value from 349 bp up to 2,074 bp. Next, these 29,706 contigs were analyzed by Blast2GO to identify homology to vertebrate RefSeq entries with an E-value cutoff of 10^{-3} (Supplemental Table 3). The majority of these contigs (56% or 16,542/29,706) could be matched to 6,695 distinct vertebrate orthologues (Supplemental Figure 1; Supplemental Table 4).

Analyzing these matching contigs further, we were able to identify matches with 30% of the contigs (4,910/16,542) against 2,233 *A. carolinensis* RefSeq proteins. The remaining 70% (11,632/16,542) mapped with highest scores to other vertebrate species (Supplemental Figure 1). Contigs that matched *A. carolinensis* RefSeq proteins but failed to align to the Anocar2.0 assembly represent genes with insufficient

overlap with the current genome sequences. The incomplete state of the *A. carolinensis* RefSeq libraries may lead to higher matches with other vertebrate species. Gaps in the Anocar2.0 assembly contribute to genes missing from *A. carolinensis* annotations. Misassembly in genome scaffolds would interrupt contiguous alignments of transcripts at the contig boundaries. Given these observations, additional sequencing to increase coverage of the *A. carolinensis* genome would improve future annotations.

Conclusions

With the release of the *A. carolinensis* genome, along with a first generation annotation provided by NCBI and Ensembl, a growing foundation of genomic resources are available for the anole reptilian model. Furthermore, genome annotations of this key reptilian model provide a valuable resource for genomic comparison with mammals, such as mice and humans. Using RNA-Seq, we have improved the genome annotation for *A. carolinensis*, which includes 59,373 transcript isoforms, many of which are complete with UTR sequences. *De novo* transcriptome assembly also identified 16,542 transcripts that are not well represented on the current Anocar2.0 genome build. This revised genome annotation and available transcriptomic sequences provide a resource for vertebrate comparative and functional studies. This work also highlights the need for additional genomic sequencing of *A. carolinensis* to fill in gaps and extend

scaffolds, as well as further transcriptomic sequencing of additional tissues.

Chapter 3

SOMITOGENESIS IN THE ANOLE LIZARD AND ALLIGATOR REVEALS A TRANSITION IN EVOLUTION OF THE SEGMENTATION CLOCK

Abstract

The axial skeleton is a defining feature of vertebrates and is patterned during somitogenesis. Cyclically expressed members of the notch and other signaling pathways, described as the 'segmentation clock', regulate the formation of somite boundaries. Comparisons among vertebrate model systems have revealed fundamental shifts in the regulation of expression among critical genes in the notch pathway. However, insights into the evolution of these expression differences have been limited by the lack of information from non-avian reptiles. We analyzed the segmentation clock of the first Lepidosaurian reptile sequenced, the green anole lizard, Anolis carolinensis, for comparison with avian and mammalian models. Using genomic sequence, RNA-Seq transcriptomic data, and in situ hybridization analysis of somite-stage embryos, we carried out comparative analyses of key genes and found that the anole segmentation clock displays features common to both amniote and anamniote vertebrates. Shared features with anamniotes, represented by Xenopus laevis and Danio rerio, include an absence of lunatic fringe (lfng) expression within the presomitic mesoderm (PSM), a hes6a gradient in the PSM not observed in the chicken or mouse, and EGF repeat structure of the divergent notch ligand, dll3. The anole and

mouse share cycling expression of dll1 ligand in the PSM. To gain insight from an Archosaurian reptile, we analyzed LFNG and DLL1 expression in the American alligator. LFNG expression was absent in the alligator PSM, like the anole but unlike the chicken. In contrast, DLL1 expression does not cycle in the PSM of the alligator, similar to the chicken but unlike the anole. Thus, our analysis yields novel insights into features of the segmentation clock that are evolutionarily basal to amniotes versus those that are specific to mammals, Lepidosaurian reptiles, or Archosaurian reptiles.

Background

One of the defining features of the vertebrates is the segmented spine, which is patterned during the developmental process of somitogenesis (Dequéant & Pourquié, 2008; Gibb, Maroto, & Dale, 2010). Molecular analysis of somitogenesis has primarily focused on analysis of four vertebrate model systems: the zebrafish, *Danio rerio* (a teleost); the African clawed frog, *Xenopus laevis* (an anuran amphibian); the chicken, *Gallus gallus* (an avian reptile); and the mouse, *Mus musculus* (a eutherian mammal). From studies in these species, a model of genetic regulation of somitogenesis, described as the 'segmentation clock' has been developed, based on the theoretical framework of the 'clock and wavefront' model (Cooke & Zeeman, 1976). In the contemporary revision of this model, information from genes expressed in a gradient along the

rostral-caudal axis of the unsegmented presomitic mesoderm (PSM) is integrated with expression of genes that are cyclically transcribed to shift cells between permissive or nonpermissive states for segment formation at the determination front. A new somite boundary is determined based on the periodic interaction of the cycling genes and these gradients. Comparative analysis has revealed conserved features of the segmentation clock in these four model systems, including: gradients of FGF8 and WNT3a proteins (Takada et al., 1994; Crossley & Martin, 1995; Beck & Slack, 1998; Reifers et al., 1998; Dubrulle, McGrew, & Pourquie, 2001; Pera et al., 2002; Aulehla et al., 2003; Draper, Stock, & Kimmel, 2003; G. Chapman, Sparrow, Kremmer, & Dunwoodie, 2011; Thorpe, Weidinger, & Moon, 2005), cyclical expression of genes in the notch signaling pathway such as the hairy/enhancer of split (hes and her) genes (Holley, 2007; Sparrow, 2009; Dequéant & Pourquié, 2008), and expression of *mesp2* orthologues at the determination front (Saga, Hata, Koseki, & Taketo, 1997; Buchberger, Seidl, Klein, Eberhardt, & Arnold, 1998; Sawada et al., 2000; Hitachi et al., 2009).

Comparative studies have also revealed divergence in the segmentation clock among the vertebrate models. Not all PSM gradient genes are conserved. As an example, the notch effector hes6 orthologues are present in a gradient in the anamniotic models, *X. laevis*, and zebrafish, but not in the amniotic models, chicken or mouse (Pissarra, Henrique, & Duarte, 2000; Cossins, Vernon, Zhang, Philpott, & Jones,

2002; Fior & Henrique, 2005; Sieger, Ackermann, Winkler, Tautz, & Gajewski, 2006). Furthermore, the orthologous genes do not necessarily share cycling expression patterns in the PSM, dividing the vertebrates studied so far into two major groups. In zebrafish and *X. laevis*, notch is cyclically activated by the delta ligand (*deltaC/X-Delta-2*) orthologous to mouse *Dll3* (Holley, 2007; Jen, Gawantka, Pollet, Niehrs, & Kintner, 1999; Jen, Wettstein, Turner, Chitnis, & Kintner, 1997; Jen et al., 1997; 1999). In mouse and chicken, which are amniotes, notch is cyclically inhibited by lunatic fringe, a glycosyltransferase expressed in the Golgi complex that modifies the ability of notch receptor to bind to the delta ligand. In addition, dynamic expression of delta-like 1 (*Dll1*) has been reported in the mouse, but not in other vertebrates to date.

There are major gaps in the representation of phylogenetic groups in the study of the evolution of the segmentation clock. In particular, reptiles are a diverse class and current classifications include the birds, together with crocodilian reptiles such as the alligator, within the infraclass Archosauromorpha. Analysis of the segmentation clock in the Lepidosauromorpha, the other main infraclass within the diapsids that includes the lizards and snakes, would be instructive given that avian reptiles display many convergent developmental features with mammals, such as a four-chambered heart and endothermy. However, without full genomic resources, molecular analysis of the segmentation clocks of reptiles has been limited.

Recently, the first genome sequence of a non-avian reptile was reported for the green anole lizard, *Anolis carolinensis* (Alföldi et al., 2011). Using this genomic data, combined with transcriptome sequence generated by our group, we sought to analyze coding sequences and dynamic gene expression patterns, which may reflect non-coding regulatory changes. Our aim was to test whether the segmentation clock in the green anole would share greater commonality with the amniote models mouse and chicken, e.g., display lunatic fringe cycling expression in the PSM, or have unexpected similarities with the anamniote representatives, *X. laevis* and zebrafish. This analysis would give insights into regulatory changes in the segmentation clock associated with vertebrate evolution. Our results show that the segmentation clock in the green anole displays transitional features of both amniotes and anamniotes, pointing to a major shift in the segmentation clock associated with the divergence of amniotes.

Materials and Methods

Anole and alligator embryos. *Anolis carolinensis* lizards were housed at 70% humidity (14 hours at 28°C daylight and 10 hours at 22°C night) conditions. All animals were maintained according to Institutional Animal Care and Use Committee guidelines. Eggs were typically laid in the soil of a potted plant at a 25-30 somite pair stage. Embryos were

dissected from eggs in PBS, fixed in 4% paraformaldehyde, dehydrated and stored in methanol at -30°C.

Eggs from the American alligator, *Alligator mississippiensis*, were collected from a wild nest by staff of the Louisiana Department of Wildlife and Fisheries from the Rockefeller Wildlife Refuge. Eggs were collected approximately 2-3 days after laying and incubated until 10 days post laying. At that point, embryos were dissected from these eggs stored in RNAlater (Qiagen) for transit and later fixed in 4% paraformaldehyde, dehydrated and stored in methanol at -30°C.

RNA-Seq transcriptome analysis of anole embryos. To carry out our analysis of the segmentation clock in the lizard, we needed to improve annotation for somitogenesis genes from the *A. carolinensis* draft genome assembly (Anocar2.0) and cDNA sequences (Alföldi et al., 2011). Embryos at 28 and 38 somite pair stages were collected for extraction using the total RNA protocol of the miRVana kit (Ambion). Total RNA samples were prepared using the Ovation RNA-Seq kit (NuGEN) to generate double stranded cDNA, and Illumina reagents were used to generate paired end sequencing libraries following manufacturer protocols. Sequencing was carried out on a HiSeq 2000 (Illumina) using paired end chemistry with read lengths of 104 base pairs. Reads were mapped to the *Anolis carolinensis* genome using Bowtie and TopHat as described previously (Langmead et al., 2009). Based on the number of reads aligned to each transcript, Cufflinks was used to generate an estimation of transcript

abundance as Fragments Per Kilobase of exon per Million fragments mapped (FPKM). The FPKM estimation is generated by determining the likelihood for the abundances of the set of transcripts, based on the mapped fragments and reporting the abundancies with the maximum likelihood (Roberts, Pimentel, Trapnell, & Pachter, 2011a). RNA-Seq analysis of *A. carolinensis* embryos allowed us to identify coding sequences for orthologues of key somitogenesis genes, based on synteny conservation comparisons and sequence alignments as outlined by the *Anolis* Gene Nomenclature Committee (personal communication, committee members KK and JWR; Table 1 and Supplemental Table 5; Supplemental Figs. 2 & 3). Together with the *A. carolinensis* second genome draft (Anocar2.0; Alföldi et al., 2011), gene annotations were established and primers were selected for RT-PCR derived RNA probe synthesis.

Comparative analysis of coding sequences. Phylogenetic analysis was carried out using MEGA5 (Tamura et al., 2011). Amino acid sequences were aligned using MUSCLE (Edgar, 2004). Evolutionary history was inferred using the Neighbor-Joining method (Saitou & Nei, 1987). Bootstrap tests with 500 replicates were used to estimate the confidence in each branch point, which are displayed as the percentage of replicates in which the associated taxa cluster together (Felsenstein, 1985). Evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid

substitutions per site (Zuckerkandl & Pauling, 1965). Sequence identifiers (NCBI Reference Sequence IDs unless otherwise noted) for comparative analysis are presented in Supplemental Table 6.

Anolis carolinensis hes6b, *hes7a* and *hes7b* predicted sequences were determined through blast analysis of orthologues in mouse, identification of open reading frames and possible splice junction sequences. Sequences for *Anolis hes6b* and *hes7a* were further validated with RNA-sequencing data and RT-PCR for ISH probe DNA template generation. Annotated sequences used for MEGA analysis were as follows:

>A. carolinensis hes6b

MSAQIDVMVKSAYYQFQLCPFLELDDGSTDTGQVQWITMTATTLASGVPK
LPNPKEERKLRKPLIERKRRERINNCLDQLKETVVGAFHLDQSKLEKADI
LEMTVKHLQNIQTSKNVADSTTGLEAQQRYSTGYIQCMHEVHNLLLTCEW
MDKTLGARLLNHLLKSLPRSSEETSKADVNPSTTRSAKGMTTELNPSQDP
FYATEDRQGFKKPFQPHIVGTHCSQRKTSPPSQTLQPHFAHNGISMGSLD
MWRPW

>A. carolinensis hes7a

MEFWSKEVYHFVETVVDDGGFSSQTWTLKKPKVEKAEILEIAVGYLREMA
SAKSQGADFSEDRTLQTCFRVGFRECLLGLAAFLQQAHPSKIWNEPEPLR
PNPDPPCGSAGQHGDQPREAQGNADTTNKRIPLPPPAFWRPWP

>A. carolinensis hes7b

MEKRRRDRMNQSLDRLRVLLFEATQDEDSRSLTRNETKHKEEAFLQRYRS
GYRECLTQATHFLRGNSGLCQGKKAYLMEHICHCMEKIAASPRAETHQPP

STASSPGYGDLQQRYSPDVFASCSPALGGAPYVLHPPPAGCPIRQGLQAS

RMNGLGQPNGCSRPSSQQSKLSETRNPVTAQNPQALNVWRPWP

Anolis carolinensis mesp2 predicted sequences were determined through blast analysis of orthologues in mouse, identification of open reading frames and possible splice junction sequences. Sequence for *A. carolinensis mesp2* was further validated with RNA-sequencing data and RT-PCR for ISH probe DNA template generation. Annotated sequences used for MEGA analysis was as follows:

>*A. carolinensis mesp2*

MQHIGGLSWGAEGAGGVTVLLASPWEGPAGTLSPGVSWAPRGSLRMAPCT

LPRPESGPRRSASQREKLRMRRLARALRDLRRFLPPSLAPPARPLTKLQT

LRLAARYIAHLGDLLRLDHGALEGRGAGAGGGGLACGWEAGSPPETQWTW

QTPAPAPEEEAPQVSREASKGTPVSLWKRPPKAIQSDPSAIKTSLGPKAV

FILLELEETPKGHPV

In situ hybridization expression analysis. For generation of antisense probes for *in situ* hybridization, the T7 conserved sequence and clamp sequences (5'-GCGTAATACGACTCACTATAGGGAGA-3') were added to the 5' end of each reverse primer sequence. Primers were designated as forward when in the sense strand, and reverse when corresponding to antisense sequence. Whole mount *in situ* hybridization was carried out as described previously (Sewell, Sparrow, Smith, & Gonzalez, 2009).

***Anolis carolinensis*.** Forward and reverse primers for each lizard gene investigated are listed are listed below:

axin2 (5′-GTCAAGGCCAATGGTCAAGT-3′, 5′-GTTCCACCCCTTTTGAGTGA-3′)

dll1 (5′-TTGTTTCAATGGTGGGACCT -3′, 5′-ACATTTGCTGCGTTCCTCTT-3′)

dll3 (5′-GGTCCCCTTCCATTTCAAGT-3′, 5′-CCAAGTGCTTCTCAATGCAA-3′)

dll4 (5′-GAAGACAGGTGCACCAACAA-3′, 5′-AATCCCTTGGGGAGCATATT-3′)

dusp6 (5′-CAGCTCTCCCATCCAAACTC-3′, 5′-GGGGGAAATGTTGGATTTCT-3′)

fgf8 (5′-TGCACTTGTTTGCTTTCTGC-3′, 5′-GACCACTCCCTGTTAGTGCC-3′)

hes4 (5′-CATTCCAAGCTGGAGAAAGC -3′, 5′-AAGGCCTCCAGACCGAGT-3′)

hes6a (5′-AAAAGCCGCGAGGAAGAG -3′, 5′-CCAGGGTCTCCACACAGATT-3′)

hes7a (5′-TCTTGATTGTTCCCCAGTCC-3′, 5′-GGGGAATCCTTTTGTTGGT-3′)

hey1 (5′-AATGTTGCACACAGCAGGAG-3′, 5′-ATCTCAGTCCCCCAAGGTCT-3′)

hey2 (5′-AGGGTTCGACTCGTCTCTCA-3′, 5′-TGGTCGGTAGGGCTTACTGT-3′)

lfng (5′-TATCTTCACGGACGGAGAGG -3′, 5′-GAATGGAAGAGGTTGCTTCG-3′)

mesp2 (5′-CCTCTTCCCCATCTTCTTCC -3′, 5′-GGAGGAGGTCTCCGAGGT-3′)

notch1 (5′-ACCGAGTCCAGCAAGAAGAA-3′, 5′-ATACAGTCGGCGTCGATTTC-3′)

spry2 (5′-GTCAGCATGACAGTGGGAGA-3′, 5′-AAGGGTTATCGGCACAGTTG-3′)

tbx6 (5′-GGAATACGGATGAGCTTGGA -3′, 5′- TCATTTGGGTGATCTGTGGA-3′)

wnt3a (5′-CACCAGGGAGTCAGCCTTTG-3′, 5′-GCAGTGGCACTTCTCTTTCC-3′)

Alligator mississippiensis. Primers for RT-PCR derived RNA probes for alligator were selected based on sequence conservation between the mouse, lizard and chick lunatic fringe genes. Forward and reverse primers generated are:

LFNG (: 5′-CTTCACGGATGGGGAGGA-3′: 5′-TGAGAGTGGAAGAGGTTGCT-3′)

DLL1 (5′-TGTGCCTCAAGCACTACCAG-3′: 5′-CAGCTTCCACCATTCTTGC-3′)

Analysis for conserved regulatory elements. Genomic sequence of delta-like 1 and lunatic fringe orthologues from *A. carolinensis*, *X.*

tropicalis, zebra finch, chicken, and mouse, including the coding region plus 10 kb of both 5' and 3' flanking sequence, were analyzed by MUSSA for potential N and E-box regulatory elements (Kuntz et al., 2008). Elements were mapped to the delta-like 1 genomic sequence that met the criterion of sequence identity for 16/18 nucleotides in the window of analysis for N and E-box consensus sequences from the TRANSFAC database (Matys et al., 2006).

Genomic sequences of lunatic fringe orthologues from mouse, chick and *A. carolinensis* up to 5 kb upstream of the start codon were analyzed for N and E consensus sites using Cister (Frith, Hansen, & Weng, 2001). Analysis was carried out using TRANSFAC N and E-box consensus sequences and a specific E-box sequence with TC nucleotides in the central position, which has been demonstrated to be functionally required for cycling expression of *Lfng* in the mouse (Cole et al., 2002; Morales et al., 2002), shown below:

E-box TC

XX					
PO	A	C	G	T	
01	1	97	1	1	C
02	97	1	1	1	A
03	5	5	5	85	T
04	5	85	5	5	C
05	1	1	1	97	T
06	1	1	97	1	G
XX					

E-box

XX					
PO	A	C	G	T	
01	0	100	0	0	C
02	100	0	0	0	A
03	25	25	25	25	N
04	25	25	25	25	N
05	0	0	0	100	T
06	0	0	100	0	G
XX					

N-box CACNAG

XX					
PO	A	C	G	T	
01	1	97	1	1	C
02	97	1	1	1	A
03	1	97	1	1	C
04	20	20	40	20	N

N-box CTCNTG

XX					
PO	A	C	G	T	
01	1	97	1	1	C
02	1	1	1	97	T
03	1	97	1	1	C
04	20	20	40	20	N

```
05   97    1    1    1    A    05    1    1    1   97    T
06    1    1   97    1    G    06    1    1   97    1    G
XX                             XX
```

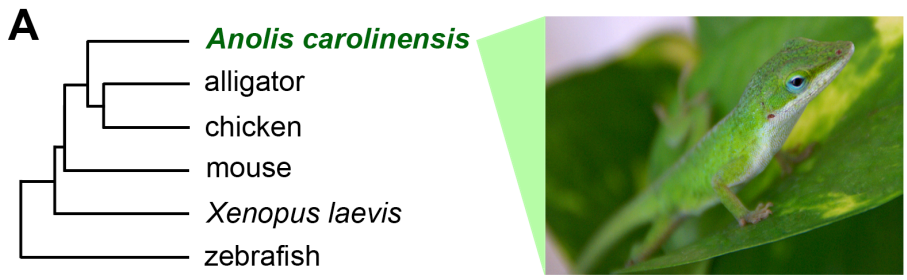
Accession numbers. Genomic and cDNA data described in this paper have been deposited into GenBank under accession numbers 28S RNA-Seq data, GSM848765; 38S RNA-Seq data, GMS848766; for *A. carolinensis* (*axin2*, JQ303083; *dll1*, JQ303084; *dll3*, JQ303085; *dusp6*, JQ303086; *hes6a*, JQ303087; *hes7a*, JQ303088; *hey1*, JQ303089; *hey2*, JQ303090; *lfng*, JQ303091; *notch1*, JQ303092; *spry2*, JQ303093; *tbx6*, JQ303094; *wnt3a*, JQ303095) and *A. mississippiensis* (*DLL1*, JQ303096; *LFNG*, JQ303097). RNA-Seq data are available from the NIH Gene Expression Omnibus (GEO) record series GSE34415.

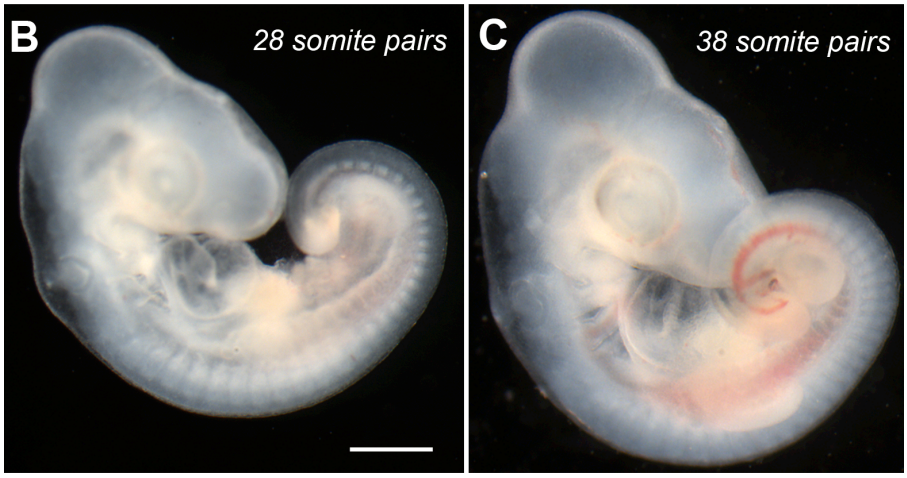
Results

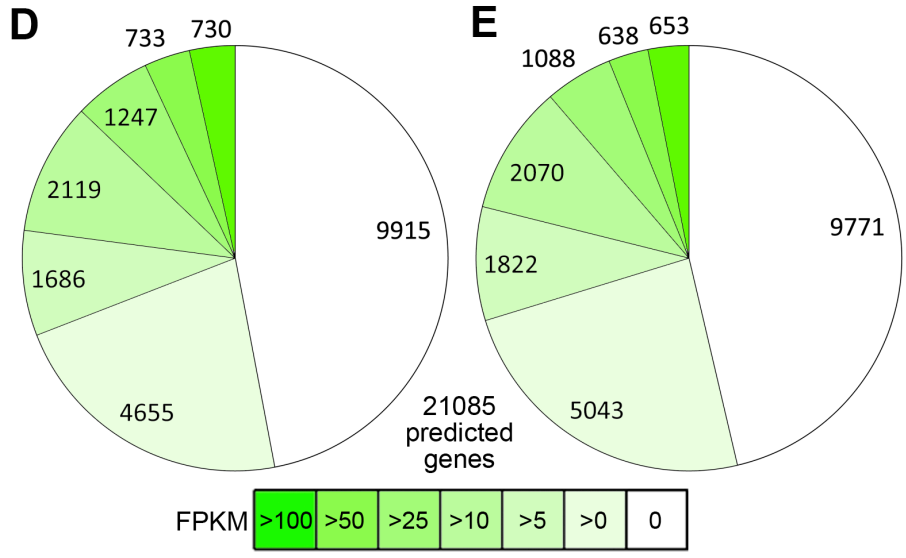
Transcriptomic analysis of anole embryos identifies segmentation clock genes. To carry out analysis of the segmentation clock in the green anole lizard, an essential first step was to identify orthologous genes. Public databases have generated predicted gene models based on the *A. carolinensis* Anocar2.0 genome assembly (Ensembl, 2011); however, these predictions were incomplete and data from available cDNA sequences were not incorporated. Furthermore, segmentation clock genes are not well represented in the current cDNA libraries sequences, which derive mostly from adult tissues, so we carried out RNA-Seq transcriptome analysis focused on somite-stage embryos. We collected

two *A. carolinensis* embryos from newly laid eggs, at 28 and 38 somite pair stages (Fig. 4B,C). RNA-Seq analysis determined that 53% of the total predicted genes were expressed at this developmental stage, i.e., 11,170 and 11,314 transcripts from the 28 and 38 somite stage embryo samples, respectively, displayed a range of non-zero Fragments Per Kilobase transcript per Million reads (FPKM) values (Fig. 4D,E; Table 4; Supplemental Table 5). The FPKM value provides an estimate of mRNA quantity, based on a probability distribution of transcript abundance derived from the number of aligned sequencing reads. Transcripts with zero FPKM represented genes that were expressed at extremely low levels in the whole embryo. These two embryos were both collected at mid-somitogenesis stages, and as might be expected, the FPKM levels for all transcripts had a very high Pearson's correlation coefficient of 0.99534. We were able to refine annotations for key segmentation clock genes in the anole, based on transcriptomic and genomic data (Table 4). Assignments of orthology were confirmed by analysis of synteny conservation (Supplemental Fig. 2).

