UNIVERSIDADE DE SÃO PAULO

FFCLRP - DEPARTAMENTO DE BIOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA COMPARADA

# Associations between *Hox* genes molecular evolution and the evolution of morphological diversity in Squamata and Marsupialia

## Associações entre a evolução molecular dos genes *Hox* e a evolução da diversidade morfológica em Squamata e Marsupialia

Sarah Ribeiro Milograna

Tese apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da USP, como parte das exigências para a obtenção do título de Doutora em Ciências, Área: Biologia Comparada.

Ribeirão Preto/SP

2015

UNIVERSIDADE DE SÃO PAULO

FFCLRP - DEPARTAMENTO DE BIOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA COMPARADA

# Associations between *Hox* genes molecular evolution and the evolution of morphological diversity in Squamata and Marsupialia

## Associações entre a evolução molecular dos genes *Hox* e a evolução da diversidade morfológica em Squamata e Marsupialia

Sarah Ribeiro Milograna

Orientadora: Profª. Drª. Tiana Kohlsdorf

Tese apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da USP, como parte das exigências para a obtenção do título de Doutora em Ciências, Área: Biologia Comparada.

VERSÃO CORRIGIDA

Ribeirão Preto/SP

2015

**Ficha Catalográfica**

Milograna, SR

Associations between *Hox* genes molecular evolution and the evolution of morphological diversity in Squamata and Marsupialia

202 p. : Il. ; 29,7 cm

Bibliografia: p. 183- 202

Tese de Doutorado apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto/USP – Área de Concentração: Biologia Comparada.

Orientadora: Kohldorf, Tiana.

1. *HOX* genes. 2. Evolution 3. Evolutionary developmental biology. 4. Bioinformatics 5. Snake 6.Amphisbaenia 7. Marsupial 8. *CsB*. 9. *Island I* 10.*CNS65* 11. Regulatory genes 12. Limbs 13.TFBS 14. Anterior-posterior body axis

To my parents and grandparents, my fundamental inspiration for all battles.

To Professor John McNamara, my scientific father.

To Jason Scealy, the co-author of my dreams and achievements.

**“It is not the strongest or the most intelligent who will survive but those who can best manage change.”**

-Leon C. Megginson

## SUMMARY OF CONTENTS

## ACKNOWLEDGEMENTS

I am very thankful to...

My supervisor Prof. Dr. Tiana Kohlsdorf, firstly, for her unconditional support regarding my carrier choices; she accepted my transference from another post-graduation program where I started my PhD and stimulated my international research internship. Thanks for having accepted and trusted me, and for always giving thoughtful advice in every step of this journey. Thanks for your strong example of hard, dedicated, passionate and tireless work, optimism and confidence. Thank you for your critical and insightful evaluation and editing in all reports, Thesis, manuscripts and presentations.

To the Post-graduation program in Comparative Biology, represented by the Coordinator Prof. Dr. Ricardo Macedo Corrêa and Castro and by Vice-Coordinator Prof. Dr. Tiana Kohlsdorf, for receiving me as a post-graduate student. To Prof. Márcia M. G. Bitondi for the careful revisions of my annual reports. To the secretary Vera Cassia Cicilini de Lucca, for all the kind and good-hearted support in administrative issues.

To the Biology Department of Faculdade de Filosofia, Ciências and Letras de Ribeirão Preto, USP, headed by Prof. Dr. Fernando Luis Medina Mantellato and by his surrogate Prof. Dr. Pietro Ciancaglini**,** and to all the helpful staff.

To Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP and for the financial support, by providing me PhD regular (process #2012/13165-5) and research internship abroad (#2014/06503-7) scholarships.

To the Herpetological Collection from the Laboratory of Tetrapod Evolution, the Yale Peabody Museum of Natural History of Yale University (New Heaven, EUA), and the authors of Vargas *et al.* (2008), for gently providing genomic DNA samples. To the Herpetological Collections at *Universidade Federal do Mato Grosso* (UFMT, Cuiabá, Brazil), *Universidade Federal do Alagoas* (MUFAL, Maceió, Brazil), and *Universidade de Brasília* (CHUNB, Brazil); to Museum of Science and Technology at *Pontifícia Universidade Católica do Rio Grande do Sul* (MCP, Porto Alegre, Brazil); Zoology Museum at USP (MZUSP, São Paulo, Brazil); and to *Universidade do Estadual de São Paulo* (UNESP, Rio Claro, Brazil) for providing tissues from Squamata for DNA extraction. Genomic DNA and tissues were used in the research project of Chapter I.

To Prof. Dr. Pedro E. M. Guimarães for having performed the Bioinformatics analyses of Transcription Factor Binding Sites prediction from Chapter I, using a pipeline

developed by himself, and for his insights on interpreting data in the light of their biological meaning. To Prof. Igor Schneider and Acácio Freitas, for contributing with the gene report expression assays that supported our data from Chapter I.

To Prof. Igor Schneider, for his sapient contribution with ideas and guidance through the experiments performed in the research project correspondent to the Chapter II. For kindly receiving me in his laboratory at Universidade Federal do Pará, providing all expertise and material to construct the vectors used in transgenic assays. To himself and his adorable family for kindly welcoming me in their home and taking me to taste hundreds of ice cream flavors made from tropical amazonic fruits. To Gabriela Lima, who dedicated several weeks to accompany me during the experiments.

To Prof. Dr. Marilyn Renfree and Prof. Andrew Pask for having welcomed me in their research groups at the University of Melbourne-Australia, and guided me to accomplish the research corresponding to Chapter III. For having assisted me despite the many challenges that they had to face through the period I was there, and with all the paperwork required to make this possible.

To the Department of Biosciences from the University of Melbourne, for having received me there as a visitor researcher, providing every resource needed for my internship.

To Dr. Hongshi Yu, for technical supervising, his dedicated guidance through analyses and experiments. To Helen Clark and Elisabeth Pharo, for their support in and out of the laboratory. To my colleagues from the Walgroup and Pask-lab, for all the help with animal manipulation, experiments and ideas. Specially, to Jennifer Hetz, Yu Chen and Danielle Hickford for patiently taking me through the qPCR technique, and to Karen Hansen for sharing her hardly-learned experience with in situ hybridization. To Brandon and Helen, for their help in a hard moment. Thanks to all of them for welcoming me, for the friendship and emotional support.

To Prof. Dr. John C. McNamara, my former academic supervisor, for his unconditional support on my carrier choices and friendship, for never having stopped being my mentor through all this years, and for his tireless editing of our manuscripts. He never stopped contributing to my academic development.

To Tania Defina and Andrea C. Quiapim for sharing their expertise with sequencing protocols and for having prepared themselves many of the samples sequenced for Chapter I.

To the lovely technician in the Laboratory of Tetrapod Evolution (LET), Paula dos Santos, for her amazing support with laboratory logistic and maintenance, and for having saved me a few times with paperwork support, especially when I was abroad.

To Prof. Dr. Tiago C. Pereira, Prof. Dr. Milton Groppo, Prof. Dr. John C. McNamara, Prof. Dr. Maria Helena S. Goldman, Prof. Dr. Maria L. T. M. Polizeli, Prof. Dr. Zilá L. P. Simões, and their students for having always kindly shared thermo-cyclers and -80°C freezers when we faced technical issues with our equipment.

To all my colleagues and friends from LET whom with I have been sharing space, ideas, opinions, experiences, problems, and life through all these years. To Fernando Andrade for having pioneered the study of *CsB* element in our laboratory, and to him and Aline C. Dragalzew for guiding me on my first steps in the laboratory.

To Mariana B. Grizante and Marina E. Singarette, for so much…to start with our endless discussions about snakelike evolution mysteries, experimental and logistical support, co-authorships and friendship. Thanks to Mariana for helping me to organize myself when I left for my sandwich PhD in Australia, and for kindly X-raying the snake and the amphibaenian from Fig 1.1 in Chapter I. Thanks to Marina for helping me with Brazilian paperwork when I was abroad. For all the love and a lifetime's sisterhood!

To my parents, for having prepared me with education, support, stimulus and unconditional love that prepared me to handle undreamed challenges. For always having believed more than myself in my potential. To my sister, Julia, for always leading by example with her incredible strength. To all my family and friends for a lot of stimulus.

To my partner Jason Scealy, for all his patience, strength, support and love through these years and especially when he was my family in Australia. To all his family that welcomed, supported and made me feel at home there.

Finally, thanks to all the people that directly or indirectly contributed to this study. Without these people, this PhD Thesis could have never became a reality.

## ABSTRACT

*Hox* genes pattern the vertebrate body during embryonic development, and understanding their evolution may unravel genetic mechanisms subjacent to morphological evolution. Molecular evolution of *Hox* genes entails signatures in regulatory regions that potentially affect gene expression, such as the *cis*-regulatory elements (CREs) that surround the HoxD cluster and its noncoding RNAs (ncRNAs). In this PhD Thesis, I have explored regulatory evolution of *HoxD* genes engaged in the development of appendicular and anterior-posterior body (AP) axes in amniotic lineages that exhibit homoplastic morphological peculiarities: snakelike squamates (Chapters I and II) and diprotodontid marsupials (Chapter III). In Chapter I, I investigated in snakes and amphisbaenians, whether equivalent regulatory signatures were registered in the Conserved Sequence B (*CsB*), a centromeric *Hoxd10-13* CRE, during evolution of snakelike morphologies. Using lizards and other tetrapods to represent the lacertiform morphology, conserved regions within *CsB* were sequenced from 38 squamate species, and transcription factor binding sites (TFBS) were predicted and compared among groups. Both snakelike lineages carry divergent and convergent regulatory signatures not identified in lizards; the convergence located in one *CsB* segment comprised loss of limb-related TFBS and gain of a binding site for PBX1. This convergent regulatory signature registered along two independent processes of snakelike evolution may relate to body elongation and limb loss, and evidences a role of *CsB* for AP axis development. In Chapter II, I investigated whether a telomeric (*CNS65*) and a centromeric (*Island I*) *Hoxd* enhancer that regulate gene expression respectively at proximal and distal regions of developing limbs retain their regulatory capacities in Serpentes. Gene reporter expression of these CREs from snakes were performed in transgenic mice and revealed that their regulatory activities were abrogated in limb buds. Comparison of predicted TFBS in these elements between snakes and limbed tetrapods revealed limb-related TFBS apparently lost in snakes, and pointed to one potential stilopodium/zeugopodium-specific element in *CNS65* and three likely autopodium-specific elements in *Island I*. Limb loss in snakes registered signatures in *Hoxd CREs* that possibly contributed for their functional impairment, putatively indicating limb-specific modules. Finally, in the chapter III, I studied ncRNAs from HoxD cluster in the context of hindlimb morphological evolution and heterochrony between fore and hindlimb development in the tammar wallaby *Macropus eugenii.* The ncRNAs mapped to HoxD cluster were selected from transcriptome performed using tammar embryo limbs at days 23 (d23) and 25 (d25) of

pregnancy, and their conservation, transcriptional profiles and expression patterns were explored. Comparison with orthologous sequences in other mammals revealed five ncRNAs conserved among mammals, and three transcripts apparently exclusive to marsupials. Transcriptional profiles of *HOXD10-13* and HoxD ncRNAs were mostly equivalent. *XLOC46* expression patterns resembled those of mouse and tammar terminal *HOXD* genes, whereas *XLOC52* and *XLOC53* showed identical expression patterns to those genes of tammar, except for *XLOC53* low expression at d25. The ncRNAs intergenic/intronic to *HOXD9-12* may regulate expression of terminal *HOXD* genes in mammals, and *XLOC52* and *XLOC53* are suitable for investigation regarding limb evolution in marsupial. This PhD Thesis demonstrates how studies of evolutionary footprints in regulatory elements of developmental genes contribute for elucidating specific processes during lineages divergence as well as functional aspects of these genes during development.

## RESUMO

Os genes *Hox* padronizam o corpo dos vertebrados durante o desenvolvimento embrionário, e a compreensão de sua evolução pode elucidar mecanismos genéticos subjacentes à evolução morfológica. A evolução molecular dos genes *Hox* imprime assinaturas em regiões regulatórias, as quais potencialmente afetam sua expressão gênica, como os elementos *cis*-regulatórios (CREs) que ladeiam o cluster D de *Hox* e seus RNAs não-codificantes (ncRNAs). Essa Tese de Doutorado enfoca a evolução regulatória de genes *HoxD* envolvidos no estabelecimento dos eixos corpóreos axial ântero-posterior (AP) e apendiculares em linhagens de aminiotas que exibem características morfológicas homoplásticas peculiares: os squamatas serpentiformes (Capítulos I e II) e os marsupiais Diprotodontia (Capítulo III). No Capítulo I investigou-se, em serpentes e anfisbênias, se assinaturas regulatórias envolvidas no estabelecimento das morfologias serpentiformes foram impressas na Sequência Conservada B (*Conserved Sequence B*, *CsB*), um CRE centromérico de *Hoxd10-13*. Usando lagartos e outros tetrápodes como referência para a morfologia serpentiforme, regiões conservadas de *CsB* foram sequenciadas em 38 espécies de Squamata, cujos TFBS foram preditos e comparados. Ambas linhagens serpentiformes exibem assinaturas regulatórias divergentes e convergentes ausentes em lagartos; a convergência localizou-se em um segmento de *CsB* que concentra perda nas linhagens serpentiformes de diversos TFBS com funções no desenvolvimento de membros e a aquisição de um sítio de ligação para PBX1. Essa assinatura convergente impressa durante evoluções independentes da morfologia serpentiforme pode estar relacionada à elongação corpórea e à perda dos membros, evidenciando um papel do *CsB* no desenvolvimento do eixo AP. No Capítulo II, foi investigado se um CRE telomérico (*CNS65*) e um centromérico (*Island I*) de *Hoxd*, os quais regulam respectivamente regiões proximais e distais dos membros tetrápodes em desenvolvimento, retêm suas capacidades regulatórias em Serpentes. Expressões de gene repórter desses CREs de serpentes foram realizadas em camundongo transgênico, revelando deficiência de suas atividades regulatórias nos brotos de membro. A comparação dos TFBS preditos nesses elementos entre serpentes e outros tetrápodes revelou que TFBS relacionados ao desenvolvimento dos membros foram perdidos nas sequências das serpentes. Ainda, essa comparação indicou um elemento em *CNS65* potencialmente envolvido especificamente na regulação da formação de estilopódio/zeugopódio, e três elementos na *Island I* exclusivamente reguladores do desenvolvimento autopodial. A perda de membros em

serpentes aparentemente imprimiu assinaturas nesses *CREs* de *Hoxd* que possivelmente contribuíram para sua degeneração funcional, putativamente indicando módulos específicos de regulação nos membros. No Capítulo III, ncRNAs do cluster D de Hox foram estudados no contexto da evolução morfológica do autopódio posterior e heterocronia entre o desenvolvimento de membros anteriores e posteriores em *Macropus eugenii.* Os ncRNAs mapeados sobre o cluster D de Hox foram selecionados a partir de transcritoma de membros de embriões de *M. eugenii* nos dias 23 (d23) e 25 (d25) de gravidez, e sua conservação, perfis transcricionais e padrões de expressão foram explorados. A comparação com sequências ortólogas de outros mamíferos revelou cinco ncRNAs conservados em mamíferos, e três aparentemente exclusivos dos marsupiais. Os perfis transcricionais de genes *HOXD10-13* e dos ncRNAs do cluster D de Hox foram predominantemente equivalentes. Os padrões de expressão de *XLOC46* foi similar aos dos genes *HOXD* terminais de camundongo e *M. eugenii*, enquanto que *XLOC52* e *XLOC53* apresentaram expressão idêntica à desses genes em *M. eugenii*, exceto pela baixa expressão de *XLOC53* no d25. Os ncRNAs intergênicos/intrônicos aos genes *HOXD9-12* possivelmente regulam a expressão de genes *HOXD* terminais em mamíferos, enquanto que *XLOC52* e *XLOC53* constituem bons candidatos para investigação relacionada à evolução dos membros de marsupiais. Esta Tese demonstra como estudos de assinaturas regulatórias na evolução de genes do desenvolvimento contribuem para o entendimento das histórias evolutivas de divergência entre linhagens e dos aspectos funcionais desses genes do desenvolvimento.

# *GENERAL BACKGROUND*

The present work corresponds to a PhD Thesis conceived in the Evolutionary Developmental Biology (EvoDevo) framework. The Thesis comprises three chapters focusing correlations between *HoxD* genes regulatory evolution and the evolution of morphological variation in tetrapods. In summary, I have explored whether and to what extent the evolution of snakelike and autopodial morphologies leaves footprints in diverse regulatory elements located externally and intergenically to HoxD cluster. The present section aims to provide a general theoretical background that introduces some concepts that permeate the three chapters, together with an explanation about the connections between *HoxD* gene expression during embryo development and the establishment of the adult phenotype. The section is divided in two items: first (named '*Hox genes roles on embryonic development*'), is a sub-section permeating *Hox* genes organization, functions and evolutionary mechanisms, where their roles and regulation on appendicular and axial axes are also introduced. Then, a second item (named '*Mechanism of morphological evolution associated with Hox regulatory systems'*) is dedicated to detail the scenarios where associations of morphological and *Hox* molecular evolution were investigated. In the end of this theoretical background, the major goal shared by the three chapters of this Thesis was explained.

*Hox genes roles on embryonic development*

*'Hox'* is the name attributed to a group of several genes that comprise extremely conserved DNA-binding domains, termed homeobox, and operate as transcription factors regulating key developmental pathways widespread through vertebrate and invertebrate embryos (Lewis 1978; Carroll 1995; Spitz 2010), maintaining also several important functions in adults (Taylor et al. 1997; James & Kazenwadel 1991; Chen & Capecchi 1999; Takahashi

et al. 2004). In genomes of the prototypic Vertebrata and Cephalochordata, these genes were named after their relative positions in clusters, which usually reflect their patterns of spatiotemporal collinear expression in specific gradients of anterior-posterior axial (AP) and appendicular body axes during embryonic development (Carroll 1995; Spitz 2010). This collinear expression, initially described in model system organisms such as mice and drosophila, originated the concept of "Hox code" (Kessel & Gruss 1991), later improved after discovery of the "posterior prevalence" property (Duboule & Morata 1994; Duboule 2007), explaining how positional information is provided by *Hox* genes to axial and paraxial tissues, establishing the identity of several structures (Deschamps & Van Nes 2005; Spitz 2010). Along evolutionary history, *Hox* genes in ancestral vertebrates first encompassed one single cluster, which after two rounds of whole genome duplications resulted in four Hox clusters (HoxA to HoxD) that can be identified in most extant vertebrates (Panopoulou & Poustka 2005; Peer et al. 2009) except for bony fish, a lineage that experienced one additional genome duplication and exhibits eight Hox clusters (Crow & Wagner 2006). Hox clusters are spread among distinct chromosomes that vary according to the species, and yet, at least in most vertebrates, they conserve their typical arrangement inside each particular cluster, which in these animals are classified as 'organized clusters' or 'type O clusters' (Lemons & McGinnis 2006; Duboule 2007). *Hox* gene expression is fine-tuned by complex regulatory networks that favor their co-option in diverse contexts, and the similarity of developmental pathways by them weaved does not prevent highly specific functional outputs (Zakany & Duboule 2007). Such features characterize a robust but flexible genic system that allows evolutionary mechanisms to succeed in a way that little variation in *Hox* genes nucleotide sequences and on their interaction with downstream targets may elicit phenotypic changes (Carroll 2005; Montavon et al. 2011; Guerreiro et al. 2013). Consequently, *Hox* genes become very seductive for

scientists trying to elucidate evolution of phenotypic diversity and its relationships with embryonic development.

In vertebrate embryos, *Hox* genes are essential for patterning structures across the AP body axis, which is composed by mesodermal modules, the somites, sequentially established in both sides of the neural tube during embryonic development in a process that constitutes the axial skeleton and its associated muscular apparatus, contributing also with cranial elements (Deschamps & van Nes 2005; Woltering 2012). Terminal *Hox* genes, which are also frequently named 'distal *Hox* genes', comprise those genes located closer to 5' positions in their respective clusters, which are transcription factors often crucial for patterning structures in posterior domains of embryo AP body axis (Roberts et al. 1995; Yokouchi et al. 1995; Warot et al. 1997). Examples of structures located in such posterior domains are the caudal somites, posterior regions of the notochord, neural tube and intestine, the cloaca and external genitalia (Griffith et al. 1992; Gajovic et al. 1993; Catala et al. 1995; Knezevic et al. 1998; Santa Barbara & Roberts 2002; Scott et al. 2005).

In addition to establishing the AP body axis, *Hox* genes contribute to the induction and patterning of limb buds, as well as other appendices in non-tetrapod vertebrates (Archambeault et al. 2014), which originate as a small protraction from the body wall (Capdevila & Belmonte 2001). During limb development, two main embryonic regions are established in this structure, the apical ectodermal ridge (AER), an inductor of proximal-distal growth, and the zone of polarizing activity (ZPA), which patterns the stylopodium (humerus and femur), the zeugopodium (radius/ulna and tibia/fibula) and the autopodium (carpi/tarsi and digits) (Fromental-Ramain et al. 1996), as illustrated in Figure 0.1. During development of the structures aforementioned, terminal *HoxA* and *HoxD* genes are collinearly expressed, assuming domains that are equivalent but not identical to the expression of these genes during trunk formation and extension (Zakany & Duboule 2007). Transcription of *HoxD* genes

during development of proximal and distal limb segments occurs in two independent phases (Nelson et al. 1996; Tarchini & Duboule 2006). The initial transcription phase determines stylopodium and zeugopodium patterning from sequential and homogeneous activation of *HoxD1-9* genes (Zakany & Duboule 2007; Montavon et al. 2011). Development of the most distal region in the limb is regulated by the expression of *HoxD10-13* genes, which occurs in inverted spatiotemporal collinearity up to the final stages of autopodial development, configuring the second *HoxD* transcriptional wave (Fromental-Ramain et al. 1996; Montavon et al. 2008; Montavon et al. 2011).

The two transcriptional phases described during limb development are regulated by several elements, which have not been completely elucidated yet, among which are *cis*-regulatory elements (CRE), noncoding RNAs (ncRNAs) (Lemons & McGinnis 2006; Soshnikova 2014) and IRES (viral internal ribosome entry sites)-like elements (Xue et al. 2015). Among these, some elements have an internal location (the intergenic *cis*-regulatory elements), whereas other ones are located externally (global *cis*-regulatory elements) to Hox clusters (Di-Poï et al. 2010; Montavon et al. 2011). The DNA segments engaged in several regulatory activities of long range, interacting with *Hox* genes from inside a remote regulatory complex, characterize the so-called global control regions (GRCs), which define chromosomic regulatory landscapes (Deschamps 2007; Spitz 2010). In these domains, genes share regulatory specificities (Spitz et al. 2003; Zuniga et al. 2004) that coordinate gene expression in different developing structures (Duboule 1998; Spitz et al. 2001; Spitz et al. 2005). The HoxD cluster encompasses GCRs which regulatory mechanism in traditional experimental organisms (e.g. mouse) are so far the best understood ones among *Hox* genes. The collinear proximal-distal transcription of *HoxD* genes during limb development experiences a transition between two opposed regulatory landscapes, one located centromeric and the other telomeric from the cluster (see Fig. 0.1), which partially overlap but correspond

to two independent topological domains (Dixon et al. 2012; Nora et al. 2012; Andrey et al. 2013). Each *HoxD* gene exhibits its specific tropism in the interaction with these domains, which depends of their relative genomic position in the cluster (Andrey et al. 2013). In the three chapters of this PhD Thesis I have investigated evolutionary aspects of the *HoxD* cluster regulatory components in the context of morphological evolution, as detailed in the following section.

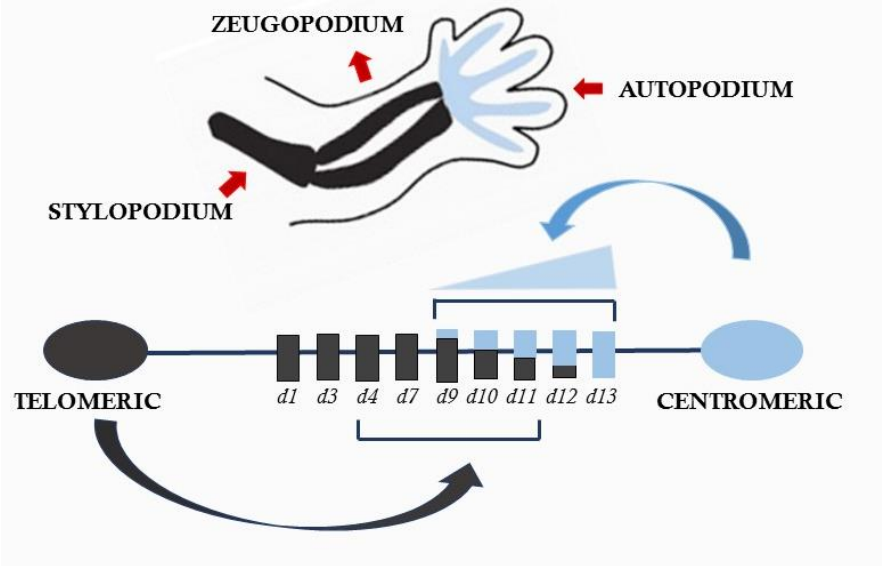


**Figure 0.1.** Mouse *Hoxd* genes are regulated by transcriptional enhancers located remotely from genes (in gray and blue), at opposed sides of the cluster. Enhancers that regulate initial transcription in proximal limb domains (gray) are located in the telomeric landscape of mouse chromosome 2, whereas enhancers regulating late expression at distal domains (blue) are located in a centromeric position. Adapted from Deschamps & van Nes (2005).