Using these refined gene annotations for *A. carolinensis*, we carried out comparative sequence analysis of segmentation clock genes (Fig. 4F; Supplemental Fig. 3). Compared with housekeeping genes such as *gapdh* and *actn1*, the segmentation clock genes varied considerably in ClustalW scores (Fig. 4F; Larkin et al., 2007) and multiple sequence alignment based analysis by MEGA (Supplemental Fig. 3). Predicted protein


A
Anolis carolinensis
alligator
chicken
mouse
Xenopus laevis
zebrafish


B
28 somite pairs
C
38 somite pairs


D
733
730
1247
2119
9915
1686
4655
21085
predicted
genes
E
1088
638
653
2070
9771
1822
5043
FPKM
>100
>50
>25
>10
>5
>0
0

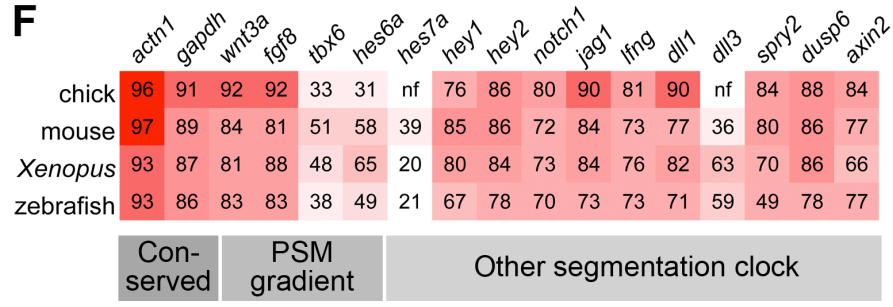

F
actn1
gapdh
wnt3a
fgf8
tbx6
hes6a
hes7a
hey1
hey2
notch1
jag1
lfng
dll1
dll3
spry2
dusp6
axin2
chick 96 91 92 92 33 31 nf 76 86 80 90 81 90 nf 84 88 84
mouse 97 89 84 81 51 58 39 85 86 72 84 73 77 36 80 86 77
Xenopus 93 87 81 88 48 65 20 80 84 73 84 76 82 63 70 86 66
zebrafish 93 86 83 83 38 49 21 67 78 70 73 73 71 59 49 78 77
Con-served
PSM gradient
Other segmentation clock

Figure 4. Next-gen transcriptome sequencing identifies segmentation clock genes in the green anole lizard, *Anolis carolinensis.* **A**, Consensus analysis of coding sequence variation places the anole most closely related to the chicken among other developmental model organisms. An adult female green anole is shown on the right. **B,C,** Total RNA from 28 (**B**) and 38 (**C**) somite stage green anole embryos were analyzed by RNA-Seq next-generation transcriptome sequencing to refine gene model predictions of the *A. carolinensis* genome. Distribution by level of gene expression for the 21,085 predicted genes in the *A. carolinensis* genome, by fragments per kilobase exon model (FPKM) scores from the 28 somite (**D**) and 38 somite (**E**) embryos. 53% (11,170/21,085 at 28 somite stage and 11,314/21,085 at 38 somite stage) embryos displayed greater than zero FPKM values. **F,** Using predicted gene models confirmed by RNA-Seq data, the amino acid sequences of green anole segmentation clock genes, as well as conserved housekeeping genes *actn1* and *gapdh*, were compared to orthologous genes in chicken, mouse, frog and zebrafish. Similarity scores generated by ClustalW (Larkin et al., 2007) are shown in a heat map.

Table 4.

RNA expression levels of Anolis carolinensis orthologues of key somitogenesis genes.

Gene	Chromo-some/scaffold	5' end start bp	3' end finish bp	28 Somite (FPKM)	38 Somite (FPKM)
axin2	GL343260.1	882504	926506	11.5	12.1
dll1	1	223341965	223355331	43.9	36.0
dll3	GL343635.1	268700	251452	1.0	0.6
dll4	GL343264.1	67781	53723	2.3	1.6
dusp6	5	32945289	32954893	2.3	2.6
fgf8	GL343239.1	1277501	1254864	3.9	13.1
hes4	GL343334.1	1044217	1038209	9.6	7.4
hes6a	3	28075541	28070611	1.1	2.9
hes7a	GL343400.1	571697	574062	0.0	0.0
hey1	4	25153734	25159192	4.3	2.5
hey2	1	180536752	180558526	4.3	5.3
jag1	1	134088765	134141003	33.1	19.5
lfng	GL343340.1	209497	250732	1.8	1.5
mesp2	GL344154.1	13650	5835	0.0	0.0
notch1	AAWZ02036232	2523	29071	0.4	0.4
spry2	3	96689658	96690599	2.5	3.2
tbx6	GL343279.1	365052	356983	0.0	0.6
wnt3a	6	3407698	3441632	0.0	3.0

sequence similarity scores of the notch effector genes *hes6a* and *hes7a* and the notch ligand *dll3* were particularly divergent (Fig. 4F; Supplemental Fig 3) compared with other segmentation clock genes such as *notch1, lfng*, and *axin2*.

hes6a is expressed in a gradient in the PSM of the anole.

Somite boundaries are determined through gradients of gene expression within the PSM interacting with genes in the determination front. Anole orthologues of three genes, *wnt3a*, *fgf8,* and *tbx6*, displayed expression in the posterior PSM (Fig. 5A-C), similar to reports in other vertebrate models (Beck & Slack, 1998; Crossley & Martin, 1995; Larkin et al., 2007; Reifers et al., 1998; Takada et al., 1994; Dubrulle et al., 2001; Pera et al., 2002) (Aulehla et al., 2003; Draper et al., 2003; G. Chapman et al., 2011; Thorpe et al., 2005). Similarly, the anole orthologue of *mesp2*, a key gene in the determination front that regulates the prepatterning of somite boundaries (Morimoto et al., 2007; Pissarra et al., 2000; Takahashi, Yasuhiko, Kitajima, Kanno, & Saga, 2007), was expressed in the –I region of the PSM (Fig. 5G). In contrast, the notch pathway hairy-enhancer of split (hes) gene 6 is expressed in the posterior PSM in *Xenopus* and zebrafish but not in mouse and chicken (Holley, 2007; Fior & Henrique, 2005; Pissarra et al., 2000). The *hes6a* orthologue was expressed in the posterior PSM (Fig. 5D,E) and in a band in the rostral PSM of the anole, unlike the chicken and mouse. Surprisingly, *hes6a* is a PSM gradient gene

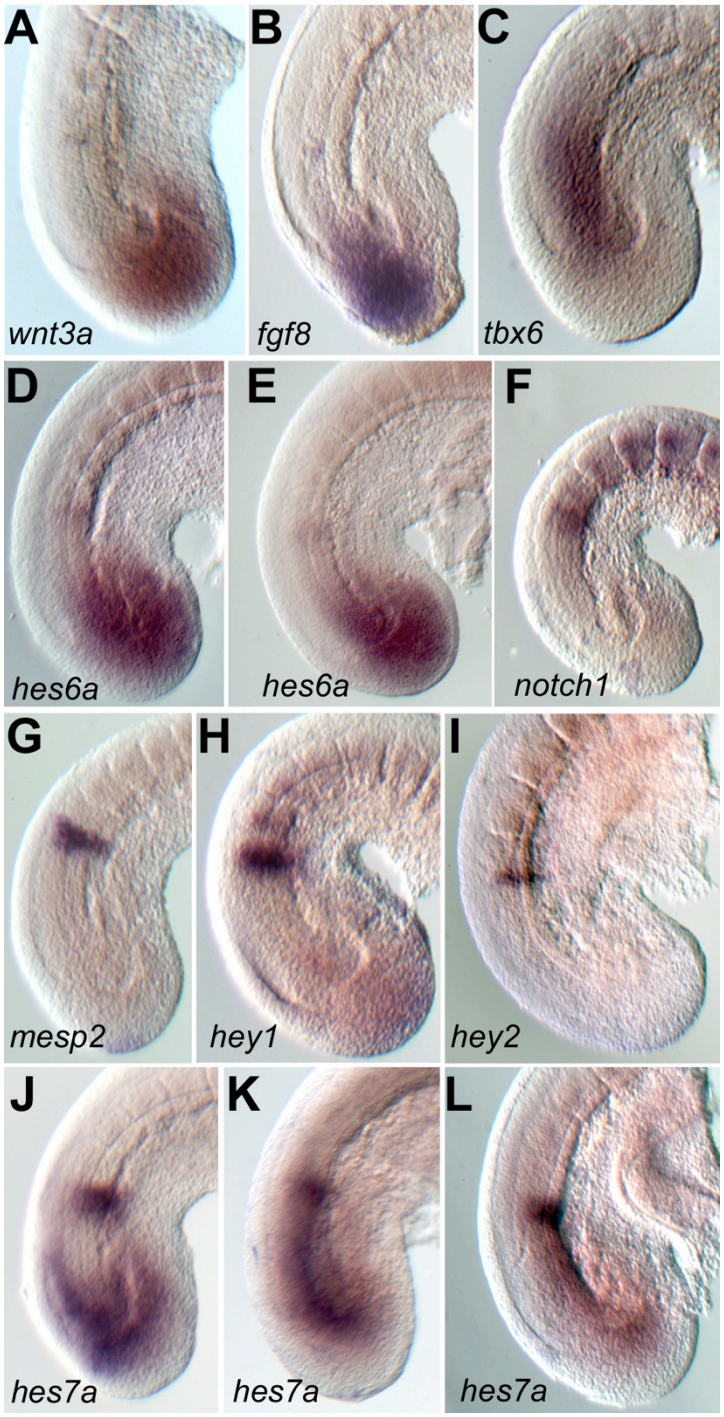

A
wnt3a
B
fgf8
C
tbx6
D
hes6a
E
hes6a
F
notch1
G
mesp2
H
hey1
I
hey2
J
hes7a
K
hes7a
L
hes7a

Figure 5. PSM gradient and determination front gene expression is conserved in the green anole, but *hes6a* PSM gradient expression is unique among amniotes. Expression of segmentation clock genes in anole embryos analyzed by whole mount *in situ* hybridization. **A-E**, Gradients of expression of *wnt3a* (n=3, **A**), *fgf8* (n=3, **B**), *tbx6* (n=3, **C**) and *hes6a* (n=6, **D-E**) are observed in the PSM of the anole. In addition to the PSM gradient, *hes6a* is expressed in a band in the rostral PSM (**D-E**). **F**, The expression of *notch1* is observed in the rostral PSM and in the somites (n=3). **G,** The expression of the determination front gene *mesp2* is observed as a rostral band at somite-I level (n=19). **H,I,** Hairy/enhancer of split genes *hey2* (n=7, **H**) and *hey1* (n=2, **I**) are expressed in stage-specific rostral bands in the PSM. **J-L**, The hes gene *hes7a* displays cycling expression in the PSM (n=9, **J,** phase I; **K,** phase II; and **K,** phase III).

for anoles, similar to *Xenopus* and zebrafish, suggesting that squamate reptiles have transitional features in common with anamniotes.

hes7a is a cycling gene in the PSM of the anole. The hes7 orthologue is a primary molecular oscillator driving cyclical expression of other notch pathway cycling genes in vertebrates (Bessho et al., 2001; Niwa et al., 2007). We identified two tandem duplications of hes7 in the anole, and the *hes7a* orthologue displayed cycling expression (Fig. 5J-L). Other hairy/enhancer of split genes such as *hey1* and *hey2*, which display cycling expression in the mouse (Leimeister et al., 2000), were not evidently dynamic in anole embryos (Fig. 2H,I). The anole *hes4* gene is an orthologue of the mouse cycling gene *Hes1* (Jouve et al., 2000) but was not expressed within the PSM (Supplemental Fig. 3C). The segmentation clock regulates the activation of the notch receptor, and consistent with this role, *notch1* is expressed in the determination front and the newly formed somites (Fig. 5F). The anole *jag1* orthologue is expressed in a similar pattern to the mouse notch ligand *Jag1*, which is seen in a static band in the anterior PSM and in the tailbud (Xue et al., 1999); Supplemental Fig. 4D). Wnt and FGF pathway genes have been identified with cyclical expression in the mouse and chicken (Krol et al., 2011; Gibb et al., 2009; Aulehla et al., 2003), with no evidence for oscillatory expression of these genes in the corn snake, *Pantherophis guttatus* (Gomez et al., 2008). The expression of the anole orthologues of the wnt pathway gene *axin2* and the FGF gene *dusp6* were not dynamic in

expression (Supplemental Fig. 4A,B). Thus, *A. carolinensis* shares components of the hes-driven segmentation clock, such as *hes7a*, but the cyclical expression of other genes are not conserved with the mouse and chicken.

Lunatic fringe is not expressed in the PSM of anole embryos and lacks enhancer regions found in mouse and chicken. Given the cyclical expression of the modulator lunatic fringe in both chick and mouse, this gene was thought to be a key dynamic notch regulator in amniotes (Barrantes et al., 1999; Evrard, Lun, Aulehla, Gan, & Johnson, 1998; Forsberg et al., 1998; Serth, Schuster-Gossler, Cordes, & Gossler, 2003). However, we observed that anole *lfng* was not expressed in the PSM and instead localized to the rostral compartment of the first new somites (Fig. 6A). As in the mouse and the chicken, anole *lfng* is also expressed in the developing neural tube (Fig. 6B,C). Thus, *lfng* is not a cycling gene in the anole, suggesting either that PSM-specific expression was lost in ancestors of the anoles, or that there is potential convergence between birds and mammals. Amino acid sequences of *A. carolinensis* fringe genes were compared between the other vertebrate orthologues using MEGA 5 (Tamura et al., 2011). As expected, *A. carolinensis* fringe genes shared the most sequence similarity with chicken fringe genes (Fig. 6D), so coding sequence divergence is not a likely mechanism for the differences observed in the anole.

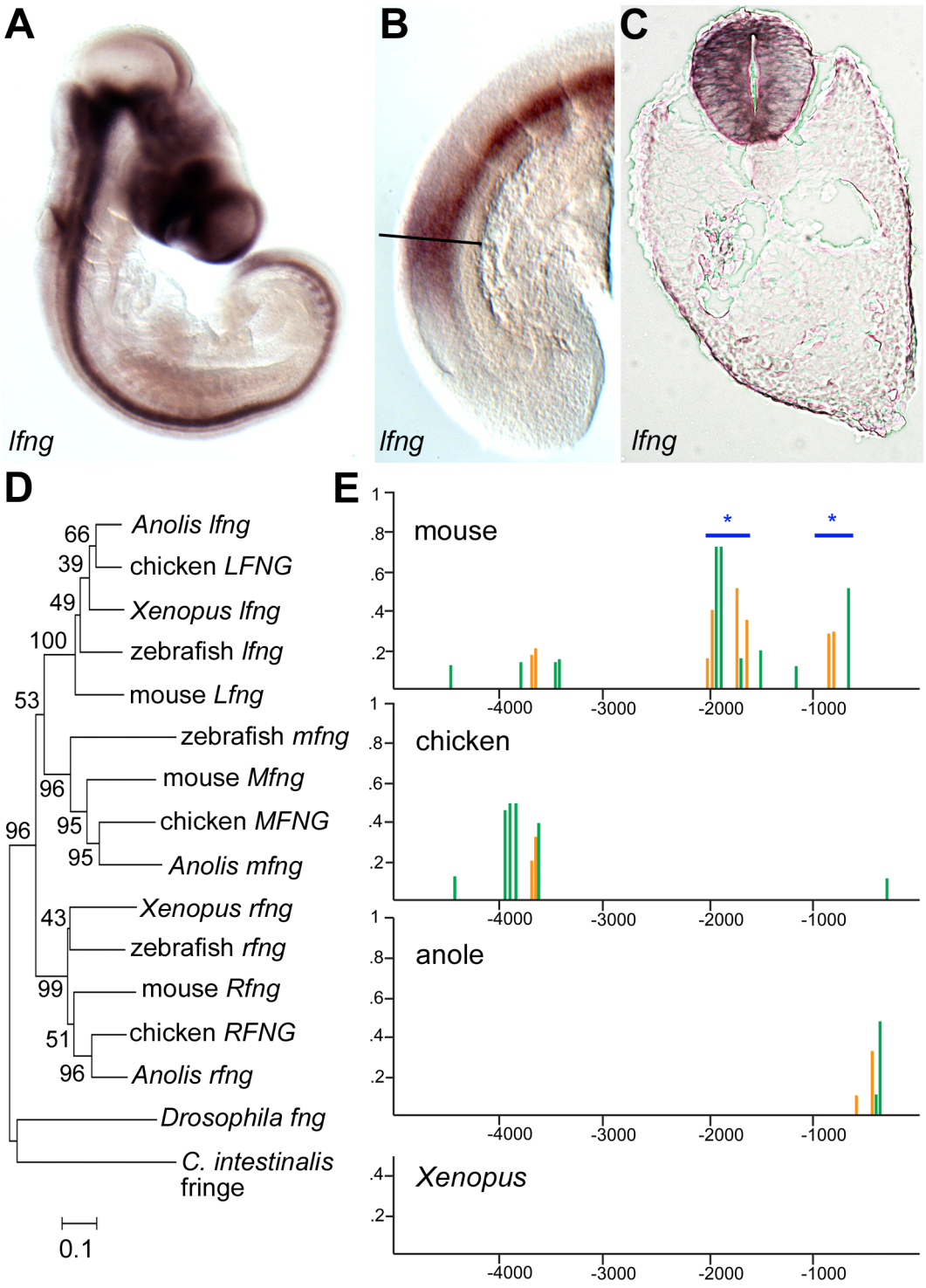

A
lfng
B
lfng
C
lfng
D
Anolis lfng
chicken LFNG
Xenopus lfng
zebrafish lfng
mouse Lfng
zebrafish mfng
mouse Mfng
chicken MFNG
Anolis mfng
Xenopus rfng
zebrafish rfng
mouse Rfng
chicken RFNG
Anolis rfng
Drosophila fng
C. intestinalis fringe
66
39
49
100
53
96
95
95
96
43
99
51
96
0.1
E
mouse
chicken
anole
Xenopus
-4000
-3000
-2000
-1000

Figure 6. Lunatic fringe is not expressed in the PSM of green anole embryos. **A,** Expression of *lfng* in the lizard embryo is localized primarily to the newly formed somites and the neural tube, as shown by whole mount in situ hybridization (n=7). **B,** Lateral view of the PSM and newly formed somites shows expression in the ventral regions of the neural tube rostrally and broadly in the neural tube adjacent to the PSM. No expression is observed in the PSM. Line in the PSM indicates level of section analyzed in **C**. **C,** The expression of *lfng* is restricted to the neural tube as confirmed by transverse cryosection. **D,** Comparison of amino acid coding variation of *A. carolinensis* fringe genes with other vertebrates, as well as *Drosophila* and *Ciona intestinalis*, was inferred using the Neighbor-Joining method displaying the optimal tree using MEGA 5 (Saitou & Nei, 1987; Tamura et al., 2011). Values shown are the percentage of replicate trees in which the associated taxa clustered together using the bootstrap test with 500 replicates. Evolutionary distances were estimated using the Poisson correction method and are in units of amino acid substitutions per site (Zuckerkandl & Pauling, 1965). **E,** The distribution of conserved N (orange) and E-box (green) noncoding elements within 5 kb 5' of *A. carolinensis lfng* compared with mouse, chicken and *Xenopus tropicalis*. Height reflects the probability of element functionality as determined by Cister (Frith et al., 2001). Mouse N and E-boxes in the -1 to -2 kb regions have been demonstrated to be required for cycling expression (blue lines and asterisks; Cole et al., 2002; Morales et

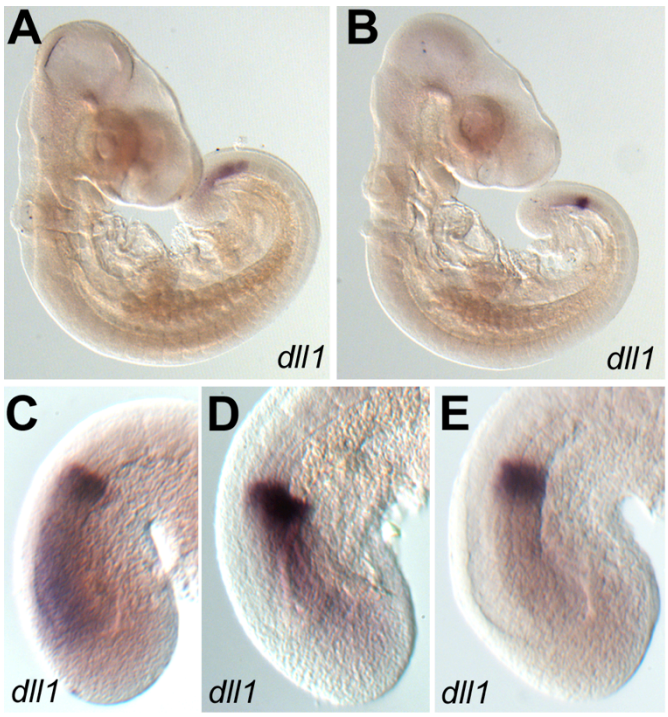
al., 2002). There is a cluster of N and E-box sequences -3 to -4 kb upstream of chicken *LFNG*, which also displays cycling expression in the PSM. *A. carolinensis lfng* is not expressed in the PSM, and N and E-box clusters of comparable size were not identified in this region. Similarly, no N and E-box clusters where found in *Xenopus* sequences.

Cyclical expression in the PSM is driven by regulatory feedback loops involving binding of hes proteins such as HES7 to non-coding regulatory elements (J.-F. Chen et al., 2005; Cole et al., 2002; Morales et al., 2002). These hes bHLH transcription factors bind to regulatory sequences, termed N and E-boxes, which are conserved among metazoans (reviewed in Davis and Turner, 2001). To test whether changes in the number or distribution of N and E-box elements could account for the convergence of lunatic fringe expression in the mouse and the chicken, we compared genomic regions up to 10 kb upstream and downstream of these genes. Changes in sequences associated with lunatic fringe cycling expression were observed. Specifically, a large cluster of N and E-box elements was identified in -3 to -4 kb upstream region in the chicken *LFNG*, comparable to a -1 to -2 kb upstream cluster in mouse *Lfng* that has been demonstrated to be required for cycling expression (Cole et al., 2002; Morales et al., 2002). No comparable enhancer cluster was identified in anole *lfng* (Fig. 3E). Thus, the difference between the -3 to -4 kb upstream N and E-box cluster in the chicken vs. the -1 to -2 kb cluster in the mouse and lack of a major cluster in the anole may reflect convergent molecular evolution of noncoding elements required for lunatic fringe cycling expression.

Delta-like 1 is a cycling gene in the anole. Dynamic expression of delta-like 1 has been reported in the mouse PSM (Maruhashi et al., 2005); however, given expression of *Dll1* throughout the mouse PSM, it is

difficult to detect clearly sweeping bands of expression within this region. In the anole, *in situ* hybridization analysis showed that *dll1* was limited in expression within the PSM and displayed clear phases of cycling expression (Fig. 7A-E). The *dll1* expression pattern can be categorized into three distinct phases of expression, with expression shifting from the caudal to the rostral PSM (Fig. 7C-E). Comparative analysis of delta-like protein sequence did not reveal any unexpected similarity between the mouse and anole orthologues (Fig. 7F).

Delta-like 3 does not display cycling expression in the anole but conserves EGF repeat structure with anamniotes orthologues. Amniotes and amphibians have three paralogues of delta-type notch ligands, and the second delta group, which includes frog *X-Delta-2,* mouse *Dll3*, and zebrafish *deltaC*, is divergent in domain structure and cyclical expression (Fig. 8D). Anole *dll3* is expressed in stationary, non-cycling bands within the rostral PSM and tailbud (Fig. 8A-C), but its localization to the caudal somite compartment (Fig. 8C) differs from mouse *Dll3* localization to the rostral compartment (Dunwoodie, Henrique, Harrison, & Beddington, 1997; Kusumi et al., 1998). The zebrafish *deltaC* and the *Xenopus laevis X-Delta-2* are cycling genes within the PSM (Holley, 2007; Holley, Geisler, & Nüsslein-Volhard, 2000; Holley, Jülich, Rauch, Geisler, & Nüsslein-Volhard, 2002; Jen et al., 1997; 1999; Jiang et al., 2000). In contrast, the mouse orthologue *Dll3* is expressed throughout the PSM


A
dll1
B
dll1
C
dll1
D
dll1
E
dll1

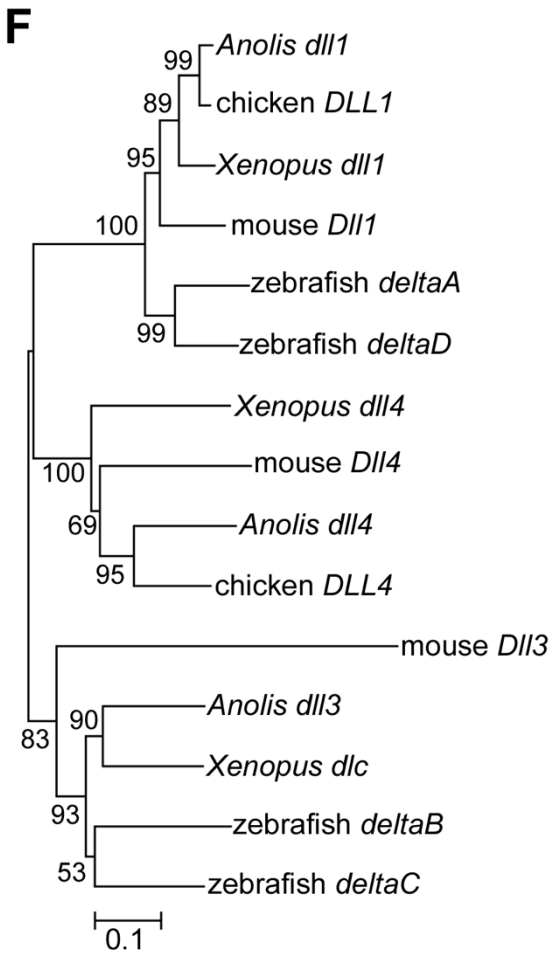

F
Anolis dll1
99
chicken DLL1
89
Xenopus dll1
95
mouse Dll1
100
zebrafish deltaA
zebrafish deltaD
99
Xenopus dll4
mouse Dll4
100
Anolis dll4
69
chicken DLL4
95
mouse Dll3
Anolis dll3
90
Xenopus dlc
83
zebrafish deltaB
93
zebrafish deltaC
53
0.1

Figure 7. *Anolis carolinensis dll1* displays cycling expression in the presomitic mesoderm. **A,B**, Expression of *dll1* is localized primarily in the lizard presomitic mesoderm (PSM), where dynamic shifts are observed. **C-E** Dynamic expression of *dll1* can be categorized into three phases characteristic of cycling genes (n=19). **F,** The evolutionary history comparing amino acid coding variation of *A. carolinensis* delta genes to other vertebrates was inferred using the Neighbor-Joining method displaying the optimal tree using MEGA 5 (Saitou & Nei, 1987; Tamura et al., 2011). The asterisk indicates the jagged/serrate ligand group illustrated in Supplemental Fig. 2. Values shown are the percentage of replicate trees in which the associated taxa clustered together using the bootstrap test with 500 replicates. Evolutionary distances were estimated using the Poisson correction method and are in units of amino acid substitutions per site (Zuckerkandl & Pauling, 1965).

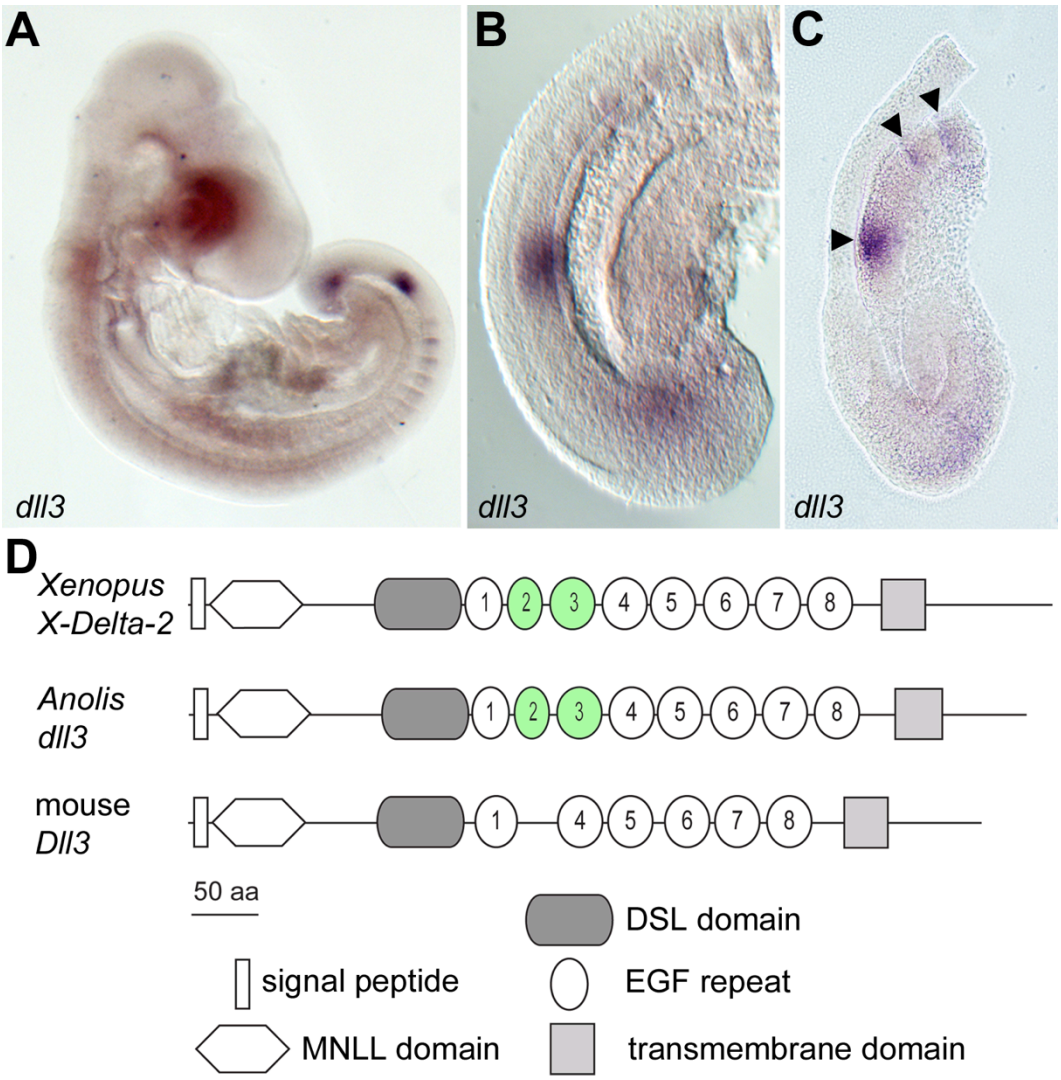

A
dll3
B
dll3
C
dll3
D
Xenopus
X-Delta-2
Anolis
dll3
mouse
Dll3
1
2
3
4
5
6
7
8
50 aa
DSL domain
signal peptide
EGF repeat
MNLL domain
transmembrane domain

Figure 8. *Anolis carolinensis dll3* is expressed in the presomitic mesoderm, and has a domain organization conserved with anamniotes. **A,** Expression of *dll3* in the lizard embryonic PSM and kidneys by whole mount *in situ* hybridization (n=7). **B,C,** Expression of *dll3* in the presomitic mesoderm localizes to a rostral band a somite-I level and the caudal PSM (**B**) and in the caudal compartment of the newly formed somites by sagittal section (**C**, black arrowheads). **D**, Diagram of functional domains of *Xenopus X-Delta-2*, lizard *dll3*, and mouse *Dll3* illustrate the loss of EGF repeats 2 and 3 in the mouse gene (highlighted in green). MNLL, N-terminal domain; DSL, delta-serrate-lag-2; EGF, epidermal growth factor.

(Dunwoodie et al., 1997; Kusumi et al., 1998), has lost two EGF repeat domains, and functions as an inhibitor of notch signaling in somitogenesis (Fig. 6D; G. Chapman et al., 2011; Geffers et al., 2007). A delta-like 3 orthologue has not been discovered in the three available bird genomes and may have been deleted during avian evolution (Supplemental Fig. 2). The sequence of the second delta ligand in the anole was most similar to *X. laevis X-Delta-2* and *X. tropicalis dlc* (Fig. 7F). Therefore, while the domain structure of anole *dll3* shares similarity with anamniotes, the lack of cycling in anole and mouse may reflect an ancestral change in non-coding regulatory function prior to amniote divergence.

In the alligator, lunatic fringe is not expressed in the PSM and DLL1 does not display cycling expression. To further investigate the divergence of lunatic fringe and delta-like 1 expression among amniotes, particularly in Archosaurian evolution, we examined the expression of *LFNG* in the American alligator, *Alligator mississippiensis* using *in situ* hybridization. Embryos collected at 10 days after laying in the alligator were comparable to the stages of newly laid anole eggs (Fig. 9B). Birds and crocodilian reptiles such as the alligator are both classified in the division Archosauria, and commonality of gene expression between chicken and alligator would suggest that the regulatory changes occurred prior to the divergence of this group. Intriguingly, alligator *LFNG* was expressed in the neural tube and somites (Fig. 9C), but not in the PSM of somite-stage embryos (Fig. 9C-D), like the anole but unlike the chicken.

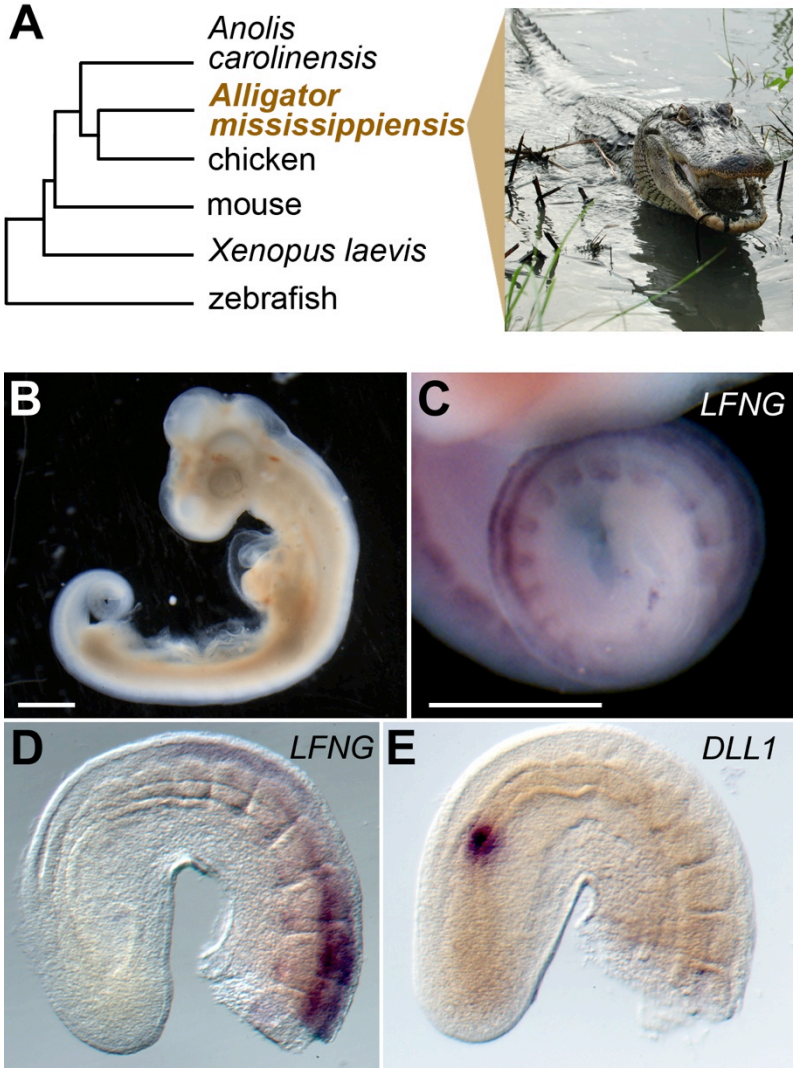

A
Anolis carolinensis
Alligator mississippiensis
chicken
mouse
Xenopus laevis
zebrafish
B
C
LFNG
D
LFNG
E
DLL1

Figure 9. *Alligator mississippiensis* embryos do not display *LFNG* expression in the PSM, in a comparable pattern with *A. carolinensis* and unlike the mouse and the chicken, and expression of *DLL1* in the alligator PSM. **A,** *A. mississippiensis* embryo at approximately day 10 of development. Scale bar is 1mm. **B,** *A. mississippiensis LFNG* is expressed in the neural tube and somites but expression is not observed in PSM (**C**; n=5). **D,** Alligator *DLL1* expression is observed as band in the rostral PSM that does not display any evident dynamic shift in expression (n=5).

Furthermore, alligator *DLL1* was expressed in a band within the rostral PSM similar to that observed in the green anole, but unlike the lizard, no evidence of cycling expression was observed in 15 embryos examined (Fig. 9E). Thus, the expression of *LFNG* in the alligator was more comparable to the anole than the chicken, but both the chicken and alligator do not demonstrate *DLL1* cycling expression in somitogenesis.

Discussion

Our analysis of somitogenesis in the green anole lizard, *Anolis carolinensis,* and the American alligator, *Alligator mississippiensis*, provides novel insights into the evolution of vertebrate somitogenesis and the segmentation clock in amniotes. Shared features between anoles and the chicken, mouse, *Xenopus*, and zebrafish include the fgf8 and wnt3a PSM gradients, mesp2 expression at the determination front, and cyclical expression of the notch pathway hes genes. However, we have identified four divergent features from other vertebrate models, which suggest major shifts in regulation of the segmentation clock associated with the evolution of the amniotes: 1) retention of a *hes6a* PSM gradient in the anole, suggesting this is a basal feature of the amniotes lost in the mouse and chicken; 2) loss of cyclical expression of dll3/deltaC orthologues in the PSM in amniotes, including the anole; 3) cyclical or dynamic expression of dll1 orthologue in the anole and mouse but not the alligator, suggesting this oscillatory expression arose in the amniotes but was lost in the

Archosaurian ancestor, and 4) cyclical expression of lunatic fringe in the mouse and chicken but not the anole and alligator. These changes are associated with divergence in coding and noncoding sequences that has arisen during the evolution of vertebrate somitogenesis.

The evolution of coding versus noncoding sequences

developmental genes. The segmentation clock is driven by the expression of genes with tightly regulated spatial and temporal patterns of expression. Amino acid conservation of the *dll1*, *wnt3a*, *fgf8*, and *lfng* match expectations between the anole and other vertebrates (Fig. 4F). The hes genes, including hes6 and hes7 orthologues (Fig. 4F) have high divergence when comparing full amino acid sequence, but are much more conserved in the functional basic and helix-loop-helix domains. The standout exceptions include the delta-like 3 orthologue, which is highly divergent in the mouse due both to loss of two EGF repeat domains in this ligand and general sequence divergence (Fig. 8D). The divergence of mammalian *Dll3* has also been associated with a change in cellular localization to the Golgi complex and shift in functional role from a trans-activatory ligand to primarily a cis-inhibitory factor (G. Chapman et al., 2011; Geffers et al., 2007; Ladi et al., 2005). In addition, there has been no avian orthologue of dll3 identified (Supplemental Fig. 1C). Within the vertebrates, there is divergence in both the localization of the dll3 expression within the PSM and whether there is cycling expression. In the anole, *Xenopus*, and zebrafish, the dll3 orthologues are expressed in two

regions, i.e., the tailbud and the rostral PSM; however, this pattern is static in anoles but cycling in *X. laevis* and zebrafish (Fig. 8A-C). Like the anole, *Dll3* in the mouse is static in expression, but unlike the anole, the expression extends to the entire PSM (Dunwoodie et al., 1997; Kusumi et al., 1998). Thus, the similarity in the PSM expression of dll3 orthologues between the anole and *X. laevis/*zebrafish suggests an ancestral amniote pattern, with potential regulatory changes leading to a loss of cycling expression.

To display cycling expression, segmentation clock genes first must be expressed in the PSM, which requires transcriptional activation specific to these cells. Paraxial mesoderm-specific enhancers have been identified for mouse *Dll1* (Beckers, Schlautmann, & Gossler, 2000; White & Chapman, 2005; White, Farkas, & Chapman, 2005). For lunatic fringe, there is no PSM expression in the anole and alligator but there is cycling expression in the chicken and mouse. PSM-specific enhancers have been identified in mouse *Lfng* by deletion analysis (Cole et al., 2002; Morales et al., 2002). Potential N and E-box binding sites were not identified -500 to -2,000 bp 5' flanking sequence in the chicken, anole or *X. laevis* (Fig. 6E), but intriguingly, the chicken 5' flanking region contains a more distant cluster of N- and E-boxes that are potential hes binding sites. Expression of lunatic fringe has been reported for the corn snake (Gomez et al., 2008). The PSM of the snake appears to have undergone an axial extension, with expansion of the region rostral to the determination front.

This has resulted in a region rostral to the determination front with multiple lunatic fringe bands of expression, which are dynamic. However, lunatic fringe in the corn snake does not appear to be a cycling gene, i.e., a gene whose expression oscillates between the caudal-to-rostral PSM caudal to the determination front. For other segmentation clock genes such as hes6 or dll3, enhancers required for PSM expression have not yet been identified.

Oscillatory expression of the hes genes have been demonstrated to derive from auto-inhibitory negative feedback loops, with HES hetero- and homo-dimers binding to N- and E-box binding sites. The presence of these binding sites have also been identified in key enhancers required for downstream cycling genes, such as has been shown for lunatic fringe in the mouse (Cole et al., 2002; Morales et al., 2002). While the enhancers driving paraxial mesoderm specific expression have been identified for delta-like 1 (Beckers et al., 2000; White et al., 2005; White & Chapman, 2005), the elements required for cyclical expression have yet to be defined. The delta-like 3 orthologues, *deltaC* and *X-Delta-2,* display cycling expression in the zebrafish and *Xenopus*, respectively, but analysis for regulatory sites directing this oscillation have also not been reported. The amniote orthologues of *dll3* do not display cycling expression, and analysis of 10 kb 5' flanking regions have been inconclusive.

Insights from analysis of anole and alligator somitogenesis into the evolution of the segmentation clock. Since developmental and molecular tools are currently available for only a limited number of organisms, each class or infraclass has been typically represented by a single species at best. Based on our analysis in the anole and alligator, we can generate new hypotheses about the evolution of the segmentation clock. With the rapidly decreasing cost of next-generation sequencing and ability to test many additional vertebrates, these hypotheses can help us to select the most informative species for further analysis.

Predicted segmentation clock components of the ancestral vertebrate. Positional information along the rostral-caudal axis of the PSM was likely established by gradients of soluble ligands WNT3A and FGF8 with gene expression gradients of *tbx6* and *hes6* (Fig. 10). Expression of the hes6 orthologue was then later lost in both the mammalian and avian radiations. The determination front is regulated by mesp orthologues, interacting with cyclical patterns of gene expression driven by hes genes (hes1 or hes7 orthologues) and their downstream targets. The cycling gene regulating notch activation was the delta-like 3/deltaC/X-Delta-2 orthologue. While it is possible that all three genes–dll3, dll1, and lfng–displayed cycling expression in the PSM of an ancestral vertebrate, cycling expression would have to have been lost for *dll1* and *lfng* in both amphibian and teleost lineages.

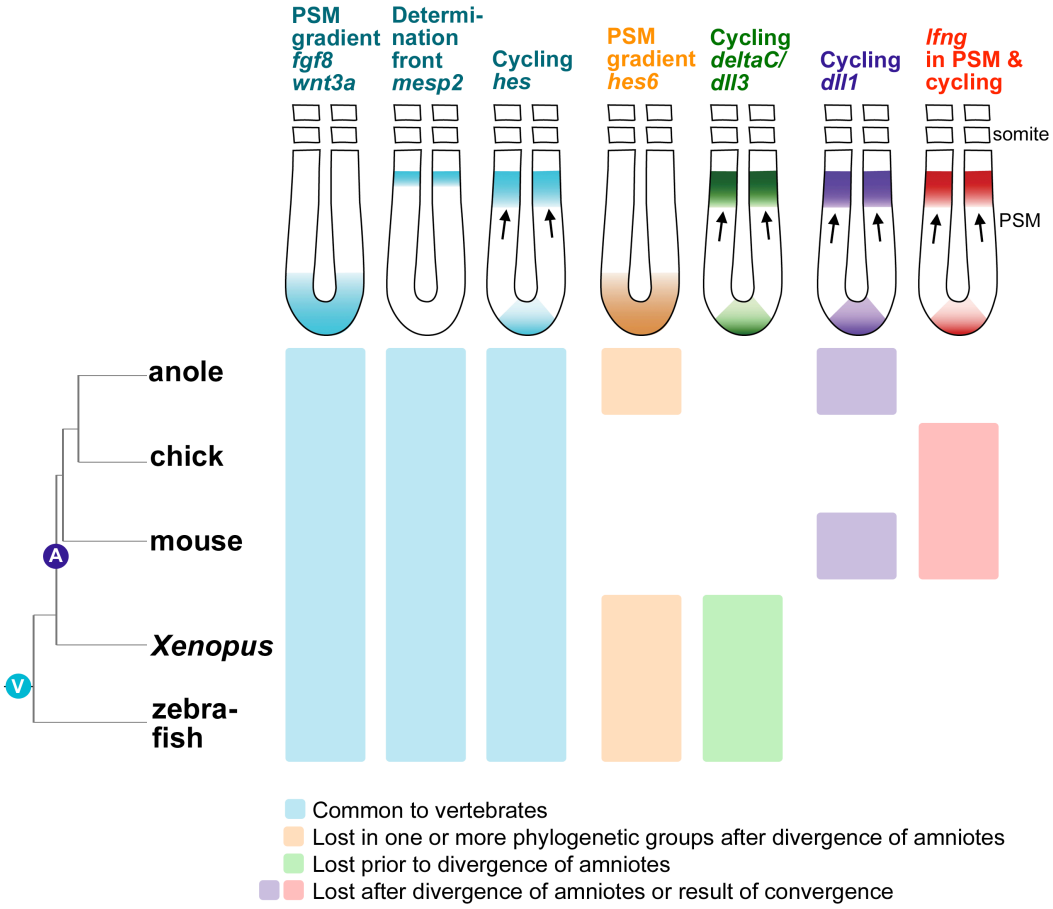

PSM gradient fgf8 wnt3a
Determi-nation front mesp2
Cycling hes
PSM gradient hes6
Cycling deltaC/ dll3
Cycling dll1
Ifng in PSM & cycling
somite
PSM
anole
chick
mouse
Xenopus
zebra-fish
A
V
Common to vertebrates
Lost in one or more phylogenetic groups after divergence of amniotes
Lost prior to divergence of amniotes
Lost after divergence of amniotes or result of convergence

Figure 10. Diagram comparing how the phylogenetic relationships of the zebrafish, frog *Xenopus laevis*, lizard *Anolis carolinensis*, chicken and mouse (Hedges & Kumar, 2009) with conserved and divergent features of the segmentation clock. Hypothetical vertebrate ancestor (light blue VA) and amniote ancestors (purple AA) are indicated. Shared among all these vertebrate developmental models is the cycling expression of hairy/enhancer of split (hes) genes, *mesp2* expression at the determination front for prepatterning of somite boundaries, and *fgf8* and *wnt3a* gradients (light blue blocks; Aulehla et al., 2003; Aulehla & Johnson, 1999; Dequéant & Pourquié, 2008; Gibb et al., 2010; Holley, 2007; Sparrow, 2009). Notch receptor activation in zebrafish and frog is regulated via cyclical expression of the ligand *deltaC/X-Delta-2* (orange blocks; (Holley, 2007; Jen et al., 1997; 1999). A gradient of *hes6* orthologue expression is observed in zebrafish, *Xenopus* and anole, but not in mouse and chicken (green blocks; Fior & Henrique, 2005; Holley, 2007; Koyano-Nakagawa, Kim, Anderson, & Kintner, 2000; Pissarra et al., 2000). A key oscillatory gene in the anole is the *dll1* ligand, which has also been reported to display dynamic expression in mouse PSM (purple blocks; (Krol et al., 2011; Maruhashi et al., 2005). Notch receptor activation in chicken and mouse is regulated by cyclical expression of the modulator *Lfng*, which is not expressed in the PSM of the anole or alligator (red blocks; Cole et al., 2002; McGrew et al., 1998; Morales et al., 2002).