*Mechanism of morphological evolution associated with Hox regulatory systems*

The present PhD Thesis approaches regulation of *HoxD* expression in the context of morphological evolution using two chordate clades: Squamata (Sauria, Tetrapoda; chapters I and II) and Metatheria (Mammalia, Tetrapoda; chapter III). Squamata comprises lizards, snakes and amphisbaenians (Pough 1998), and represents a very intriguing biological system

for disclosing associations between molecular evolution and morphological divergence due to the recurrence in multiple lineages of remarkable morphological transitions involving limb reduction (and complete loss) associated with trunk elongation (Wiens & Slingluff 2001; Caldwell 2003). Several origins of snakelike morphologies occurred in this group, many of them associated with the occupancy of new environments imposing selective pressures that weaved or alternatively benefited from changes in developmental pathways (Coates & Ruta 2000; Wiens & Slingluff 2001; Woltering et al. 2009; Woltering 2012). However, *which* were these crucial changes in developmental pathways that produced new phenotypes, and *how* or *where* they have been settled in the genomes of these species, are issues that have sharpened the curiosity of scientists for decades, nurturing intense debate. In this context, it has been claimed that shifts in *Hox* gene expression patterns and associated downstream developmental networks might be strong candidates for connecting development with snakelike evolution, due to their consolidated roles in patterning both axial and appendicular axes (Cohn & Tickle 1999; Woltering et al. 2009; Di-Poï et al. 2010; Woltering 2012; Guerreiro et al. 2013), although evidences are still controversial (see Sanger & Gibson-Brown 2004; Head & Polly 2015).The anterior expansion of *HoxC6* and *HoxC8* gene expression domains along python paraxial and lateral plate mesoderm has been accepted for a long time as a developmental explanation for concomitant forelimb loss and trunk elongation (Cohn & Tickle 1999), although recent evidence suggests that functional changes in primaxial *Hox* genes did not cause axial de-regionalization in snakes (Head and Polly 2015). Other authors, conversely, suggest that alternative downstream interpretations of *Hox* codes in the pre-caudal region seems to determinate vertebral de-regionalization in caecilians and snakes (Woltering et al. 2009), which in snakes led to expansion of the rib cage (Guerreiro et al. 2013).

Beyond interest in the causal mechanisms that establish snakelike morphologies, questions concerning the evolution of genes essential for development of vertebrate structures

that no longer exist in snakelike lineages (i.e. the limbs) remain relegated in EvoDevo studies. The scarce information available for evolution of *Hox* genes in snakelike squamates is restricted to Serpentes and often focus on coding sequences or gene expression patterns (Cohn & Tickle 1999; Di-Poï et al. 2010; Woltering 2012; Head & Polly 2015; but see Singarette et al. 2015 for some exceptions), so comparisons of developmental processes among snakelike squamate lineages are practically inexistent, especially considering non-coding regions. In the first two chapters of this PhD Thesis (Chapters I and II), I have investigated molecular signatures of snakelike morphologies in *Hoxd* regulatory elements; I did not seek particularly for cause-effect correlations between molecular and morphological evolution. Because *Hox* genes are involved in multiple functions (Mallo et al. 2010; Montavon et al. 2011; Archambeault et al. 2014) and changes in their coding regions likely involve pleiotropy (Sivanantharajah & Percival-Smith 2015), variation in expression patterns associated with morphological evolution seem to concentrate on evolution of CREs (Warren et al. 1994; Averof & Akam 1995; Carroll 1995; Burke et al. 1995). Mechanisms of CRE evolution involve mutational loss and loss in the number, affinity and topology of transcription factor binding sites (TFBS; see Jeong et al. 2006; Wang & Chamberlin 2002; Gompel et al. 2005), besides the co-option of TFBS in pre-existing CREs (Wang & Chamberlin 2002; Gompel et al. 2005). Even single insertions or deletions among TFBS have the potential to affect function of regulatory elements and expression levels of genes regulated by them (Zinzen et al. 2006; Carroll 2008). Based on this scenario, in Chapter I, I have investigated whether independent evolution of snakelike morphologies in Serpentes and Amphisbaenia l mmolecular signatures in the Conserved Element B (*CsB*), a *HoxD* centromeric regulatory region. In Chapter II, I have evaluated whether elements that regulate *HoxD* gene expression [one centromeric (*Island I*) and one telomeric (*CNS65*) element] and are specially relevant for tetrapod limb

development retained in limbless organisms (snakes) their functional ability to regulate gene expression in limb buds.

The second taxonomic group used to evaluate regulation of *HoxD* expression in the context of morphological evolution was Metatheria. Macropodidae marsupials, such as the tammar wallaby *Macropus eugenii*, exhibit several peculiar limb characteristics in comparison with eutherian mammals. Their limb differences appear further beyond the adult morphology, being also registered in the heterochrony observable during development of forelimbs and hindlimbs. Forelimbs develop considerably early and allow the young newborn to climb its mother´s body right after birth, encountering the pouch by himself while still blind and holding fetal hindlimbs (Sears 2009; Keyte & Smith 2010; Chew et al. 2014). Fascination does not end here because, after birth, hindlimb development surpasses forelimb growth, and they establish as very long and strong limbs specialized for hopping; yet, they present a digital formula that is unique to macropodids but includes an homoplastic component (Chew et al. 2012). Syndactylous, reduced and skin-bound digits II and III have evolved independently in two marsupial orders (Nilsson et al. 2004; Phillips et al. 2006), recapitulating a similar scenario to the aforementioned recurrent evolution of snakelike morphologies. However, information about the evolution of developmental genetic systems underlying the origin of such heterochronic processes and the establishment of these morphologies remains scarce for Macropodidae marsupials (Chew et al. 2012; Deakin 2012). Therefore, in the third chapter of this PhD Thesis (Chapter III) I have examined ncRNAs from the *HoxD* cluster present in *M. eugenii* developing limbs, aiming to disclose unknown regulatory elements that potentially emerged in the marsupial lineage that might have contributed to evolution of their peculiar limb morphological and developmental characteristics.

The core of this PhD Thesis settles on unraveling links between molecular evolution of *HoxD* cluster regulatory regions and the evolution of morphological diversity using two vertebrate groups as targets, squamates and marsupials.

# *CHAPTER I*

The independent evolution of snakelike squamate lineages registered both convergent and divergent regulatory signatures in the HoxD cluster enhancer *CsB*.

## 1.1 ABSTRACT

Snakelike morphologies evolved independently several times within Squamata. Evolution of morphological diversity is often coupled with variation in developmental patterns of *Hox* gene expression, which may reside in nucleotide substitutions at regulatory regions. The genes *Hoxd10-13* are particularly relevant in this evolutionary context because they play essential roles during the development of structures remarkably modified in snakelike lineages: autopodia and axial anterior-posterior (AP) axis. We compared between Serpentes and Amphisbaenia the molecular signatures of a regulatory *Hoxd* element [Conserved Sequence B (*CsB*)], using lizards and other tetrapods as reference for the lacertiform morphology. Conserved regions within *CsB* (*CP1* [725 bp] and *CP2* [423 bp]) were sequenced from 38 squamate and 1 alligator species; our database was complemented by sequences from GenBank for chicken, turtle and mammals. Transcription factor binding sites (TFBS) were predicted by a pipeline that merges binding sites preferences to functional and evolutionary clustering. Variation within Squamata concentrated within *CP2*, whereas differences between squamates and non-squamates appeared also in *CP1*. Regulatory haplotypes of amphisbaenians seemed more similar to those of other lizards than to those of snakes. However, both snakelike lineages shared specific regulatory signatures not identified in lacertiform lizards, which were located in a short segment of *CP2* (initial positions 841-849 in our alignment) comprised by: 1) lack of TFBS for C-MAF, HIC1, NEUROD or TTF-1, AP-4 and XPF-1 or LBP-1, and 2) replacement at position 846 of a TFBS for HIC1, present in lizards, by one for PBX1. Molecular evolution of *CP2* in Squamata left convergent regulatory signatures during the independent origins of Serpentes and Amphisbaenia that may relate to body elongation and limb loss, and suggests a possible function of this element on AP axis in the developing vertebrate embryo, so far undescribed.

## 1.2 INTRODUCTION

Independent origins of equivalent morphologies are identified when a given phenotype is observed in relatively close taxonomic lineages but is absent in representatives of the group including their most recent common ancestral (Hall 2012). These origins may represent either parallel (based on the same developmental genetic mechanisms) or convergent (based on distinct developmental genetic mechanisms) processes (Wake et al. 2011). Patterns of homoplastic evolution are driven by specific mechanisms, such as natural selection, which depend on their respective evolutionary context (Grant 1999; Rundle et al. 2000). However, equivalence in a given morphological transition repeated along different lineages might have been settled through similar changes in developmental pathways (Sucena et al. 2003; Shapiro et al. 2006; Prud´homme et al. 2006; Abouheif 2008). In a scenario where the developmental bases of morphological similarity remain partially obscure, referring to such processes as 'independent origins of equivalent morphologies' is more accurate than classifying phenotypes as 'convergent' or 'parallel'. This consideration does not apply however to the genetic level, where nucleotide signatures that are common or exclusive to compared lineages may be respectively referred to as 'convergent' or 'divergent' molecular patterns. The conceptual counterpoint between these two biological levels is particularly relevant given that possible developmental bases of independent origins of equivalent morphologies have been poorly explored using empirical data, so the correspondence between phenotypic and genetic transitions during processes resulting in multiple independent origins of equivalent morphologies remains uncertain.

Among several examples of independent evolution of equivalent phenotypes, the multiple origins of snakelike morphologies within Squamata have received considerable attention in the past years. The snakelike body shape is characterized by an elongated body axis composed of 'de-regionalized' and numerous vertebrae (when compared to lacertiform

morphologies; see (Woltering 2012) and reduced or absent limbs (Wiens & Slingluff 2001; Caldwell 2003). In Squamata, snakelike morphologies evolved independently from lacertiform ancestors at least 26 times (Shapiro et al. 2003; Wiens et al. 2006), being observed in the families Amphisbaenidae, Anguidae, Cordylidae, Dibamidae, Gerrhosauridae, Gymnophthalmidae, Pygopodidae, Scincidae and in the Serpentes clade (Wiens et al. 2006; Uetz 2007). In some lineages the evolution of this conspicuous morphology has occurred more than once (Pellegrino et al. 2001; Brandley et al. 2008), and it is widely accepted that the snakelike phenotype has evolved independently in snakes and amphisbaenians (Vidal & Hedges 2005; Vidal & Hedges 2009; Conrad 2008; Losos et al. 2012; Pyron et al. 2013), as illustrated in Figure 1.1. During these multiple transitions from lacertiform to snakelike morphologies, body elongation is recurrently associated with limb reduction, which suggests the presence of developmental constraints acting on limb development in the elongated taxa (Woltering 2012). Such constraints likely involve *Hox* genes expression because these pattern the anterior-posterior (AP) axis in the developing vertebrate embryo, and especially those from clusters A and D were apparently co-opted during the evolution of limb (Duboule 1992; Zákány & Duboule 1999; Tschopp & Duboule 2011) and genitalia (Dollé et al. 1989; Zákány & Duboule 1999; Tickle 2006; Spitz 2010; Tschopp et al. 2014) in tetrapods.

As mentioned in the General Background section, *Hoxd* genes are essential for development of both AP body axis (caudal regions of somites, neural tube, notochord, intestine, cloaca and external genitalia) (Griffith et al. 1992; Gajovic et al. 1993; Catala et al. 1995; Knezevic et al. 1998; Santa Barbara & Roberts 2002; Scott et al. 2005; Mallo et al. 2010) and limbs (Capdevila & Belmonte 2001; Zakany & Duboule 2007). Among these genes,

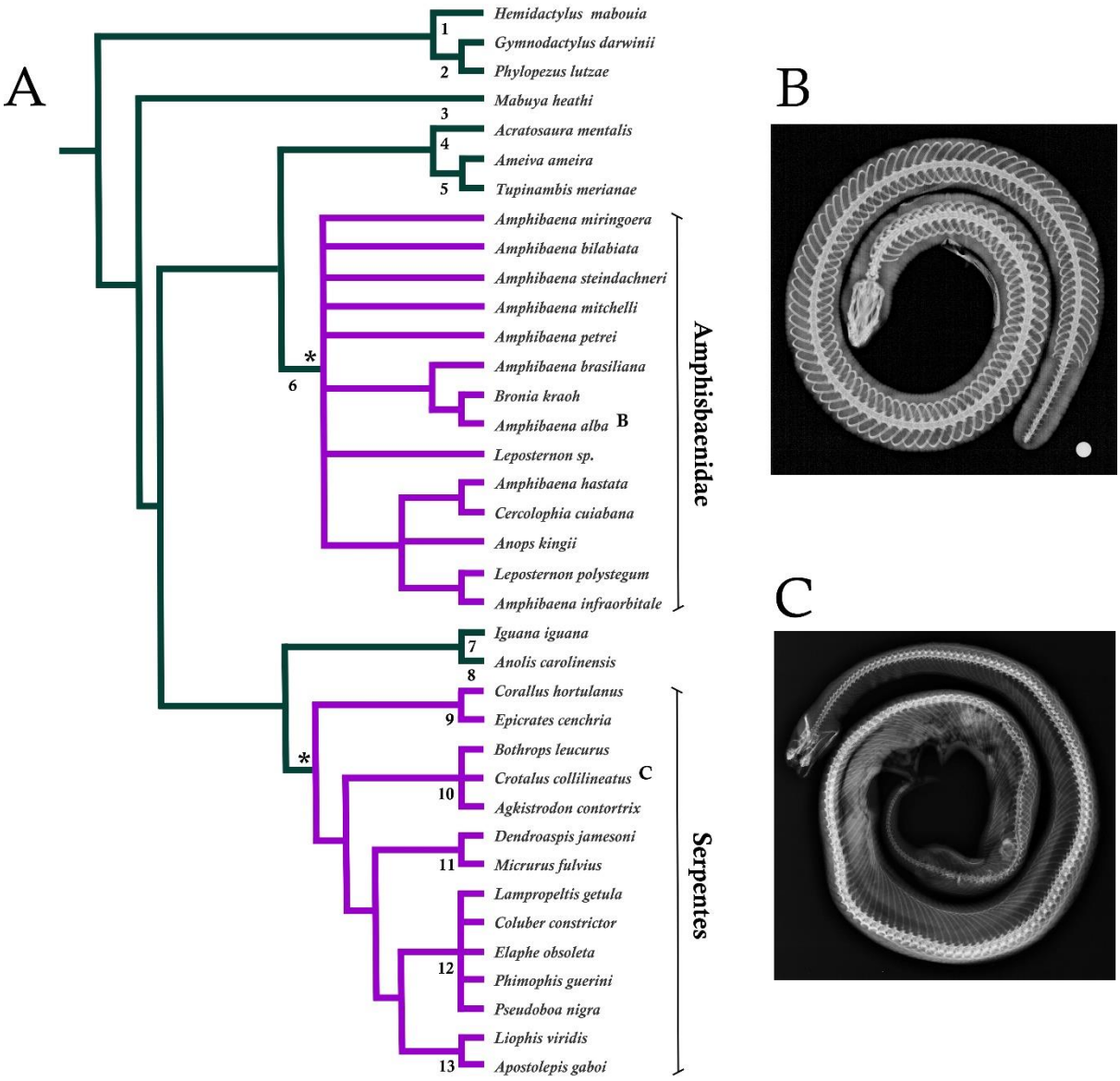


**Figure 1.1.** Squamata phylogenetic relationships and the snakelike morphology. (A) Topology illustrating phylogenetic relationships among the squamate taxa from which *CsB* fragments were sequenced, according to Pyron et al. (2013) and Mott & Vieites (2009), coupled with x-rays illustrating the similar morphologies of amphisbaenians (B), represented by *Amphisbaena alba*, and snakes (C), represented by *Crotalus collilineatus*. In (A) the asterisk highlights snakelike lineages, represented by purple branches. Green branches indicate lineages with a lacertiform morphology. Numbers correspond to squamate families as follows: 1. Gekkonidae, 2. Phyllodactylidae, 3. Scincidae, 4. Gymnophthalmidae, 5. Teiidae, 6. Amphisbaenidae, 7. Iguanidae, 8. Dactyloidae, 9. Boidae, 10. Viperidae, 11. Elapidae, 12. Colubridae, 13. Dipsadidae. X-rays gently given by Mariana Bortoletto Grizante.

terminal ones from the HoxD cluster represent good candidates for a genetic correspondence with phenotypic transitions during the independent origins of snakelike morphologies because they are involved in the establishment of both the AP axis and the limbs. During development of the axial skeleton, *Hoxd10* is expressed in mouse lumbar and sacral vertebrae and triggers interruption of ribs formation, inducing thoracic-lumbar transition, whilst *Hoxd11* is involved in the lumbar-sacral transition (Davis & Capecchi 1994; Burke et al. 1995; Wellik & Capecchi 2003; Carapuço et al. 2005). Two other genes, *Hoxd12* and *Hoxd13*, are expressed in mouse caudal vertebrae, the latter being involved in the mechanism that terminates somitogenesis, which modulates the final number of caudal vertebrae (Economides 2003; Kmita & Duboule 2003; Soshnikova & Duboule 2009). Limb development also involves *Hoxd* gene expression (Zakany & Duboule 2007; Nelson et al. 1996; Tarchini & Duboule 2006), and the same terminal genes aforementioned (*Hoxd10, Hoxd11, Hoxd12* and *Hoxd13*) are essential for canonical autopodium and zeugopodium formation (Tickle 2006; Montavon et al. 2008; Montavon et al. 2011).

Expression of *Hox* genes in vertebrates is mainly regulated by *cis-trans* transcriptional mechanisms, although a translational control is also involved (Kondrashov et al. 2011; Montavon & Soshnikova 2014). The regulatory machinery modulating expression patterns of terminal *Hoxd* genes during limb development has been very well characterized in the mouse: a landscape located telomeric to HoxD cluster regulates gene expression during stylopodium and zeugopodium formation (Montavon et al. 2011; Andrey et al. 2013), whereas autopodium development relies on a second transcriptional wave mastered by a centromeric regulatory landscape that coordinates *Hoxd13* to *Hoxd10* inverted collinear expression (Tickle 2006; Montavon et al. 2008; Montavon et al. 2011). This centromeric regulatory landscape comprises several enhancers which complex interaction modulates expression of terminal *Hoxd* genes in the embryo in the so-called 'distal phase' (DP) expression pattern. This pattern

has been mostly established based on developmental processes of digit formation (Tickle 2006; Montavon et al. 2008; Montavon et al. 2011), but it was recently shown to regulate several other distally elongated structures in vertebrates, representing an ancestral feature co-opted to pattern diverse body plan features (Archambeault et al. 2014). The Conserved Sequence B (*CsB*), in particular, belongs to a global control region (GCR; see (Spitz et al. 2003; Gonzalez et al. 2007) that regulates *Hoxd10-13* genes expression during development of autopodia, external genitalia and the dorsal neural tube (Spitz et al. 2003; Gonzalez et al. 2007; Schneider & Shubin 2013). Among vertebrates, three domains within *CsB* (hereafter described as conserved peaks *CP1*, *CP2* and *CP3*) are extremely conserved (Gonzalez et al. 2007; Schneider & Shubin 2013), and the first two encompass tetrapod-specific elements (nominated *B1* and *B2* by (Gonzalez et al. 2007).

Comparative analyses focusing on the *CsB* element seem especially promising for evaluating the likelihood of molecular convergence coupled to the independent origins of snakelike morphologies because the relevance of this region for *Hoxd* gene expression has been already well characterized, and there is indication of variation in the expression patterns of *Hox* genes during snake development (e.g. Di-Poï et al. 2010; Woltering 2012). For example, *Hoxd10* and *Hoxd11* are expressed more posteriorly in the AP body axis of snakes than in mouse and lizard, a pattern likely related to the elongated trunks lacking the transition from rib-less to rib-bearing dorsal vertebrae regions (equivalent to thoracic-lumbar transition in mammals; see (Dequéant & Pourquié 2008; Woltering et al. 2009; Di-Poï et al. 2010; Woltering 2012). Moreover, *Hoxd12* gene is not found in the genome of snakes whereas *Hoxd13* expression is weak in the tail region, opposed to their strong expression in the lizard tail (Di-Poï et al. 2010); such differences apparently contribute to a delay in somitogenesis termination and likely contribute to tail elongation in Serpentes (Burke et al. 1995; Di-Poï et al. 2009; Woltering 2012). Despite recent discussions regarding developmental mechanisms

underlying the evolution of a snakelike body shape (e.g. (Cohn & Tickle 1999; Kohlsdorf & Wagner 2006; Di-Poï et al. 2009; Woltering 2012; Head & Polly 2015), to our knowledge the topic remains unexplored under a comparative phylogenetic framework where regulatory footprints are addressed in the scenario of independent origins of snakelike morphologies in Squamata [but see Singarete et al. (2015) for a recent study on protein evolution]. The present study aims to establish this connection between genotypic and phenotypic dimensions in the context of multiple origins of snakelike morphologies. Specifically, I have performed comparative analyses of a *CsB* fragment comprising the first two conservation peaks (*CP1* and *CP2*), and focused on regulatory haplotypes that are shared by all species of Serpentes or Amphisbaenia evaluated and *exclusive* to these respective groups (hereafter referred to as 'divergent signatures') and those that are present in all species of both lineages but absent in lacertiform species (hereafter referred to as 'convergent signatures'). I tested the following hypothesis: 1) some regulatory signatures in the *CsB* fragment regarding predicted transcription factor binding sites (TFBS) would be exclusive to Serpentes or to Amphisbaenia, a pattern that likely reflects the independent evolutionary histories of these two lineages; 2) concomitantly, it would be possible to identify equivalent regulatory signatures in the *CsB* that seem coupled to the independent evolution of snakelike morphologies in these two lineages, providing evidence for possible developmental constraints acting during the different origins of such phenotypes. Both divergent and convergent regulatory signatures within *CP2* fragment in snakes and amphisbaenians were found, suggesting that they may have been registered by axial elongation and limblessness evolution. The results obtained here were complemented with a functional dissection of *CsB* through mice transgenic assays, performed by Prof. Dr. Igor Schneider, to explore whether the fragment analyzed here would alone cover the regulatory capacity held by the entire *CsB* element in limbs and AP axis.

## 1.3 METHODS

*Sequencing CsB fragments*

Fragments of *CsB* were amplified and sequenced from 14 species of Serpentes, 14 species of Amphisbaenia, 9 lizard species and one alligator (see Fig. 1.1). Voucher numbers of tissue samples and corresponding GenBank accession numbers for the sequences are provided in the supplementary material (see appendix for Table A1.1). Genomic DNA for sequencing *CsB* fragments was obtained from: 1) aliquots given by Yale Peabody Museum of Natural History from Yale University (New Heaven, EUA) and the authors of Vargas et al. (2008) (*A. mississippiensis*) 2) tissue samples available at Herpetological Collection from Laboratory of Tetrapod Evolution (*Laboratório de Evolução de Tetrápodes*, LET) at *Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto* (FCLRP-USP), Ribeirão Preto, Brazil; *Conselho de Gestão do Patrimônio Genético* (CGen) registry n° 02000.002450/2011-23), and 3) tissue samples gently given by Herpetological Collections at *Universidade Federal do Mato Grosso* (UFMT, Cuiabá, Brazil), from *Universidade Federal do Alagoas* (MUFAL, Maceió, Brazil), and from *Universidade de Brasília* (CHUNB, Brazil); from Museum of Science and Technology at *Pontifícia Universidade Católica do Rio Grande do Sul* (MCP, Porto Alegre, Brazil); from Zoology Museum at USP (MZUSP, São Paulo, Brazil); and from *Universidade do Estadual de São Paulo* (UNESP, Rio Claro, Brazil).

I extracted DNA from samples previously preserved in 95% ethanol using the DNAEasy Tissue kit (Qiagen), following protocol provided by the manufacturer. Regions *CP1* and *CP2* were amplified by polymerase chain reaction (PCR) using the following primers: region *CP1* (approximately 652-699 bp) - forward 5' GTCCTTCTGTCTAACTAATAATTGC and reverse 5' GCTTCTATTATGATCTCTTG;

region *CP2* (approximately 403-407 bp) - forward 5' CAGATCAAGAGATCATAATAG and reverse 5' TTGTAATGCTAACAGGCAAT. Annealing temperatures used in the PCR ranged between 50 and 54°C (see appendix for Table A1.2).PCRs efficacy was verified by 1% agarose gel electrophoresis, subsequently repeated with doubled volume (50 µl) and ran in 2% agarose gel. Bands of interest, presenting size correspondent to fragments amplification, were purified using *Gel Purification* kit (*Qiagen*), according to macnufacturer´s protocol. Purified PCR products were ligated to pGEM-t vector (Promega) using T4 ligase enzyme, incubated overnight at 16°C, and used in heat-shock transformation of thermo-competent *E. coli* (DH5α or Top10 lineages). Transformed cells were plated in solid agar containing ampicillin, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) and Isopropyl β-D-1-thiogalactopyranoside (IPTG), to ensure double selection of clones containing insert. Six white colonies from each Petri dish, corresponding to clones expected to encompass the gene of interest, were incubated in Lysogeny broth (LB) for 14 hours at 37°C. Concomitantly, PCRs using colonies smear as templates and T7/SP6 primers (given at the pGEM-t vector/Promega protocol) or respective specific primers were performed to ensure that the colonies selected contained inserts of expected sizes. Cultures containing the insert were purified using the *PureYield™ Plasmid Miniprep System Protocol* kit (Promega), following the protocol provided by the manufacturer, but one fraction of each clone (~500 µl) was stored in 50% glycerol. Multiple clones (from 3 to 5 for each species) were sequenced in both directions (forward and reverse, using respectively the primers T7 and SP6. This stage consisted in performing PCR reactions containing 350 µg of plasmid DNA, *primers* T7 or SP6 (5 pM), *Byg Dye Terminator* v.3 and 5X *Big Dye* buffer. PCR products were precipitated using isopropanol (65%) and ethanol (60%) before plates were sent to sequencing.