Predicted segmentation clock components in the ancestral amniote. For cycling genes in amniotes, analysis of additional reptiles models will help to distinguish between two alternate models for both lunatic fringe and delta-like 1 (Fig. 10).

For *lfng*, the amniote ancestor could have displayed cycling expression, which was subsequently lost in the anole and alligator, or cycling expression of lfng in the PSM could have arisen independently in birds and mammals. For *dll1*, the amniote ancestor could have displayed cycling expression as seen in the anole, and mouse, with loss of cycling expression in Archosauria, or this cycling could have arisen independently in mammalian and squamate evolution.

In addition to the components detailed above, the ancestral amniote could have lost cyclical expression of *dll3* in the PSM. If the *dll3* orthologue was no longer an essential cycling activator of notch signaling, replaced by either or both lunatic fringe and delta-like 1 in the amniote ancestor, the loss or rapid divergence of the gene became possible. In mammals, the *Dll3* gene could have undergone rapid divergence in coding sequence (Fig. 7F and 8), changes in intracellular localization (G. Chapman et al., 2011; Geffers et al., 2007; Ladi et al., 2005), and functional shift from trans-activation to cis-inhibition of notch signaling. In birds, the gene may have been completely deleted.

Conclusion

Through this study we have identified several key features of the segmentation clock that are both divergent and convergent between vertebrates that would not have been possible without the investigation of non-avian reptiles such as the lizard or alligator. This study has also demonstrated that expression of genes in the segmentation clock, particularly cycling genes, is a sensitive read-out of the changes in regulatory networks during vertebrate evolution. These findings highlight the challenge of integrating evolutionary studies of developmental networks with sequence-based phylogenetic analysis (Fig. 10). Further analysis of chelonian and additional crocodilian reptiles (Mansfield & Abzhanov, 2010), urodele amphibians, paleognath birds, and monotreme mammals could yield further insights into the complex evolution of divergent and convergent processes in vertebrate development.

Chapter 4

DEEP SEQUENCING IDENTIFIES NOVEL MICRORNAS LINKED TO STEM CELL PROLIFERATION AND DIFFERENTIATION DURING TAIL REGENERATION IN THE LIZARD ANOLIS CAROLINENSIS

Abstract

Squamate lizards, such as the green anole, *Anolis carolinensis*, have evolved the ability to regenerate their tail after self-amputation, or autotomy. During this regeneration process lizards are capable of regrowing *de novo* muscle groups, hyaline cartilage, a spinal cord and skin. MicroRNAs (miRNAs) inhibit gene expression through either repressing translation of or signaling the degradation of their mRNA transcript targets and have been implicated to play key roles in stem cell proliferation and differentiation in regeneration. Through deep sequencing of small RNAs we investigated the role of miRNA regulation of the tail regeneration process in the *A. carolinensis* lizard. Through this work we identified 350 novel miRNAs in the lizard and 11 differentially expressed miRNAs between the growing tip of the regenerating tail and the more differentiated base. We also correlated miRNA expression with the expression their mRNA targets in the regenerating tail and identified several key biological pathways relating to the proliferation and differentiation of stem cells under the regulation of miRNAs during tail regeneration. As the first small RNA sequencing in the lizard, these results have aided in annotating miRNAs and suggest miRNAs play a critical role

during tail regeneration. Combined with the recent complete genomic sequencing of *Anolis carolinensis*, this work has helped establish the lizard model system for regeneration with the end goal of translating findings towards improving regeneration humans.

Background

Similar to many lizard species, *Anolis carolinensis* has the ability to regenerate its tail after autotomy, or self-amputation (Alibardi, 2010; Bellairs & Bryant, 1985; R. E. Fisher et al., 2012; Goss, 1992; Maginnis, 2006; Reichman, 1984; Ritzman et al., 2012; Vitt, 1981). The regenerated lizard tail contains a spinal cord with an ependymal cell core surrounded by a hyaline cartilage tube. Bundles of muscle fibers are found superficial to the cartilage tube and have attachments to the cartilage tube, as well as to other muscle fibers in the regenerating tail (R. E. Fisher et al., 2012). With the recent release of the complete genomic sequence for *A. carolinensis* (Alföldi et al., 2011), combined with a revised genome annotation based on 14 deep transcriptomes (Eckalbar et al., in review), it is now possible to study the molecular processes regulating tail regeneration at the genomic level.

The tail regeneration process in lizards can be divided into four general stages. The first stage, wound healing, lasts up to 10 days post-autotomy (dpa). The second stage is marked by the formation of a cone of mesenchymal and ependymal cells during 10-15 dpa. The third stage, tail

growth, begins at 15 dpa, were the regenerated muscle fibers, cartilage and ependymal cells are observed in the base, and a loose mesenchyme of less differentiated cells is observed in the tip (Hutchins, et al., in preparation). The final stage is maturation from approximately 60 dpa and later (R. E. Fisher et al., 2012; Ritzman et al., 2012). In other model organisms for regeneration, such as the axolotl limb regeneration, it has been proposed that linage restricted progenitor cells organize this regenerative process (Kragl et al., 2009; Tanaka and Reddien, 2011). Recently, efforts have been made to investigate gene regulatory networks involved in this regenerative process in the second and third stages of tail regeneration utilizing high-density next generation RNA sequencing (RNA-Seq; Hutchins et al., in preparation). While these studies have yielded critical insights into the genetic mechanisms controlling tail regeneration, RNA-seq of this kind primarily reflects the state transcriptional regulation and does not address methods of post-transcriptional regulation, such that by microRNAs (miRNAs).

MicroRNAs are short, approximately 22 base pair (bp), single stranded RNAs, which act to inhibit the translation of and/or signaling the degradation of their mRNA targets (Bartel, 2009; Sun et al., 2010). Targeting of miRNA inhibition is governed through imperfect complimentary base pairing in the 3’ UTR of mRNA transcripts, and it is estimated that each miRNA may target several hundred transcripts Using these mechanisms of inhibiting gene expression, miRNAs have been

shown to play critical roles in regulating a variety of biological processes, including the proliferation and differentiation of stem cells during development (Stefani & Slack, 2008). For example, the role of miRNA in regulating muscle development and repair has been extensively studied (A. H. Williams, Liu, van Rooij, & Olson, 2009). In particular, miR-1 has been found to play a critical role in inhibiting key genes regulating the differentiation of muscle from its progenitor stem cells, including *Hdac4* (J.-F. Chen et al., 2005), *Hand2* (Zhao, Samal, & Srivastava, 2005; Zhao & Srivastava, 2007), *Dll1* (Muth, Tsuchihashi, McManus, & Schwartz, 2007), *Mef2* (Simon, Madison, & Conery, 2008) and *Pax7* (J.-F. Chen et al., 2010). Additionally, miR-133a, 133b, and 206 have been shown to play key roles in regulating the differentiation of muscle tissues (C. Anderson, Catoe, & Werner, 2006; H. K. Kim, Lee, Sivaprasad, Malhotra, & Dutta, 2006; Koutsoulidou, Mastroyiannopoulos, Furling, Uney, & Phylactou, 2011). miR-133 has also been shown to regulate the tail regeneration process in zebrafish. Specifically, miR-133 is highly expressed in the injured tail, but strongly down regulated in the regenerating tail. It was also shown that addition of miR-133 inhibited the tail regeneration process in the zebrafish fin, while inhibiting the expression miR-133 through knockdown with an antisense morpholino improved tail regeneration (Yin et al., 2008). miRNAs have also been shown to play an important role in neurogenesis. Specifically, miR-184 has been shown to promote the proliferation of adult neural

stem/progenitor cells (C. Liu et al., 2010). To highlight the ability of miRNAs to regulate multiple biological processes, miR-133b has also been found to be involved in recovery from spinal cord injury through promoting axon growth (Yu et al., 2011), as well as regulating the balance between cell proliferation and apoptosis by inhibiting antiapoptotic genes (Patron et al., 2012).

Despite a growing understanding of the rolemiRNAs play in regulating a multitude of biological processes, few studies have attempted genome-wide miRNA profiling of stem cell proliferation and differentiation *in vivo*. To investigate the role of miRNAs in the tail regeneration process in the lizard we utilize small-RNA, next generation sequencing on the Illumina Genome Analyzer IIx platform. Specifically, we sequence miRNAs in the growing tip of the regenerating tail of the *Anolis carolinensis* lizard and the more differentiated base of the regenerating tail. To aid in annotation of novel miRNAs present in *A. carolinensis,* and for comparison with the regenerating tail, we also sequenced miRNAs in the adult brain and skeletal muscle. From this sequencing data we identifed differentially expressed miRNAs between the tip and base of the regenerating tail that may play important roles in regulating stem cell proliferation and differentiation during regeneration. Furthermore, in this study we predict the 3' UTR targets of lizard miRNAs and correlate the expression of these miRNA with the expression of their mRNA targets determined previously (Hutchins et al., in preparation). Through analyzing the mRNA targets of

miRNAs expressed in the tip and base of the regenerating tail we also identify specific biological processes that are likely under the regulation of miRNAs as the tail regenerates. This new information can now serve as the groundwork for understanding how methods of transcriptional and post-transcriptional regulation combine to coordinate mechanisms of stem cell proliferation and differentiation during the tail regeneration process in the lizard.

Materials and methods

Animal care, tissues collection and sample preparation for Illumina sequencing. Adult *A. carolinensis* lizards were purchased from Charles D. Sullivan, Inc. (Nashville, TN) or Marcus Cantos Reptiles (Fort Myers, FL). Lizards were housed in Percival incubators at 70% humidity and 14 hours at 28°C daylight and 10 hours at 19°C night. Lizards were fed crickets supplemented with Rep-cal calcium and Vitamin D every two days. Lizards were given water through daily misting of artificial plant surfaces. Care for all lizards were maintained in accordance to the Institutional Animal Care and Use Committee guidelines at Arizona State University. Autotomy was induced by firmly holding a point on the tail 5 cm from the base, while the lizard was otherwise allowed to move on a flat surface.

Regenerated tails were collected through inducing a subsequent autotomy at 25 dpa. A total of nine 25 dpa regenerating tails were cut into

3 sections. These sections were then split into pools of three tip and base sections. This led to three replicates of three pooled tails for each tip and base. Brain and muscle tissues were collected from lizards after euthanasia with sodium pentobarbital. Small RNAs were extracted following the miRVana kit protocol (Ambion). The 8 small RNAs were then barcoded for multiplexed sequencing on two Illumina GAIIx lanes, generating single end 40 base pair reads, through services provided by LC Sciences.

miRNA annotation. Raw sequencing reads from the resulting small RNA libraries were demultiplexed through services provided by LC Sciences. The adapters used for sequencing (TGGAATTCTCGGGTGCCAAGG) were then trimmed from the demultiplexed reads, while keeping only reads 18bp or greater, using the FASTX-Toolkit (available at http://hannonlab.cshl.edu/fastx_toolkit/). Also using the FASTX-Toolkit, the trimmed reads where quality filtered by removing all sequencing reads with less than 80% of the bases with at least an Illumina predicted quality score of 20, which is an estimated error rate of 1% per base.

The adapter trimmed and quality filtered reads for each for the samples were then mapped to the Anocarol2.0 repeat masked genome as available from Ensembl (Ensembl Build 67) using the miRDeep2 package (Friedlander et al., 2011). Specifically, the mapper.pl script was run with the following options: d, e, h, i, j, m. This generated a collapsed set of non-

redundant reads while retaining read counts along with the genomic location of the mapped reads.

To aid in annotation of miRNAs present in the tissues sequenced, we obtained predicted miRNAs from miRBase for the *A. carolinensis* lizard (Griffiths-Jones, 2006; Griffiths-Jones, Saini, & Van Dongen, 2008; S. G. Jones, 2004; Kozomara & Griffiths-Jones, 2011). Also, the miRBase miRNAs datasets for human, mouse, chicken, frog and zebrafish were obtained to aid in assigning orthology to novel miRNA in *A. carolinensis.* The miRDeep2 package was then used to annotate novel miRNA in *A. carolinensis*, as well as validate predicted miRNAs from miRBase. This was done by passing the read alignments the miRDeep2.pl wrapper, along with the predicted *A. carolinensis* miRNAs, as well as the miRNA sequences from human, mouse, chicken, frog and zebrafish. Novel miRNA genes predicted by miRDeep2 are assigned a score based on read support and secondary structures consistent with the biogenesis of miRNAs. Novel miRNAs predicted by miRDeep2 were then retained for further analysis if they had a miRDeep2 score of 5 or above, corresponding to an estimated false discovery rate of 6%.

Statistical analysis of miRNA expression. To determine miRNA expression levels the collapsed reads from the mapper.pl step were first aligned to the miRBase miRNAs and novel miRNAs predicted by miRDeep2 using the quantifier.pl script as part of the miRDeep2 package. This step produced a raw counts file that was then used as input into the

DESeq R/Bioconductor package for further statistical analysis (Anders & Huber, 2010). Differential expression tests in DESeq were conducted only for those miRNA genes with at least 10 reads of support in the regenerating tail. Differential expression tests were conducted with DESeq parameters of fitType="local" and sharingMode="fit-only".

miRNA target prediction. The mRNA targets of the known miRBase and novel miRNAs were predicted using RNAhybrid and miRanda against the 3' UTR sequences extracted from the ASU Acar v2.2.0 gene annotation (Betel, Wilson, Gabow, Marks, & Sander, 2007; Krüger & Rehmsmeier, 2006; Eckalbar et al., in review). The RNAhybrid prediction started by first calibrating the location and scale parameters of the extreme value distribution for each miRNA by using the RNAcalibrate tool against the 3' UTR sequences. This step was done to improve the p-value calculations for each target prediction for each specific miRNA. These calibrated parameters were then used as input for the d-option for the final RNAhybrid prediction step. Additionally, the minimum free energy parameter was set to -20 kcal/mol and p-value cut off to .01. The set of miRanda miRNA target predictions was generated by also setting the minimum free energy to -20 kcal/mol and requiring no mismatch in the seed region. Only overlapping miRNA target predictions from both RNAhybrid and miRanda were retained. Additionally, miRNA targets were filtered for transcripts, which were the target of two or more miRNAs.

Comparison of miRNA expression and mRNA target expression. Expression of miRNAs in the regenerating tail was then compared to the expression of their mRNA targets. As outlined above, DESeq was used to determine the expression levels of the known and novel miRNAs (Anders & Huber, 2010). Transcript expression levels in the tip and base of the tip (section 1) and base (section 5) regenerating tail were determined first through mapping RNA-Seq reads from a single regenerating tail (animal termed A81; described further in Hutchins, et al., in preparation) to the repeat masked Anocar2.0 genome and the ASU Acar v2.2.0 annotations (Eckalbar et al., in review) using Tophat version 2.0.4 (Langmead and Salzberg et al., 2012; Trapnell et al., 2009; Langmead et al., 2009). Cufflinks version 2.0.2 was then used to estimate transcript abundances based on the number of reads aligned to each transcript resulting in Fragments per Kilobase of exon per Million fragments mapped (Langmead et al., 2009; Trapnell et al., 2009). Differential expression of ASU Acar v2.2.0 transcripts between section 1 and section 5 was tested for using cuffdiff (Roberts et al., 2012; Roberts, Pimentel, Trapnell, & Pachter, 2011a; Trapnell et al., 2012).

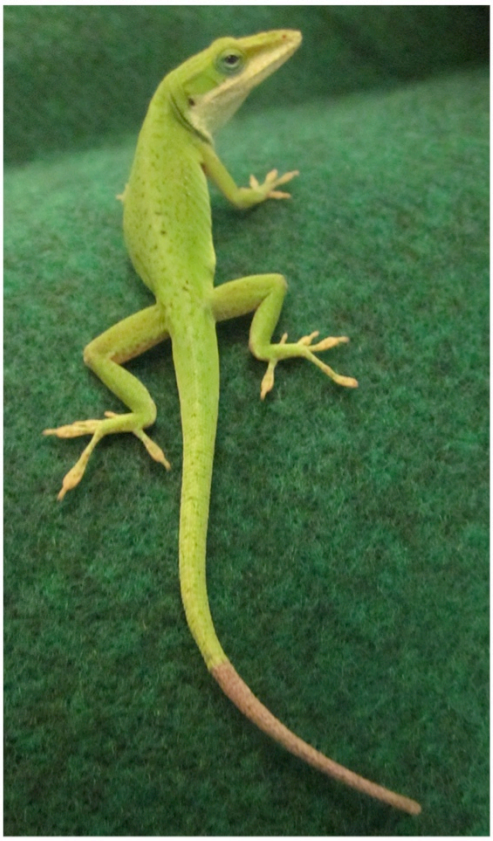
Transcript-miRNAs interactions were then filtered for co-expression of both the miRNA and mRNA in either the tip or base of the regenerating tail. All one or greater DESeq normalized values for expression of miRNAs were retained. Similarly, transcripts were required to have at least a

cufflinks estimated FPKM of 1 or greater in at least one section of the regenerating tail to be retained for further analysis.

Identification of biological processes under miRNA regulation. After filtering for co-expression of miRNAs and their transcript targets, biological processes under differential regulation by miRNAs in the tip or base of the regenerating were identified. This was done by further filtering the co-expressed miRNAs and targets by fold change from tip to base of the miRNAs. Specifially, all miRNA-target interactions that displayed either a two-fold change from tip to base or visa versa were retained. The targets differentially regulated by miRNAs were then subject to cluster based on Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway using DAVID (Aoki-Kinoshita & Kanehisa, 2007; Botstein et al., 2000; Dennis, Sherman, Hosack, & Yang, 2003; Gene Ontology Consortium, 2004; Kanehisa, Araki, Goto, & Hattori, 2008).

Results

Sequencing of lizard small RNAs. In order to investigate the role of miRNAs in regulating the tail regeneration process in the *A. carolinensis* lizard, we carried out small RNA sequencing with the Illumina GAIIx on the tip and base of a 25 days post-autotomy (dpa) regenerating tails. The 25 dpa tails were cut producing a 1mm section of the most distal tip and 1mm section of the most proximal base of the regenerating portion of the tail (Figure 11). A total of nine tip and base sections of the regenerating tails



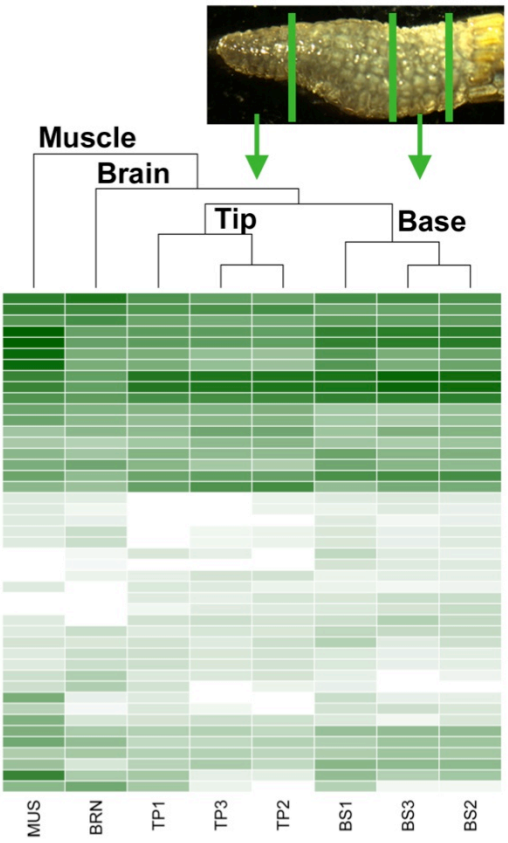

Muscle
Brain
Tip
Base
MUS
BRN
TP1
TP3
TP2
BS1
BS3
BS2

Figure 11. An adult Male *Anolis carolinensis* lizard with a regenerating tail and the experimental design for small RNA sequencing. Tissues extracted for small RNA sequences included the skeletal muscle, brain, as well as the tip and base of the regenerating tail (abbreviated MUS, BRN, TP and BS, respectively). For the regenerating tail samples, a total of nine tails were sectioned to obtain an approximate 1 mm section of the most distal tip (TP) and proximal base of the regenerating tail. In order to obtain sufficient quantity of small RNAs for sequencing, three samples of either tip or base were pooled together to create three sets of three pooled replicates for each the tip and base (TP1, TP2, TP3, BS1, BS2, BS3). Expression of novel and known miRNA was then determined through first annotation of miRNAs with miRDeep2, followed analysis with DESeq. The top 40 differentially expressed miRNAs between the tip and the base by adjusted p-values are represented as a heat map.

were collected. Due to the low yield of small RNAs from each tip and base section, three sections of each type were pooled together for small RNA extractions. This resulted in three sets of three pooled regenerating tail replicates of both tip and base for small RNA sequencing. To aid in annotating novel RNAs in the lizard, as well as for comparison with the regenerating tail, we sequenced small RNAs from the skeletal muscle and brain of an adult *A. carolinensis*. The brain and skeletal muscle were chosen based on the demonstrated diversity in transcript expression (Eckalbar et al., in review) and for comparison with miRNAs that may related to neural and muscle differentiation. In total, 58,931,365 raw sequencing reads from the small RNAs were obtained from the Illumina sequencing (Table 5). After adapter trimming and quality filtering for reads with at least 80% of bases with a quality score of 20 (corresponding to a 1% error rate) or better, a total of 48,210,322 sequencing reads remained (Table 5).

Annotation of novel miRNAs. In order to both confirm miRBase miRNA predictions and annotate novel miRNAs in *A. carolinensis*, we utilized the miRDeep2 pipeline, which was specifically designed to identify novel miRNAs from next generation sequencing data sets (Friedlander et al., 2011; Griffiths-Jones, 2006; Griffiths-Jones et al., 2008; S. G. Jones, 2004; Kozomara & Griffiths-Jones, 2011). Inside the miRDeep2 pipeline identical adapter trimmed, quality filtered sequencing reads were collapsed to a unique set of reads while retaining read counts. This unique

Table 5.

Overview of small RNA sequencing for annotation and quantification of miRNAs in the A. carolinensis in skeletal muscle, brain, as well as the tip and base of the regenerating tail.