A consensus sequence for each species was obtained from the sequences confirmed in BLASTn (NCBI) as corresponding to the *CsB* element. For each one of the species

aforementioned, a consensus was first obtained for each cloned sequenced bidirectionally using T7 and SP6 primers, and then a consensus of at least three clones was produced. Consensus sequences obtained for *CP1* and *CP2* regions were concatenated into a single *CsB* fragment. Additional *CsB* sequences were retrieved from the *UCSC Genome Browser* database using the BLAST tool searching genomes of *Gallus gallus, Pelodiscus sinensis, Anolis carolinensis, Mus musculus* and *Homo sapiens* (see appendix for Table A1.1). All sequences were aligned using the Clustal-V algorithm (Higgins et al. 1992) implemented in the software BioEdit, and the resulting alignment was used in the subsequent analyses, which aimed to find regulatory signatures registered in the snakelike lineages compared (see below). Analyses were performed having a major focus on comparisons between snakes and amphisbaenians, and can be grouped in two categories: 1) cluster analyses, and 2) estimation of TFBS.

*Cluster analyses*

Cluster analyses based on Maximum Likelihood were implemented in JModelTest 2 (Darriba et al. 2012), a program that also calculates the most suitable model of evolution prior to tree estimation. The base of tree search chosen for computing likelihood was Best, which automatically selects the most appropriate parameter between NNI and SPR for each analysis. The model of evolution HKY+I+G was estimated using BIC (Bayesian Information Criterion). Heuristic search estimation was set for a maximum of 5,000 trees. In order to reassure our results, we also performed cluster analyses using Parsimony, implemented in PHYLIP 3.695 (Felsenstein 2005), and 1,000 bootstrap replicates were applied to confer statistical confidence to nodes (Felsenstein 1985; Hillis & Bull 1993). Heuristic searches performed in both analyses were unrooted.

*Prediction of Transcription Factor Binding Sites*

Prediction of TFBS was performed individually for each sequence of the alignment by an in-house developed pipeline that uses the Match algorithm (Kel 2003) and TRANSFAC vertebrate databases, in collaboration with Prof. Dr. Pedro E. M. Guimarães from *Universidade Federal de Uberlândia*, Brazil. In order to avoid the prone to predict false positive TFBS with little or none biological value attached, very common in sequence analysis, we implemented a series of steps to improve the reliability of our data. After comparing results using different minimum detection thresholds, we settled the value for this study as 0.8 for both the matrix core and the complete matrix profiles. Lower values were too permissive and increased noise in the analyses, while higher values were too strict and resulted in loss of biological relevant binding sites (data not shown). TFBS matrices were clustered into families of functional and evolutionary related transcription factors, based on the hierarchical classification provided by TRANSFAC database. The family clustering procedure accommodates the expected TFBS sequence variation associated with the evolutionary history of each genome evaluated, avoids the by-product sequence changes deriving from relaxation of selective constraints and reduces the prediction of overlapping binding sites (Balhoff & Wray 2005; Akhtar & Veenstra 2011). A total of 1106 vertebrate matrices were clustered into 72 TFBS families of functional and evolutionary related transcription factors. Consequently, the possibility of analyzing regulatory sequences across species was expanded and comparisons of genomes phylogenetically as distant as human and the coelacanth were enabled. The predicted binding sites were transformed into strings of TFBS families, respecting the prediction order and recording the relative position of the regulatory elements. The relative position was used to identify TFBS signatures constrained during evolution, an approach recognized as *phylogenetic footprinting*, based on the principle that functional regulatory sequences show lower mutation rates than non-functional. Using the prediction

pipeline associated with the control steps described above we searched for TFBS that are conserved in all sequences from predefined sets of species, at the same relative position in the alignment, but are not predicted in the same position in the other groups compared. Estimated TFBS often extend for 5-8 bp, so we settled the initial relative position in the alignment of a given predicted TFBS as the reference point for comparing positional differences among groups. Aware of the restrictions of predictions based algorithms and the necessity of experimentally validate the data, we used the above mentioned approaches to identify gains and losses of reliable TFBS during the evolution of *CsB* across tetrapods.

## 1.4 RESULTS

*CsB* sequences were compared among the lineages studied using bioinformatics revealed a molecular convergence in regulatory haplotypes between Serpentes and Amphisbaenia, besides the presence of regulatory signatures exclusive to Squamata and also other patterns present only in one or another of the two snakelike lineages focused here. The first evidence supporting the existence of molecular signatures common to both snakelike lineages was provided by cluster analyses based on the *CP1/CP2* alignment. Although these analyses have partially recovered some relationships predicted by phylogenetic studies [see Pyron et al. (2013) for Squamata and Benton (1990) for Tetrapoda], many snakelike species that belong to different lineages were grouped together. Specifically, the *CsB* sequences grouped two well-supported clades of pentadactyl lizards (cluster B and E, Fig. 1.2), except for the teiid *Ameiva ameiva*, which was placed among the snakelike squamates (cluster D, Fig. 1.2). The *CsB* sequences of most amphisbaenians were comprised into a single group that also included one of the two groups clustering snake sequences (cluster D, Fig. 1.2), but *Amphisbaenia infraorbitale* was included in the other group that clustered snake sequences (cluster A, Fig. 1.2). *CsB* sequences of mammals (*Homo sapiens* and *Mus musculus*) and those from the three major non-squamate lineages of Reptilia (the turtle *Pelodiscus sinensis*, the bird *Gallus gallus* and the alligator *Alligator mississippiensis*) were clustered together in a separate group (cluster C, Fig. 1.2).

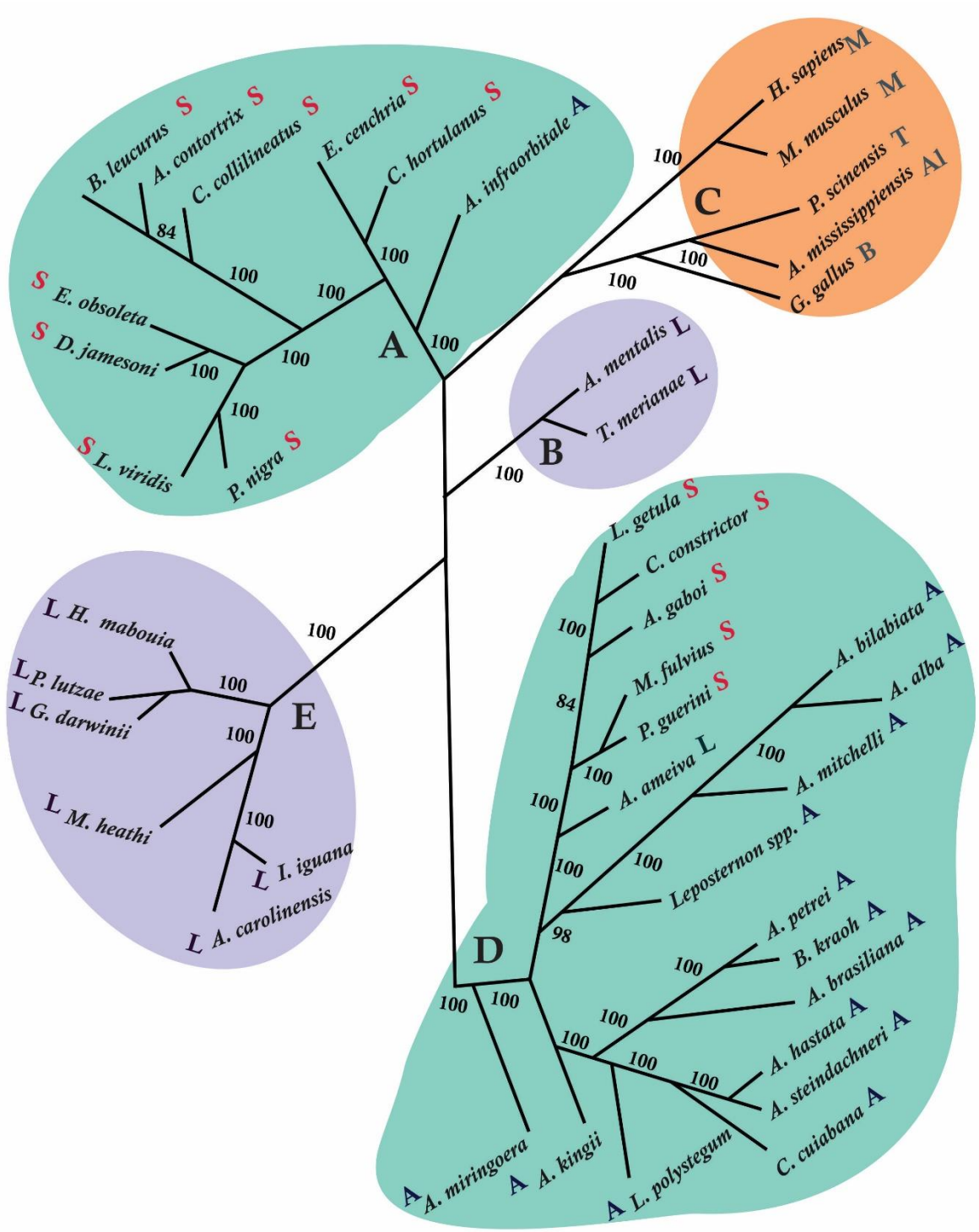


**Figure 1.2.** Topology from cluster analyses based on *CsB* sequenced fragments, implemented using Maximum Likelihood. Relationships were also supported when analyses based on Parsimony were performed. Cluster A comprises *A. infraorbitale* as a sister group of snakes: Boidae species (*E. cenchria* and *C. hortullanus*) are placed as a sister group of Viperidae (*B.leucurus, A. contortrix* and *C. collulineatus*), which form a sister group of *L. viridis, P. nigra, E. obsoleta* and *D. jamesoni*. Cluster B includes the lizards *A. mentalis* and *T. merianae*. Cluster C encompasses mammals (*H. sapiens* and *M. musculus*) in one

branch and the non-squamate Reptilians (*G. gallus*, *P. scinensis* and *A. mississippiensis*) in another branch. Cluster D comprises all species of amphisbaenians, and the snakes *C. constrictor*, *L. getula*, *A. gaboi*, *M. fulvius* and *P. guerini* are situated in a branch as a sister group of the lizard *A. ameiva* and, together, those species are a sister group of the amphisbaenians *A. bilabiata*, *A. alba*, *A. mitchelli* and *Leposternon spp.* Cluster E includes only lizards, *H. mabouia* (Gekkonidae) as a sister group of Phyllodactilydae (*G. darwinii* plus *P. lutzae*), which together are a sister group of *M. heathi* (Scincidae), a sister group of the Iguania branch (*I. iguana* and *A. carolinensis*). The clustering of species based on *CsB* sequences supports some molecular convergence of snakelike species and contradicts the phylogenetic hypothesis of (Pyron et al. 2013), where Iguania is a sister group of snakes, and *A. mentalis, A. ameiva* and *T. merianae* seem more closely related to amphisbaenians (see Fig.1.1). Initials close to species names indicate taxonomic groups: **A** (amphisbaenians); **B** (birds); **Al** (alligator); **L** (lizards); **M** (mammals); **S** (snakes); **T** (turtles).

The second evidence of molecular convergence in the *CsB* sequences of different snakelike lineages derived from estimation of TFBS. Despite the presence of regulatory signatures in this fragment that are exclusive either to snakes or to amphisbaenians (see below), in both lineages an estimated binding site for HIC1 (C2H2_BTB/POZ family) present at position 846 in all lizards has been replaced by a predicted TFBS for PBX1 (PBX family, see Fig. 1.3). Moreover, the *CsB* sequences of pentadactyl lizards exhibit six exclusive TFBS not identified in both snakelike lineages (synthesized in Fig. 1.3): C-MAF (AP-1 family) at position 841, HIC1 (C2H2_BTB/POZ family) at position 843, NEUROD (TAL/TWIST/ATONAL/HEN family) or TTF-1 (NKX family) at positions 847, AP-4 (AP-4 family) at position 848, and XPF-1 or LBP-1 (both from the Grainyhead family) at position 849 in our alignment.

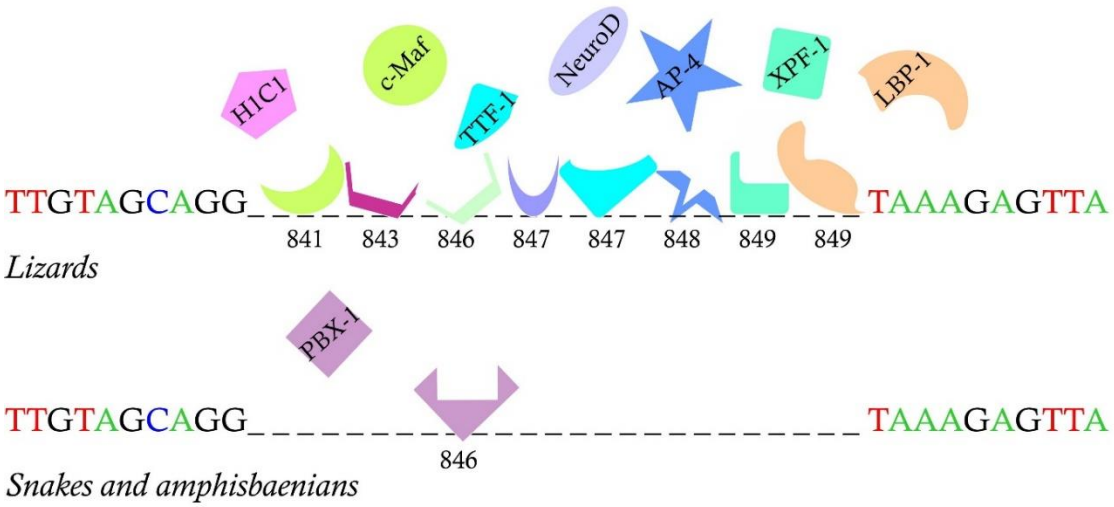


**Figure 1.3.** Positional variation in *CP2*'s predicted TFBS between lizards and lineages with snakelike phenotypes. Analyses indicate loss of binding sites for c-Maf (AP-1), HIC1 (C2H2_BTB/POZ), Neuro D (Tal/Twist/Atonal/Hen), TTF-1 (NKX), AP-4 (AP-4) and LBP-1 or XPF-1 (Grainyhead). At initial position 846 the TFBS for HIC1 in lizards has been substituted by one for PBX-1 in both Serpentes and Amphisbaenia.

Prediction of TFBS in the *CsB* sequences has also detected regulatory signatures that are exclusive to each snakelike lineage (Table 1.1). Snakes exhibit a higher amount of exclusive sites, while sequences of lizards and amphisbaenians are more similar to each other regarding predicted TFBS. In the alignment, the region between positions 756 and 1062 comprises binding sites that are exclusively predicted in snakes, whereas amphisbaenians also exhibit predicted TFBS that are not detected in the same positions in snakes and lizards; these are located in the region between positions 790 and 850 of the *CsB* alignment (Table 1.1). Snakes differ from pentadactyl lizards in a considerable amount of estimated positions for TFBS (located in the region between positions 735 and 1062 of the alignment; see appendix for Table A1.3), and both groups only exhibit four predicted TFBS having the same location that are not found in amphisbaenians: HoxA3 (initial position 845); AP-4 (initial position 848); BEN or CBF1 (initial position 850); and DBP (initial position 851). Amphisbaenians also differ from pentadactyl lizards in some estimated positions for TFBS (located in the region between positions 790 and 850 of the alignment; see appendix for Table A1.4).

**Table 1.1** Estimated initial positions of TFBSs in the *CP2-CsB* sequence exclusive to snakes or to amphisbaenians.

| TF Family | SNAKES | AMPHISBAENIANS |
|---|---|---|
| C2H2_BTB/POZ | BCL6: (+756, 762)<br>KAISO (+789) | |
| GATA | GATA4 (+756) | |
| PBX_fam | PBX1 (+759) | PBX1 (+842) |
| Interferon-reg | IRF-1 (+760) | |
| SMAD_fam | SMAD (+846)<br>SMAD1 (+760)<br>SMAD4 (+849) | |
| Homeodomain | CDXA (+762) | CDXA (+847)<br>CPD (+845) |
| TBP | TBP (+763) | |
| HMGI_fam | HGMIY (+764) | HGMIY (+847) |
| T3R | T3R (+769) | |
| ETV_fam | PEA3 (+773) | |
| TFII-I | TFII-I (+774) | |
| NR0-6 | TR4 (+789) | |
| Sp/KLF | SP1 (+789) | |
| NF-1 | NF-1, NF-1A (+842) | |
| zinc_finger | REX1 (+848) | |
| AP-2 | AP-2 γ (+850, +852)<br>AP-2α (+851, +852) | |
| Cell-cycle | E2F-1 (+850) | |
| p53-like | p53 (+851) | |
| CEBP_fam | | C/EBP (+842, +843)<br>C/EBPα (+790, +842, +843)<br>C/EBPβ (+790) |
| NF-AT | | NF-AT4 (+791) |
| STAT | | PARP (+792) |
| bZIP/PAR | | DBP (+846) |
| Myb | | c-MYB (+845, +848)<br>MYB (+850) |
| RING_finger | | BRCA:USF2 (+850) |
| NKX_fam | TTF-1 (+1054) | |
| STAT | STAT3, STAT4, STAT5A, STAT6: (+1057) | |
| TCF_fam | TCF-3, LEF1 (+1060) | |
| SOX_fam | SOX10, SOX (+1062) | |

Some regulatory haplotypes in the *CsB* fragment studied that are common to all squamates but not observed in the remaining tetrapods considered here have also been identified. These involve variation in the predicted positions of specific TFBS, and are concentrated between positions 133 and 1074 of the alignment (see appendix for Table A1.5; Fig. 1.4). In contrast, some TFBS predicted in non-squamate species are absent at the same sites in all squamates (see appendix for Table A1.4; Fig. 1.4).

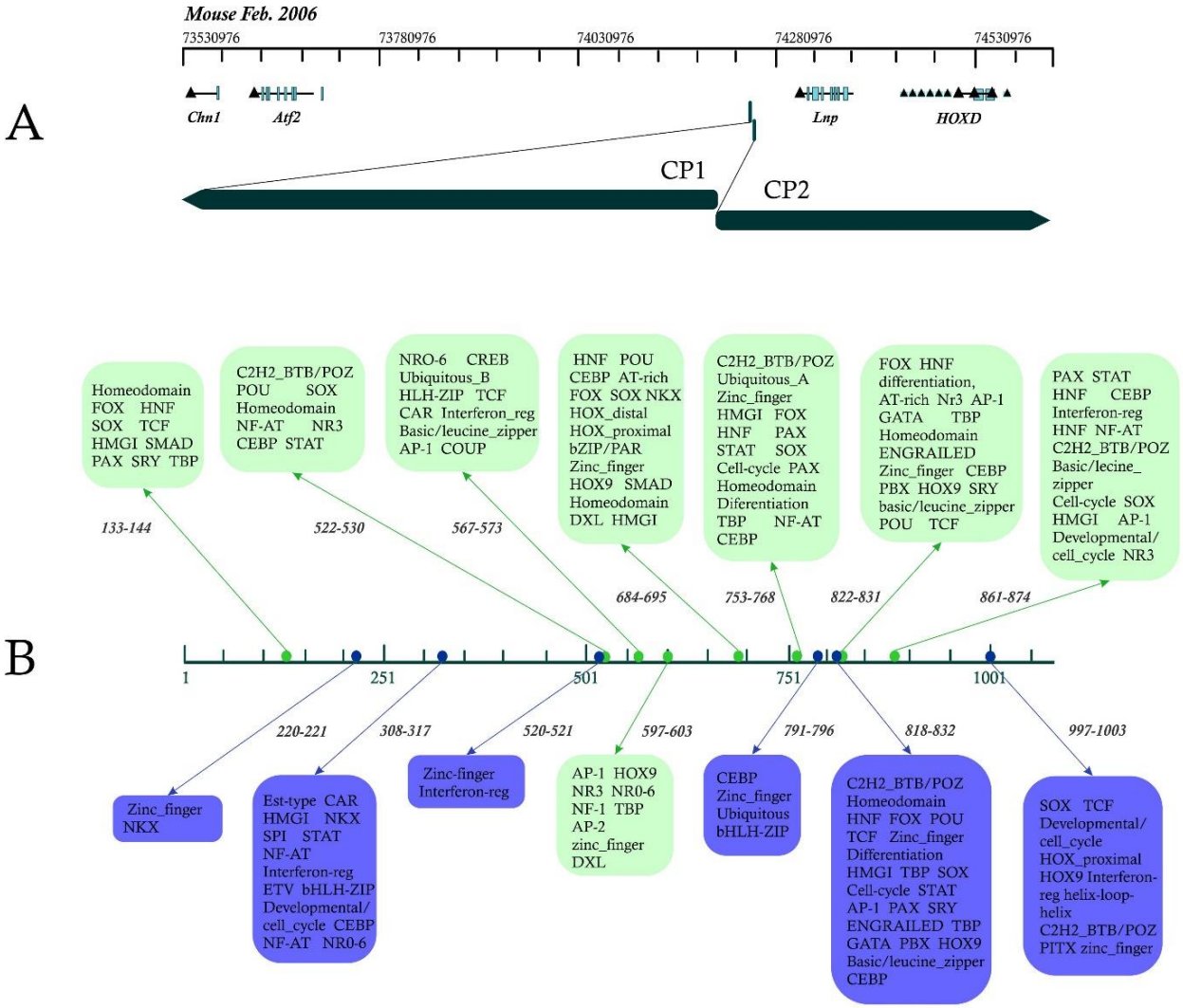


**Figure 1.4.** Positions for transcription factor binding sites (TFBS) that are exclusive to Squamata or to non-squamate vertebrates. A) Position of regions *CP1* and *CP2* in mouse chromosome II. B) Comparison of predicted TFBS families shared by Squamata species (blue groups), and those of other

vertebrates (green groups) reveals variations occurring ‘in blocks’ throughout the fragments. Region *CP2* seems to concentrate most of the variation.

## 1.5 DISCUSSION

The present study provides evidence for molecular convergence of a regulatory region during the evolution of snakelike morphologies in two distinct lineages of Squamata. Further studies based on functional assays may confirm the relevance of such convergence for the establishment of elongated limbless morphologies during embryo development. The regulatory patterns that are common to snakes and amphisbaenians were detected in a fragment of the *CsB* element, which regulates terminal *Hoxd* genes expression during embryo development (Spitz et al. 2003; Gonzalez et al. 2007). The regulatory signatures common to snakelike lineages are located within the *CP2* region of the *CsB* sequence, and foretell this region as a strong candidate for functional evaluations, given its role as a key regulatory locus in the context of snakelike morphological evolution. Regulatory signatures that were solely registered in either one of the three compared squamate groups (snakes, amphisbaenians and lizards) were also detected here, revealing divergent patterns that likely reflect independent trajectories during the evolution of different lineages from their lepidosaur ancestors. Finally, a functional dissection of *CsB* performed with transgenic assays of isolated *CR1* and *CR2* fragments, in collaboration with Prof. Dr. Igor Schneider, complements the present study by suggesting that *CR1* possibly encompasses elements related to an ancient regulatory role during AP body axis development, an activity that is observed even after *CR2* deletion (see Appendix section). Together with the regulatory signatures identified here in the two snakelike lineages, this observation raises the possibility that *CsB* could also contribute for AP axis patterning, a role so far relegated for this region. The transgenic assay also show that complete sequences of each one of the three conservation peaks of *CsB* are required for regulatory activity in the autopodium, since deletion of *CR1* or *CR2* impairs reporter gene expression in limb buds.

*Divergent regulatory signatures of CsB: how much phylogenetic signal is registered in this regulatory region?*

The *CsB* fragments studied here carry some degree of phylogenetic signal, a conclusion derived from both the cluster analyses and the identification of predicted TFBS that are exclusive to diverging lineages. Specifically, phylogenetic relationships recently proposed for the major tetrapod groups (Benton 1990) were recovered by the *CsB* fragment: mammals (*Homo sapiens* and *Mus musculus*) were grouped together, as well as the Reptilians that diverged previously to the Squamata origin (the turtle *Pelodiscus sinensis*, the bird *Gallus gallus* and the alligator *Alligator mississippiensis*; cluster C in Fig. 1.2). The squamate species were grouped in four distinct clusters (Fig. 1.2), partially recovering positions proposed in phylogenetic studies: *A. mentalis* (Gymnophthalmidae) and *T. merianae* (Teiidae) were grouped together (cluster B), and cluster E placed *H. mabouia* (Gekkonidae) as a previously divergent sister group of *G. darwinii* plus *P. lutzae* (Phyllodactilydae); *M. heathi* (Scincidae) was situated with *I. iguania* and *A. carolinensis* (Iguania) in another branch, which is congruent to (Pyron et al. 2013) (illustrated in Fig.1.1).

Predictions of TFBS along the *CsB* fragment studied reveals that evolution of Squamata has left specific regulatory signatures in *CsB* both during its divergence from other tetrapods, identified throughout almost the entire segment but concentrated in specific *CsB* segments, as well as within main squamate groups (snakes, amphisbaenians and lizards), which are detected solely in the *CP2* region. Specifically within Squamata, divergence in the *CsB* sequence can be observed in the three groups, but snakes exhibit a large amount of exclusive predicted TFBS in comparison with other squamates, while amphisbaenians and other lizards share more predicted TFBS among each other. The apparent reduction in sequence variation between amphisbaenians and the other lizards agrees with the current hypothesis proposing

that amphisbaenians diverged more recently (~115 Mya) than snakes [125 Mya; (Wiens et al. 2006), despite the controversial phylogenetic relationships within Squamata (see Losos et al. 2012; Pyron et al. 2013), and the concurrent hypothesis of a later divergence of the snakes clade (Vidal & Hedges 2009)]. The presence of phylogenetic signal in the *CsB* sequence can also be inferred from the exclusive regulatory haplotypes detected only in snakes or in amphisbaenians. These diverging patterns in *CsB* are also observed by the prediction of TFBS in positions that are exclusive to Squamata, which likely relate to the independent trajectory of this lineage after its split from an ancestor shared with other reptilian lineages. In fact, our cluster analysis recovered especially well the relationships among main tetrapod lineages proposed by phylogenetic studies (Benton 1990), suggesting the presence phylogenetic signal in the *CsB* fragment studied.

Several transcription factors which sites were predicted in exclusive positions in *CP2* of amphisbaenians or snakes play roles in model-system organisms (see Table 1.2) that are associated with development of the genitourinary system (testicle, prostate, mammary glands, ureters, kidneys), brain (neurogenesis, forebrain and cerebellum) and axial structures (AP patterning, axial skeleton, bones, cartilage, muscle, somite, spinal cord and neural tube). In *CsB* sequences from both snakes and amphisbaenians, binding sites for CDXA were predicted in different positions; CDX family of transcription factors plays indispensable roles for axial development, participating on Wnt and FGF signaling cascades during *Hox6-Hox9* expression (Akker & Forlani 2002; Fukuda & Kikuchi 2005; Young et al. 2009), and also contribute for posterior neural tube and cloacal development through interaction with Wnt signaling (van de Ven et al. 2011). Interestingly, TFBS that are specifically associated with limb development do not predominate in the *CP2* of these lineages. Therefore, the prediction of exclusive TFBS in Serpentes and Amphisbaenia support my concluding hypothesis (discussed hereafter) that the *CsB* region studied here may be inserted in a regulatory module related to AP body axis

patterning or axial skeleton development, in addition to modules related to previously described *CsB* roles.

**Table 1.2.** Phenotypic information related to developmental pathways for the TFBS estimated in exclusive positions in the *CP2-CsB* sequence of either snakes or amphisbaenians, listed in Table 1.1; information retrieved from current published literature (The UniProt Consortium 2008).

| TF family | TF | Serpentes | Amphisbaenia | Functions |
|---|---|---|---|---|
| C2H2_BTB/ POZ | BCL6 | +756, +762 | - | Neurogenesis, erythrocyte development and spermatogenesis |
| | KAISO | +789 | - | |
| GATA | GATA4 | ++756 | - | Cardiac, digestive tract, testicular development and spermatogenesis |
| PBX_fam | PBX1 | ++759 | +842 | Embryonic skeletal system, sexual and adrenal gland development; proximal/distal and anterior/posterior pattern specification; ureteric bud and embryonic limb morphogenesis and neuron differentiation |
| Interferon-reg | IRF-1 | +760 | - | Cell proliferation and differentiation, regulation of the cell cycle, and programmed cell death |
| SMAD_fam | SMAD1 | +760 | - | Bone, cartilage, brain, kidney, ureteric bud and inflammatory response development, embryonic pattern specification and gametogenesis |
| | SMAD4 | +849 | - | Somite rostral/caudal axis specification; heart, endoderm, sebaceous gland and brainstem development; ureter and kidney morphogenesis; developmental growth |
| | SMAD | +846 | - | |
| homeodomain | CDXA | +762 | +847 | Anterior/posterior pattern specification; bone morphogenesis |
| | CPD | - | +845 | - |
| TBP | TBP | +763 | - | Role during early embryogenesis, ventral embryo patterning during gastrulation |
| HMGI_fam | HGMIY | +764 | +847 | - |

| | | | | |
|---|---|---|---|---|
| T3R | T3R | +769 | - | - |
| ETV_fam | PEA3 | +773 | - | Mammary gland duct morphogenesis and stem cell differentiation |
| TFII-I | TFII-I | +774 | - | Early embryonic and cerebellum development; embryonic stem cells; spermatogenesis |
| NR0-6 | TR4 | +789 | - | - |
| Sp/KLF | SP1 | +789 | - | Embryonic skeletal system, embryonic placenta, liver lung and eye development |
| NF-1 | NF-1 | +842 | - | Spinal cord, adrenal gland, eye, cerebral cortex, heart, liver, metanephrons and muscle development |
| | NF-1[a] | +842 | - | - |
| Zinc_finger | REX1 | +848 | - | Female and male gonad development, in utero embryonic and spermatid development; meiosis |
| AP-2 | AP-2 γ | +850, +852 | - | Limb, eye, face, body wall, cerebral cortex, male gonad, sebaceous glands and neural tube development |
| | AP-2α | +851, +852 | - | - |
| Cell-cycle | E2F-1 | +850 | - | Forebrain development and spermatogenesis |
| p53-like | p53 | +851 | | Central nervous system, embryonic organ, and in utero embryonic development |
| CEBP_fam | C/EBP | | +842, +843 | - |
| | C/EBPα | - | +790, +842, +843 | - |
| | C/EBPβ | - | +790 | Embryonic placenta development, mammary gland epithelial cell, neuron and osteoblast differentiation |
| NF-AT | NF-AT4 | - | +791 | Heart and muscle development |
| STAT | PARP | - | +792 | DNA repair. |
| bZIP/PAR | DBP | - | +846 | mRNA processing, regulation of cell cycle |
| MYB | c-MYB | - | +845, +848 | Control of proliferation and differentiation of hematopoietic progenitor cells |
| | MYB | - | +850 | - |
| RING_finger | BRCA:USF2 | - | +850 | DNA repair; cell cycle arrest; post-transcriptional regulation |

| | | | | |
|---|---|---|---|---|
| NKX_fam | TTF-1 | +1054 | - | Chromatin remodeling |
| STAT | STAT3 | +1057 | - | Ventralization of embryos; mesoderm induction in early embryos |
| | STAT4 | +1057 | - | |
| | STAT5A | +1057 | - | Development of secondary sexual characteristics and of prostate and mammary glands |
| | STAT6 | +1057 | - | Mammary gland morphogenesis |
| TCF_fam | TCF-3 | +1060 | - | Neuronal and muscle cell differentiation |
| | LEF1 | +1060 | - | Somitogenesis; embryonic limb and face morphogenesis; hypothalamus, mammary and trachea glands, muscle fiber, tongue development |
| SOX_fam | SOX10 | +1062 | - | Anatomical structure, digestive tract morphogenesis; enteric and peripheral nervous system and in utero embryonic development. |
| | SOX | +1062 | - | |

*Convergent regulatory signatures of CsB: independent origins of snakelike morphologies*

Independent origins of the snakelike morphology in two separate lineages of Squamata involved molecular convergence in the fragment of *CsB* studied here. The first evidence suggesting molecular convergence in the *CsB* of snakes and amphisbaenians was provided by cluster analyses, which grouped snakes in two clades, one of them situated closer to the amphisbaenians cluster than to the other snakes' group, and also placed *Amphisbaenia infraorbitale* together with snakes, a pattern that contradicts most of the phylogenetic hypotheses for Squamata that have been proposed using molecular markers (e.g. Vidal & Hedges 2005; Conrad 2008; Vidal & Hedges 2009; Losos et al. 2012; Pyron et al. 2013). According to the very inclusive hypothesis recently proposed by Pyron et al. (2013), snakes form a sister group of the lizard clade comprising Iguania [represented here by *Iguana iguana* (Iguanidae) and *Anolis carolinensis* (Dactyloidae)], while Amphisbaenia forms a clade with

Lacertidae (not included in this work), a sister group of the clade comprising Gymnophthalmidae (represented by *Acratosaura mentalis*) and Teiidae (represented by *Ameiva ameiva* and *Tupinambis merianae*). Despite the presence of phylogenetic signal in the *CsB* fragment studied here (discussed earlier), its sequence clustered together snakes, amphisbaenians and *Ameiva ameiva* in one group and all the other pentadactyl lizards in separated groups, and also placed one amphisbaenian (*A. infraorbitale*) within some snakes. Together with the identification of TFBS that are common to snakes and amphisbaenians but not present in the remaining lizards (and *vice versa*), this result suggests that independent origins of snakelike morphologies in Squamata registered convergent signatures in this *CsB* fragment.

Prediction of TFBS in the *CsB* fragment also supports the occurrence of regulatory signatures that are common to both snakelike lineages but absent in lacertoid lizards. The convergent patterns shared by snakes and amphisbaenians are concentrated specifically between initial positions 841 and 849 of the alignment, and correspond to the apparent loss of five binding sites in specific positions and a substitution that resulted in the same relative position for the TFBS of PBX1 in both lineages. Specifically regarding the loss of TFBS, both snakelike lineages seem to have lost in specific positions of region *CP2* in *CsB* the binding sites for C-MAF, HIC1, NeuroD or TTF-1, AP-4, and XPF-1 or LBP-1. Most of these transcription factors belong to protein families involved in developmental pathways that coordinate axial skeleton or limb development. For example, HIC1 is related to GLI3 and both belong to the $H_2C_2$ family (Büscher et al. 1997), while NEUROD belongs to TWIST family. Both GLI3 and TWIST are key factors in the signaling cascade involved in the establishment of digit identity (Marigo & Tabin 1996; Kanagae et al. 1998), and the last one is also crucial for axial skeleton development (Kanagae et al. 1998; Zuniga et al. 2002; Yasutake et al. 2004; O'Rourke et al. 2002; Krawchuk et al. 2010). Another factor that participates in the development of the axial

skeleton is NKX-3.1, related to TTF-1 (NKX-2.1; see Price et al. 1992), a transcription factor which binding site was lost at initial position 847 in the *CsB* of snakes and amphisbaenians. Expression of NKX-3.1 is induced and maintained by notochord and neural tube and also contributes to formation of caudal somites (Deutsch et al. 1988; Goulding et al. 1991; Wolf et al. 1991; Kos et al. 1998). Finally, another predicted TFBS that has been apparently lost in both snakelike lineages corresponds to XPF-1 or LBP-1, which has as initial position the site 849 in the lizards sequences; these two factors are related to Grainyheadlike-3, a factor essential to the process of neural tube closure during spine development (Gustavsson et al. 2007).

A very striking convergent pattern identified in the *CsB* fragment of all species of snakes and amphisbaenians studied here is the replacement at initial position 846 of a binding site for HIC1, predicted in lizards, by one for PBX1. The substitution of HIC1 at the same position in the two snakelike lineages is particularly interesting both regarding the binding site that disappeared and the one that was acquired in its place. As aforementioned during the discussion of binding sites lost in snakes and amphisbaenians, HIC1 belongs to the same family of GLI and may play a role during autopodial development (Büscher et al. 1997), a process that no longer takes place in the establishment of limbless phenotypes. The site acquired in its position is PBX1, a co-factor of HOX proteins (Saleh et al. 2000; Burel et al. 2006; Capellini et al. 2008) that is highly expressed on the notochord, sclerotome and vertebrae, and is crucial for the establishment of vertebral identity and girdles positioning during embryo development (Selleri et al. 2001; Moens & Selleri 2006; Capellini et al. 2008; Capellini et al. 2010; Capellini, Handschuh, et al. 2011; Capellini, Zappavigna, et al. 2011), although it is also important for limb patterning (Capdevila & Belmonte 2001; Capellini et al. 2006). Interestingly, PBX-HOX sites have been recently proposed as typical pioneer motifs during evolution of new *cis*-regulatory elements, as they drive robust and specific, but

modifiable expression hints (Parker et al. 2011). Independent acquisitions of the same motifs by different lineages suggest implication in development of features that are *present*, not lost. Thus, although it is important to further empirically investigate the precise functional meaning of this noteworthy signature on *CP2*, the acquisition of PBX1 in both snakelike lineages likely indicates the involvement of this regulatory region in a developmental process related to the emergence of limbless elongated body shapes, rather than solely representing a signature of limb loss. Two separate losses, in Serpentes and Amphisbaenia, of the same six TFBS predicted in lizards, together with two different origins of a binding site for PBX1 at the same position in the snakelike lineages, likely correspond to a footprint in *CsB* molecular evolution coupled to these two morphological transitions, which involved both limb loss and axial elongation.

Mouse transgenic assays using sequences from zebrafish, coelacanth and mouse (see Appendix section), coupled with previous results from Gonzalez et al. (Gonzalez et al. 2007), support the possibility that *CsB* may play a so far unexplored regulatory role during establishment of the AP axis. These assays show that two first conserved peaks of *CsB* cannot alone regulate gene expression in the developing autopodia: transgenic assays with only *CR1* (*CP1*+*CP2*) elicit regulatory activity along the neural tube but not in limb buds. The functional dissection of *CsB* performed by Gonzalez et al. (2007) showed that deletion of *B1* and *B2* fragments, which are tetrapod-specific elements present respectively within *CP1* and *CP2*, triggers reporter expression patterns that are similar to the activity exerted by complete *CsB* in limbs, whereas expression in the neural tube is reduced [see Fig 4A in (Gonzalez et al. 2007)]; given these results, *B1* and *B2* have been previously suggested as being critical for neural expression but dispensable for limb expression (Gonzalez et al. 2007). The patterning of AP body axis in vertebrates has not been fully elucidated, so possible interferences among *Hox* developmental pathways arising from different axial tissues remain obscure; *cis*-regulatory

integration of such information is even less understood (Bel-Vialar et al. 2002). However, some signaling pathways and transcription factors driving axial elongation or presomitic mesoderm segmentation and controlling *Hox* genes expression, such as CDX and *Wnt* (Akker & Forlani 2002; Aulehla et al. 2003; van de Ven et al. 2011; Montavon & Soshnikova 2014; Denans et al. 2015), are indeed known to take place both in vertebrates neural tube and somite domains (McMahon et al. 1998; Sela-Donenfeld & Kalcheim 2000). Somite morphogenesis processes are fed by signals from neural tube, floor plate and notochord (Münsterberg & Lassar 1995), and the absence of a neural tube prevents somite maturation (Packard & Jacobson 1976). Although the original literature describes *CsB* as a regulatory element of terminal *Hoxd* gene expression during limb, external genitalia and neural tube development (Spitz et al. 2003; Gonzalez et al. 2007; Montavon et al. 2011; Schneider et al. 2011), the results described here and in. Gonzalez et al. (2007) pave a way for hypothesizing that *CP1* and *CP2* may also play an indirect role on the establishment of AP axis in the vertebrate embryo.

## 1.6 CONCLUSIONS

Prediction of the same six consecutive TFBS losses and the one TFBS acquisition evolving separately in the *CP2-CsB* fragment of two elongated limbless lineages point out to this element as a possible main candidate for connecting the genotypic and the phenotypic dimensions that converged during the evolution of snakelike morphologies. Together with TFBS predictions in the *CP2* fragment elucidated in this study, the expression patterns of region *CR1-CsB* from zebrafish, coelacanth and mouse in the neural tube (see Appendix) suggests a possible ancient role of this fragment for the establishment of structures along the AP body axis during embryo development. Evolution of this regulatory fragment of *CsB* within Squamata therefore apparently left divergent and convergent signatures during two independent origins of snakelike morphologies that may be related to the limbless elongated body shape characteristic of snakes and amphisbaenians.

## 1.7 APPENDIX

*Gene reporter assays performed by Prof. Dr. Igor Schneider (UFPA)*

**METHODS: gene reporter assays - plasmid generation and mouse transgenesis**

A functional dissection of *CsB* was performed to access the independent regulatory capacity of regions *CP1* and *CP2*. The regulatory roles of these two regions were explored here using a set of gene reporter assays in transgenic mice. First, the following fragments of *CsB* from mice were amplified (see Fig. A1.1): A) the complete *CsB* (2.5 Kb); B) *CR1*, a 1.2 Kb fragment of *CsB* comprising the two conserved fragments that were compared among Squamata (*CP1* and *CP2*), therefore excluding *CP3*; C) *CR2*, a 1.3 Kb fragment of *CsB* including *CP3* and excluding *CR1*. The complete *CsB* and the fragment *CR2* were PCR-amplified from mouse genomic DNA using the following primers sequences: *CsB*-F, 5'-GAGACTCTGCCACGTTTTCAG-3'; *CsB*-R, 5'-ACCCAGAGCTTAGGACCACAT-3'; *CR2*-F, 5'-GAGAAGGAGCAGGGAATGTCT-3'. Regulatory activity of the *CR1* fragment was also scanned in a broader evolutionary scenario, isolating the corresponding genomic DNA fragments using the Expand High Fidelity PCR System (Invitrogen) from one representative of each of three major Gnathostomata lineages: Actinopterygii (zebrafish), Sarcopterygii (coelacanth) and Tetrapoda (mouse).

DNA oligonucleotide sequences used to amplify the CsB element in the three Gnathostomata main lineages aforementioned were as follows:

1) zebrafish*CSB*-F: 5'-TGACTCTTGTGTTGTCGCAGA-3';
2) zebrafish*CsB*-R: 5'-ATCAGTTGGTCAGTTGGTTCG-3';
3) mouse*CsB*-F: 5'-GAGACTCTGCCACGTTTTCAG-3';
4) mouse*CsB*-R: 5'-TCCTTCTCAAGAGCAGAGCAG-3';

5) coelacanth*CsB*-F: 5'-ACTAAGCCACCAACACCCTCT-3';

6) coelacanth*CsB*-R: 5'-ACTAGACTGTTGCCCCCTGAT-3'.

Fragments were cloned into an entry vector (pENTR/D-TOPO) and transferred via the Gateway System (Invitrogen) into the Gateway-Hsp68-LacZ vector (kind gift from Dr. Marcelo Nobrega, University of Chicago, USA). Plasmid DNA was purified using the Wizard Plus SV miniprep Kit (Promega), and 50 µg of each plasmid was digested with SalI to excise the vector backbone. Following a gel purification step using the QIAquick Gel Extraction Kit (Qiagen), the DNA to be injected was further purified using a standard ethanol precipitation and diluted to a concentration of 2 ng/µL. Purified, linearized plasmid DNA was then used for pronuclear injections of CD1 mouse embryos in accordance with standard protocols approved by the University of Chicago Transgenic Mouse Facility. Mouse embryos were harvested, stained and fixed according to standard protocols (Schneider et al. 2011). Each construct resulted in at least 3 LacZ-positive embryos displaying similar reporter activity, which were verified by genotyping. All work was performed in compliance with and approved by the Ethics Committee of the Universidade Federal do Pará (Brazil) and according to standard protocols approved by The University of Chicago.

## RESULTS

In order to investigate their autonomous regulatory capacities, gene reporter expression of the two first peaks of conservation, *CP1* and *CP2* (comprised within the fragment termed *CR1* within, see Fig. A1.1A) were examined. Regulatory activity held by *CsB* after deletion of *CR1*, using a fragment termed *CR2* (Fig. A1.1A), which contains *CP3,* was also investigated. Assays using the entire *CsB* element were performed as positive control. As previously detailed, *CsB*, *CR1* and *CR2* were cloned into vectors containing the LacZ gene and

the HSP68 minimal promoter (Amemiya et al. 2013), which were then used in mouse transgenic assays. Whereas the complete *CsB* is capable of driving reporter gene expression in neural tube, forebrain, and limbs (Fig. A1.1B, left), a pattern previously reported by (Spitz et al. 2003), deletion of the *CR1* region (*CR2* assay) eliminates most, if not all of *CsB* regulatory activity (Fig. A1.1B). Interestingly, gene reporter expression of *CR1* shows that deletion of *CR2* from *CsB* culminates in loss of regulatory activity in the autopodium but not the neural tube (Fig. A1.1C). This result highlights that the fragment comprising *CP1* and *CP2* (but not *CP3*) is not sufficient to exert the regulatory activity of *CsB* in the limbs. On the other hand, intense activity is retained across the AP axis even after deletion of *CR2*, so the region *CR1* seems to allocate enhancers that wield regulatory functions for *Hoxd* gene expression during development of forebrain and neural tube, possibly also playing a role on AP axis establishment. Such regulatory activity seems ancestral in Gnathostomata because gene expression was promoted along the neural tube in the three key species used here (Fig. A1.1): zebrafish (Actinopterygii), coelacanth (Sarcopterygii) and mouse (Tetrapoda).

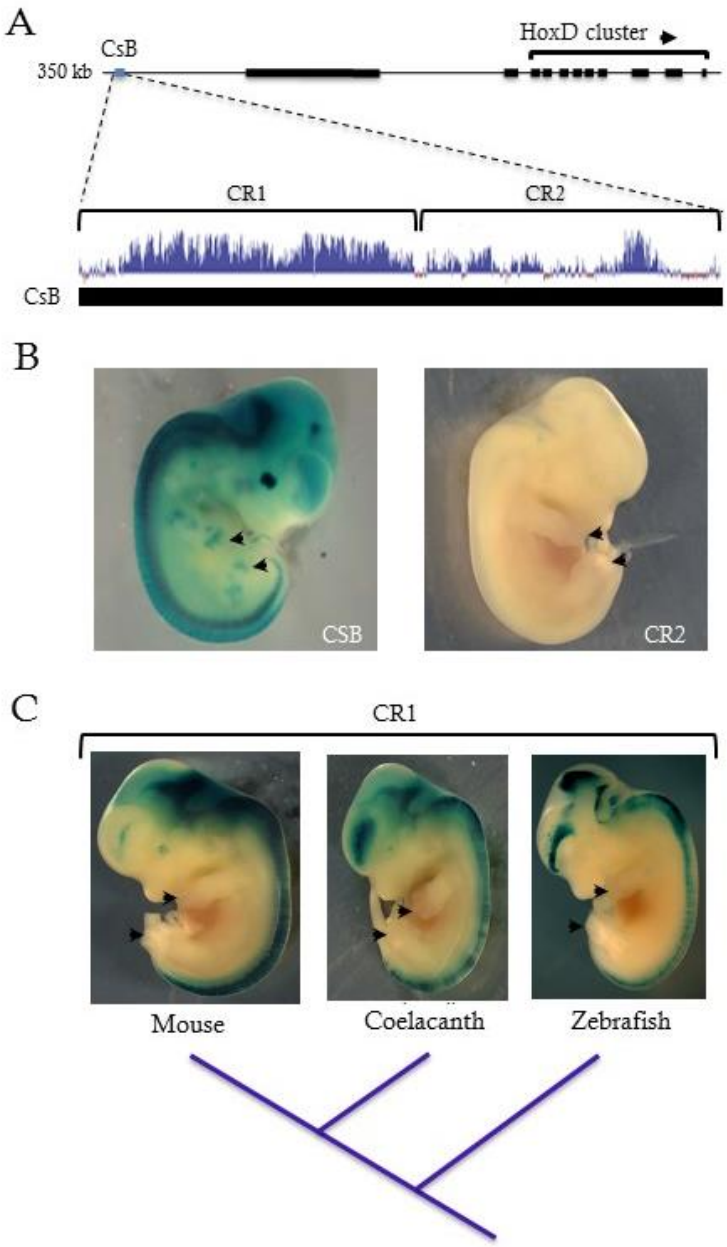


**Figure A1.1.** Patterns of regulatory activity elicited by constructions containing *CsB* fragments on mouse transgenic assays. The top panel (A) represents location of *CsB* in relation to HoxD cluster and an mVista alignment (mVISTA program, homology threshold set to 70 percent) showing the level of conservation (in blue) of *CsB* regions among 60 vertebrates species. On (B), left assay shows regulatory activity of the complete *CsB* – staining observed in the forebrain, neural tube and limb buds; right assay suggests that the *CR2* fragment lacks autonomous regulatory role (reporter expression is not observed). Bottom panel (C) presents embryos where regulatory activity is induced by *CR1* fragment containing *CP1+CP2* but not *CP3*. The patterns suggest the regulatory role of *CsB* in neural tube is ancestral in Gnathostomata and mainly concentrated in *CR1*. Arrowheads highlight limb buds in the images.