Sample	Illumina Reads	Adapter Trimmed Reads	Quality Filtered Reads	Unique Reads	Mapped Unique Reads
Tip 1	6896312	4911787	4638573	267572	80047
Tip 2	8771826	7690607	7073991	213808	67955
Tip 3	8738345	8054177	7339012	205521	84089
Base 1	6905196	6084203	5763610	317605	134040
Base 2	9398842	8815680	8181644	245564	87691
Base 3	5898914	5514428	5107890	157094	62711
Muscle	3510208	2890930	2744587	124822	48387
Brain	8811722	7673990	7361015	172585	77664
Total	58931365	51635802	48210322	NA	NA

set of reads from each sample was then mapped to the AnoCar2.0 genome. After mapping reads to the genome, the miRDeep2 pipeline uses several lines of evidence to score the probability that a putative miRNA is a true positive: quantity of sequencing reads, stability of the pre-miRNA hairpin structure as predicted by RNAfold (Hofacker & Stadler, 2006; Hofacker et al., 1994), and homology to known miRNAs from related species. The miRDeep2 pipeline annotated a total of 817 novel miRNAs with a score of 0 or greater (corresponding to a 38% false discovery rate), and confirmed the presence of 228 of the 275 (83%) miRBase predicted miRNAs in the sequencing data (Supplemental Table 7). To limit the expansion of false discovery rates in further analyses, only novel miRNA predicted by miRDeep2 with a score of five or higher were retained. This corresponds to a predicted false positive rate of 6%). This resulted in a more conservative list of 350 novel miRDeep2 predicted miRNAs for subsequent analysis (Table 6, Supplemental Table 7).

As part of the miRDeep2 pipeline, novel miRNAs with seed regions identical to known miRNAs from closely related species are identified as possible orthologues. A total of 140 of the 350 novel miRDeep2 predicted miRNAs with a score of 5 or higher had 100% identity to at least one miRNA from miRNAs that are part of the miRBase databases for human, mouse, chicken, frog, or zebrafish. Additionally, potential orthologues to novel miRNAs were identified using reciprocal BLAST also with the miRBase databases from human, mouse, chicken, frog and zebrafish.

Table 6.

Summary of the top scoring novel miRNAs predicted by miRDeep2.

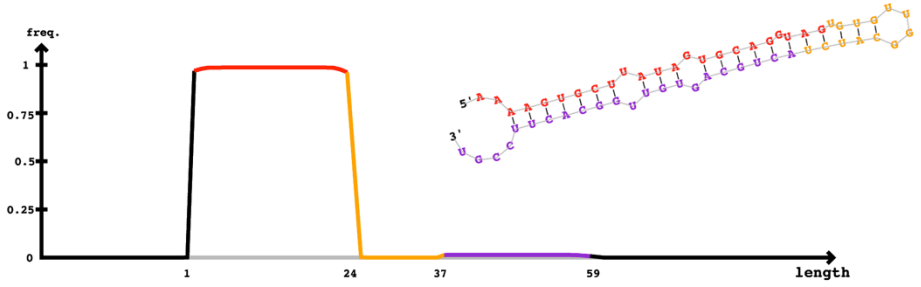
Novel miRNA id	miRDeep2 score	Total read count	Mature miRNA Sequence
GL344214.1_1034	8834.9	17320	aaaagugcuuauagugcagguag
2_14171	3341.1	6552	auaaccaauguauggacaugcacu
2_14152	1635.2	3205	ccggugggaucgagagcaggg
1_16347	1175.6	2297	uuuggucccccuuuaaccagcu
2_12428	1174.7	2302	ucagcucacaggauuccuggcc
GL343253.1_6342	1163.1	2281	cauggaacccaguuacagagcu
1_17442	1163.1	2281	cauggaacccaguuacagagcu
1_17444	1162.5	2280	cauggaacccaguuacagagcu
4_11170	1032.3	2022	cuggugggacgagagacagg
3_15258	773.5	1515	gugggacgagagacagggccu
2_13093	668.6	1312	ucugcuuugaacuggguuaucu
3_15047	663.6	1300	ucugcuuugaacuggguuaucu
GL343364.1_4363	406.1	802	ucucucacucgucccccaccaacu
1_17191	369.1	733	uauauuccaguucaaagcagaaa
GL343270.1_6155	365.4	717	uauauuccaguucaaagcagaaa
6_8745	364.5	706	cuggaggaccuagaaaaugccuagc
GL343471.1_3255	364	711	ugggaacacgagacagggccuucu
1_16432	361.3	718	uauauuccaguucaaagcagaaa
GL344014.1_1368	358.2	700	cugucucucgucccccaccagcc
2_13994	354.6	693	cugucucucgucccccaccagcc

Two of the more highly expressed miRNAs (GL344214.1_1034 and 1_16347) were matched with potential orthologues in other vertebrate species (Supplemental Table 8). Further analysis of using clustalW alignments of orthologues and paralogous miRNAs from other species showed that GL344214.1_1034 is likely the *A. carolinensis* orthologue of miR-106, and 1_16347 is likely the orthologue of miR-133b (Figure 12).

Expression analysis of miRNAs in the regenerating tail. The expression of miRNAs in the skeletal muscle, brain, as well as the tip and base of the regenerating tail were determined by mapping the set of unique reads to the miRBase predicted miRNAs and novel miRNAs using the miRDeep2 quantifier script (Supplemental Table 9). Differential expression of miRNAs in the tip and base of the regenerating tail was determined by using the read counts mapped to each miRNA as input into DESeq. DESeq is an R/Bioconductor package, which utilizes a negative binomial distribution model to test for differential expression in next generation sequencing datasets. miRNAs with less than a total of 10 reads summed across all the regenerating tail replicates were excluded from differential testing in DESeq (Supplemental Table 10). Of the 632 novel and miRBase predicted miRNAs, a total of 453 miRNA were tested for differential expression. Overall, the miRNA expression in the tip and base of regenerating tail is highly correlated with a Spearman's rank correlation coefficient of greater than .968 (Figure 13). The linear regression of miRNA expression in the tip versus the base of the

A.

GL344214.1_1034



Mature **Star**

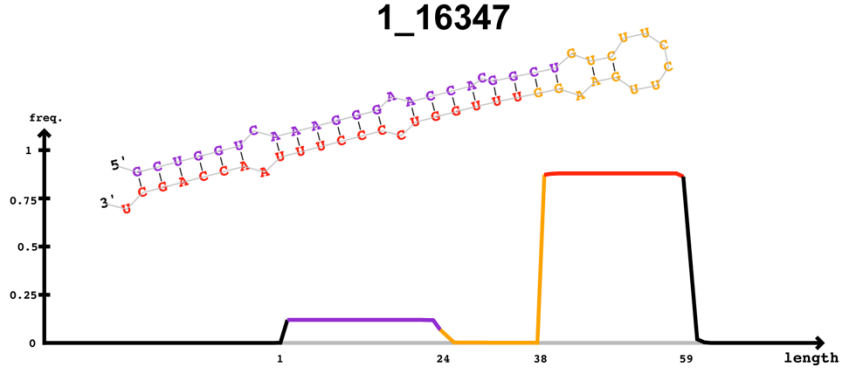
5'- ggggaaagccaucggggugcaaaagugcuuauagugcagguaguguguuggcaucuacugcaguguuggcacuuccgugccacgauggcgaugcgcgucaggugcaacugug -3'

Score Total	: 8834.9
Score from Reads	: 8824.2
Score from MFE	: 2.2
Score from randfold	: 1.6
Score from Cons. Seed	: 3
Total Read Count	: 17320

miRNA	Mature sequence	
hsa-miR-106b-5p	TAAAGTGCTGACAGTGCAGAT---	21
mmu-miR-106b-5p	TAAAGTGCTGACAGTGCAGAT---	21
hsa-miR-106a-5p	AAAAGTGCTTACAGTGCAGGTAG-	23
gga-miR-106	AAAAGTGCTTACAGTGCAGGTA--	22
xtr-miR-106	AAAAGTGCTTATAGTGCAGGTAGA	24
GL344214.1_1034	AAAAGTGCTTATAGTGCAGGTAG-	23
	******** * ******* *	

B.

1_16347



Star **Mature**

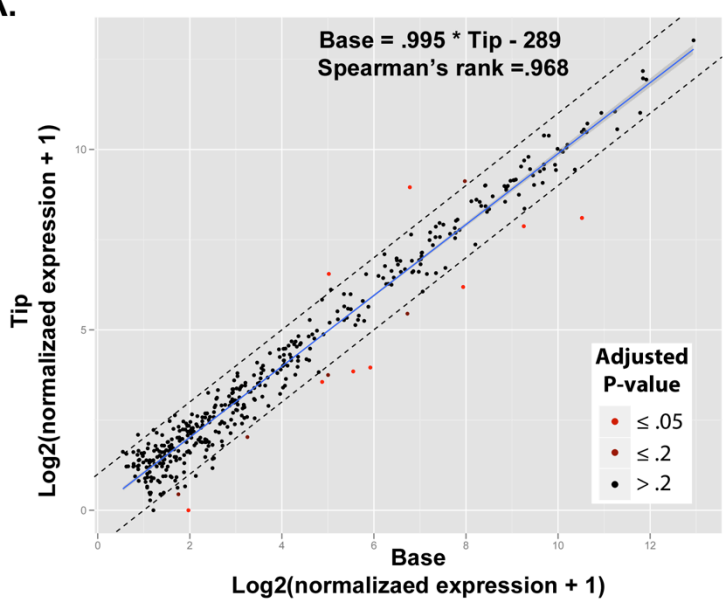
5'- gccacuugcagguagaacuugaguccugcucuggcuggucaaagggaaccacggcugucuuccuugaagguuuggucccccuuuaaccagcuacagcaguccugauaucag -3'

Score Total	: 1175.6
Score from Reads	: 1165.1
Score from MFE	: 2
Score from randfold	: 1.6
Score from Cons. Seed	: 3
Total Read Count	: 2297

miRNA	Mature sequence	
gga-miR-133b	-TTGGTCCCCTTCAACCAGCTA-	21
xtr-miR-133b	-TTGGTCCCCTTCAACCAGCTA-	21
hsa-miR-133b	TTTGGTCCCCTTCAACCAGCTA-	22
dre-miR-133b	TTTGGTCCCCTTCAACCAGCTA-	22
mmu-miR-133b-3p	TTTGGTCCCCTTCAACCAGCTA-	22
1_16347	TTTGGTCCCCTTTAACCAGCT--	21
	********** ********	

Figure 12. The annotation and orthology assignment of two novel miRNAs by miRDeep2: GL344214.1_1034 and 1_16347. **A,** GL344214.1_1034 was identified to be the *A. carolinensis* orthologue of miR-106 by reciprocal blast and clustalW analysis and had a high miRDeep2 score of 8834.9, which was largely attributed to the high number (17,320) of reads supporting it. **B,** 1_16347 was identified to be the *A. carolinensis* orthologue of miR-133b by reciprocal blast and clustalW analysis and had a high miRDeep2 score of 1175.6, which was largely attributed to the high number (2,297) of reads supporting it.

A.


Base = .995 * Tip - 289
Spearman's rank =.968
Tip
Log2(normalizaed expression + 1)
Base
Log2(normalizaed expression + 1)
Adjusted P-value
≤ .05
≤ .2
> .2

B.

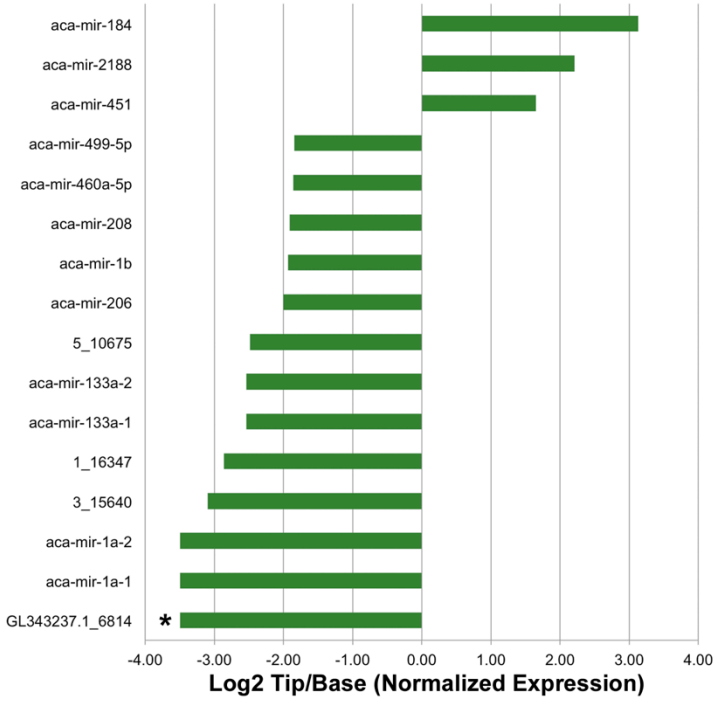

aca-mir-184
aca-mir-2188
aca-mir-451
aca-mir-499-5p
aca-mir-460a-5p
aca-mir-208
aca-mir-1b
aca-mir-206
5_10675
aca-mir-133a-2
aca-mir-133a-1
1_16347
3_15640
aca-mir-1a-2
aca-mir-1a-1
GL343237.1_6814
*
Log2 Tip/Base (Normalized Expression)

Figure 13. MicroRNA expression in the tip and base of the regenerating tail at 25 dpa. **A,** Expression of miRNA in the tip versus the base of the regenerating tail are highly correlated, with a beta-value (slope) of .995, and Spearman's rank equal to .968. B, Differentially expressed miRNAs in the tip and base with adjusted p-values < .2. Asterisk indicates a division by zero.

regenerating tail also shows that expression is nearly a 1-to-1 relationship on average with a beta-value (slope) of .995 (Figure 13a). Despite this high degree of correlation, however, 11 miRNAs were significantly differentially expressed (adjusted p-value < .05) between the tip and base of the regenerating tail, and an additional five miRNAs were nearly significantly differentially expressed (adjusted p-value between .05 and .2; Figure 13b). Additionally, 59 miRNA display at least a 2-fold increase in expression in the base and 36 miRNA display at least a 2-fold increase in expression in the tip of the regenerating tail (Table 7; Supplemental Table 10).

Prediction of miRNA targets in the lizard. Each miRNA may regulate several hundred genes through either inhibition of translation or causing mRNA degradation. This gives miRNAs the ability to substantially influence a broad range of molecular processes. To determine which biological processes were under the regulation of miRNAs, we predicted the 3' UTR targets of both known and novel miRNAs, and associated miRNA expression with the expression of their predicted targets in the regenerating tail. Targets of novel and known miRNAs were computationally predicted using RNAhybrid and miRanda against the 3' UTRs for each transcript in the ASU Acar v2.2.0 gene annotation (Betel et al., 2007; Krüger & Rehmsmeier, 2006; Eckalbar et al., in review; Hutchins et al., in preparation), and only miRNA-mRNA transcription interaction predicted by both were retained. RNAhybrid predicted a total of 1,570,395

Table 7.

Summary of the top differentially expressed miRNAs between the tip and base of the regenerating tail.

Top miRNAs down regulated in the tip						
miRNA id	**Base**	**Tip**	**Fold Change**	**Log2 Fold Change**	**P-value**	**Adjusted P-value**
aca-mir-1a-1	37218.2	3301.3	0.09	-3.49	7.45E-09	1.69E-06
aca-mir-1a-2	37155.4	3296.4	0.09	-3.49	7.44E-09	1.69E-06
3_15640	4.8	0.6	0.12	-3.10	5.21E-03	0.16
1_16347	373.7	51.3	0.14	-2.87	1.03E-05	8.16E-04
GL343703.1_2182	2.1	0.3	0.15	-2.75	0.11	0.63
aca-mir-133a-1	2815.8	485.7	0.17	-2.54	1.08E-05	8.16E-04
aca-mir-133a-2	2815.8	485.7	0.17	-2.54	1.08E-05	8.16E-04
5_10675	257.6	45.9	0.18	-2.49	1.23E-04	6.98E-03
1_17246	4.7	0.9	0.19	-2.43	0.02	0.36
GL343312.1_5257	4.7	0.9	0.19	-2.43	0.02	0.36
5_9840	2.9	0.6	0.19	-2.38	0.05	0.53
aca-mir-5458	3.2	0.6	0.20	-2.36	0.11	0.63
aca-mir-5412-5p	11.2	2.2	0.20	-2.33	0.01	0.26
aca-mir-499-3p	7.5	1.6	0.21	-2.23	0.03	0.46
aca-mir-206	10511.2	2625.9	0.25	-2.00	2.09E-04	0.01
aca-mir-1b	130.6	34.1	0.26	-1.94	1.12E-03	0.05
aca-mir-208	25.0	6.6	0.27	-1.91	4.67E-03	0.15
aca-mir-460a-5p	839.0	231.3	0.28	-1.86	5.89E-03	0.17
aca-mir-499-5p	148.7	41.4	0.28	-1.84	4.40E-03	0.15
6_8731	11.5	3.2	0.28	-1.83	9.80E-03	0.25
GL343237.1_6814	6.2	0.0	0.00	IND	1.14E-04	6.98E-03
6_8590	2.4	0.0	0.00	IND	0.02	0.36
Top miRNAs up regulated in the tip						
miRNA id	**Base**	**Tip**	**Fold change**	**Log2 fold change**	**P-value**	**Adjusted P-value**
aca-mir-451	2918.1	9156.1	3.14	1.65	2.91E-03	0.11
aca-mir-5425	1.2	4.0	3.42	1.77	0.05	0.53
aca-mir-5446	2.4	8.3	3.49	1.80	0.03	0.48
2_13248	0.9	3.8	4.38	2.13	0.09	0.63
aca-mir-2188	151.1	699.5	4.63	2.21	5.87E-04	0.03
aca-mir-449c	1.4	7.1	5.07	2.34	0.01	0.34
aca-mir-217	0.7	4.1	5.56	2.47	0.02	0.36
aca-mir-184	883.3	7743.9	8.77	3.13	3.03E-07	4.57E-05

miRNA-mRNA transcript interactions and miRanda predicted a total of 624,994 miRNA-mRNA interaction. This resulted in an overlap of 79,642 interactions (Figure 14). Because gene transcripts are typically regulated by multiple miRNAs, gene transcripts were only considered if the mRNA was the target of two or more miRNAs. This left a total of 66,282 miRNA-mRNA transcript interactions, stemming from 7,206 genes and 17,539 transcript isoforms (Figure 14; Supplemental Table 11).

To gain a further understanding of which biological processes were being regulated by miRNAs in the regenerating tail, predicted target mRNA transcripts were also filtered for coexpression in the regenerating tail with their respective miRNAs. Expressional values for gene transcripts in the tip and base of a 25 dpa regenerating tail where obtained from RNA-sequencing data as described previously (Hutchins et al., in preparation). RNA-sequencing reads from the regenerating tail sections 1-5 (tip to base) were first mapped to the AnoCarol2.0 genome, and expression values for each gene and differentially expressed genes were obtained using Cufflinks (Langmead et al., 2009; Roberts, Pimentel, Trapnell, & Pachter, 2011a; Roberts, Trapnell, Donaghey, Rinn, & Pachter, 2011b; Trapnell et al., 2009; 2010). To consider a transcript to be expressed in the regenerating tail we required a cuffdiff derived FPKM (Fragments per Kilobase exon model per Million reads) of 1 or greater in either the tip or base of the regenerating tail. These coexpression criteria reduced the number of total miRNA-mRNA transcript interactions to

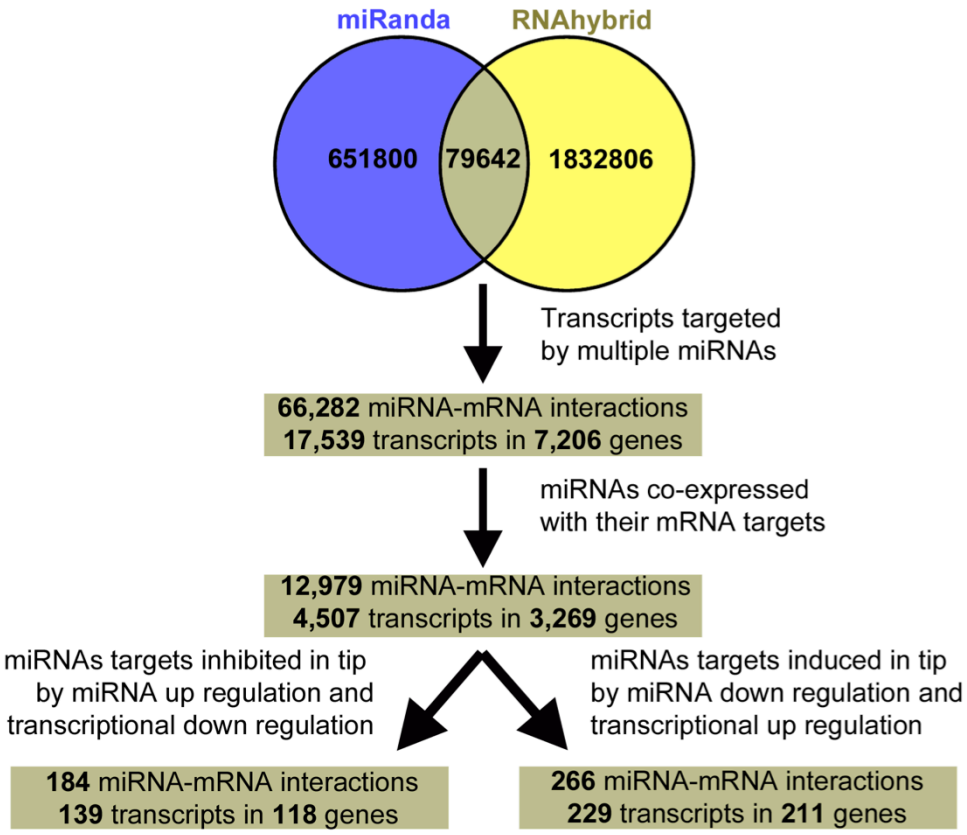

miRanda
RNAhybrid
651800
79642
1832806
Transcripts targeted by multiple miRNAs
66,282 miRNA-mRNA interactions
17,539 transcripts in 7,206 genes
miRNAs co-expressed with their mRNA targets
12,979 miRNA-mRNA interactions
4,507 transcripts in 3,269 genes
miRNAs targets inhibited in tip by miRNA up regulation and transcriptional down regulation
miRNAs targets induced in tip by miRNA down regulation and transcriptional up regulation
184 miRNA-mRNA interactions
139 transcripts in 118 genes
266 miRNA-mRNA interactions
229 transcripts in 211 genes

Figure 14. Overview of the analysis of the predicted mRNA targets of A. carolinensis miRNAs. Prediction of known and novel miRNA targets by miRanda and RNAhybrid identified a total of 731442 and 1912448 potential miRNA-mRNA interactions, respectively. Of these interactions, 79,642 predictions were shared between miRanda and RNAhybrid, and 66,282 miRNA-mRNA interactions in 17,539 transcripts of 7,206 genes were the target of multiple miRNAs. A total of 4,507 transcripts of 3,269 genes were coexpressed in the regenerating tail with their respective miRNA. Additionally, miRNA-mRNA interactions were filtered for coordinated inhibition or induction of gene expression in the tip of the regenerating tail. A total of 139 and 229 transcripts were identified as induced or inhibited by coordinated transcriptional and miRNA regulation, respectively.

12,979, stemming from 4,507 transcript isoforms in 3,269 genes (Supplemental Table 12). Additionally, targets of miRNAs in the regenerating tail were filtered into up and down-regulated transcripts based on two-fold expressional changes. This resulted in 782 genes with transcripts down-regulated in tip of the regenerating tail, and 981 genes with transcript expression up-regulated in the tip. The predicted targets of miRNAs with two-fold expression changes were also separated into up and down-regulated genes in the tip of the regenerating tail. This resulted in 531 genes targeted by miRNA up-regulated in the tip of the regenerating tail, and 757 genes targeted by miRNAs down-regulated in the tip of the regenerating tail. We next correlated both transcriptional and the post-transcription regulation by miRNAs of gene expression. Genes with transcripts with both down-regulated expression and targets of up-regulated miRNAs in the tip of the regenerating tail totaled 268. Similarly, genes with transcripts with both up-regulated expression and targets of down-regulated miRNAs in the tip of regenerating tail totaled 177 (Supplemental Table 12).