**Table A1.1.** Species which *CP1* and *CP2* fragments were sequenced or obtained from GenBank. Voucher numbers (snakes, amphisbaenians, lizards and crocodile) and GenBank accession numbers are provided.

| Squamata Lineage | Family | Species | Voucher | GenBank accession numbers |
|---|---|---|---|---|
| Serpentes | Elapidae | *Dendroaspis jamesoni* | YPM 13380 | KP412493 |
| | Viperidae | *Agkistrodon contortrix* | YPM 13252 | KP412495 |
| | Colubridae | *Lampropeltis getula* | YPM 13574 | KP412499 |
| | Colubridae | *Coluber constrictor* | YPM 13574 | KP412498 |
| | Elapidae | *Micrurus fulvius* | YPM 14096 | KP412494 |
| | Colubridae | *Elaphe obsoleta* | YPM 13086 | KP412500 |
| | Viperidae | *Crotalus collilineatus* | CHRP, didactic collection. | KP412496 |
| | Dipsadidae | *Liophis viridis* | CHRP 694 | KP412505 |
| | Dipsadidae | *Apostolepis gaboi* | CHRP 296 | KP412506 |
| | Boidae | *Corallus hortulanus* | MZUSP 17291-2/12 | KP412503 |
| | Viperidae | *Bothrops leucurus* | MUFAL 8797-1/2 | KP412497 |
| | Boidae | *Epicrates cenchria* | MUFAL 8933-1/10 | KP412504 |
| | Colubridae | *Phimophis guerini* | MUFAL 3665-1/3 | KP412501 |
| | Colubridae | *Pseudoboa nigra* | MUFAL 9701-1/7 | KP412502 |
| Amphibaenians | Amphisbaenidae | *Amphisbaena brasiliana* | UFMT-R 6110 | KP412507 |
| | Amphisbaenidae | *Amphisbaena mitchelli* | UFMT 8881 | KP412508 |
| | Amphisbaenidae | *Amphisbaena alba* | UFMT-R 5478 | KP412509 |
| | Amphisbaenidae | *Amphisbaena steindachneri* | UFMT-R 4151 | KP412510 |
| | Amphisbaenidae | *Leposternon sp.* | CHRP 561 | KP412514 |
| | Amphisbaenidae | *Amphisbaena infraorbitale* | UFTM-R 8864 | KP412511 |
| | Amphisbaenidae | *Amphisbaena bilabiata* | UFMT-R 4767 | KP412512 |
| | Amphisbaenidae | *Amphisbaena miringoera* | UFMT-R 9234 | KP412513 |
| | Amphisbaenidae | *Amphisbaena hastata* | CHRP 281 | KP412515 |
| | Amphisbaenidae | *Amphisbaena petrei* | MUFAL 9054-1/2 | KP412516 |
| | Amphisbaenidae | *Anops kingii* | MCP14722 | KP412520 |
| | Amphisbaenidae | *Leposternon polystegum* | CHUNB 30669 | KP412517 |
| | Amphisbaenidae | *Cercolophia cuiabana* | TMOTT 204 | KP412518 |
| | Amphisbaenidae | *Bronia kraoh* | CHUNB 30767 | KP412519 |
| Lizards | Gekkonidae | *Hemidactylus mabouia* | CHRP 848 | KP412521 |
| | Gymnophthalmidae | *Acratosaura mentalis* | CHRP 555 | KP412522 |
| | Phyllodactylidae | *Gymnodactylus darwinii* | MUFAL 4617-1/7 | KP412523 |
| | Phyllodactylidae | *Phyllopezus lutzae* | MUFAL 2836-1/4 | KP412524 |
| | Teiidae | *Ameiva ameiva* | MUFAL 3930-1/8 | KP412525 |

| | | | | |
|---|---|---|---|---|
| | Teiidae | *Tupimambis merianae* | MUFAL 10610-1/5 | KP412526 |
| | Scincidae | *Mabuya heathi* | CHRP 687 | KP412527 |
| | Iguanidae | *Iguana iguana* | *tissue from UNESP* | KP412528 |
| | Dactyloidae | *Anolis carolinensis* | - | NW003338903.1 |
| Crocodile | Alligatoridae | *Alligator mississippiensis* | *tissue from Vargas et al. (2008)* | KP412529 |
| Bird | | *Gallus gallus* | - | NC006094.3 |
| Turtle | | *Pelodiscus scinensis* | - | NW005854041.1 |
| Mouse | | *Mus musculus* | - | NC000068.7 |
| Human | | *Homo sapiens* | - | NC000002.12 |

**Table A1.2.** Species which fragments *CP1* and *CP2* of *CsB* were amplified, identified by collection numbers (when available) and respective specific PCR conditions.

| Clade | Species | PCR conditions | |
|---|---|---|---|
| | | ***CP1*** | ***CP2*** |
| Serpentes (Elapidae) | *Dendroaspis jamesoni* (YPM 13380) | 50°C; 1mM $MgCl_2$ | 50-54°C; 2,5mM $MgCl_2$ |
| Serpentes (Viperidae) | *Agkistrodon contortrix* (YPM 13252) | 50°C; 2,5 mM $MgCl_2$ | 50-54°C; 2,5mM $MgCl_2$ |
| Serpentes (Colubridae) | *Lampropeltis getula* (YPM 13574) | 50°C; 1mM $MgCl_2$ | 50-54°C; 2,5 mM $MgCl_2$ |
| Serpentes (Colubridae) | *Coluber constrictor* (YPM 13574) | 54°C; 1mM $MgCl_2$ | 50-54°C; 2,5mM $MgCl_2$ |
| Serpentes (Elapidae) | *Micrurus fulvius* (YPM 14096) | 50°C; 1mM $MgCl_2$ | 50-54°C; 2,5mM $MgCl_2$ |
| Serpentes (Colubridae) | *Elaphe obsoleta* (YPM 13086) | 50-54°C; 1mM $MgCl_2$ | 50-54°C; 2,5mM $MgCl_2$ |
| Serpentes (Viperidae) | *Crotalus collilineatus*. (LET, didactic collection) | 50°C; 1mM $MgCl_2$ | 54-56°C; 1mM $MgCl_2$ |
| Serpentes (Dipsadidae) | *Liophis viridis* (CHRP 694) | 50-54°C; 1mM $MgCl_2$ | 50-56°C; 1mM $MgCl_2$ |
| Serpentes (Dipsadidae) | *Apostolepis gaboi* (CHRP 296) | 54 a 56°C; 1mM $MgCl_2$ | 54-56°C; 1mM $MgCl_2$ |
| Serpentes (Boidae) | *Corallus hortulanus* (MZUSP 17291-2/12) | 50°C; 2,5mM $MgCl_2$ | 50°C; 2,5mM $MgCl_2$ |
| Serpentes (Viperidae) | *Bothrops leucurus* (MUFAL 8797-1/2) | 50°C; 1mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |
| Serpentes (Boidae) | *Epicrates cenchria* (MUFAL 8933-1/10) | 50°C; 1mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |
| Serpentes (Colubridae) | *Phimophis guerini* (MUFAL 3665-1/3) | 50°C; 2,5mM $MgCl_2$ | 50°C; 2,5mM $MgCl_2$ |
| Serpentes (Colubridae) | *Pseudoboa nigra* (MUFAL 9701-1/7) | 50°C; 1mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |
| Amphisbaenian (Amphisbaenidae) | *Amphisbaena brasiliana* (UFMT-R 6110) | 50°C; 1mM $MgCl_2$ | 54-56°C; 1mM $MgCl_2$ |
| Amphisbaenian (Amphisbaenidae) | *Amphisbaena mitchelli* (UFMT 8881) | 50°C; 2,5mM $MgCl_2$ | 54-56°C; 1mM $MgCl_2$ |
| Amphisbaenian (Amphisbaenidae) | *Amphisbaena alba* (UFMT-R 5478) | 50°C; 2,5 mM $MgCl_2$ | 56°C; 1mM $MgCl_2$ |
| Amphisbaenian (Amphisbaenidae) | *Amphisbaena steindachneri* (UFMT-R 4151) | 50°C; 1mM $MgCl_2$ | 54-56°C; 1mM $MgCl_2$ |
| Amphisbaenian (Amphisbaenidae) | *Amphisbaena infraorbitale* (UFTM-R 8864) | 50°C; 1mM $MgCl_2$ | 54°C; 1mM $MgCl_2$ |
| Amphisbaenian (Amphisbaenidae) | *Amphisbaena bilabiata* (UFMT-R 4767) | 50°C; 1mM $MgCl_2$ | 54°C; 1mM $MgCl_2$ |
| Amphisbaenian (Amphisbaenidae) | *Amphisbaena miringoera* (UFMT-R 9234) | 50°C; 1mM $MgCl_2$ | 56°C; 1mM $MgCl_2$ |
| Amphisbaenian (Amphisbaenidae) | *Leposternom sp.* (CHRP 561) | 50°C; 1mM $MgCl_2$ | 50-56°C; 1mM $MgCl_2$ |
| Amphisbaenian (Amphisbaenidae) | *Amphisbaena hastata* (CHRP 281) | 50°C; 1mM $MgCl_2$ | 54-56°C; 1mM $MgCl_2$ |
| Amphisbaenian Amphisbaenidae) | *Amphisbaena petrei* (MUFAL 9054-1/2) | 50°C; 1mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |

| | | | |
|---|---|---|---|
| Amphisbaenian (Amphisbaenidae) | *Anops kingii* (MCP14722) | 50°C; 1mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |
| Amphisbaenian (Amphisbaenidae) | *Leposternom polystegum* (CHUNB 30669) | 50°C; 1mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |
| Amphisbaenian (Amphisbaenidae) | *Cercolophia cuiabana* (TMOTT 204) | 50°C; 2,5mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |
| Amphisbaenian (Amphisbaenidae) | *Bronia kraoh* (CHUNB 30767) | 50°C; 1mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |
| Lizard (Gekkonidae) | *Hemidactylus mabouia* (CHRP 848) | 50°C; 1mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |
| Lizard (Gymnophthalmidae) | *Acratosaura mentalis* (CHRP 555) | 50°C; 1mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |
| Lizard (Leiosauridae) | *Enyalius bibronii* (MUFAL 9179-1/3) | 50°C; 2,5mM $MgCl_2$ | 50°C; 2,5mM $MgCl_2$ |
| Lizard (Phyllodactylidae) | *Gymnodactylus darwinii* (MUFAL 4617-1/7) | 50°C; 1mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |
| Lizard (Phyllodactylidae) | *Phyllopezus lutzae* (MUFAL 2836-1/4) | 50°C; 1mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |
| Lizard (Teiidae) | *Ameiva ameiva* (MUFAL 3930-1/8) | 50°C; 2,5mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |
| Lizard (Teiidae) | *Tupimambis merianae* (MUFAL 10610-1/5) | 50°C; 1mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |
| Lizard (Scincidae) | *Mabuia heathi* (CHRP 687) | 50°C; 1mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |
| Lizard (Tropiduridae) | *Eurolophosaurus divaricatus* (CHRP 298) | 50°C; 1mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |
| Lizard (Iguanidae) | *Tropidurus torquatus* (CHRP 639) | 50°C; 1mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |
| Lizard (Iguanidae) | *Iguana iguana* (UNESP Rio Claro) | 50°C; 1mM $MgCl_2$ | 50°C; 1mM $MgCl_2$ |
| Alligator (Alligatoridae) | *Alligator mississippiensis* (cedido por *Vargas et al., 2008)* | 50 - 54°C; 1mM MgCl2 | 54°C; 1mM MgCl2 |

**Table A1.3.** Differences in predicted TFBS in the *CsB* of snakes and lizards.

| TF family | SNAKES | LIZARDS |
|---|---|---|
| C2H2_BTB/POZ | BCL6 (+735, +741, +747, +754, +756, +762, +993)<br>KAISO (+789)<br>HIC1 (+985) | HIC1 (+843, +846)<br>BCL6 (+1047) |
| Interferon-reg | IRF (+740)<br>IRF-1 (+739, +760)<br>IRF-4 (+746)<br>IRF-8 (+745) | |
| TCF_fam | TCF-3 (+739, +1060)<br>LEF1 (+1060) | LEF1 (+1057, +1059)<br>LEF-1, TCF1, TCF-3, TCF-4 (+1058) |
| SMAD_fam | SMAD (+846)<br>SMAD1 (+741, +760)<br>SMAD3 (+763, +765)<br>SMAD4 (+849) | SMAD (+990)<br>SMAD1 (+1057)<br>SMAD3 (+850) |
| NF-AT | NF-AT (+741)<br>NF-AT1 (+745)<br>NF-AT3 (+744) | |
| Zinc finger | DMRT4, FAC1 (+741)<br>REX1 (+848) | |
| CEBP_fam | C/EBP (+739, +846, +984)<br>C/EBPα (+739, +846, +984)<br>CEBPß (+742, +847, +987) | CEBP (+846)<br>C/EBPγ (+749)<br>C/EBPα (+1058) |
| HNF_fam | HNF3α (+742)<br>HNF3A (+744) | HNF4α (+1052) |
| TBP | TBP (+743, +763) | TBP (+1060) |
| FOX_fam | FOXQ1, FOXT1 (+743)<br>FOXD3 (+744)<br>FOXO1 (+745)<br>FOXO3A (+745) | FOXO1 (+1060)<br>FOXO4 (+1058) |
| Homeodomain | CDXA (+744, +762) | NANOG (+1053)<br>CDXA (+1059) |
| HMGI_fam | HMGIY (+744, +763, +764) | HMGIY (+1061) |
| PAX_fam | PAX-2 (+744) | PAX2 (+1057) |
| Cell-cycle | E2F-1 (+744, +745, +850) | E2F6 (+1058) |
| STAT | STAT1 (+745, +990)<br>STAT3 (+1057)<br>STAT4 (+1057)<br>STAT5A (+1057)<br>STAT6 (+1057)<br>PARP (+744, +989) | STAT1 (+1058)<br>PARP (+1060) |
| GATA | GATA-4 (+756) | |
| PBX_fam | PBX1 (+759, +846) | |
| NR0-6 | TR3 (+769) | |

| | | |
|---|---|---|
| | TR4 (+789)<br>SF1 (+854, +991)<br>LRH1 (+989) | |
| ETV_fam | PEA3 (+773) | |
| | TFII-I (+774) | |
| SP1_fam | SP1 (+789) | |
| Ubiquitous_bhlh-zip | USF2 (+789)<br>SREBP (+845)<br>SREBP1 (+845)<br>SREBP2 (+850) | |
| Basic/helix-loop-helix | RBP-Jkappa (+792) | LMAF (+847) |
| PHD_finger | AIRE (+838) | |
| NF-1 | NF-1 (+841, +842, +851)<br>CTF (+851)<br>NF-1A (+842) | |
| Hetero_CCAAT | NY-F (+847) | |
| Ungrouped | CAAT box (+848)<br>cap (+853, +991) | GEN-INI (+850)<br>Cap (+852, +1057) |
| NKX_fam | TTF-1(NKX2-1) (+849, +850, +988, +1054) | TTF-1 (+741, +847) |
| AP-2 | AP-2 γ (+850, +852)<br>AP-2α (+851, +852) | |
| Developmental/cell_cycle | GKLF (+850)<br>GFI1 (+852) | WT1 , WZF1, HELIOSA, WT1, KROX, CACBINDPROTEIN, EGR, GKLF, KLF15 (+764)<br>IK-2, CACBINDINGPROTEIN, KLF15, WT1, MZF1, GKLF, IK-1, EKLF (+764) |
| p53-like | p53 (+851) | |
| RING-finger | RUSH-1α (+851, +989)<br>BRCA1:USF2 (+854, +992) | RUSH-1α (+743, +1055)<br>BRCA1:USF2 (+1056) |
| | CTF/NF1 (+851) | |
| HOX_proximal_fam | HOXA7 (+853) | |
| | BRCA1:USF2 (+854, +992) | |
| SRY | SRY (+855) | SRY (+1057) |
| SOX_fam | SOX (+1060)<br>SOX10 (+990, +1060) | SOX (+851, +1058)<br>SOX2 (+1054)<br>SOX9 (+850, +1056)<br>SOX10 (+1059)<br>SOX17 (+1060) |
| bZIP/PAR | | DBP (+748) |
| AP-1 | | c-MAF (+841) |
| TAL/TWIST/ATONAL/HEN | | NEUROD (+847) |
| DNA_wing | | RFX (+847) |
| AP-4 | | AP-4 (+848, +849) |
| Grainyhead | | XPF-1, LBP-1 (+849) |
| NR3 | | AR (+849), GR (+1051) |

**Table A1.4.** Differences in predicted TFBS in the *CsB* of amphisbaenians and lizards.

| TF family | AMPHISBAENIANS | LIZARDS |
|---|---|---|
| CEBP | C/EBP (+842, +843)<br>C/EBPα (+790, +842, +843)<br>CEBPß (+790) | |
| AP-1 | | c-MAF (+841) |
| C2H2_BTB/POZ | | HIC1 (+843, +846) |
| HOX | | HOXA3 (+845) |
| Tal/Twist/Atonal/Hen | | NEUROD (+847) |
| Grainyhead | | XPF-1, LBP-1 (+849) |
| Homeodomain | CPD (+845)<br>CDXA (+847) | |
| HMGI_fam | | |
| AP-4 | | AP-4 (+848) |
| NF-AT | NF-AT4 (+791) | |
| STAT | | |
| PBX | PBX1 (+842) | |
| bZIP/PAR | DBP (+846) | DBP (+851) |
| MYB | c-MYB (+846, +848),<br>MYB (+850) | |
| RING_finger | BRCA:USF2 (+850) | |
| NKX_fam | | TTF-1 (+847) |
| Other | | CBF1, BEM (+850) |

**Table A1.5.** Differences in TFBSs families between Squamata and the remaining tetrapods.

| TF family | SQUAMATES | NON-SQUAMATES |
|---|---|---|
| AP-1 | +(829, 832) | +(572, 573, 597, 828, 831, 843, 846, 850, 868) |
| AP-2 | | +(602) |
| AREB_fam | | +(997) |
| AT-rich | | +(687, 690, 825) |
| ATF | | +(847) |
| Basic/leucine_zipper | +(832) | +(572, 831, 846, 869, 1074) |
| bZIP/PAR | +(1002) | +(690) |
| C2H2_BTB/POZ | +(303, 316, 818, 832, 1001) | +(223, 522, 753, 817,867, 874) |
| CAR | | +(569) |
| CEBP_fam | +(315, 316, 791, 832) | +(526, 687, 767, 768, 830, 864, 1072) |
| Cell-cycle | +(827, 829) | +(762, 870, 871) |
| COUP | | +(572) |
| CREB_fam, | | +(573, 845, 847, 850) |
| Developmental/cell_cycle | +(314, 974, 998) | +(871) |
| Differentiation | +(826, 832) | +(140, 141, 633, 825, 827) |
| DLX_fam | | +(603, 694) |
| ENGRAILED_fam | +(830) | +(313, 692, 829) |
| ETS-like | +(308, 311, 312, 313, 314, 315, 316) | +(656) |
| ETV_fam | +(314, 315) | |
| FOX_fam | +(823, 829) | +(135, 136, 138, 140, 688, 758, 822) |
| GATA_fam | +(831) | +(829, 847, 848, 849) |
| Helix-loop-helix | (1001) | |
| High-mobility | +(974) | |
| HNF_fam | +(821, 823, 828, 831) | +(135, 136, 139, 141, 142, 684, 687, 758, 759, 761, 762, 824, 827, 830, 864, 865) |
| HMGI_fam | +(309, 313 826, 831, 1002) | +(139, 144, 695, 757, 762, 765, 766, 871) |
| Homeodomain | +(821, 828, 830 ) | +(133, 141, 142, 308, 310, 313, 524, 689, 694, 695, 763, 764, 829, 843) |
| HOX_9_fam | +(831, 999, 1000) | +(597, 602, 690, 830, 874) |

| | | |
|---|---|---|
| HOX_distal_fam | +(870) | +(688, 692, 694) |
| HOX_proximal_fam | +(999, 1001) | +(694) |
| Interferon-reg | +(314, 521, 1001) | |
| ISL_fam | | +(692) |
| LIM_fam, | | +(692) |
| LSF/CP2 | | +(811) |
| MYB | +(526) | +(998, 1000) |
| NF-1 | | +(599) |
| NF-AT | +(313, 315, 316) | +(525, 765, 867, 870, 871) |
| NKX_fam | +(220, 311) | +(691, 693) |
| NR0-6 | +(317) | +(567, 573, 597, 661, 662, 994) |
| NR3 | +(860) | +(525, 570, 597, 602, 827, 871) |
| Other | +(824) | +(685) |
| PAX_fam | +(829) | +(140, 141, 692, 761, 762, 763, 861, 870) |
| PITX_fam | +(1001) | +(997) |
| PBX_fam | +(831) | +(830, 845) |
| POU | +(823, 832) | +(523, 524, 687, 691, 831, 843) |
| RING_finger | | +(692, 1000) |
| RUNT | +(868) | |
| SMAD_fam | | +(139, 140, 690, 842, 1066) |
| SOX_fam | +(827, 829, 865, 997) | +(137, 139, 530, 691, 693, 762, 871) |
| SPI_fam | +(311, 314, 316) | |
| SRY | +(829, 831) | +(140, 141, 830, 1000) |
| STAT | +(205, 313, 314, 317, 828) | +(526, 693, 762, 766, 856, 864, 867, 869) |
| TBP | +(827, 831) | +(141, 142, 143, 600, 692, 764, 765, 825, 830) |
| TCF_fam | +(823, 832, 997, 998) | +(138, 571, 569, 831) |
| Ubiquitous_bHLH-ZIP | +(314, 796) | +(568, 569, 754 ) |
| Ungrouped | +(796, 831, 865, 996) | +(693, 695, 830, 847, 869) |
| Zinc_finger | +(221, 520, 793, 795, 826, 831, 832, 973, 1003) | +(136, 137, 603, 690, 755, 829, 831, 994) |

# *CHAPTER II*

Limb loss in snakes and degeneration of *CNS65* and *Island I* regulatory capacities: morphological evolution left footprints on *Hoxd* genes enhancers.

## 2.1 ABSTRACT

The Serpentes clade comprises elongated squamates either entirely limbless or exhibiting vestigial hindlimbs. Current literature recently incorporated evidence for associations between the origin of snakelike morphologies and evolution of *Hox* genes, but information about *Hox* enhancers during Serpentes evolution remains scarce. Regulation of *Hoxd* gene expression is of particular interest due to their relevance for limb development. Here I combine transgenic assays with bioinformatics to explore footprints of snake limb loss in *Hoxd* enhancers essential for developing the tetrapod limb. Reporter expression of telomeric [*CNS65*] and centromeric [*Island I*] *Hoxd* enhancers from snakes in transgenic mice were performed to investigate if these retained their regulatory capacities. Activities of both enhancers from snakes were abrogated in the limbs of most mouse transgenic embryos, which contrasts with the strong regulatory signals reported in the literature for assays performed with both enhancers from mouse, and also fish for *Island I*. Regulatory signatures in *CNS65* and *Island I* were evaluated through prediction of transcription factor binding sites (TFBS). Comparisons between snakes and 'limbed reptiles' (lizard, alligator, turtles and birds) revealed that 43-52% of TFBS predicted in all 'limbed reptiles' except snakes were limb-related; when snake enhancers were compared with those from other limbed amniotes (reptiles and mammals), 44-70% among the predicted TFBS diverging between the two groups were limb-related. I identified one segment that potentially represents a stilopodium/zeugopodium-specific element in *CNS65* and three possible autopodium-specific elements in *Island I*, comprised by several binding sites for limb-related transcription factors estimated in all limbed amniotes that seem lost by snakes. In conclusion, loss of *CNS65* and *Island I* regulatory capacities in snakes suggests existence of regulatory segments specifically dedicated to limb development in these enhancers that apparently degenerated after limb loss. Regulatory signatures suggest that snakes lost at least four segments that in limbed amniotes are likely enriched by binding sites for transcription factors essential for limb development. Our results suggest that these specific regulatory signatures could represent the major sites which degeneration caused by relaxation during evolution of Serpentes impaired ability of these enhancers to regulate *Hoxd* expression on developing limbs, providing clues on the specific mechanisms of *Hox* genes molecular evolution linked to the origin of snakes. Additionally, they reveal segments within *CNS65* and *Island I* that possibly belong to regulatory modules strictly associated with vertebrate limb development.

## 2.2 INTRODUCTION

The snakelike morphology characteristic of Serpentes remarkably diverges from standard vertebrate body shapes due to the lack of limbs and extreme elongation of the trunk; this phenotype evolved from lacertoid lizard ancestors exhibiting limbs (Coates & Ruta 2000; Tchernov et al. 2000; Bejder & Hall 2002) not only in this clade but also in many other squamate lineages (see Sanger & Gibson-Brown 2004; Wiens et al. 2006; Brandley et al. 2008; Woltering et al. 2009; Kohlsdorf et al. 2010). All snakes lack forelimbs and scapular girdles (Cohn & Tickle 1999; Woltering 2012). Colubroids, commonly referred as "advanced" snakes, are entirely limbless and have axial skeletons composed by numerous rib-bearing vertebrae; these trunks seem more homogeneous than those of lineages having earlier origins (scolecophidians and booids), which retained rudimentary pelvic girdles and a vestigial femur (Bellairs & Underwood 1950; Cohn & Tickle 1999). Evolution of a phenotype so remarkably divergent from its ancestor both regarding structures gain (e.g. number of vertebrae) and traits loss (i.e. limbs) has intrigued scientists for decades, being subject of intense debate focusing not only on the historical circumstances of snake origins (see Caldwell & Lee 1997; Coates & Ruta 2000; Tchernov et al. 2000; Rieppel et al. 2003; Vidal & Hedges 2004; Lee 2005; Apesteguía & Zaher 2006), but also on the developmental variations that resulted in such elongated limbless morphologies (see Cohn & Tickle 1999; Wiens & Slingluff 2001; Woltering et al. 2009; Woltering 2012).

Developmental changes identified as possible candidates for explaining the origin of Serpentes have placed *Hox* genes at the focal point of this discussion. As detailed in the General Background section, current literature sustains that shifts in *Hox* gene expression patterns and downstream developmental pathways are the likely connection between morphological transitions and developmental pathways in the context of snake origins, as

these genes play essential roles in patterning both axial and appendicular axes (Cohn & Tickle 1999; Woltering et al. 2009; Di-Poï et al. 2010; Woltering 2012; Guerreiro et al. 2013, but see Head and Polly 2015; Sanger and Gibson-Brown 2004 for controversial opinions). Snake evolution has registered specific signatures in the sequence of *Hoxa13* first exon (Kohlsdorf et al. 2008), a pattern apparently absent in the sequence of other snakelike lineages, as amphisbaenians and caecilians (Singarette et al. 2015). Together, these examples reinforce the idea that the way *Hox* genes are expressed and how they interact with other molecules ought to have been modified during the evolution of snakes. Paradoxically, our knowledge about *Hox* enhancers in snakes remains scarce. The present study aims to fill this gap by testing whether *Hox* enhancers involved in the development of tetrapod limbs remain regulatory functional after ~100 million years of limb loss in Serpentes (Brandley et al. 2008; Vidal et al. 2010; Pyron et al. 2013; Reynolds et al. 2014; Head 2015). This question is particularly intriguing given that *Hox* genes are often involved in multiple functions during embryo development (e.g. Fromental-Ramain et al. 1996; Deschamps & van Nes 2005; Scott et al. 2005; Zakany & Duboule 2007; Archambeault et al. 2014), and therefore sequence changes in these genes likely involve considerable pleiotropy (Sivanantharajah & Percival-Smith 2015).

In the context of limbless morphologies evolution, regulation of *Hoxd* genes expression is of particular interest due to their relevance for development of the autopodium, a feature absent in Serpentes. One can now study in limbless organisms some of the key regulatory elements involved in tetrapod limb development because the regulatory dynamics of *Hoxd* gene expression has been recently very well characterized in mice limb buds (Gonzalez et al. 2007; Rinn et al. 2007; Montavon et al. 2008; Montavon et al. 2011; Tschopp & Duboule 2011; Tschopp et al. 2011; Delpretti et al. 2013; Soshnikova 2014). During limb development (as well as when other vertebrate appendages

are developing; see Archambeault et al., 2014), *Hoxd* gene expression occurs in two independent phases that will establish proximal and distal segments (Nelson et al. 1996; Tarchini & Duboule 2006). The initial transcription phase is a homogeneous and sequential expression of *Hoxd1-11,* which together with other genes coordinate patterning of stilopodium and zeugopodium (Zakany & Duboule 2007; Montavon et al. 2011). The second transcriptional phase defines autopodium formation and involves expression of *Hoxd9-13* genes in an inverted spatio-temporal collinear mode (Tickle 2006; Montavon et al. 2008; Montavon et al. 2011). These transcriptional phases are regulated by several overlapping elements including noncoding RNAs (Soshnikova 2014) and cis-regulatory elements (CREs), among which are intergenic and global enhancers (Di-Poï et al. 2009; Montavon et al. 2011; Andrey et al. 2013). Regulatory landscapes comprise many CREs that are located far away from the genes they regulate (Deschamps 2007; Spitz 2010; Montavon et al. 2011; Andrey et al. 2013): regulation of the first transcriptional wave during stylopodium/zeugopodium development engages elements from the telomeric landscape (downstream of *HoxD* genes; see Tarchini & Duboule 2006; Andrey et al. 2013; Woltering et al. 2014), while autopodium development involves a shift between regulatory domains and is controlled by a centromeric region located upstream of *Hoxd13* (Dixon et al. 2012; Nora et al. 2012; Andrey et al. 2013).

The essential roles for limb development performed by some CREs located in these regulatory landscapes may guide our view towards regulatory elements likely modified during the evolution of limbless phenotypes. For example, *Hoxd* gene expression along the anterior-posterior limb bud axis is modulated by the *early limb control region* (ELCR) located at the HoxD telomeric regulatory landscape (Zákány et al. 2004; Deschamps & van Nes 2005; Tarchini & Duboule 2006; Schneider et al. 2011), which comprises several conserved noncoding sequences (CNS) assigning enhancers for limb development (Andrey et al.

2013). Among these, one fragment named *CNS65* induces the initial transcription of *Hox* genes in a broad proximal domain of developing limb buds and, when deleted, reduces in about 30-40% the physiological levels of *Hoxd8-11 mRNA*s at the most proximal portion of mice limb buds (Andrey et al. 2013). The centromeric landscape also comprises several regulatory elements involved in the development of limb distal portions and the central neural system (Spitz et al. 2003; Gonzalez et al. 2007; Montavon et al. 2011), and a 600 Kb region so-called genic desert located between *Lnp* and *Atp5g3* contains a 'regulatory archipelago', comprised by *Island I* to *V*, essential for autopodium development (Montavon et al. 2011). The complete transcriptional activation of terminal *Hoxd* genes requires interactions among islands I and IV, and other enhancers such as the ones in the global control region (*GRC)* and *Prox*; mouse *Island I* reporter expression shows that its activity occurs from the proximal extremity of the bud to the top of the paddle across the region of digits III and IV in E12.5 mouse embryos (Montavon et al. 2011). Enhancers located in these two regulatory regions, such as the *CNS65* (ELCR at the HoxD telomeric regulatory landscape) and *Island I* (genic desert at the HoxD centromeric regulatory landscape), represent therefore proper candidates for testing if the regulatory machinery involved in tetrapod limb development has degenerated in Serpentes. Differences in *Hox* gene expression associated to morphological evolution likely entail variation in CREs (Warren et al. 1994; Averof & Akam 1995; Carroll 1995; Burke et al. 1995) because these genes are often involved in multiple functions (Sivanantharajah & Percival-Smith 2015) and a complex, fine-tuned association of several Hox regulatory elements confers both specificity for gene expression in particular structures and evolutionary flexibility (Duboule 2007; Gonzalez et al. 2007; Deschamps 2007; Montavon et al. 2011). Consequently, sequence variation affecting, for example, transcription factor binding sites (TFBS) of a given enhancer may impose a functional loss (Zinzen et al. 2006) that in *Hox* genes could affect

only one developmental function or involve considerable pleiotropy, depending on the specificity of the regulatory module (see Tschopp et al. 2011). This matter as approached in the context of Serpentes evolution by focusing on one element from HoxD telomeric (*CNS65*) and another one from the centromeric (*Island I*) regulatory landscapes to test if regulatory capacities of limb-specific enhancers degenerated in limbless lineages. Specifically, I combined bioinformatics predicting TFBS with reporter expression of *CNS65* and *Island I* from snakes in transgenic mouse. I retrieved genome sequences for *Phyton bivittatus* (Phytonidae, Boidae) (Fig. 2.1A), a snake presenting femoral vestiges, and *Ophiophagus hannah* (Elapidae, Colubroidae) (Fig. 2.1B), a completely limbless snake, to construct plasmids for transgenic assays; the sequences were also compared with limbed amniotes regarding predicted TFBS. I predict that evolution of Serpentes involved functional impairment of *CNS65* and *Island I* regulatory capacities in developing limbs, a degeneration of these sequences characterized by the loss of TFBS involved in limb developmental pathways.

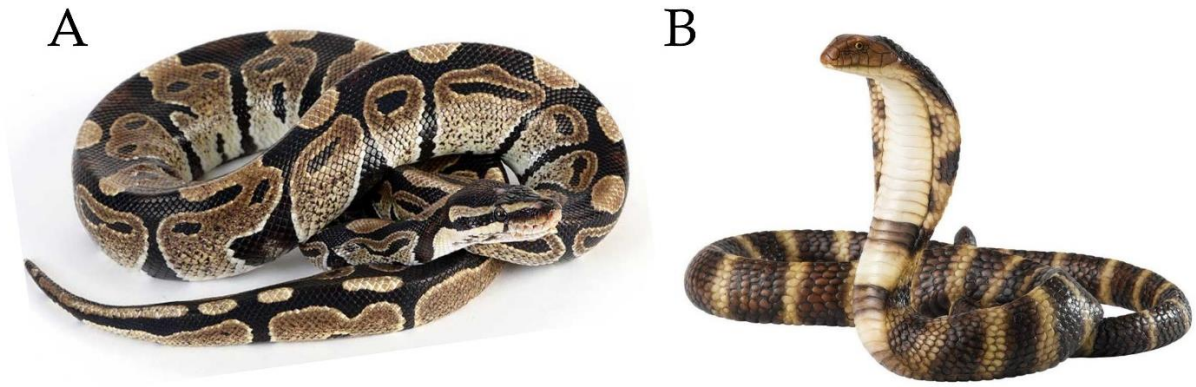


**Figure 2.1.** Images of representatives from *Phyton molurus bivittatus* (Phytonidae, Boidae), bearer of femoral vestige (A), and *Ophiophagus hannah* (Elapidae, Colubroidae), completely limbless (B).

## 2.3 METHODS

*Gene reporter assays: plasmid generation and mouse transgenesis*

The fragments studied were retrieved from *Whole-genome shotgun contigs (wgc)* deposits on Genbank by their similarity with mouse sequences gently given by Prof. Igor Schneider, using BLASTn tool: *CNS65* from *Phyton mulurus bivittatus* (1,168 bp; AEQU02037710.1) and *Ophiophagus hannah* (1,178 bp; AZIM01000055.1); *Island I* from *P. bivittatus* (1,060 bp; AEQU02066876.1). The sequences were synthesized and inserted in pUC57 vector by FASTBIO (Ribeirão Preto-BR). The primers pUC77-F, 5'-GTAAAACGACGGCCAGTG-3'; pUC77-F, 5'-GGAAACAGCTATGACCATG-3', were used to amplify those sequences from snakes by PCR, using conditions of 48-50°C annealing temperature, 1mM $MgCl_2$, and extension time varied from 1 minute for *CNS65* and 1.5 min for *Island I*. When amplyfication was succesfull, PCR were repeated using doubled volum (50 µl), which were run on 2% agaroses gel electroforeses and purified using *QIAquick Gel Extraction Kit* (*Qiagen)*. Products were ligated to PCR8-Topo TA (Invitrogen) and cloned into *Top-10* or *DH5α* thermo-competent bacteria according to manufacturers' instructions. Plasmids were extracted using *PureYield™ Plasmid Miniprep System* (Promega) and sequenced to confirm insert. PCR8-Topo TA was used because it can be easily ligated to DNA fragments by TOPOR cloning system and contains attL sites from lambda bacteriophage, specific for rapid and highly efficient recombination with any vector containing attR sites. PCR8-Topo TA plasmid DNA carrying sequences of *CNS65* from *P. bivittatus* and *O. hannah*, and of *Island I* from *P. bivittatus*, were cleaved using the enzime EcoRI as a second confirmation of inserts. *CNS65* clones 1 and 4 from *P. bivittatus* and clone 3 from *O. hannah* (Fig.2.2), were picked for LR recombination reaction. All of the *Island I* clones from *P. bivittatus* contained insert (Fig. 2.3).

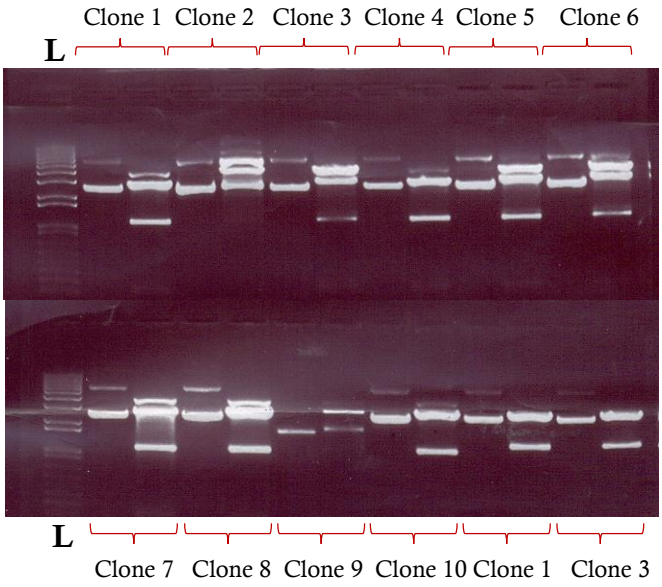


**Figure 2.2.** Agaroses gel for diagnosing clones containing insert of expected size cloned into vector PCR8. Clones 1 to 6 (top) and 7 to 10 (bottom) indicated the presence of *P. bivittatus* (Pmb) *CNS65* insert. Clones 1 to 3 (bottom) correspond to *O. hannah* (Oh) *CNS65*. First well in each clone corresponds to intact plasmid DNA, and second well corresponds to plasmid cleaved with EcoRI enzime. Clones 1 and 4 of Pmb *CNS65* and clone 3 of Oh *CNS65* were picked for LR reactions. L=ladder.

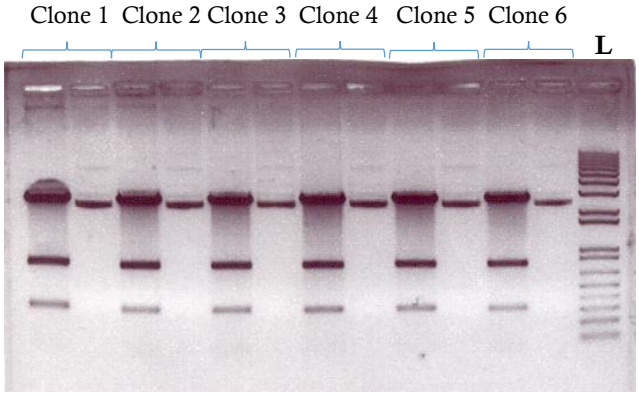


**Figure 2.3.** Agaroses gel for diagnosing presence of *P. bivittatus Island I* insert into PCR8. Each clone cleaved with EcoRI enzime, and their intact DNA, respectively, were run in adjacent wells for each clone (1 to 6). All clones apparently contained expected insert. L=ladder.

LR recombination reaction from Gateway System (Invitrogen) was subsequently performed in order to transfer inserts into the Gateway-Hsp68-LacZ vector, according to

manufacturer's protocol. 50-150 ng (1-7 µl) of PCR8 plasmid DNA containing respective inserts were added to 1 µl of Hsp68-LacZ vector, topping with TE buffer, pH 8.0, up to 8 µl. Every sample mix prepared was homogenized, briefly centrifuged, and added to 2 µl LR Clonase™ II enzimatic mix previously left resting on ice for 2 minutes and vortexed twice for 2 seconds each. Final mixtures were again homogenized, vortexed twice and quickly centrifuged. Reactions were incubated at 25°C for 3-4 hours. Subsequently, 1 µl of proteinase K solution was added to each sample to stop enzymatic reaction, following up with vortexing and incubating at 37°C for 10 minutes. Products of LR recombination reaction were cloned in thermo-competent *Top-10* bacteria, and 6-8 colonies were inoculated in liquid LB overnight at 37°C. Plasmids were extracted using QIAprep Spin Miniprep Kit (Qiagen) and subjected to enzimatic cleavages and sequencing to confirm inserts. Band patterns expected after enzimatic cleavages were predicted using the software *NEBcutter V2.*0 (www.neb.com/external-links/nebcutter), which performs *in silico* cleavage of sequences, in case they present binding sites for the chosen enzymes. For instance, the *Island I* fragment studied has a binding site for EcoRI enzyme (Fig. 2.4).

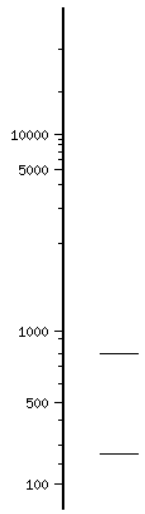


**Figure 2.4.** Pattern of bands expected after cleaving *Island I* from *P. bivivtatus* with EcoRI enzime, estimated by NEBcutter V2.0 software.

Besides sequences themselves, HSP68-lacZ vector also comprises binding sites for EcoRI and EcoRV enzymes. The patterns revealed in the agarose gel were compared with the patterns estimated for each sequence. For example, combining information obtained from enzymatic digestions with both enzymes (EcoRV, Figs 2.5 and 2.7; EcoRI, Figs 2.6 and 2.8), clones 1, 2, 3, 5, 6, 7, 8 of *P. bivittatus CNS65*, and clones 1, 2, 4 and 6 of *O. hahhah CNS65* apparently contained respective inserts of interest. Estimation of HSP68-lacZ clones containing *P. bivittatus Island I* as inserts were confirmed using the same method. Finally, the selected clones were sequenced for ensuring that they were defenitively suitable for transgenic assays. Most of the aforementioned experiments were accomplished in Prof. Igor Schneider's laboratory at *Universidade Federal do Pará*.

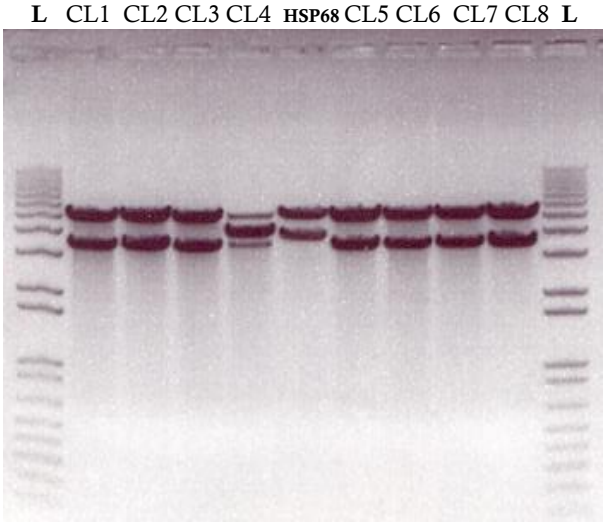


**Figure 2.5.** Cleavage with EcoRV for diagnosing *P. bivittatus CNS65* enhancer cloned into HSP68-lacZ vector. Aparently, clone 4 was the only one lacking insert. L=ladder.

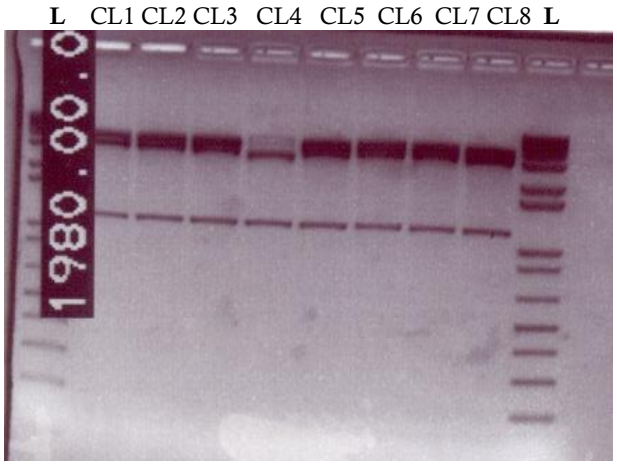


**Figure 2.6.** Cleavage using EcoRI for diagnosing *P. bivittatus CNS65* insert cloned into HSP68-lacZ vector. This cleavage confirms information obtained by cleavage with EcoRV, showing that clone 4 lacks insert. L=ladder.

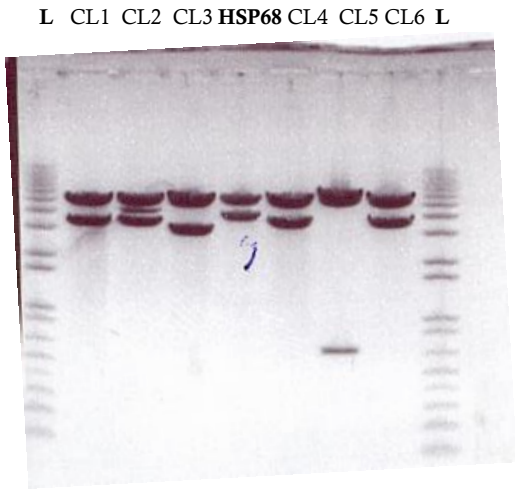


**Figure 2.7.** Cleavage using EcoRV for diagnosing *O. hannah CNS65* insert cloned into HSP68-lacZ vector. Apparently, clones 1, 2, 4 and 6 contained expected inserts. L=ladder.

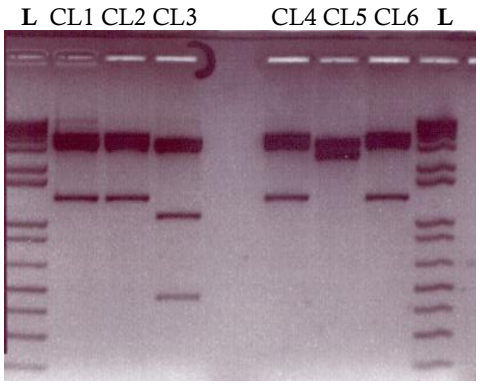


**Figure 2.8.** Cleavage with EcoRI enzyme for diagnosing *O. hannah CNS65* insert cloned into HSP68-lacZ vector. Apparently, clones 1, 2, 4 and 6 contained expected inserts. L=ladder.

The vector backbone was excised by digesting 50 µg of each plasmid with SalI enzyme, purified using *QIAquick Gel Extraction Kit* (Qiagen), and the DNA to be injected was further purified using a standard ethanol precipitation and diluted to 2 ng/µL, which was performed at the University of Chicago, by Prof. Igor Schneider´s group. Pronuclear injections of CD1 mouse embryos, followed by embryos harvesting, staining and fixing, were performed by *Cyagen* (Chicago-EUA). Each construct resulted in at least 3 LacZ-positive embryos displaying similar reporter activity.

*Estimation of Transcription Factor Binding Sites*

Orthologous sequences of *CNS65* and *Island I* from nine limbed amniotes were retrieved from Genbank for comparisons with snake sequences (accession numbers are provided in Table A2.1): the lizard *Anolis carolinensis*, the alligator *Aligator mississippiensis,* the avian *Gallus gallus* and *Columba livea,* the turtles *Pelodiscus scinensis* and *Chrysemis picta,* and the mammals *Mus musculus, Rattus norvergicus e Homo sapiens.* Alignments were produced using ClustalW algorithm (Thompson et al. 1994) implemented in the software BioEdit sequence alignment editor, for adjusting 3'-5' orientation and assuring equivalence of fragments compared. Transcription Factor Binding Sites (TFBS) were predicted individually for each sequence of the alignment and also for different groups of species, using MultiTF/Mulan online software (Ovcharenko et al. 2005). I performed multiple TFBS comparisons by contrasting different groups of species, defined as follows: 1) snakes (*P. bivittatus* and *O. hannah*); 2) squamates (snakes and the lizard *Anolis carolinensis*); 3) snakes and the alligator *Alligator mississippiensis*; 4) 'reptiles' (snakes, lizard, alligator, the turtles *Pelodiscus scinensis* and *Chysemys picta*, the birds *Gallus gallus* and *Columba livia*); 5) 'reptiles' and mammals (all species listed in 4 and *Homo sapiens*, *Mus musculus* and *Rattus norvegicus*); 6) 'reptiles' excluding the snakes; 7) mammals and 'reptiles' excluding snakes.

Results obtained using different minimum detection thresholds were compared and we settled the value for complete matrix profiles as 0.85, which allowed prediction of a considerable amount of TFBS while avoiding noise in the analyses resulting from excessive permissiveness. I also performed an 'optimized for function' prediction in order to narrow down comparisons and access information on a smaller group of likely more relevant candidate TFBS that were lost or acquired in Serpentes. MultiTF online software does no process pre-aligned sequences nor sequences containing gaps, performing instead inside alignments prior to TFBS prediction. Therefore, it was not possible to have the same interval positions for equivalent TFBS in all groups compared in every comparison. In groups including snakes, I used *P. bivittatus* TFBS positions as a reference for data description. For groups excluding snakes, the reference used was *A. mississippiensis*. The TFBS predicted were plotted in Microsoft Excel, where I searched for TFBS conserved in sequences from predefined sets of species that were not predicted in equivalent positions in snake sequences. I also annotated TFBS predicted in snake sequences that were not identified in the regulatory elements from the limbed groups. This approach enabled identification of TFBS gains and losses during evolution of *CNS65* and *Island I* in Serpentes, although our discussion focuses on TFBS losses possibly related to limb loss.