Pathways regulated by miRNAs in the regenerating tail. To determine biological processes potentially regulated by miRNAs in the regenerating tail we functionally annotated genes predicted to be targets of miRNAs. Genes targets coexpressed with the miRNAs in the regenerating tail were subject to Gene Ontology (GO) and KEGG pathway clustering analysis using DAVID (Supplemental Table 13; Aoki-Kinoshita &

Kanehisa, 2007; Botstein et al., 2000; Dennis et al., 2003; Gene Ontology Consortium, 2004; Kanehisa et al., 2008). Additionally, this functional annotation process was used to identify biological processes specifically regulated by miRNAs either up or down-regulated in the tip of regenerating tail. A total of 42 clusters of GO terms were related to genes inhibited by miRNAs up-regulated in the tip of the regenerating tail, while 40 clusters were identified as related to genes down-regulated in the tip of the regenerating tail (Table 8 and 9; Supplemental Table 14). KEGG pathway analysis revealed six pathways with predicted targets of miRNAs down-regulated in the tip of the regenerating tail. Only one of these, the insulin signaling pathway, was statistically significant when adjusted multiple testing (P-value < .05). KEGG pathway analysis also revealed 19 pathways with predicted targets of miRNAs up-regulated in the tip of regenerating tail. This included two statistically significant pathways, again the insulin signaling pathway and the ErbB pathway (Supplemental Table 15). To further elucidate the biological processes in regenerating tail we functionally annotated genes with transcripts with both down-regulated expression and targets of up-regulated miRNAs in the tip of the regenerating tail. This analysis identified 12 clusters of GO terms (Table 10). Notably, the most over represented terms relate to nerve and muscle cell function, as well as bone development. Similarly, we functionally annotated genes with transcripts stimulated by both up-regulated transcription and down-regulated miRNAs in the tip (Table 11). KEGG

Table 8.

Functional annotation clusters of GO biological processes predicted to be the targets of miRNAs up regulated in the tip of the regenerating tail.

Cluster	GO terms (top 4 representatives per cluster)	Count	Fold Enrichment	P-value
Cluster 1 ES=3.94	GO:0051603~proteolysis involved in cellular protein catabolic process	38	2.158	1.53E-05
	GO:0044257~cellular protein catabolic process	38	2.147	1.70E-05
	GO:0043632~modification-dependent macromolecule catabolic process	36	2.137	3.30E-05
Cluster 2 ES=2.52	GO:0044449~contractile fiber part	12	3.901	2.47E-04
	GO:0030017~sarcomere	11	4.123	3.21E-04
	GO:0006941~striated muscle contraction	8	5.926	3.49E-04
Cluster 3 ES=2.50	GO:0043233~organelle lumen	72	1.453	8.41E-04
	GO:0070013~intracellular organelle lumen	70	1.445	0.001
	GO:0031974~membrane-enclosed lumen	72	1.425	0.001
Cluster 4 ES=2.36	GO:0000267~cell fraction	46	1.560	0.003
	GO:0005626~insoluble fraction	37	1.620	0.004
	GO:0005624~membrane fraction	35	1.589	0.007
Cluster 5 ES=2.29	GO:0044437~vacuolar part	8	4.591	0.002
	GO:0005774~vacuolar membrane	7	4.591	0.004
	GO:0005773~vacuole	14	2.041	0.020
Cluster 6 ES=2.28	GO:0044429~mitochondrial part	30	1.852	0.002
	GO:0031980~mitochondrial lumen	14	2.265	0.009
	GO:0005759~mitochondrial matrix	14	2.265	0.009
Cluster 7 ES=2.25	GO:0043232~intracellular non-membrane-bounded organelle	95	1.344	0.001
	GO:0043228~non-membrane-bounded organ.	95	1.344	0.001
	GO:0005856~cytoskeleton	46	1.223	0.112
Cluster 8 ES=2.24	GO:0030198~extracellular matrix organization	11	3.604	9.30E-04
	GO:0030199~collagen fibril organization	5	5.875	0.010
	GO:0043062~extracell. structure organization	11	2.300	0.022
Cluster 9 ES=2.02	GO:0005201~extracellular matrix structural	9	3.431	0.005
	GO:0031012~extracellular matrix	18	1.916	0.013
	GO:0005578~proteinaceous extracell. matrix	17	1.951	0.014
Cluster 10 ES=2.02	GO:0060048~cardiac muscle contraction	6	10.761	1.74E-04
	GO:0006941~striated muscle contraction	8	5.926	3.49E-04
	GO:0003015~heart process	6	8.889	4.56E-04
Cluster 11 ES=1.86	GO:0005201~extracellular matrix structural	9	3.431	0.005
	GO:0005581~collagen	5	5.247	0.014
	GO:0005583~fibrillar collagen	3	9.182	0.041

Table 9.

Functional annotation clusters of GO biological processes predicted to be the targets of miRNAs down regulated in the tip of the regenerating tail.

Cluster Number	GO terms (top 5 representatives per cluster)	Count	Fold Enrichment	P-value
Cluster 1 ES=2.82	GO:0031974~membrane-enclosed lumen	62	1.636	7.21E-05
	GO:0070013~intracellular organelle lumen	59	1.624	1.43E-04
	GO:0043233~organelle lumen	59	1.588	2.63E-04
	GO:0031981~nuclear lumen	45	1.520	0.004
	GO:0005730~nucleolus	23	1.614	0.027
Cluster 2 ES=2.62	GO:0016192~vesicle-mediated transport	25	2.032	0.001
	GO:0010324~membrane invagination	13	2.766	0.003
	GO:0006897~endocytosis	13	2.766	0.003
	GO:0016044~membrane organization	18	2.211	0.003
Cluster 3 ES=1.96	GO:0045184~establishment of protein localization	30	1.826	0.002
	GO:0015031~protein transport	29	1.781	0.003
	GO:0046907~intracellular transport	25	1.781	0.007
	GO:0008104~protein localization	31	1.645	0.007
	GO:0006886~intracellular protein transport	16	2.003	0.014
Cluster 4 ES=1.90	GO:0000267~cell fraction	37	1.673	0.002
	GO:0005626~insoluble fraction	27	1.576	0.021
	GO:0005624~membrane fraction	25	1.513	0.041

Table 10.

Functional annotation clusters of GO biological processes predicted to be the targets of miRNAs up regulated and transcriptionally down regulated in the tip of the regenerating tail.

Cluster Number	GO terms (top 4 representatives per cluster)	Count	Fold Enrichment	P-value
Cluster 1 ES=1.78	GO:0042391~regulation of membrane potential	5	8.01	3.29E-03
	GO:0006873~cellular ion homeostasis	7	4.02	0.01
	GO:0055082~cellular chemical homeostasis	7	3.96	0.01
	GO:0050801~ion homeostasis	7	3.68	0.01
Cluster ES=1.50	GO:0006942~regulation of striated muscle contraction	3	29.28	4.50E-03
	GO:0006937~regulation of muscle contraction	3	8.95	0.04
	GO:0044057~regulation of system process	4	2.78	0.17
Cluster 3 ES=1.48	GO:0001501~skeletal system development	7	4.71	3.33E-03
	GO:0001503~ossification	3	5.60	0.10
	GO:0060348~bone development	3	5.24	0.11
Cluster 4 ES=1.44	GO:0004857~enzyme inhibitor activity	6	4.24	0.01
	GO:0004866~endopeptidase inhibitor activity	4	5.27	0.04
	GO:0030414~peptidase inhibitor activity	4	4.99	0.04
Cluster 5 ES=1.39	GO:0030036~actin cytoskeleton organization	5	4.75	0.02
	GO:0030029~actin filament-based process	5	4.45	0.02
	GO:0007010~cytoskeleton organization	5	2.46	0.14
Cluster 6 ES=1.30	GO:0010033~response to organic substance	10	2.98	0.01
	GO:0001541~ovarian follicle development	3	15.71	0.02
	GO:0048511~rhythmic process	4	6.71	0.02
	GO:0009725~response to hormone stimulus	6	3.51	0.03
Cluster 7 ES=1.23	GO:0000267~cell fraction	11	2.03	0.04
	GO:0005626~insoluble fraction	9	2.14	0.05
	GO:0005624~membrane fraction	8	1.97	0.10
Cluster 8 ES=1.15	GO:0005578~proteinaceous extracellular matrix	6	3.74	0.02
	GO:0031012~extracellular matrix	6	3.47	0.03
	GO:0005201~extracellular matrix structural constituent	3	6.66	0.07
	GO:0007155~cell adhesion	7	2.15	0.10
Cluster 9 ES=1.12	GO:0043292~contractile fiber	4	6.60	0.02
	GO:0030017~sarcomere	3	6.11	0.08
	GO:0030016~myofibril	3	5.40	0.10
	GO:0044449~contractile fiber part	3	5.30	0.11

Table 11.

Functional annotation clusters of GO biological processes predicted to be the targets of miRNAs down regulated and transcriptionally up regulated in the tip.

Cluster Number	GO terms top 5 representatives per cluster	Count	Fold Enrichment	P-value
Cluster 1 ES=1.75	GO:0000267~cell fraction	18	2.31	1.49E-03
	GO:0005624~membrane fraction	15	2.58	1.66E-03
Cluster 2 ES=1.65	GO:0000166~nucleotide binding	31	1.60	0.01
	GO:0030554~adenyl nucleotide binding	24	1.76	0.01
	GO:0001883~purine nucleoside binding	24	1.74	0.01
Cluster 3 ES=1.64	GO:0031090~organelle membrane	18	2.28	0.00
	GO:0044429~mitochondrial part	11	2.57	0.01
	GO:0005739~mitochondrion	15	1.92	0.02
Cluster 4 ES=1.64	GO:0043632~modification-dependent macromolecule catabolic process	12	2.42	0.01
	GO:0019941~modification-dependent protein catabolic process	12	2.42	0.01
	GO:0043161~proteasomal ubiquitin-dependent protein catabolic process	5	5.67	0.01
Cluster 5 ES=1.33	GO:0016887~ATPase activity	9	3.12	0.01
	GO:0042626~ATPase activity, coupled to transmembrane movement of substances	5	5.32	0.01
	GO:0043492~ATPase activity, coupled to movement of substances	5	5.27	0.01
Cluster 6 ES=1.23	GO:0016192~vesicle-mediated transport	12	2.41	0.01
	GO:0006887~exocytosis	4	4.02	0.08
	GO:0015031~protein transport	11	1.67	0.12
Cluster 7 ES=1.16	GO:0012501~programmed cell death	11	2.08	0.04
	GO:0006915~apoptosis	10	1.92	0.07
	GO:0008219~cell death	11	1.77	0.09
Cluster 8 ES=1.13	GO:0042981~regulation of apoptosis	13	1.87	0.04
	GO:0043067~regulation of programmed cell death	13	1.85	0.05
	GO:0010941~regulation of cell death	13	1.84	0.05

pathway analysis for genes transcripts with expression up-regulated in the tip and the target of miRNAs down-regulated in the tip found only the insulin signaling pathway to modulated. For genes with transcripts down-regulated in the tip and the target of miRNAs only two KEGG pathways were identified, ECM-receptor interaction and focal adhesion, with adjusted p-values of .25 and .24, respectively.

Discussion

To further elucidate the role of miRNAs in regulating the proliferation and differentiation of progenitor cells *in vivo*, we utilized small-RNA deep sequencing to profile miRNA expression in the growing tip and the more differentiated base of the regenerating tail in the *A. carolinensis* lizard. In addition to this, we sequenced miRNAs from the lizard brain and skeletal muscle (Figure 11). Subsequent analysis of the sequencing reads predicted 350 novel miRNAs with a false discovery rate of 6% or less, in the lizard (Table 5; Supplemental Table 7). Furthermore, we detected 11 significantly differentially expressed miRNAs between the tip and base of the regenerating tail (Figure 13; Table 7). In order to gain a further understanding of the role of miRNAs in regulating tail regeneration, we also predicted the targets of both novel and known miRNA in the lizard and identified several pathways specifically regulated by miRNAs.

As the first study to conduct deep sequencing of small RNAs from the *A. carolinensis* lizard, a high number of novel miRNAs were identified

as part of this project (Table 6, Supplemental Table 7). MiRBase had previously predicted 275 conserved miRNAs orthologues in the lizard, and our sequencing efforts in combination with mirDeep2 analysis confirmed 228 of those miRNAs. Of the remaining 47 miRBase predicted miRNAs, 33 of them had at least 1 sequencing read of support, and several miRNAs had read support into the thousands (Supplemental Table 7). It is likely that this discrepancy is due to the differing miRNA folding prediction algorithms between miRBase and the miRDeep2 pipeline, which utilizes RNAfold (Bonnet, Wuyts, Rouze, & Van de Peer, 2004). It is also possible that sequencing of more tissue types in the lizard would increase the read support of several of the known miRNAs and allow for their detection by miRDeep2. Finally, it is possible that several of the miRBase predicted miRNAs are incorrectly annotated, and that the correct orthologue assignment is among the novel miRNAs identified by miRDeep2.

Through the miRDeep2 pipeline, potential orthologues of the novel miRNAs were identified based on an exact nucleotide match in the seed region of either human, mouse, chicken, frog or zebrafish miRBase miRNAs. Of the 350 novel miRNAs with a miRDeep2 score of 5 or higher, 40% (140) were assigned potential orthologues using this method (Supplemental Table 7). To further improve orthology assignments, both lizard novel and known miRNAs were analyzed using reciprocal blast against known miRNAs in the reference species of human, mouse, chicken, frog and zebrafish. A total of 13 novel miRNAs found a one-to-

one match in at least one of the five reference species. Specifically, this improved the orthology assignment for one of the most highly expressed novel miRNAs, GL344214.1_1034. GL344214.1_1034 was previously predicted to be the orthologue of zebrafish miR-17a, however miR-17a was already annotated by miRBase and confirmed by our miRDeep2 pipeline. GL344214.1_1034 was instead found to be the orthologue of miR-106 through reciprocal blast and was confirmed with clustalW (Figure 12a, Supplemental Table 8). Additionally, another of the most highly expressed novel miRNAs, 1_16347 was predicted to be the orthologue of zebrafish miR-133a. However, miR-133a was also among the confirmed miRBase predictions, but additional analysis utilizing clustalW revealed 1_16347 to instead be the orthologue of miR-133b (Figure 12b). This analysis highlights the difficulty of assigning orthology to miRNAs, which do not benefit from conserved amino acid sequences, as with mRNAs, nor long nucleotide sequence length for comparison with related species, as with ncRNAs.

To investigate the potential role of miRNAs in regulating the tail regeneration process in lizards, novel and known miRNA were examined for differential expression using DESeq. Of the significantly differentially expressed miRNAs between the growing tip and more differentiated base of the regenerating tail, eight were up-regulated in the base, and two were up-regulated in the tip (Figure 13; Table 7). Several of the most highly up-regulated miRNAs in the base of the regenerating tail have previously

been shown to be key regulators of the differentiation of muscle stem cells, such as the adult muscle stem cell, satellite cells, in other model organisms (Chen et al., 2006; Zhao et al., 2005; Zhao et al., 2007; Ivey et al., 2008; Simon et al., 2008; Chen et al., 2010; Koutsoulidou et al., 2011; Anderson et al., 2006; Kim et al., 2006). This includes aca-miR-206, aca-miR-1a and aca-miR-1b, as well as members of the miR-133 family, 1_16347 (which was identified as mir-133b) and aca-miR-133a. Supporting this was the fact that expression in of these miRNAs was higher in skeletal muscle than base of the regenerating tail, with the exception of miR-206 and 1_16347 (Supplemental Table 9). This data suggests that previously identified roles of miRNAs in regulating the differentiation of muscle tissues are conserved in the lizards and at work in the regenerating tail. Additionally, the two novel miRNAs 5_10675 and GL343237.1_6814, for which orthology could not be determined, displayed significant upregulation in the base of the regenerating tail. In contrast, only two miRNAs displayed significant upregulation in the growing tip of the regenerating tail, aca-miR-2188 and aca-miR-184. Interestingly, miR-184 was recently shown to induce proliferation of adult neural stem/progenitor cells (Liu et al., 2010) and was most highly expressed in the tip of the regenerating tail relative to the brain and skeletal muscle (Supplemental Table 9). This suggests that aca-miR-184 could be playing a key role in regulating the growth of the neural-ependymal tube in the tip of the regenerating tail. Additionally, miR-2188

has been shown to regulate vascular development in zebrafish embryos (Soares et al., 2012), which suggests that aca-miR-2188 may regulate the vascularization of the regenerating tail. These data support the theory that tail regeneration in lizard may be organized by proliferation and differentiation of tissue specific progenitor cells located in the growing tip of the regenerating tail. Furthermore, expression of several key miRNA may regulated these choices of either cell proliferation or differentiation of progenitor cells during regeneration.

Since miRNAs exert their biological function through inhibiting the translation of and/or signaling the degradation of their mRNA targets, we further extended our investigation to predict the mRNA target of lizard miRNAs. Our study predicted that a total of 7,206 genes in the lizard may be regulated by miRNAs and that 3,269 of those genes may be targeted in the regenerating tail based on coexpression of the miRNA and its mRNA target (Hutchins et al., in preparation). Analysis of over-represented GO and KEGG pathways in coexpressed target genes identified biological processes under the regulation of miRNAs in the regenerating tail (Supplemental Table 12). However, to obtain a deeper understanding of which biological processes were specifically targeted by miRNA in the proliferative tip or the more differentiated base of the regenerating tail, we filtered for mRNA targets of miRNAs displaying at least two-fold increase in either the tip or the base. Through this analysis we identified several key biological processes likely critical to regeneration that are the target of

miRNAs. The GO pathway analysis identified groups of genes targeted by miRNAs up-regulated in the base of the regenerating tail relating to muscle function and development, as well the extra cellular matrix and the cytoskeleton (Table 8). This GO pathway analysis of groups of targeted by miRNAs up-regulated in the tip of the regenerating tail identified clusters relating primarily to cellular functions, such as membrane enclosed lumen, vesicle transport or protein transport (Table 8). However, the two KEGG pathways were identified as targets of miRNAs up-regulated in the tip with statistical significance. These two were the insulin signaling pathway and the ErbB signaling pathway, which has been shown to regulate cell proliferation and differentiation, and has become the a focal point of research in the treatment of cancer (Supplemental Table 14; Citri & Yarden, 2006). This could suggest that miRNA regulation of the ErbB pathway plays a critical role in the tail regeneration process in lizards.

To gain a further understanding of the regulation of gene expression through transcriptional and post-transcriptional regulation by miRNAs, we expanded our analysis to include target mRNA transcripts that display differential expression in the regenerating tail. Using this analysis we identified biological processes under coordinated transcriptional and miRNA regulation in both the proliferative tip of the regenerating tail and the more differentiated base of the regenerating tail. Of specific interest were miRNAs that displayed increased expression in the tip and whose target genes were also transcriptionally down-regulated

in the tip. Our GO pathway analysis of this group found clusters of genes related to the differentiation of several tissue types observable in the regenerating tail, including those involved in neuron function, muscle function, bone development, and the extracellular matrix. To identify genes induced in the tip of the regenerating tail, we filtered for the set of genes for those transcriptionally up-regulated in the tip and the target of miRNAs with expression down up-regulated in the tip. The GO pathway analysis for this subset of genes identified several over represented clusters involved in molecular processes such as ATPase activity, nucleotide binding, and cellular transportation. However, through this analysis we did identify two clusters related to apoptosis. These results suggest that while miRNAs in the tip of the regenerating tail are likely regulating the cell proliferation and programmed cell death, they are also required for proper growth and later differentiation of the regenerating tail.

Conclusion

With the recent release of the *A. carolinensis* genome, it has become possible to investigate molecular processes in the lizard on a genomic scale. In this study we present the first deep sequencing efforts to investigate miRNAs in *A. carolinensis*, and as such we have identified many novel miRNAs in lizard. This study has also identified several miRNAs differentially expressed during the tail regeneration process in the lizard, as well as provided an index for miRNA expression in the lizard

brain and skeletal muscle. Furthermore, as part of this study we have identified miRNAs as regulators of key genes involved regulating the proliferation and differentiation of progenitor cells of muscular, nervous and cartilaginous tissues. The miRNAs we identified as differentially expressed make excellent candidates for follow up studies to demonstrate their *in vivo* regulatory roles in regenerative process in lizard, with the ultimate aim translating these finding to regenerative medical research.

Chapter 5

CONCLUSION

The release of the *A. carolinensis* genome has provided a key resource for developing a reptilian model organism. Furthermore, the genome annotations initially provided by Ensembl and NCBI have made a valuable contribution to the genomic resources for lizard, which are instrumental for comparisons with mammals, such as mouse and human. However, these initial genome annotations were largely based on protein alignments from other species, such as chicken and human, and thus have difficulty identifying novel or highly divergent genes in the lizard, as well as UTR sequences. As part of this work, we have substantially improved these gene annotations through RNA-Seq from 14 different adult and embryonic tissues. Through analysis of the RNA-Seq we identified a large number of transcript that are not well represented on the current Anocar2.0 genome build. This identifies a clear need for additional genomic sequencing to supplement Anocar2.0 by closing gaps and improving the scaffolding structure of the assembly. These improvements would aid in the annotation of genes in currently poorly assembled regions, as well as help in the analysis of gene regulatory elements in non-coding regions of the genome. Additional improvements in the genome annotation could be made through sequencing of more RNA-Seq of different tissue types. Also, utilizing longer insert size libraries for RNA-Seq would allow for increased confidence in genes which have multiple

alternative splice forms involving regions of the gene long distances from each other. However, despite these remaining challenges, this genome annotation has substantially improved the genomic resources for the lizard. This will then aid comparative studies between vertebrates, as well as functional studies using *Anolis carolinensis* as a model system.

To demonstrate the utility of a reptilian model organism for comparisons of development and evolution, we investigated the developmental process of somitogenesis in the lizard, as well as in the alligator. Through these studies we identified both convergent and divergent features of the segmentation process among the vertebrates. More specifically, we identified a large shift in the regulation of somitogenesis that occurred concurrent with the evolution of the amniotes and adaptation to terrestrial life. However, it remains unclear which secondary oscillator controlling cyclical Notch expression in the amniote common ancestor. To answer this question would require further investigation other vertebrate orders, such as the turtles, marsupials and monotremes. This study also did not thoroughly investigate the expansion of gene oscillatory expression in the WNT and FGF pathways that has been observed between zebrafish, chicken and mammals (Krol et al., 2011). Additional expressional studies in the lizard would be needed to determine if lizards are similar to birds, with few oscillating genes in the WNT and FGF pathway, or more similar to mammals with a large number of oscillatory genes in these pathways. While this study has advanced the

lizard as a developmental model organism, several experimental resources are not yet available for the lizard. The development of methods to create transgenic lizards would be particularly helpful for developmental studies. Additionally, explant culture methods would greatly improve the ability to conduct functional experiments of lizard development.

As part of this study we also used *A. carolinensis* to investigate a process that is unique to non-avian reptiles among the amniotes, tail regeneration. Squamate lizards, including *A. carolinensis*, provide the most closely related species to humans with the regenerative capacity to regrow *de novo* muscle groups, hyaline cartilage, as well as motor and sensory neurons. Molecular studies of this significant regenerative capacity in the lizard are now possible with the release of *A. carolinensis* genome, along the revised genome annotation (Eckalbar et al., in review). Recent efforts have made significant progress towards describing the state of transcriptional regulation during the tail regeneration process (Hutchins, et al., in preparation). Here we have expanded those studies to include the post-transcriptional regulation through sequencing, annotation and expressional analysis of miRNAs. This study has provided a critical link between the transcriptional regulation of the tail regeneration process and the true gene expression as their protein products. However, the advancement of proteomic approaches in the lizard will provide the next step towards understanding the interactions of gene transcripts and

miRNAs. The continued efforts to create cell culture models for the lizard would allow for experimental validation of findings coming from large transcriptomic, miRNA and proteomic screens. Furthermore, these studies of regeneration in the lizard provide the bases for future studies that may eventually be translated toward improving the regenerative capacity in mammals.

As complete genomic sequences are becoming available for an ever-increasing number of species, a major bottleneck will in the subsequent analysis will be the generation of high quality gene annotations. Here we have shown that a deep transcriptomic sequencing for several adult and embryonic tissues can improve gene annotations to qualities approaching those of well-established model systems, such as the mouse. However, it is unlikely the required funds will be available for large transcriptomic sequencing projects to complement the genomic sequencing efforts for most newly sequenced species. Instead, what is likely to occur is small amounts of transcriptomic sequencing will be done to complement specific studies in a particular species. This will allow researchers the ability to complement predictive annotations, based on transcript alignments from closely related species, with specific sequencing to improve the annotation of genes of the highest interest in their particular study.

This work has provided a guide for methods in adopting new model organisms in the era of next-generation sequencing. Additionally, based

on this work, we can make several suggestions to researchers attempting to utilize a new model organism for studies of evolution, development and genomics. First, while RNA-Seq is a powerful tool to annotate and determine the expression of genes of interest in specific tissues, lowly expressed genes will often not assemble well, leaving the annotation largely up to alignments from other species and *ab initio* prediction methods. To help combat this problem, we highly recommend the use of strand-specific RNA-Seq, which can yield higher quality assemblies in low coverage situations. Second, *ab initio* prediction methods are greatly improved with a set of high quality transcripts for training. Thus, even if a particular gene may have high enough coverage for *de novo* assembly, it may benefit from well trained *ab initio* predictions. Third, when *de novo* transcript assembly fails it may be possible to utilize a reference based assembly method to improve gene annotations. This process often requires not just a genomic reference, but also an initial genome annotation to guide assembly.

Incorporating all the information that goes into a genome assembly is still a challenging problem. However, we recommend a multistep process of genome annotation, including merging information from other species transcript and protein alignments, trained *ab initio* predictions, and *de novo* assembled transcripts. This first round annotation will hopefully prove a high quality annotation for reference-guided assembly of the RNA-Seq data, which will be used to create the second round annotation.

Finally, we also recommend using the *de novo* assembled transcripts in a third round of genome annotation to identify alternative splice forms. However, we emphasize being as conservative as possible in this final step, as RNA-Seq data may contain many pre-mRNAs and uneven coverage leading to inaccurate assemblies. Strand-specific, poly-A library preparations can help combat these issues, and again we highly recommend utilizing such protocols for the first RNA-Seq projects in a new model organism.

Establishing a new model organism goes far beyond genomic sequences and genome annotations, however. While issues such as relatively small size and short reproductive cycles aids in the adaptation of a novel model organism, ultimately the researchers' willingness to tackle experimental and general laboratory challenges posed by the new species will make the most difference in determining the success of the project. As researchers expand into non-traditional model organisms, lessons gained from those working in closely related species should expedite adoption of further model organisms. Thus, maintaining a network of connections not only between other groups in the same species, but also between those in closely related species, will likely provide a key resource for accelerating the adoption of novel model organisms.

As an emerging model system, many of the resources and techniques available in other model species are not yet available for *Anolis carolinensis*. However, with this thesis we have advanced the

genomic and transcriptomic resources, including those for miRNAs, for the lizard. Through demonstration of the application of the lizard model organism we have also made substantial progress towards adapting the lizard as a model system for development, evolution and regeneration. Furthermore we have provided the foundation for future studies of these topics in *A. carolinensis*, as well as a guide for utilizing next-gen sequencing to accelerate the adoption of a novel model organism.

REFERENCES

Alföldi, J., di Palma, F., Grabherr, M., Williams, C., Kong, L., Mauceli, E., Russell, P., et al. (2011). The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature, 477(7366), 587–591. doi:10.1038/nature10390

Alibardi, L. (2010). Morphological and Cellular Aspects of Tail and Limb Regeneration in Lizards. Adv Anat Embryol Cell Biol., 207:iii, v-x, 1-109.

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. 11(10):R106. doi:10.1186/gb-2010-11-10-r106

Anderson, C., Catoe, H., & Werner, R. (2006). MIR-206 regulates connexin43 expression during skeletal muscle development. Nucleic acids research, 34(20), 5863–5871. doi:10.1093/nar/gkl743

Aoki-Kinoshita, K. F., & Kanehisa, M. (2007). Gene annotation and pathway mapping in KEGG. Methods In Molecular Biology, 396, 71-91. doi: 10.1007/978-1-59745-515-2_6

Aulehla, A., & Johnson, R. L. (1999). Dynamic expression of lunatic fringe suggests a link between notch signaling and an autonomous cellular oscillator driving somite segmentation. Developmental biology, 207(1), 49–61. doi:10.1006/dbio.1998.9164

Aulehla, A., Wehrle, C., Brand-Saberi, B., Kemler, R., Gossler, A., Kanzler, B., & Herrmann, B. G. (2003). Wnt3a plays a major role in the segmentation clock controlling somitogenesis. Developmental cell, 4(3), 395–406. doi:10.1016/S1534-5807

Barrantes, I. B., Elia, A. J., Wünsch, K., Hrabe de Angelis, M. H., Mak, T. W., Rossant, J., Conlon, R. A., et al. (1999). Interaction between Notch signalling and Lunatic fringe during somite boundary formation in the mouse. Current biology : CB, 9(9), 470–480. doi: 10.1016/S0960-9822

Bartel, D. P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. Cell, 136(2), 215–233. doi:10.1016/j.cell.2009.01.002

Baxter, L. R. (2001). Brain mediation of Anolis social dominance displays. III. Differential forebrain 3H-sumatriptan binding in dominant vs. submissive males. Brain, behavior and evolution, 57(4), 202–213.

Beck, C. W., & Slack, J. M. (1998). Analysis of the developing Xenopus tail bud reveals separate phases of gene expression during determination and outgrowth. Mechanisms of development, 72(1-2), 41–52. doi:10.1016/S0925-4773

Beckers, J., Schlautmann, N., & Gossler, A. (2000). The mouse rib-vertebrae mutation disrupts anterior-posterior somite patterning and genetically interacts with a Delta1 null allele. Mechanisms of development, 95(1-2), 35–46. doi:10.1016/S0925-4773

Bessho, Y., Hirata, H., Masamizu, Y., & Kageyama, R. (2003). Periodic repression by the bHLH factor Hes7 is an essential mechanism for the somite segmentation clock. Genes & development, 17(12), 1451–1456. doi:10.1101/gad.1092303

Bessho, Y., Miyoshi, G., Sakata, R., & Kageyama, R. (2001). Hes7: a bHLH-type repressor gene regulated by Notch and expressed in the presomitic mesoderm. Genes to Cells, 6(2), 175-185. doi:10.1046/j.1365-2443.2001.00409.x

Betel, D., Wilson, M., Gabow, A., Marks, D. S., & Sander, C. (2007). The microRNA.org resource: targets and expression. Nucleic acids research, 36(Database), D149–D153. doi:10.1093/nar/gkm995

Birol, I., Jackman, S. D., Nielsen, C. B., Qian, J. Q., Varhol, R., Stazyk, G., Morin, R. D., et al. (2009). De novo transcriptome assembly with ABySS. Bioinformatics, 25(21), 2872–2877. doi:10.1093/bioinformatics/btp367

Bonnet, E., Wuyts, J., Rouze, P., & Van de Peer, Y. (2004). Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. Bioinformatics, 20(17), 2911–2917. doi:10.1093/bioinformatics/bth374

Botstein, D., Cherry, J. M., Ashburner, M., Ball, C. A., Blake, J. A., Butler, H., Davis, A. P., et al. (2000). Gene Ontology: tool for the unification of biology. Nature genetics, 25(1), 25–29. doi:10.1038/75556

Buchberger, A., Seidl, K., Klein, C., Eberhardt, H., & Arnold, H. H. (1998). cMeso-1, a novel bHLH transcription factor, is involved in somite formation in chicken embryos. Developmental biology, 199(2), 201–215. doi:10.1006/dbio.1998.8919

Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B.,

Holt, C., et al. (2007). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome research, 18(1), 188–196. doi:10.1101/gr.6743907

Caprioli, A., Goitsuka, R., Pouget, C., Dunon, D., & Jaffredo, T. (2002). Expression of Notch genes and their ligands during gastrulation in the chicken embryo. Mechanisms of development, 116(1-2), 161–164. doi: 10.1016/S0925-4773.

Castoe, T. A., de Koning, J. A. P., Hall, K. T., Yokoyama, K. D., Gu, W., Smith, E. N., Feschotte, C., et al. (2011). Sequencing the genome of the Burmese python (Python molurus bivittatus) as a model for studying extreme adaptations in snakes. Genome biology, 12(7), 1–8. doi:10.1186/gb-2011-12-7-406

Chapman, G., Sparrow, D. B., Kremmer, E., & Dunwoodie, S. L. (2011). Notch inhibition by the ligand DELTA-LIKE 3 defines the mechanism of abnormal vertebral segmentation in spondylocostal dysostosis. Human molecular genetics, 20(5), 905–916. doi:10.1093/hmg/ddq529

Chen, J.-F., Mandel, E. M., Thomson, J. M., Wu, Q., Callis, T. E., Hammond, S. M., Conlon, F. L., et al. (2005). The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. Nature genetics, 38(2), 228–233. doi:10.1038/ng1725

Chen, J.-F., Tao, Y., Li, J., Deng, Z., Yan, Z., Xiao, X., & Wang, D.-Z. (2010). microRNA-1 and microRNA-206 regulate skeletal muscle satellite cell proliferation and differentiation by repressing Pax7. The Journal of Cell Biology, 190(5), 867–879. doi:10.1083/jcb.200911036

Citri, A., & Yarden, Y. (2006). EGF–ERBB signalling: towards the systems level. Nature Reviews Molecular Cell Biology, 7(7), 505–516. doi:10.1038/nrm1962

Clark, E. C., & Baxter, L. R. (2000a). Mammal-like striatal functions in Anolis. I. Distribution of serotonin receptor subtypes, and absence of striosome and matrix organization. Brain, behavior and evolution, 56(5), 235–248. doi:10.1159/000047207

Clark, E. C., & Baxter, L. R., Jr. (2000b). Mammal-Like Striatal Functions in Anolis. Brain, behavior and evolution, 56(5), 235–248. doi:10.1159/000047207

Cole, S. E., Levorse, J. M., Tilghman, S. M., & Vogt, T. F. (2002). Clock regulatory elements control cyclic expression of Lunatic fringe during

somitogenesis. Developmental cell, 3(1), 75–84.

Conesa, A., & Götz, S. (2008). Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. International journal of plant genomics, 2008, 1–12. doi:10.1155/2008/619832

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics, 21(18), 3674–3676. doi:10.1093/bioinformatics/bti610

Cooke, J., & Zeeman, E. C. (1976). A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. Journal of theoretical biology, 58(2), 455–476. doi:10.1016/S0022-5193

Cossins, J., Vernon, A. E., Zhang, Y., Philpott, A., & Jones, P. H. (2002). Hes6 regulates myogenic differentiation. Development, 129(9), 2195–2207. http://dev.biologists.org/content/129/9/2195.long

Crossley, P. H., & Martin, G. R. (1995). The mouse Fgf8 gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. Development, 121(2), 439–451. http://dev.biologists.org/content/121/2/439.long

Dale, J. K., Malapert, P., Chal, J., Vilhais-Neto, G., Maroto, M., Johnson, T., Jayasinghe, S., et al. (2006). Oscillations of the Snail Genes in the Presomitic Mesoderm Coordinate Segmental Patterning and Morphogenesis in Vertebrate Somitogenesis. Developmental cell, 10(3), 355–366. doi:10.1016/j.devcel.2006.02.011

Dale, J. K., Maroto, M., Dequéant, M. L., Malapert, P., McGrew, M., & Pourquie, O. (2003). Periodic notch inhibition by lunatic fringe underlies the chick segmentation clock. Nature, 421(6920), 275–278. doi:10.1038/nature01244

Denli, A. M., Tops, B. B. J., Plasterk, R. H. A., Ketting, R. F., & Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature, 432(7014), 231–235. doi:10.1038/nature03049

Dennis, G., Jr, Sherman, B. T., Hosack, D. A., & Yang, J. (2003). DAVID: database for annotation, visualization, and integrated discovery. Genome Biology, 4(P3). doi:10.1186/gb-2003-4-5-p3

Dequéant, M.-L., & Pourquié, O. (2008). Segmental patterning of the

vertebrate embryonic axis. Nature reviews. Genetics, 9(5), 370–382. doi:10.1038/nrg2320

Doench, J. G., & Sharp, P. A. (2004). Specificity of microRNA target selection in translational repression. Genes & development, 18(5), 504–511. doi:10.1101/gad.1184404

Draper, B. W., Stock, D. W., & Kimmel, C. B. (2003). Zebrafish fgf24 functions with fgf8 to promote posterior mesodermal development. Development, 130(19), 4639–4654. doi:10.1242/dev.00671

Dubrulle, J., McGrew, M. J., & Pourquie, O. (2001). FGF Signaling Controls Somite Boundary Position and Regulates Segmentation Clock Control of Spatiotemporal Hox Gene Activation. Cell. 106(2), 219–232. 10.1016/S0092-8674

Dunwoodie, S. L., Clements, M., Sparrow, D. B., Sa, X., Conlon, R. A., & Beddington, R. S. P. (2002). Axial skeletal defects caused by mutation in the spondylocostal dysplasia/pudgy gene Dll3 are associated with disruption of the segmentation clock within the presomitic mesoderm. Development, 129(7), 1795-806. http://dev.biologists.org/content/129/7/1795.long

Dunwoodie, S. L., Henrique, D., Harrison, S. M., & Beddington, R. S. (1997). Mouse Dll3: a novel divergent Delta gene which may complement the function of other Delta homologues during early pattern formation in the mouse embryo. Development, 124(16), 3065–3076. http://dev.biologists.org/content/124/16/3065.long

Eckalbar, W. L., Lasku, E., Infante, C. R., Elsey, R. M., Markov, G. J., Allen, A. N., Corneveaux, J. J., et al. (2012). Somitogenesis in the anole lizard and alligator reveals evolutionary convergence and divergence in the amniote segmentation clock. Developmental biology, 363(1), 308–319. doi:10.1016/j.ydbio.2011.11.021

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC bioinformatics, 5, 113. doi:10.1186/1471-2105-5-113

Evrard, Y. A., Lun, Y., Aulehla, A., Gan, L., & Johnson, R. L. (1998). lunatic fringe is an essential mediator of somite segmentation and patterning. Nature, 394(6691), 377–381. doi:10.1038/28632

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 39(4), 783-791.

http://www.jstor.org/stable/2408678

Fior, R., & Henrique, D. (2005). A novel hes5/hes6 circuitry of negative regulation controls Notch activity during neurogenesis. Developmental biology, 281(2), 318–333. doi:10.1016/j.ydbio.2005.03.017

Fisher, R. E., Geiger, L. A., Stroik, L. K., Hutchins, E. D., George, R. M., DeNardo, D. F., Kusumi, K., et al. (2012). A Histological Comparison of the Original and Regenerated Tail in the Green Anole, Anolis carolinensis. The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology, 295(10), 1609–1619. doi:10.1002/ar.22537

Flicek, P., Amode, M. R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., et al. (2011). Ensembl 2012. Nucleic acids research, 40(D1), D84–D90. doi:10.1093/nar/gkr991

Forsberg, H., Crozet, F., & Brown, N. A. (1998). Waves of mouse Lunatic fringe expression, in four-hour cycles at two-hour intervals, precede somite boundary formation. Current biology : CB, 8(18), 1027–1030. doi:10.1016/S0960-9822(07)00424-1

Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W., & Rajewsky, N. (2011). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic acids research, 40(1), 37–52. doi:10.1093/nar/gkr688

Frith, M. C., Hansen, U., & Weng, Z. (2001). Detection of cis-element clusters in higher eukaryotic DNA. Bioinformatics, 17(10), 878–889. doi:10.1093/bioinformatics/17.10.878

Fujita, M. K., Edwards, S. V., & Ponting, C. P. (2011). The Anolis Lizard Genome: An Amniote Genome without Isochores. Genome biology and evolution, 3(0), 974–984. doi:10.1093/gbe/evr072

Geffers, I., Serth, K., Chapman, G., Jaekel, R., Schuster-Gossler, K., Cordes, R., Sparrow, D. B., et al. (2007). Divergent functions and distinct localization of the Notch ligands DLL1 and DLL3 in vivo. The Journal of Cell Biology, 178(3), 465–476. doi:10.1083/jcb.200702009

Gene Ontology Consortium. (2004). The Gene Ontology (GO) database and informatics resource. Nucleic acids research, 32(90001), 258D–261. doi:10.1093/nar/gkh036

Genome 10K Community of Scientists. (2009). Genome 10K: A Proposal

to Obtain Whole-Genome Sequence for 10 000 Vertebrate Species. Journal of Heredity, 100(6), 659–674. doi:10.1093/jhered/esp086

Gibb, S., Maroto, M., & Dale, J. K. (2010). The segmentation clock mechanism moves up a notch. Trends in cell biology, 20(10), 593–600. doi:10.1016/j.tcb.2010.07.001

Gibb, S., Zagorska, A., Melton, K., Tenin, G., Vacca, I., Trainor, P., Maroto, M., et al. (2009). Interfering with Wnt signalling alters the periodicity of the segmentation clock. Developmental biology, 330(1), 21–31. doi:10.1016/j.ydbio.2009.02.035

Glinka, A., Wu, W., Delius, H., Monaghan, A. P., Blumenstock, C., & Niehrs, C. (1998). Dickkopf-1 is a member of a new family of secreted proteins and functions in head induction. Nature, 391(6665), 357–362. doi:10.1038/34848

Gomez, C., Ozbudak, E. M., Wunderlich, J., Baumann, D., Lewis, J., & Pourquié, O. (2008). Control of segment number in vertebrate embryos. Nature, 454(7202), 335–339. doi:10.1038/nature07020

Goss, R. J. (1992). The evolution of regeneration: Adaptive or inherent? Journal of theoretical biology, 159(2), 241–260. doi:10.1016/S0022-5193(05)80704-0

Gotz, S., Garcia-Gomez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., et al. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic acids research, 36(10), 3420–3435. doi:10.1093/nar/gkn176

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology, 29(7), 644–652. doi:10.1038/nbt.1883

Grassa, C. J., & Kulathinal, R. J. (2011). Elevated Evolutionary Rates among Functionally Diverged Reproductive Genes across Deep Vertebrate Lineages. International journal of evolutionary biology, 2011, 1–9. doi:10.4061/2011/274975

Gregory, R. I., Yan, K.-P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., & Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature, 432(7014), 235–240. doi:10.1038/nature03120

Griffiths-Jones, S. (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic acids research, 34(90001), D140–D144. doi:10.1093/nar/gkj112

Griffiths-Jones, S., Saini, H. K., & Van Dongen, S. (2008). miRBase: tools for microRNA genomics. Nucleic acids research, 36, D154-D158. doi:10.1093/nar/gkm952

Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D. L., et al. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell, 106(1), 23–34.

Haas, B. J. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic acids research, 31(19), 5654–5666. doi:10.1093/nar/gkg770

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome biology, 9(1), R7. doi:10.1186/gb-2008-9-1-r7

He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., et al. (1998). Identification of c-MYC as a target of the APC pathway. Science, 281(5382), 1509–1512. doi:10.1126/science.281.5382.1509

Hicks, C., Johnston, S. H., diSibio, G., Collazo, A., Vogt, T. F., & Weinmaster, G. (2000). Fringe differentially modulates Jagged1 and Delta1 signalling through Notch1 and Notch2. Nature cell biology, 2(8), 515–520. doi:10.1038/35019553

Hitachi, K., Kondow, A., Danno, H., Nishimura, Y., Okabayashi, K., & Asashima, M. (2009). Molecular analyses of Xenopus laevis Mesp-related genes. Integrative zoology, 4(4), 387–394. doi:10.1111/j.1749-4877.2009.00110.x

Hofacker, I. L., & Stadler, P. F. (2006). Memory efficient folding algorithms for circular RNA secondary structures. Bioinformatics, 22(10), 1172–1176. doi:10.1093/bioinformatics/btl023

Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., & Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. Monatshefte für Chemie Chemical Monthly, 125(2), 167–

188. doi:10.1007/BF00818163

Holley, S. A. (2007). The genetics and embryology of zebrafish metamerism. (O. Pourquié & J. F. Fallon, Eds.)Developmental Dynamics, 236(6), 1422–1449. doi:10.1002/dvdy.21162

Holley, S. A., Geisler, R., & Nüsslein-Volhard, C. (2000). Control of her1 expression during zebrafish somitogenesis by a delta-dependent oscillator and an independent wave-front activity. Genes & development, 14(13), 1678–1690. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC316735/pdf/X12.pdf

Holley, S. A., Jülich, D., Rauch, G.-J., Geisler, R., & Nüsslein-Volhard, C. (2002). her1 and the notch pathway function within the oscillator mechanism that regulates zebrafish somitogenesis. Development, 129(5), 1175–1183. http://dev.biologists.org/content/129/5/1175.long

Holt, C., & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC bioinformatics, 12(1), 491. doi:10.1186/1471-2105-12-491

Huang, X., & Madan, A. (1999). CAP3: A DNA sequence assembly program. Genome research, 9(9), 868–877.

Huang, Y., Niu, B., Gao, Y., Fu, L., & Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics, 26(5), 680–682. doi:10.1093/bioinformatics/btq003

Hutvágner, G., McLachlan, J., Pasquinelli, A. E., Bálint, E., Tuschl, T., & Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science (New York, N.Y.), 293(5531), 834–838. doi:10.1126/science.1062961

Janes, D. E., Chapus, C., Gondo, Y., Clayton, D. F., Sinha, S., Blatti, C. A., Organ, C. L., et al. (2011). Reptiles and Mammals Have Differentially Retained Long Conserved Noncoding Sequences from the Amniote Ancestor. Genome biology and evolution, 3(0), 102–113. doi:10.1093/gbe/evq087

Jen, W. C., Gawantka, V., Pollet, N., Niehrs, C., & Kintner, C. (1999). Periodic repression of Notch pathway genes governs the segmentation of Xenopus embryos. Genes & development, 13(11), 1486–1499.