## 2.4 RESULTS

In order to investigate whether the regulatory element *CNS65* from the limbless lineage Serpentes retains the regulatory capacity in limb buds that has been reported for mouse (Andrey et al. 2013), I cloned this enhancer from the snakes *P. bivittatus* and *O. hannah* into vectors containing the LacZ gene and the HSP68 minimal promoter (Amemiya et al. 2013) and examined their gene reporter expression in mouse embryos. For the same purpose, the regulatory element *Island I* from the snake *P. bivittatus* was used in similar transgenic assays, compared with mouse (Montavon et al. 2011) and the nonteleostean spotted gar fish, *Lepisosteus oculatus* (Gehrke et al. 2014). The sequence of *Island I* from *O. hannah* available in Genbank was too fragmented and therefore could not be tested in transgenic assays. *CNS65* and *Island I* gene reporter expression assays show that both CREs in Serpentes have apparently lost their regulatory capacities over gene expression in limb buds, as no activity was identified in developing limbs of mice embryos when snake sequences were injected in mouse pro-nucleous. In the transgenic assays performed with *CNS65* from *O. hannah*, reporter expression was absent in three among six positive embryos and strongly reduced in two among six positive embryos (Fig. 2.9A). Degeneration of snake *CNS65* regulatory capacity in limb buds was even more conspicuous for the sequence from *P. bivittatus CNS65*: reporter expression was absent in all of the eight positive embryos (Fig. 2.9B). Similarly, lack of regulatory activity in limb buds was also observed for *Island I* sequences from *P. bivittatus*: all four positive embryos did not exhibit any reporter expression in developing limbs (Fig. 2.9D). These weak or absent expression patterns remarkably differ from the strong activities reported by assays using corresponding fragments from mouse [Andrey et al. (2013) for *CNS65* (Fig. 2.9C) and Montavon et al. (2011) for *Island I*] (Fig. 2.9F) and gar [Gehrke et al., (2015) for *Island I* (Fig. 2.9E)]. Results from transgenic assays corroborate thus our hypothesis that evolution of Serpentes

involved functional impairment of *CNS65* and *Island I* regulatory capacities in developing limbs, a finding that places these two elements as limb-specific regulatory regions passive to sequence degeneration when the process they are involved no longer takes place in these limbless tetrapods.

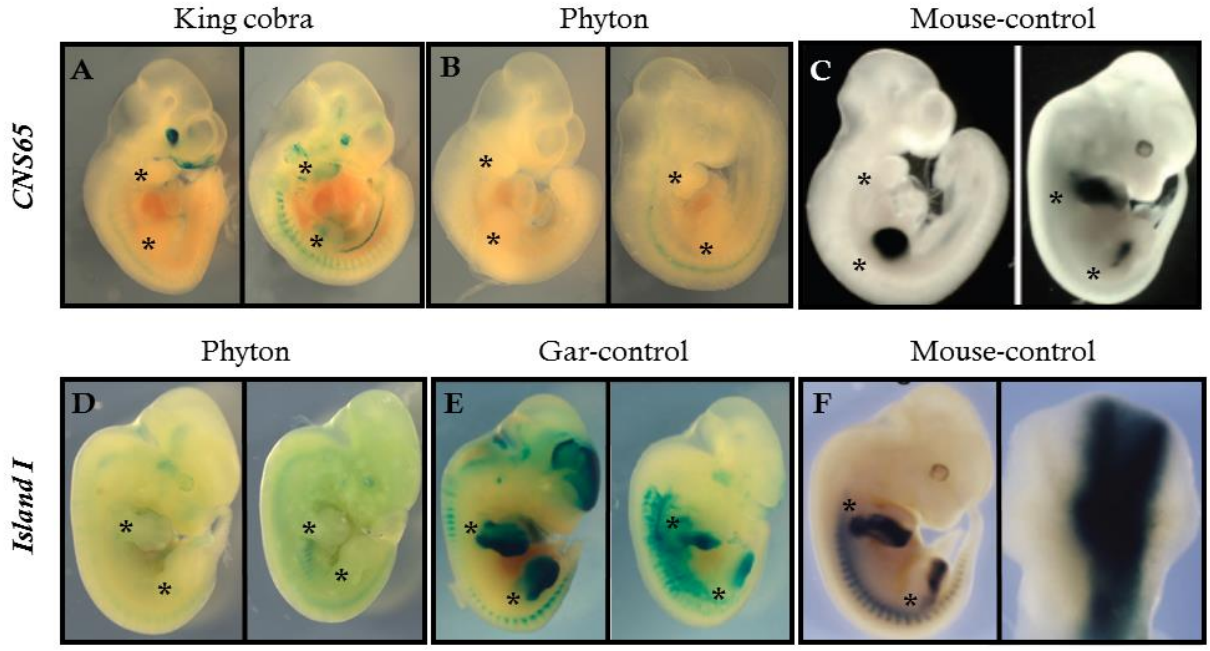


**Figure 2.9.** Gene reporter expression assays reveal that *CNS65* and *Island I* enhancers from snakes have their activities reduced or abrogated in transgenic E 11.5 (*CNS65*) and 12.5 (*Island I* ) mouse embryos. Superior left panel (**A**) illustrates two representative sample of reporter expression of *CNS65* from the king cobra *Ophiophagus hannah*, where expression was absent in 3/6 embryos (left) and strongly reduced in 2/6 embryos (right). Middle left panel (**B**) shows reporter expression of *CNS65* from the snake *Phyton bivittatus*, where in 8/8 embryos there was no reporter expression. *CNS65* mouse-control (**C**) on the top right corresponds to Fig. 2B in Andrey et al., (2013), exhibiting gene reporter expression of mouse *CNS65* in mouse, where strong activity is registered in either fore or hindlimbs, representing a sample of 23/28 embryos. Bottom left panel (**D**) illustrates reporter expression of *Island I* from the snake *P. bivittatus*, where in 4/4 embryos the reporter expression is absent, contrasting to the *Island I* gar-control (**E**) on bottom middle, performed using *Island I* from a nonteleostean spotted gar fish, *Lepisosteus oculatus,* and **(F, left)** in transgenic mouse, where strong activity is reported in both fore and hindlimbs (3/7 embyos; see Fig4 and FigS3 in Gehrke et al., (2015). In **(F, right)**, activity in the autopodia is strongly distributed from proximal to distal extremities of the paddle over the region of digits III and IV.

In face of the evidence that the elements *CNS65* and *Island I* have lost their regulatory capacities for limb development in snakes, I tested the derived prediction that functional degeneration of these elements involved loss of estimated TFBS for genes involved in limb developmental pathways. Specifically, I searched for regulatory signatures, aiming to identify segments that putatively experienced high evolutionary rates after limb loss in Serpentes, an investigation that would allow prediction of limb-specific regulatory regions within *CNS65* and *Island I*. This search was based on *in silico* TFBS predictions that compared snake sequences with those from other limbed amniotes. Aiming to infer those TFBS estimated exclusively in the *CNS65* and *Island I* snake sequences, I predicted the TFBS that were common to: 1) both snake species; 2) snakes and the alligator *A. mississippiensis*; 3) snakes and the lizard *A. corolinensis*; 4) snakes and all other 'reptiles' (including two birds); and 5) snakes and all limbed amniotes considered here. Descriptions of TFBS 'exclusive' to a given group, as well as those that were 'gained' or 'lost' in one specific lineage, refer to individual TFBS in each position specified, and do not mean absence or exclusiveness of that TFBS along the entire sequence of the group. After annotating predicted TFBS exclusive to snakes (in the positions listed in Tables A2.2 and A2.3), I predicted those lost in snake sequences by accessing sites predicted in all 'reptiles' except snakes and also those predicted in all amniotes except snakes.

The segments 1-196, 242-305, 676-691, 855-902, 990-1012, 1045-1054 in *CNS65* sequence from *P. bivitattus* correspond to regions in this element that are poorly conserved among species compared. The *CNS65* snake sequence has apparently lost some TFBS predicted in all other 'reptiles' (Table A2.2): PEA3 (position 202-208); NKX62, CDXA, OTX or STAT4, GATA3 or GATA6, CEBPB (247-267); FOXO1, SRY, GATA, GATA1, CDC5, GATA1 or GATA6, GATA (331-349), some of which may have been substituted by CAP (327-334) and PITX2 (330-340) in snakes; STAT5A (354-361); NKX25B or

NKX25, HMGIY, CDXA, CEBPB, CEBP, NFAT, NFAT (368-385); HELIOSA (375-385); CEBP or CT1, CDXA, HNF3ALPHA, NKX62, XFD1, XFD2 or HFH8, LPOLYA, TBP, TBP, CDXA, STAT4, CDXA, PBX1, POU6F1, LPOLYA, TBP, CDXA (376-393); CDXA, NKX25B or NKX25 (387-395); RUSH1A, BRCA (394, 403); RUSH1A (413-422); STAT4 (416-423); CREBP1CJUN (451-458); AP2REP or CDXA, CAP (494-500); ETS1 (506-520); GFI1, AP2REP, IK2, RBPJK (521-535), possibly substituted by CAP and PITX2 (327-340) in snakes; OTX, CDXA or NKX25 (539-546); (614-621); SMAD, CBF, CEBP, TEF1, AML1 (616-638); CDXA (670-676); CDXA (717-723); STAT1 or STAT4 or STAT5A (752-759); STAT4 or STAT6 (921-928); FOXD3 (997-1008); and NKX25 (1047-1053); P53, TBP, CDXA (1088-1113), possibly substituted by CAP and EN1 (1085-1094) in snakes. When the 'optimized for function' parameter was set, searches of TFBS narrowed investigation towards functional relevance, and the sites predicted to all 'reptiles' except for snakes were: GATA3, SRY, CDC5, NKX, FD1, HFH8, POLYA POLYA (250-401); PAX2 (443-454); and ETS1 (506-520).

The TFBS predicted in all the other vertebrates compared in our sample that were possibly lost by snakes *CNS65* are: FOXO1, PAX2 (383-392); YY1 (395-403); TBX5, CAP (418-427); CAP (493-500); CP2 (517-527); CAP, MEIS1, TGIF, CREB (559-568); GATA3 (566-575); CDXA (567-573); TBP (569-576); TBP (570-576), possibly substituted by AP3 (562-569) in snakes; TCF4 (572-579); STAT4 (979-986) (Table A2.2). Among those, the segment 559-579 is a strong candidate for functional relevance because 5 (MEIS, TGIF, CREB, GATA3, TCF4) among 8 TFBS lost by snakes present limb development-related functions (Fig. 2.10). The 'optimized for function' parameter only estimated three TFBS of *CNS65* sequence that are common to all the vertebrates of our study except for snakes: OCT-1 (422-434); SRF (478-495); SMAD (596-604).

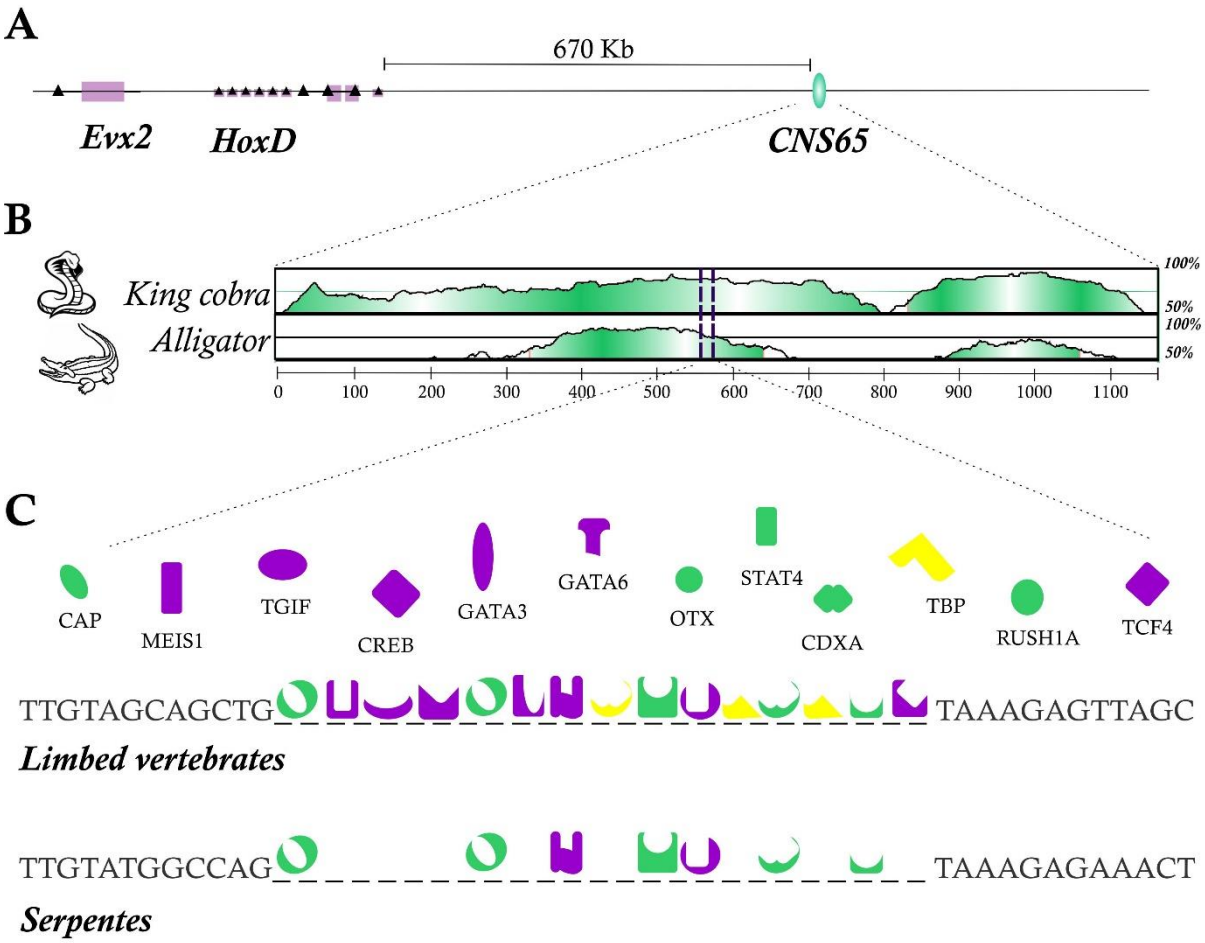


**Figure 2.10.** (**A**) Representation of *CNS65* chromosomic location in the telomeric domain upstream to the HoxD cluster. (**B**) Topologic graphic designed on mVista online software illustrating the percentage of similarity of the *CNS65* studied fragment from the king cobra *Ophiophagus hannah* and *Alligator mississippiensis* with the snake *Phyton bivittatus*. Note that the sequence from *O. hannah* presents a much higher degree of similarity in relation to that of *P. bivittatus* than the alligator's. A putative signature of limb loss was predicted in a conserved region of the *CNS65* fragment (559-579 in *A. mississippiensis*), marked between purple bars in **B**. This signature is illustrated in **C**, showing that binding sites for CAP, MEIS1, TGIF, CREB, GATA3, CDXA, TBP and TCF4 that appear in *CNS65* of all limbed vertebrates studied were putatively lost in correspondent segments of *P. bivitattus* and *O. hannah*. TFBS represented in purple correspond to those functionally related to limb development according to the literature, whereas the one represented in yellow and green correspond to those presenting other functions. The green ones mark the TFBS that were retained in the *CNS65* sequences from snakes, while the non-related to limb development TFBS that were apparently lost are highlighted in yellow.

*Island I* is low conserved among all the compared species in the segments correspondent to positions1-518, 793-875, 894-936, 1099-1145, 1246-1283 in *Island I* sequence of *P. bivitattus*. The TFBS predicted in *Island I* sequences of all `reptiles´ except for snakes are: MYOGENIN (585-592); CAP, CAP, CAP, GATA6, SRY, PBX1, PAX2, STAT1, HGMIY, CDXA (611-639); TFIIA, PR (660-672); GR (666-673); CDXA (701-707); CAP (706-713); OCT-1 (711-724); CAP (712-719); CDXA (750-756); FOXM1 (751-759); MYOGENIN, NKX25B or NKX25 or MYC, CAP, NKX25, USF2 (875-882), possibly substituted by p53 (879-888) in snakes; MSX1 (922-930), possibly substituted by IPF1 or NKX62 (944-955) in snakes; CEBPA (925-938) possibly substituted by S8 (944-959) in snakes; CEBPΔ (926-937), possibly substituted by CDXA or NKX25L (944-959) in snakes; CAP or TST1 (939-946); ETS (939-953); HSF1, PAX2, CDXA or LEF1B, RUSH1A, LEF1, CETS1P54 (1021-1037), possibly substituted by CAP, NKX25L or CDXA, STAT1 or STAT3 or STAT4 or STAT5A (1042-1050) in snakes; CETS1P54 (1026-1035); CEBPΔ (1027-1038) possibly substituted by CDXA and NKX25L (1046-1057) in snakes; TEF1 (1033-1038); CEBP (1082-1093); ΔEF1 (1139-1149); BRCA, RFX (1142-1152); HOXA4, AREB6 (1154-1169); CAP (1160-1167); AML1, EN1 (1172-1179) (Table A2.3). Based on the 'optimized for function' parameter, the sites predicted in common to all 'reptiles' except for snakes are: SRY (735-741); PAX8, PBX1, CDXA (751-768); ER, T3R (769-779); HNF4 (772-786); COUP (773-786); NKX25B (876-882); PXR (928-939); RORA2 (932-944); PPARG, TFIII, PEA3 (933-949); LEF1, LEF1TCF1, LEF1B (970-981); CREB (995-1006); HSF1, ELK1, CEBP, TEF1 (1021-1038); TATA (1048-1057); HMGIY (1069-1083); LEF1B, NCX (1116-1130); TBX5, IF1, RFX (1140-1160).

The TFBS possibly lost by snakes *Island I* sequences, which are predicted in all the other vertebrates compared are: NKX25, ΔEF1, LMO2COM (568-595); E2A, USF2 (587-

594); GATA6, CAP, GATA3, CAP, ERR1 (652-673); T3R, HNF4 (665-673); GATA3, GATA6, GATA (685-693); CAP, GATA6 (718-730), possibly substituted by STAT6 (723-730); NKX62 (750-761), possibly substituted by CEBP (756-769); OTX, STAT1 or STAT4, CDXA, CEBPγ, LPOLYA, POU1F1, CEBPΓ, TBP, PBX1 (752-762); CDXA, CDXA, PBX1 (756-765); PEA3 (939-945); AP3, NKX25 or CDXA, STAT4 or STAT6 (942-952); CDXA (959-965); AP1, CREBP1 or CREB, AP1, CREB (997-1005); STAT3 (1032-1039); FOXM1 (1074-1082); LEF1, CDXA or LEF1B, RUSH1A, LEF1 (1115-1125); S8, NCX, CEBP, HOXA4 (1120-1135), possibly substituted by NKX3A and CDXA (1163-1176); CEBP, NFAT, NFAT (1124-1136), possibly substituted by CHCH (1181-1186); STAT1 (1128-1135); STAT4 (1128-1135); STAT6 (1128-1135); TBX5, TBX5 (1140-1151); CEBP (1144-1156); CEBP, OSF2 or AML1 (1149-1156); TBX5 (1160-1171); TBX5 (1162-1169) (Table A2.3A). The segments especially rich in TFBS related to limb developmental roles that were apparently lost by this enhancer in snakes are 568-592 (NKX25, ΔF1, LMO2COM, E2A, USF2), 752-765 (STAT1/4, CEBPγ, PBX1) and 1115-1154 (LEF1, S8, CEBP, NFAT, STAT1/4/6, TBX5, CEBP, OSF2 and AML1) (Fig. 2.11). The TFBS predicted in the *Island I* by the 'optimized for function' parameter as common to all the vertebrates of our study except for snakes are: E47 (582-597); ERR1 (660-673); OTX (684-691); PBX1, CDXA (754-762); CEBPΔ (926-937); PPARG, TFIII (933-949); CREB (995-1006); PAX3, LEF1B (1092-1112); NCX (1121-1130); TBX5 (1140-1151); MIF1 (1143-1160). All the TFBS that are exclusive of snakes in both *CNS65* and *Island I* conserved fragments are listed on Tables A2.2 (*CNS65*) and A2.3 (*Island I*).

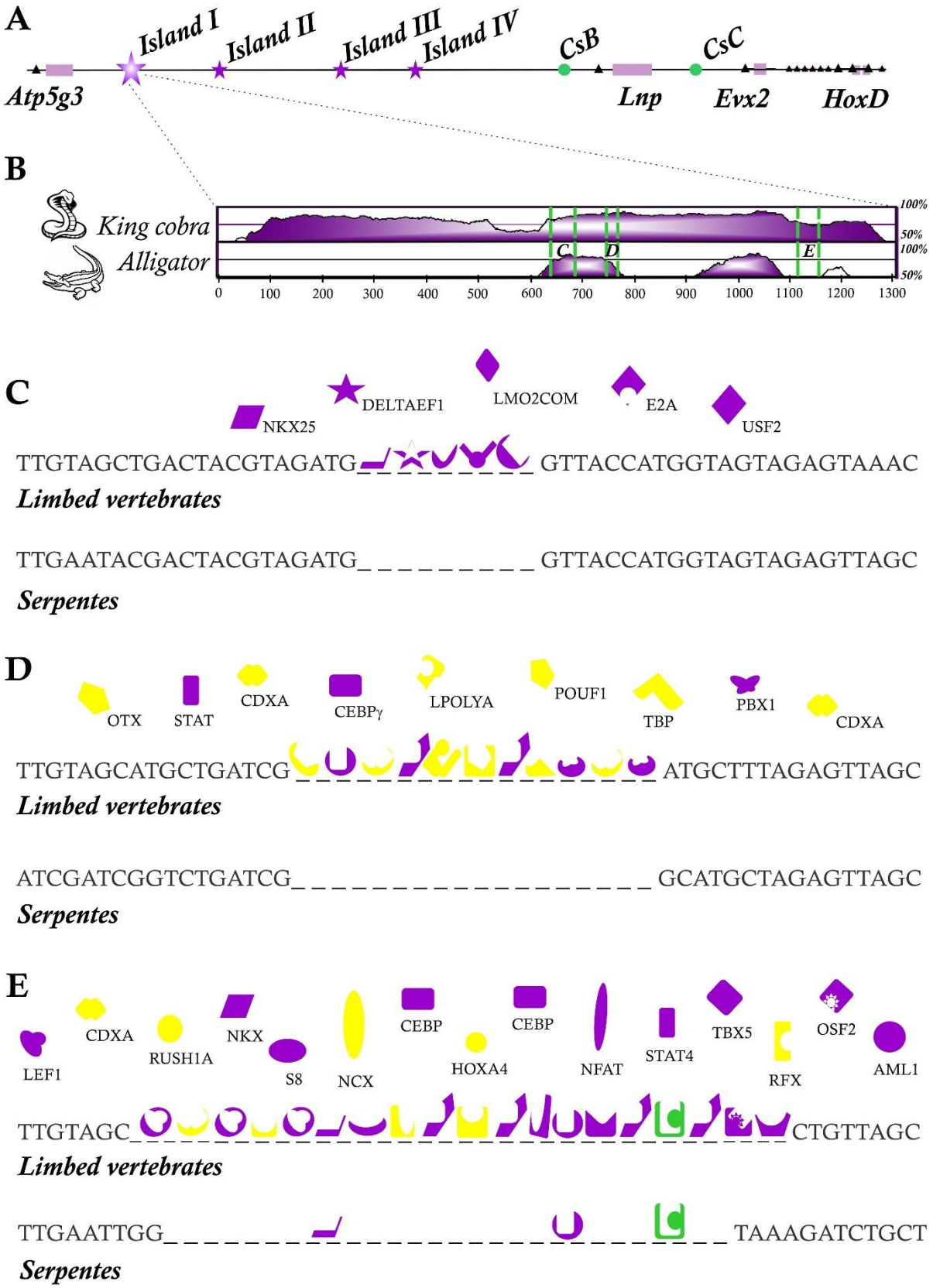


**Figure 2.11.** (**A**) Representation of *Island I* chromosomic location in the centromeric domain downstream to the *HoxD* cluster. (**B**) Topologic graphic designed on mVista online software illustrating the percentage of similarity of the *Island I* studied fragment from the king cobra

*Ophiophagus hannah* and *Alligator mississippiensis* with the snake *Phyton bivittatus*. Note that the sequence from *O. hannah* presents a much higher degree of similarity in relation to that of *P. bivittatus* than the alligator´s. Marked between green bars in **B** two putative signatures of limb loss were predicted in a conserved region of the *Island I* fragment (568-592 and 752-765 in *A. mississippiensis*), and the most remarkable one, predicted in a less conserved region (1115-1154 in *A. mississippiensis*). The signature illustrated in **C** (568-592) shows that binding sites for NKX25, ΔF1, LMO2COM, E2A and USF2 that appear in *Island I* of all limbed vertebrates studied were putatively lost in correspondent segments of *P. bivitattus* and *O. hannah*. In **D**, 5 binding sites for STAT1/4, CREB and PBX1 are among 11 TFBS apparently lost by the snakes in the second fragment of the conserved region (752-765). In **E**, the most striking signature (1115-1154) is represented, in which 13 among 19 TFBS are limb-related (LEF1, NKX, S8, CEBP, NFAT, STAT4, TBX5, OSF2, AML1). TFBS represented in purple correspond to those functionally related to limb development according to the literature, whereas the ones represented in yellow and green correspond to those presenting other functions. The green one mark a TFBS retained in the *CNS65* sequences from snakes, while the non-related to limb development TFBS that were apparently lost are highlighted in yellow.