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC316761/pdf/x12.pdf

Jen, W. C., Wettstein, D., Turner, D., Chitnis, A., & Kintner, C. (1997). The Notch ligand, X-Delta-2, mediates segmentation of the paraxial mesoderm in Xenopus embryos. Development, 124(6), 1169–1178. http://dev.biologists.org/content/124/6/1169.full.pdf

Jiang, Y. J., Aerne, B. L., Smithers, L., Haddon, C., Ish-Horowicz, D., & Lewis, J. (2000). Notch signalling and the synchronization of the somite segmentation clock. Nature, 408(6811), 475–479. doi:10.1038/35044091

John, J. S., Braun, E., & Isberg, S. (2012). Sequencing three crocodilian genomes to illuminate the evolution of archosaurs and amniotes. Genome Biology, 13(415), doi:10.1186/gb-2012-13-1-415

Johnson, M. A., & Wade, J. (2010). Behavioural display systems across nine Anolis lizard species: sexual dimorphisms in structure and function. Proceedings of the Royal Society B: Biological Sciences, 277(1688), 1711–1719. doi:10.1098/rspb.2009.2323

Johnson, M. A., Cohen, R. E., Vandecar, J. R., & Wade, J. (2011). Relationships among reproductive morphology, behavior, and testosterone in a natural population of green anole lizards. Physiology & behavior, 104(3), 437–445. doi:10.1016/j.physbeh.2011.05.004

Jones, S. G. (2004). The microRNA registry. Nucleic acids research. 32(1), D109-D111. doi:10.1093/nar/gkh023

Jouve, C., Palmeirim, I., Henrique, D., Beckers, J., Gossler, A., Ish-Horowicz, D., & Pourquie, O. (2000). Notch signalling is required for cyclic expression of the hairy-like gene HES1 in the presomitic mesoderm. Development, 127(7), 1421–1429. http://dev.biologists.org/content/127/7/1421.long

Kanehisa, M., Araki, M., Goto, S., & Hattori, M. (2008). KEGG for linking genomes to life and the environment. Nucleic acids research, 36, D480–D484. doi:10.1093/nar/gkm882

Kent, W. J. (2002). BLAT---The BLAST-Like Alignment Tool. Genome research, 12(4), 656–664. doi:10.1101/gr.229202

Kim, H. K., Lee, Y. S., Sivaprasad, U., Malhotra, A., & Dutta, A. (2006). Muscle-specific microRNA miR-206 promotes muscle differentiation. The Journal of Cell Biology, 174(5), 677–687. doi:

10.1083/jcb.200603008

Kolbe, J. J., Revell, L. J., Szekely, B., Brodie, E. D., & Losos, J. B. (2011). Convergent evolution of phenotypic integration and its alignment with morphological diversification in Caribbean Anolis ecomorphs. Evolution; international journal of organic evolution, 65(12), 3608–3624. doi:10.1111/j.1558-5646.2011.01416.x

Korf, I. (2004). Gene finding in novel genomes. BMC bioinformatics, 5, 59. doi:10.1186/1471-2105-5-59

Koshiba-Takeuchi, K., Mori, A. D., Kaynak, B. L., Cebra-Thomas, J., Sukonnik, T., Georges, R. O., Latham, S., et al. (2009). Reptilian heart development and the molecular basis of cardiac chamber evolution. Nature, 461(7260), 95–98. doi:10.1038/nature08324

Koutsoulidou, A., Mastroyiannopoulos, N. P., Furling, D., Uney, J. B., & Phylactou, L. A. (2011). Expression of miR-1, miR-133a, miR-133b and miR-206 increases during development of human skeletal muscle. BMC Developmental Biology, 11(1), 34. doi:10.1186/1471-213X-11-34

Koyano-Nakagawa, N., Kim, J., Anderson, D., & Kintner, C. (2000). Hes6 acts in a positive feedback loop with the neurogenins to promote neuronal differentiation. Development, 127(19), 4203–4216. http://dev.biologists.org/content/127/19/4203.long

Kozomara, A., & Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic acids research. doi: 10.1093/nar/gkq1027

Kragl, M., Knapp, D., Nacu, E., Khattak, S., Maden, M., Epperlein, H. H., & Tanaka, E. M. (2009). Cells keep a memory of their tissue origin during axolotl limb regeneration. Nature, 460(7251), 60–65. doi:10.1038/nature08152

Krol, A. J., Roellig, D., Dequéant, M. L., Tassy, O., Glynn, E., Hattem, G., Mushegian, A., et al. (2011). Evolutionary plasticity of segmentation clock networks. Development, 138(13), 2783–2792. doi:10.1242/dev.063834

Krüger, J., & Rehmsmeier, M. (2006). RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic acids research, 34(Web Server), W451–W454. doi:10.1093/nar/gkl243

Kuntz, S. G., Schwarz, E. M., DeModena, J. A., De Buysscher, T., Trout,

D., Shizuya, H., Sternberg, P. W., et al. (2008). Multigenome DNA sequence conservation identifies Hox cis-regulatory elements. Genome research, 18(12), 1955–1968. doi:10.1101/gr.085472.108

Kusumi, K., Kulathinal, R. J., Abzhanov, A., Boissinot, S., Crawford, N. G., Faircloth, B. C., Glenn, T. C., et al. (2011). Developing a community-based genetic nomenclature for anole lizards. BMC genomics, 12(1), 554. doi:10.1186/1471-2164-12-554

Kusumi, K., Sun, E. S., Kerrebrock, A. W., Bronson, R. T., Chi, D. C., Bulotsky, M. S., Spencer, J. B., et al. (1998). The mouse pudgy mutation disrupts Delta homologue Dll3 and initiation of early somite boundaries. Nature genetics, 19(3), 274–278. doi:10.1038/961

Ladi, E., Nichols, J. T., Ge, W., Miyamoto, A., Yao, C., Yang, L.-T., Boulter, J., et al. (2005). The divergent DSL ligand Dll3 does not activate Notch signaling but cell autonomously attenuates signaling induced by other DSL ligands. The Journal of Cell Biology, 170(6), 983–992. doi:10.1083/jcb.200503113

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods, 9(4), 357–359. doi:10.1038/nmeth.1923

Langmead, B., Trapnell, C., & Pop, M. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology, 10:R25. doi:10.1186/gb-2009-10-3-r25

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics, 23(21), 2947–2948. doi:10.1093/bioinformatics/btm404

Lee, R. C., & Ambros, V. (2001). An Extensive Class of Small RNAs in Caenorhabditis elegans. Science Signaling, 294(5543), 862. doi:10.1126/science.1065329

Leimeister, C., Schumacher, N., Steidl, C., & Gessler, M. (2000). Analysis of HeyL expression in wild-type and Notch pathway mutant mouse embryos. Mechanisms of development, 98(1-2), 175–178. doi: 10.1016/S0925-4773

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics, 22(13), 1658–1659. doi:10.1093/bioinformatics/btl158

Lim, C. H., Hamazaki, T., Braun, E. L., Wade, J., & Terada, N. (2011). Evolutionary genomics implies a specific function of Ant4 in mammalian and anole lizard male germ cells. (A. P. McGregor, Ed.)PloS one, 6(8), e23122. doi:10.1371/journal.pone.0023122.g006

Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., Bartel, D. P., et al. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature, 433(7027), 769–773. doi:10.1038/nature03315

Liu, C., Teng, Z.-Q., Santistevan, N. J., Szulwach, K. E., Guo, W., Jin, P., & Zhao, X. (2010). Epigenetic Regulation of miR-184 by MBD1 Governs Neural Stem Cell Proliferation and Differentiation. Cell Stem Cell, 6(5), 433–444. doi:10.1016/j.stem.2010.02.017

Loke, J. C. (2005). Compilation of mRNA Polyadenylation Signals in Arabidopsis Revealed a New Signal Element and Potential Secondary Structures. Plant physiology, 138(3), 1457–1468. doi:10.1104/pp.105.060541

Losos, J., Braun, E., Brown, D., & Clifton, S. (2005). Proposal to sequence the first reptilian genome: The green anole lizard, Anolis carolinensis. White Paper. http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/GreenAnoleLizardAmericanAlligatorSeq.pdf

Losos, J., Jackman, T., Larson, A., Queiroz, K., & Rodriguez-Schettino, L. (1998). Contingency and determinism in replicated adaptive radiations of island lizards. Science, 279(5359), 2115–2118. doi: 10.1126/science.279.5359.2115

Lovern, M. B., Holmes, M. M., & Wade, J. (2004). The green anole (Anolis carolinensis): a reptilian model for laboratory studies of reproductive morphology and behavior. ILAR journal / National Research Council, Institute of Laboratory Animal Resources, 45(1), 54–64. http://dels-old.nas.edu/ilar_n/ilarjournal/45_1/pdfs/v4501lovern.pdf

Maginnis, T. L. (2006). The costs of autotomy and regeneration in animals: a review and framework for future research. Behavioral Ecology, 17(5), 857–872. doi:10.1093/beheco/arl010

Mansfield, J. H., & Abzhanov, A. (2010). Hox expression in the American alligator and evolution of archosaurian axial patterning. Journal of experimental zoology. Part B, Molecular and developmental evolution, 314(8), 629–644. doi:10.1002/jez.b.21364

Mardis, E. R. (2011). A decade's perspective on DNA sequencing technology. Nature, 470(7333), 198–203. doi:10.1038/nature09796

Maruhashi, M., Van De Putte, T., Huylebroeck, D., Kondoh, H., & Higashi, Y. (2005). Involvement of SIP1 in positioning of somite boundaries in the mouse embryo. Developmental Dynamics, 234(2), 332–338. doi:10.1002/dvdy.20546

Matys, V., Kel-Margoulis, O., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., et al. (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic acids research, 34(Database issue), D108–D110. doi:10.1093/nar/gkj143

McCusker, C., & Gardiner, D. M. (2011). The axolotl model for regeneration and aging research: a mini-review. Gerontology, 57(6), 565–571. doi:10.1159/000323761

McGrew, M. J., Dale, J. K., Fraboulet, S., & Pourquie, O. (1998). The lunatic fringe gene is a target of the molecular clock linked to somite segmentation in avian embryos. Current biology, 8(17), 979–982. doi:10.1016/S0960-9822(98)70401-4

Moloney, D. J., Panin, V. M., Johnston, S. H., Chen, J., Shao, L., Wilson, R., Wang, Y., et al. (2000). Fringe is a glycosyltransferase that modifies Notch. Nature, 406(6794), 369–375. doi:10.1038/35019000

Morales, A. V., Yasuda, Y., & Ish-Horowicz, D. (2002). Periodic Lunatic fringe Expression Is Controlled during Segmentation by a Cyclic Transcriptional Enhancer Responsive to Notch Signaling. Developmental cell, 3(1), 63–74. doi:10.1016/S1534-5807(02)00211-3

Morimoto, M., Sasaki, N., Oginuma, M., Kiso, M., Igarashi, K., Aizaki, K.-I., Kanno, J., et al. (2007). The negative regulation of Mesp2 by mouse Ripply2 is required to establish the rostro-caudal patterning within a somite. Development, 134(8), 1561–1569. doi:10.1242/dev.000836

Muneoka, K., Allan, C. H., Yang, X., Lee, J., & Han, M. (2008). Mammalian regeneration and regenerative medicine. Birth defects research. Part C, Embryo today : reviews, 84(4), 265–280. doi:10.1002/bdrc.20137

Muth, A. N., Tsuchihashi, T., McManus, M. T., & Schwartz, R. J. (2007). Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in

mice lacking miRNA-1-2. Cell, 129(2), 303-317 10.1016/j.cell.2007.03.030

Niwa, Y., Masamizu, Y., Liu, T., Nakayama, R., Deng, C.-X., & Kageyama, R. (2007). The initiation and propagation of Hes7 oscillation are cooperatively regulated by Fgf and notch signaling in the somite segmentation clock. Developmental cell, 13(2), 298–304. doi:10.1016/j.devcel.2007.07.013

Novick, P. A., Smith, J. D., Floumanhaft, M., Ray, D. A., & Boissinot, S. (2011). The Evolution and Diversity of DNA Transposons in the Genome of the Lizard Anolis carolinensis. Genome biology and evolution, 3(0), 1–14. doi:10.1093/gbe/evq080

Nye, H. L. D., Cameron, J. A., Chernoff, E. A. G., & Stocum, D. L. (2003). Regeneration of the urodele limb: A review. Developmental Dynamics, 226(2), 280–294. doi:10.1002/dvdy.10236

Ozbudak, E. M., & Pourquié, O. (2008). The vertebrate segmentation clock: the tip of the iceberg. Current opinion in genetics & development, 18(4), 317–323. doi:10.1016/j.gde.2008.06.007

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., et al. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic acids research, 37(18), e123. doi:10.1093/nar/gkp596

Patron, J. P., Fendler, A., Bild, M., Jung, U., Müller, H., Arntzen, M. Ø., Piso, C., et al. (2012). MiR-133b Targets Antiapoptotic Genes and Enhances Death Receptor-Induced Apoptosis. PloS one, 7(4), e35345. doi:10.1371/journal.pone.0035345

Pera, E. M., Kim, J. I., Martinez, S. L., Brechner, M., Li, S. Y., Wessely, O., & De Robertis, E. M. (2002). Isthmin is a novel secreted protein expressed as part of the Fgf-8 synexpression group in the Xenopus midbrain-hindbrain organizer. Mechanisms of development, 116(1-2), 169–172. doi: 10.1016/S0925-4773(02)00123-5,

Pissarra, L., Henrique, D., & Duarte, A. (2000). Expression of hes6, a new member of the Hairy/Enhancer-of-split family, in mouse development. Mechanisms of development, 95(1-2), 275–278. doi: 10.1016/S0925-4773(00)00443-3

Poss, K. D., Keating, M. T., & Nechiporuk, A. (2003). Tales of regeneration in zebrafish. Developmental Dynamics, 226(2), 202–210.

doi:10.1002/dvdy.10220

Reichman, O. J. (1984). Evolution of Regeneration Capabilities. American Naturalist, 123(6), 752-763. http://www.jstor.org/stable/2460898

Reifers, F., Böhli, H., Walsh, E. C., Crossley, P. H., Stainier, D. Y., & Brand, M. (1998). Fgf8 is mutated in zebrafish acerebellar (ace) mutants and is required for maintenance of midbrain-hindbrain boundary development and somitogenesis. Development, 125(13), 2381–2395. http://dev.biologists.org/content/125/13/2381.long

Rhind, N., Chen, Z., Yassour, M., Thompson, D. A., Haas, B. J., Habib, N., Wapinski, I., et al. (2011). Comparative Functional Genomics of the Fission Yeasts. Science, 332(6032), 930–936. doi:10.1126/science.1203357

Ritzman, T., Stroik, L., & Julik, E. (2012). The Gross Anatomy of the Original and Regenerated Tail in the Green Anole (Anolis carolinensis). The Anatomical Record, 295(10), 1596-608. doi: 10.1002/ar.22524

Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols, 7(3), 562–578. doi:10.1038/nprot.2012.016

Roberts, A., Pimentel, H., Trapnell, C., & Pachter, L. (2011a). Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics, 27(17), 2325–2329. doi:10.1093/bioinformatics/btr355

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L., & Pachter, L. (2011b). Improving RNA-Seq expression estimates by correcting for fragment bias. Genome biology, 12(3), R22. doi:10.1186/gb-2011-12-3-r22

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., Mungall, K., et al. (2010). De novo assembly and analysis of RNA-seq data. Nature methods, 7(11), 909–912. doi:10.1038/nmeth.1517

Robinson, G. E., Hackett, K. J., Purcell-Miramontes, M., Brown, S. J., Evans, J. D., Goldsmith, M. R., Lawson, D., et al. (2011). Creating a Buzz About Insect Genomes. Science, 331(6023), 1386–1386. doi:10.1126/science.331.6023.1386

Saga, Y., Hata, N., Koseki, H., & Taketo, M. M. (1997). Mesp2: a novel mouse gene expressed in the presegmented mesoderm and essential

for segmentation initiation. Genes & development, 11(14), 1827–1839. doi: 10.1101/gad.11.14.1827

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution, 4(4), 406–425. http://mbe.oxfordjournals.org/content/4/4/406.long

Sanger, T. J., Revell, L. J., Gibson-Brown, J. J., & Losos, J. B. (2012). Repeated modification of early limb morphogenesis programmes underlies the convergence of relative limb length in Anolis lizards. Proceedings of the Royal Society B: Biological Sciences, 279(1729), 739–748. doi:10.1098/rspb.2011.0840

Sawada, A., Fritz, A., Jiang, Y. J., Yamamoto, A., Yamasu, K., Kuroiwa, A., Saga, Y., et al. (2000). Zebrafish Mesp family genes, mesp-a and mesp-b are segmentally expressed in the presomitic mesoderm, and Mesp-b confers the anterior identity to the developing somites. Development, 127(8), 1691–1702. http://dev.biologists.org/content/127/8/1691.long

Serth, K., Schuster-Gossler, K., Cordes, R., & Gossler, A. (2003). Transcriptional oscillation of lunatic fringe is essential for somitogenesis. Genes & development, 17(7), 912–925. doi:10.1101/gad.250603

Sewell, W., Sparrow, D. B., Smith, A. J., & Gonzalez, D. M. (2009). Cyclical expression of the Notch/Wnt regulator Nrarp requires modulation by Dll3 in somitogenesis. Developmental Biology, 329(2), 400-409. doi: 10.1016/j.ydbio.2009.02.023

Shen, Y., Ji, G., Haas, B. J., Wu, X., Zheng, J., Reese, G. J., & Li, Q. Q. (2008). Genome level analysis of rice mRNA 3'-end processing signals and alternative polyadenylation. Nucleic acids research, 36(9), 3150–3161. doi:10.1093/nar/gkn158

Sieger, D., Ackermann, B., Winkler, C., Tautz, D., & Gajewski, M. (2006). her1 and her13.2 are jointly required for somitic border specification along the entire axis of the fish embryo. Developmental biology, 293(1), 242–251. doi:10.1016/j.ydbio.2006.02.003

Simon, D. J., Madison, J. M., & Conery, A. L. (2008). The MicroRNA miR-1 Regulates a MEF-2-Dependent Retrograde Signal at Neuromuscular Junctions. Cell, 30;133(5), 903-915. doi: 10.1016/j.cell.2008.04.035

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. M., &

Birol, I. (2009). ABySS: A parallel assembler for short read sequence data. Genome Research, 19(6), 1117–1123. doi: 10.1101/gr.089532.108

Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC bioinformatics, 6, 31. doi:10.1186/1471-2105-6-31

Sparrow, D. B. (2009). Advances in Experimental Medicine and Biology. (M. Maroto & N. V. Whittock, Eds.)Somitogenesis (Vol. 638, pp. 73–94). New York, NY: Springer New York. doi:10.1007/978-0-387-09606-3_4

Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics, 19(Suppl 2), ii215–ii225. doi:10.1093/bioinformatics/btg1080

Stanke, M., Schöffmann, O., Morgenstern, B., & Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC bioinformatics, 7, 62. doi:10.1186/1471-2105-7-62

Stefani, G., & Slack, F. J. (2008). Small non-coding RNAs in animal development. Nature Reviews Molecular Cell Biology, 9(3), 219–230. doi:10.1038/nrm2347

Summers, C. H., Forster, G. L., Korzan, W. J., Watt, M. J., Larson, E. T., Overli, O., Höglund, E., et al. (2005a). Dynamics and mechanics of social rank reversal. Journal of comparative physiology. A, Neuroethology, sensory, neural, and behavioral physiology, 191(3), 241–252. doi:10.1007/s00359-004-0554-z

Summers, C. H., Watt, M. J., Ling, T. L., Forster, G. L., Carpenter, R. E., Korzan, W. J., Lukkes, J. L., et al. (2005b). Glucocorticoid interaction with aggression in non-mammalian vertebrates: reciprocal action. European journal of pharmacology, 526(1-3), 21–35. doi:10.1016/j.ejphar.2005.09.059

Sun, W., Julie Li, Y.-S., Huang, H.-D., Shyy, J. Y.-J., & Chien, S. (2010). microRNA: A Master Regulator of Cellular Processes for Bioengineering Systems. Annual Review of Biomedical Engineering, 12(1), 1–27. doi:10.1146/annurev-bioeng-070909-105314

Suriben, R., Fisher, D. A., & Cheyette, B. N. R. (2006). Dact1 presomitic mesoderm expression oscillates in phase with Axin2 in the

somitogenesis clock of mice. Developmental Dynamics, 235(11), 3177–3183. doi:10.1002/dvdy.20968

Takada, S., Stark, K. L., Shea, M. J., Vassileva, G., McMahon, J. A., & McMahon, A. P. (1994). Wnt-3a regulates somite and tailbud formation in the mouse embryo. Genes & development, 8(2), 174–189. doi: 10.1101/gad.8.2.174

Takahashi, Y., Yasuhiko, Y., Kitajima, S., Kanno, J., & Saga, Y. (2007). Appropriate suppression of Notch signaling by Mesp factors is essential for stripe pattern formation leading to segment boundary formation. Developmental biology, 304(2), 593–603. doi:10.1016/j.ydbio.2007.01.007

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular biology and evolution, 28(10), 2731–2739. doi:10.1093/molbev/msr121

Tanaka, E. M., & Reddien, P. W. (2011). The cellular basis for animal regeneration. Developmental cell, 21(1), 172–185. doi:10.1016/j.devcel.2011.06.016

Tarailo-Graovac, M., & Chen, N. (2002). Current Protocols in Bioinformatics. (A. D. Baxevanis, G. A. Petsko, L. D. Stein, & G. D. Stormo, Eds.). Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/0471250953.bi0410s25

Thorpe, C. J., Weidinger, G., & Moon, R. T. (2005). Wnt/beta-catenin regulation of the Sp1-related transcription factor sp5l promotes tail development in zebrafish. Development, 132(8), 1763–1772. doi:10.1242/dev.01733

Tollis, M., & Boissinot, S. (2011). The transposable element profile of the Anolis genome: How a lizard can provide insights into the evolution of vertebrate genome size and structure. Mobile genetic elements, 1(2), 107–111. doi:10.4161/mge.1.2.17733

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics, 25(9), 1105–1111. doi:10.1093/bioinformatics/btp120

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., et al. (2012). Differential gene and transcript expression

analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols, 7(3), 562–578. doi:10.1038/nprot.2012.016

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology, 28(5), 511–515. doi:10.1038/nbt.1621

Vitt, L. J. (1981). Lizard Reproduction: Habitat Specificity And Constraints On Relative Clutch Mass. The American Naturalist, 117(4), pp. 506-514. http://www.jstor.org/stable/2460458

Wade, J. (1999). Sexual Dimorphisms in Avian and Reptilian Courtship: Two Systems that Do Not Play by Mammalian Rules. Brain, behavior and evolution, 54(1), 15–27. doi:10.1159/000006608

Weidinger, G., Thorpe, C. J., Wuennenberg-Stapleton, K., Ngai, J., & Moon, R. T. (2005). The Sp1-related transcription factors sp5 and sp5-like act downstream of Wnt/beta-catenin signaling in mesoderm and neuroectoderm patterning. Current biology : CB, 15(6), 489–500. doi:10.1016/j.cub.2005.01.041

White, P. H., & Chapman, D. L. (2005). Dll1 is a downstream target of Tbx6 in the paraxial mesoderm. Genesis (New York, N.Y. : 2000), 42(3), 193–202. doi:10.1002/gene.20140

White, P. H., Farkas, D. R., & Chapman, D. L. (2005). Regulation of Tbx6 expression by Notch signaling. Genesis (New York, N.Y. : 2000), 42(2), 61–70. doi:10.1002/gene.20124

William, D. A., Saitta, B., Gibson, J. D., Traas, J., Markov, V., Gonzalez, D. M., Sewell, W., et al. (2007). Identification of oscillatory genes in somitogenesis from functional genomic analysis of a human mesenchymal stem cell model. Developmental biology, 305(1), 172–186. doi:10.1016/j.ydbio.2007.02.007

Williams, A. H., Liu, N., van Rooij, E., & Olson, E. N. (2009). MicroRNA control of muscle development and disease. Current Opinion in Cell Biology, 21(3), 461–469. doi:10.1016/j.ceb.2009.01.029

Williams, E. E. (1983). Ecomorphs, faunas, island size, and diverse end points in island radiations of Anolis. Lizard ecology: studies of a model organism.

Xue, Y., Gao, X., Lindsell, C. E., Norton, C. R., Chang, B., Hicks, C., Gendron-Maguire, M., et al. (1999). Embryonic lethality and vascular defects in mice lacking the Notch ligand Jagged1. Human molecular genetics, 8(5), 723–730. doi: 10.1093/hmg/8.5.723

Yandell, M., & Ence, D. (2012). A beginner's guide to eukaryotic genome annotation, 1–14. doi:10.1038/nrg3174

Yin, V. P., Thomson, J. M., Thummel, R., Hyde, D. R., Hammond, S. M., & Poss, K. D. (2008). Fgf-dependent depletion of microRNA-133 promotes appendage regeneration in zebrafish. Genes & development, 22(6), 728–733. doi:10.1101/gad.1641808

Yu, Y.-M., Gibbs, K. M., Davila, J., Campbell, N., Sung, S., Todorova, T. I., Otsuka, S., et al. (2011). MicroRNA miR-133b is essential for functional recovery after spinal cord injury in adult zebrafish. European Journal of Neuroscience, 33(9), 1587–1597. doi:10.1111/j.1460-9568.2011.07643.x

Zhao, Y., & Srivastava, D. (2007). A developmental view of microRNA function. Trends in biochemical sciences, 34(4), 189–197. doi: 10.1016/j.tibs.2007.02.006

Zhao, Y., Samal, E., & Srivastava, D. (2005). Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. Nature, 436(7048), 214–220. doi:10.1038/nature03817

Zuckerkandl, E., & Pauling, L. (1965). Molecules as documents of evolutionary history. Journal of theoretical biology, 8(2), 357–366. doi:10.1016/0022-5193(65)90083-4