## 2.5 DISCUSSION

This study showed that two regulatory elements essential for *HoxD* gene expression during tetrapod limb development have lost their regulatory capacities in a limbless lineage, the Serpentes clade. Current literature reports that mouse *CNS65* and *Island I* elicit strong gene reporter expression in developing limbs (Andrey et al. 2013; Gehrke et al. 2014) and the role of gene expression regulation in vertebrate appendices seems so ancient that *Island I* from fish already detains such regulatory capacity in limb buds (Gehrke et al. 2015). However, when mouse embryos were injected with corresponding fragments of *CNS65* and *Island I* from snakes, reporter expressions were either remarkably reduced or completely abrogated. The loss of regulatory capacity of these two enhancers in limb buds apparently involved evolution of specific regulatory signatures in their sequences, as I did not identify in Serpentes several TFBS predicted in the sequences of limbed vertebrates that relate to genes involved in tetrapod limb development. Together, these results suggest a functional impairment of both regulatory elements during evolution of limblessness in snakes, a process that might have occurred either concomitantly to limb reduction in the Serpentes ancestor or after limb loss was already established in the lineage by other causal factors.

Comparisons between snakes and limbed amniotes of the *CNS65* and *Island I* sequences based on TFBS point to specific regions as candidate segments comprising limb-specific functional elements, as in these regions many TFBS predicted in limbed amniotes that apparently disappeared from the snake enhancers have well-described roles during limb development, some of them configuring key elements in limb developmental pathways (see Table 2.1). Among the TFBS predicted in all 'reptiles' except for snakes (alligator, lizard, birds and turtles), approximately 43-52% are limb-related (63 out of 121 TFBS located in the *Island I* fragment, and 42 out of 97 located in *CNS65*). Interestingly,

sequences of *CNS65* and *Island I* exhibit common patterns regarding limb-related TFBS predicted in limbed 'reptiles' except for snakes: in both regulatory elements, the TFBS for GATAs, CEBP, SRY, PBX1, NKX25, TEF1 and AML1 are predicted in limbed species but not identified in the limbless lineage (see Table A2.2). A molecular convergence among telomeric and centromeric enhancers in specific regulatory signatures, in particular when these likely involve impairment of limb-developmental regulatory capacities, proposes a scenario implying release of selective constraints due to character loss (see Marshall et al. 1994). Accordingly, upstream changes in limb developmental pathways that resulted in limb loss might have relaxed stabilizing selection on regions previously regulating *HoxD* gene expression during development of the trait that no longer exists, accommodating some degree of sequence degeneration (see Shapiro et al. 2004; Gompel et al. 2005; Prud´homme et al. 2006) because the lack of a fine-tuned modulation of gene expression in limb buds does not affect fitness anymore.

Concentration of limb-related predicted TFBS in specific regions of *CNS65* (9/16) and *Island I* (48/68) sequences from all limbed amniotes considered (alligator, lizard, birds, turtles and mammals) that seem absent in snakes provide evidence for segments potentially essential for limb development in these regulatory elements. The *CNS65* fragment comprises one particular region (position 559 to 579) where, from the eight TFBS predicted in limbed species that were not identified in the limbless lineage, five have been described in the literature as being involved in limb-related functions: MEIS, TGIF, CREB, GATA3, TCF4 (Galceran et al. 1999; Mercader et al. 1999; Long et al. 2001; A. Chen et al. 2005; Kmita et al. 2005; Mercader et al. 2005; Lorda-Diez et al. 2009; Karamboulas et al. 2010).

This *CNS65* region can therefore be proposed as candidate segment for representing or belonging to a stilopodium/zeugopodium-specific regulatory module (Fig

2.10C) that likely degenerated during evolution of limblessness in Serpentes. This telomeric enhancer also comprises regions where snakes apparently have lost the limb-related TFBS for FOXO1 (position 383-392; see Villarejo-Balcells et al. 2011), TBX5 (position 418-425; see Agarwal et al. 2003) and STAT4 (position 979-986; see Sahni et al. 1999). The *Hoxd* centromeric enhancer *Island I* seems to have experienced even stronger erosion than *CNS65*, where estimated loss of 68 of the TFBS predicted in all limbed vertebrates, among which at least 70% have limb-related functions described in the literature, evidence the existence of limb-specific regulatory segments that apparently degenerated during snake evolution. In Squamata, limb loss evolves through a centripetal process where distal structures are usually lost prior to proximal ones (Wiens & Slingluff 2001; Caldwell 2003), a pattern that may explain the apparently stronger functional impairment of *Island I*. According to this centripetal process, the first limb structure lost during evolution comprises the distal region encompassing the autopodium, which development involves the second wave of *Hoxd* expression that is regulated by the genetic desert comprising *Island I* (Montavon et al. 2011; Dixon et al. 2012; Nora et al. 2012; Andrey et al. 2013). Assuming that limb loss in snakes was a gradual process where distal limb elements were lost before the proximal ones (see Scanlon et al. 1999; Tchernov et al. 2000; Rieppel et al. 2003; Palci & Caldwell 2007), this regulatory element may have been released from strong selection much earlier than *CNS65*. Based on the absence in snakes of TFBS predicted in limbed amniotes, I identified three segments within *Island I* likely encompassing limb-specific functions. The first two regions (positions 568-592 and 752-765 of the *A. missipiensis* sequence) comprise limb-related TFBS, including NKX25, ΔF1, LMO2COM, E2A, USF2, STAT1/4, CEBPγ and PBX1 (see Figs. 2.11C and 2.11D), while the third segment (positions 1115-1154) comprises 13 limb-related TFBS (LEF1, S8, CEBP, NFAT, STAT1/4/6, TBX5, CEBP, OSF2 and AML1; Fig 2.11E).

Interestingly, this last segment is located in a low-conserved region of *Island I* (see Fig. 2.11B), reinforcing the idea that very-conserved regions are not always the only ones functionally relevant (see Fisher et al. 2006).

Development of tetrapod limbs is established by crosstalk involving several developmental pathways. In summary, retinoic acid controls *Hox* gene expression in the lateral plate mesoderm (LPM), delimiting limb fields (Capdevila & Izpisúa-Belmonte 2001). Induction of Apical Ectodermal Ridge (AER) and limb buds is mainly attributed to interactions between fibroblast growth factors (FGFs), which induce proximal-distal growth, and the dorsal-ventral patterning canonical *Wnt*/β-catenin pathway (Parr & Mcmahon 1995; Fromental-Ramain et al. 1996; Kawakami et al. 2006). The establishment of a zone of polarizing activity (ZPA, which patterns structures recognizable in the adult limb) is mediated by sonic hedgehog (*Shh)* gene expression under regulation of GLI and BMP protein families (Park et al. 2000; Riddle et al. 1993; Kozhemyakina et al. 2014). The *Shh* developmental pathway ultimately controls *Hox* gene expression patterns in emerging digits (Capdevila & Izpisúa-Belmonte 2001), whereas BMPs are involved in anlage cartilaginous formation and establishment of the chondrogenic fate in limb cell lines (Duprez et al. 1996; Provot & Schipani 2005). Among the transcription factors which binding sites were lost in the putative *CNS65* and *Island I* signatures of limb loss in snakes, TBX5, MEIS and PBX1 stand out due to their relevance for limb development, as they represent downstream or upstream signals released through unrolling of the aforementioned developmental pathways. The first factor, TBX5, has been apparently lost in the snake *Island I* segment between positions 1115 and 1154, and refers to a primary and direct initiator of forelimb in vertebrates (Agarwal et al. 2003); without TBX5, development of forelimb buds does not occur (Rallis et al. 2003). This transcription factor acts as an upstream stimulus to *Wnt*/*Fgf* signaling pathways initiating limb primordia

(Agarwal et al. 2003). The two other factors, MEIS (position 561-572 in *A. mississippiensis CNS65*) and PBX1 (position 754-762 in *A. mississippiensis Island I*), are expressed by homeobox genes and define limb development in mouse, chicken and *Drosophila* (Mercader et al. 1999). *Meis* expression is induced by retinoic acid (Niederreither et al. 2002) and is restricted to a proximal domain (stylopodium) where it represses *Shh* signaling, specifying cell fates and differentiation patterns in the limb proximal-distal axis. These genes protect cells in proximal compartments from being distalized by AER signals (Mercader et al. 1999; Mercader et al. 2005; Kmita et al. 2005), and *Meis1* ectopic expression in chicken impairs distal limb development, triggering distal-to-proximal transformations (Mercader et al. 1999). Finally, *Pbx1* is essential for distal limb patterning, controlling *Hox* genes spatial distribution and *Shh* expression (Capellini et al. 2006). Expression of *Pbx1* in the lateral plate and early limb field mesoderm is apparently regulated by MEIS (Mercader et al. 1999).

Other TFBS that were lost by snakes in *CNS65* and *Island I* predicted regulatory signatures are relevant for limb development and interconnected through some of these same developmental pathways (see Table 2.1). For example, GATA, USF2 and AML1 connect with the *Shh via*. Snakes have apparently lost binding sites for GATA transcription factors in both *CNS65* and *Island I*; *Gata* expression is triggered by BMP proteins during early limb development (Karamboulas et al. 2010), and GATA TFs negatively regulate ectopic *Shh* expression in limb buds (Kozhemyakina et al. 2014). These molecules influence cell identity (Kozhemyakina et al. 2014) and contribute for the establishment of cartilage, perichondrium, tendon and other non-muscle cell types in limb buds (Karamboulas et al. 2010). USF2, a transcription factor having binding sites predicted in *Island I* of limbed amniotes that were not identified in snakes, negatively regulates GLI proteins (Villavicencio et al. 2002; Shahi et al. 2012), allowing *Hox* genes to be expressed

and digits to develop (Welscher et al. 2002; Tarchini & Duboule 2006). A binding site for AML1, which expression is regulated by BMP2 and organizes maturation and proliferation of chondrocytes (Yoshida et al. 2004), was predicted in *Island I* from limbed amniotes but not identified in snakes. Interestingly, a study of *Shh* regulatory elements conservation shows that several factors that appear in our predictions (specially those that were lost by snakes), occurs upstream (AP1, NF-kB, TCF, GATA1, LMO2COM, ΔEF1, MYOD, USF, NKX25, MYB, SP1, CEBPβ, NFAT, USF, RFX1, AP4), intronically (AP4, NFAT, AP1, CEBPβ, ΔEF1, LMO2COM, NKX25, TCF11, MYOD, GATA, AP2, USF, S8, HNF1, NKX25, MYB, SP1, CETS1P54, OCT1, CF11, SRY), and downstream (RFX1, NFAT, SRY, GATA, LMO2COM, CEBPβ, SOX5, NKX25, MYB, GATA, SRY, OCT1, S8) to *Shh* locus (see Lemos et al., 2004). ΔEF1 (584-594 in *Island I*) is another target of hedgehog signaling (*Ihh*) in developing limbs (Bellon et al. 2009), which absence generates short and dumpy limb phenotypes (Takagi et al. 1998).

On the other hand, CREB, TCF4, and OSF-1 play roles on the *Wnt/ β-catenin* limb *via* (Long et al. 2001; Galceran et al. 1999; Sensiate et al. 2014). CREB (563-568 in *CNS65*) is required for *Wnt*-directed myogenic gene expression in the limb mesenchyme (Chen et al. 2005). CREB also regulates proliferation of chondrocytes that will ultimately comprise mature bones; its knockout in mouse embryo causes short-limbed dwarfism due to proliferation defects and differentiation delay (Long et al. 2001). TCF4 (572-579 in *CNS65*) is a marker of undifferentiated mesenchymal cells that also play roles in bone cartilage differentiation and is expressed in a pattern similar to those of other *Wnt/β-catenin* signaling pathway molecules (Cho & Dressler 1998; Kardon et al. 2003; Sensiate et al. 2014). OSF2-1 (1115-1124 in *Island I*) regulates *Fgf8* expression and formation of the AER (Galceran et al. 1999). Last but not least, TGIF1 (TGF-interacting factor), lost in the

*CNS65* (561-571) of snakes, is regulated by TGF signaling and is expressed along with developing tendons in the limb buds (Lorda-Diez et al. 2009).

**Table 2.1.** Summary of roles on limb development related to the predicted TFBSs that were lost in snakes' *CNS65* and *Island I*.

| **TFBS** | *Island I* **Position** | *CNS65* **Position** | **Signaling pathways** | **Roles on limb development** |
|---|---|---|---|---|
| AML1 | 1220-1226 | - | BMP2 | Organizes maturation and proliferation of chondrocyte in growing limbs (Yoshida et al. 2004). |
| CEBP | 1123-1134, 1235-1247, 1144-1156, 1149-1156 | - | | Cartilage development, chondrocyte hypertrophy; deficiency causes dwarfism in mice (Hirata et al. 2009). |
| CREB | 997-1004, 1000-1005 | 563-568 | *Wnt* | Regulates proliferation of chondrocytes and maturation of bones; absence in mouse embryo causes short-limbed dwarfism (Long et al. 2001; A. Chen et al. 2005). |
| ΔEF1 | 584-594 | - | Hedgehog (Ihh) | Absence generates short and dumpy limb phenotypes (Takagi et al. 1998; Bellon et al. 2009). |
| E2A | 587-594 | - | Muscle regulatory factors (MRFs) | Activation of muscle-specific gene expression during myogenesis (Brand-Saberi 2005). |
| ERR1 | 660-673 | - | | Cartilage development; regulation of osteoblast differentiation (The UniProt Consortium 2008) expressed in forelimb bones such as the ulna, radius, and humerus (Bonnelye et al. 1997). |
| FOXO1 | - | 383-392 | EGFR, FGFR | Osteoblast and bone mass regulation; expressed in forelimb-level somites (Hosaka et al. 2004; Villarejo-Balcells et al. 2011). |
| GATA3 | 655-663, 685-693 | 566-575 | *Wnt*; BMP | Cartilage, perichondrium, tendon and other non-muscle cell types in the limb buds (Karamboulas et al. 2010). |
| GATA6 | 652-661, 685-694, 721-730 | - | BMP, TGFβ1, TGFβ2 | Chondrogenesis (Alexandrovich et al. 2008), absence in himdlimb buds causes preaxial polydactyly (Kozhemyakina et al. 2014). |
| LEF1 | 1115-1124, 1117-1122 | - | *Wnt*; BMP; *Fgf8* | Embryonic limb and face morphogenesis; muscle development; formation of AER (Galceran et al. 1999) |
| LMO2COM | 584-595 | - | SHH | - |
| MEIS1 | - | 561-572 | PAX6, PBX1, PBX2, *Hox*, SHH, RA | *S*pecifies cell fates in proximal compartments of the limb bud by inhibiting SHH signals (Mercader et al. 1999; Mercader et al. 2005; Kmita et al. 2005) |
| MYOGENIN | 585-592, 875-882 | | | Skeletal-muscle differentiation (Rudnicki & Jaenisch 1995). |
| NFAT | 1236-1247 1127-1136 | | | Osteoclasts differentiation (Hogan et al. 2003); chondrogenesis regulation; development of tendons in limbs (Ranger et al. 2000). |
| NKX25 | 568-574, 876-882, | | GATA4, PBX1, *Wnt* | Expressed in the lateral plate mesoderm (Stanfel et al. 2005). |

| | | | | |
|---|---|---|---|---|
| | 876-885, 943-949 | | | |
| OSF2 | 1149-1154 | | *Notch* | Positive regulator of chondrocyte and osteoclast differentiation, periosteal bone formation; inactivation causes severe shortening of the limbs (Stricker et al. 2002). |
| PBX1 | 754-762, 757-765 | | MEIS, HOX | Anterior/posterior and proximal/distal patterns specification and ossification in developing limb |
| PEA3 | 939-945 | | | Expressed in the LPM, in the forelimb field and in the nascent limb bud mesenchyme (Chotteau-Lelievre et al. 2001) |
| S8 | 1120-1135 | | | Cartilage development, embryonic limb morphogenesis (The UniProt Consortium 2008); regulates proliferation or differentiation in the bones of the zeugopods (Berge et al. 1998). |
| SMAD | | 614-621, 616-624 | STATs, CBP/p300 LIF, BMP2 TGF-β | Limb cartilage development (Williams 2000; Karamboulas et al. 2010) |
| SRY | 625-631 | 331-337 | BMP | Chondrogenesis and cartilage differentiation in the limb mesoderm (Chimal-Monroy et al. 2003). |
| STAT1 | 752-759, 1128-1135 | | JAK-STAT; FGF, SMADs, CBP/p300, BMP | Chondrocyte proliferation; bone development (Sahni et al. 1999; Williams 2000). |
| STAT4 | 752-759, 945-952, 1128-1135 | 979-986 | Cytokines, SMADs, CBP/p300 LIF, BMP2 | Chondrocyte proliferation; bone development (Sahni et al. 1999). |
| STAT6 | 945-952, 1128-1135 | | | Chondrocyte proliferation; bone development (Sahni et al. 1999). |
| TBX5 | 1140-1151, 1140-1149, 1160-1171, 1162-1169 | 418-425 | | Limb pattern formation; forelimb induction (The UniProt Consortium 2008). |
| TCF4 | | 572-579 | *Wnt*, *Fgf8* | Mouse limb buds development, regulates formation of AER (Galceran et al. 1999). |
| TGIF | | 561-571 | *Nodal*, FGFR, TGF. | Developing tendons in the limb buds (Lorda-Diez et al. 2009). |
| USF2 | 587-592 | | USF2, GLI, *Shh* | Pattern digits by negatively regulating GLI proteins (Villavicencio et al. 2002; Shahi et al. 2012). |

## 2.6 CONCLUSIONS

The present study suggests that the regulatory capacity of two elements that modulate *HoxD* gene expression during the development of proximal (*CNS65*) and distal (*Island I*) regions of the tetrapod limb became impaired during snake evolution. Predictions of TFBS in the sequences of limbed amniotes that were not observed in snake enhancers reveal one segment likely candidate to comprise stilopodium/zeugopodium-specific regulatory functions in *CNS65* (situated between positions 559 and 579 in *A. mississippiensis*) and three putative autopodium-specific segments in *Island I* (positions 568 to 592, 752 to 765 and 1115 to 1154 in *A. mississippiensis*). Sequence degeneration of snake *CNS65* and *Island I*, together with results confirming the impairment of their limb-developmental regulatory capacities during evolution of Serpentes, strengthen the prediction that trait loss might relax selective constraints acting on specific genomic regions (see Shapiro et al. 2004; Gompel et al. 2005; Prud´homme et al. 2006). Such prediction is particularly suitable if regulatory elements comprise trait-specific modules (Suryamohan & Halfon 2015), given that degeneration of genomic regions compromised with multiple functions likely involves pleiotropy (Sucena et al. 2003; Colosimo 2005; Prud´homme et al. 2006). In this context, our study also reveals putative candidate regulatory segments within *CNS65* and *Island I* that seem dedicated exclusively limb development in tetrapods and became degenerated in limbless lineages.

## 2.7 APPENDIX

**Table A2.1.** GenBank accession numbers for *CNS65* and *Island I* fragments.

| Lineage | Species | *CNS65* | *Island I* |
|---|---|---|---|
| Serpentes | *Python molurus bivittatus* | AEQU02037710.1 | AEQU02066876.1 |
| | *Ophiophagus hannah* | AZIM01000055.1 | AZIM01000331.1 |
| Lizard | *Anolis carolinensis* | NW_003338958.1 | W_003338924.1 |
| Crocodile | *Alligator mississippiensis* | AKHW01084356.1 | AKHW01083192.1 |
| Birds | *Gallus gallus* | NC_006094.3 | NC_006094.3 |
| | *Columba livia* | AKCR01078007.1 | AKCR01078079.1 |
| Turtles | *Pelodiscus scinensis* | AGCU01060288.1 | AGCU01060382.1 |
| | *Chrysemys picta* | AHGY02196106.1 | AHGY02196137.1 |
| Mouse | *Mus musculus* | CAAA01035426.1 | CAAA01059384.1 |
| Rat | *Rattus norvegicus* | NC_005102.3 | AABR07052547.1 |
| Human | *Homo sapiens* | NC_018913.2 | ADDF02068953.1 |

**Table A2.2.** Estimated positions of TFBS gained or lost exclusively by snakes in *CNS65* sequence conserved fragments.

| **Position** *P. bivitatus* | **Snakes** | **Position** *A. mississippiensis* | **'Reptiles'-snakes** | **All-snakes** |
|---|---|---|---|---|
| 202-209 | PU1 | 202-208 | PEA3 | - |
| 226-232 | CDXA | | | |
| 239-246 | CAP | | | |
| | | 247-258 | NKX62 | - |
| | | 248-254 | CDXA | - |
| | | 249-256 | OTX, STAT4 | - |
| | | 250-259 | GATA3, GATA6 | - |
| | | 254-267 | CEBPB | - |
| 327-334 | CAP | 331-340 | FOXO1 | - |
| 330-340 | PITX2 | 331-337 | SRY | - |
| | | 339-349 | GATA | - |
| | | 340-352 | GATA1 | - |
| | | 341-352 | CDC5 | - |
| | | 342-351 | GATA1, GATA6 | - |
| | | 343-349 | GATA | - |
| 346-354 | PAX2 | 354-361 | STAT5A | - |
| 348-354 | EN1 | | | |
| 365-372 | CAP | 368-374 | NKX25B, NKX25 | - |
| | | 369-383 | HMGIY | - |
| | | 370-376 | CDXA | - |
| | | 374-387 | CEBPB | - |
| | | 374-386 | CEBP | - |
| | | 374-383 | NFAT | - |
| | | 374-385 | NFAT | - |
| | | 375-387 | HELIOSA | - |
| | | 376-387 | CEBP , CT1 | - |
| | | 378-390 | CDXA | - |
| | | 380-386 | HNF3ALPHA | - |
| | | 380-390 | NKX62 | - |
| | | 381-392 | XFD1, XFD2 | - |
| | | 381-394 | HFH8 | - |
| | | 381-393 | LPOLYA_B | - |
| | | 382-388 | TBP | - |
| | | 383-392 | FOXO1 | FOXO1 |
| | | 383-391 | PAX2 | PAX2 |
| | | 383-390 | TBP | - |
| | | 378-384 | CDXA | - |
| | | 383-390 | STAT4 | - |
| | | 384-390 | CDXA | - |
| | | 384-392 | PBX1 | - |
| | | 384-394 | POU6F1 | - |
| | | 386-393 | LPOLYA_B | - |
| | | 386-392 | TBP | - |
| | | 387-393 | CDXA | - |
| | | 389-395 | NKX25B, NKX25 | |
| 383-389 | IRF1 | 394-403 | RUSH1A | - |
| 386-400 | HNF3B | 395-403 | YY1 | YY1 |
| 387-403 | FOXP3 | 396-403 | BRCA | - |
| 387-392 | TEF1 | | | |
| | | 413-422 | RUSH1A | |
| | | 416-423 | STAT4 | - |

| | | | | |
|---|---|---|---|---|
| | | 418-425 | TBX5 | TBX5 |
| | | 420-427 | CAP | CAP |
| 433-445 | CEBP | | | |
| 438-448 | BACH2 | | | |
| 438-449 | PBX1 | | | |
| | | 451-458 | CREBP1CJUN | - |
| 467-473 | CDXA | | | |
| | | 493-500 | CAP | CAP |
| | | 494-500 | AP2REP, CDXA | - |
| | | 494-501 | CAP | - |
| | | 506-520 | ETS1 | - |
| | | 517-527 | CP2 | CP2 |
| 514-523 | FOXO1 | 521-530 | GFI1 | - |
| 514-522 | PAX2 | 527-533 | AP2REP | - |
| 516-527 | GATA4 | 527-538 | IK2_01 | - |
| 517-523 | SRY | 528-535 | RBPJK | - |
| | | 539-546 | OTX | - |
| | | 540-546 | CDXA, NKX25 | - |
| 562-569 | AP3 | 559-566 | CAP | CAP |
| 566-573 | STAT5A, STAT6 | 561-572 | MEIS1 | MEIS1 |
| 568-574 | EN1 | 561-571 | TGIF | TGIF |
| 579-586 | AP3 | 563-568 | CREB | CREB |
| 587-604 | NF1 | 566-575 | GATA3 | GATA3 |
| 592-599 | CETS168 | 567-573 | CDXA | CDXA |
| 595-602 | BRCA | 569-576 | TBP | TBP |
| 597-606 | RUSH1A | 572-579 | TCF4 | TCF4 |
| 599-608 | MYB | | | |
| 606-618 | GATA1 | 614-621 | SMAD | |
| 607-616 | GATA1, GATA2, | 616-624 | SMAD | |
| | GATA3, GATA6 | 628-643 | CBF | |
| 608-614 | CDXA | 628-640 | CEBP | |
| 599-609 | GATA | 630-635 | TEF1 | |
| 609-615 | GATA | 633-638 | AML1 | |
| 609-620 | PBX1 | | | |
| 609-615 | TBP | | | |
| 610-618 | PAX2 | | | |
| 610-617 | TBP, STAT1, | | | |
| 610-616 | STAT4 | | | |
| 611-622 | CDXA | | | |
| 633-641 | GATA4 | | | |
| 636-647 | HOXA3 | | | |
| 636-642 | CEBP | | | |
| | CDXA | | | |
| 660-670 | MYB | 670-676 | CDXA | |
| 663-672 | MYB | | | |
| 664-672 | MYB | | | |
| 662-670 | MYB | | | |
| 665-672 | NKX25 | | | |
| 674-681 | STAT5A | | | |
| 675-683 | HOXA3 | | | |
| 675-685 | PAX4 | | | |
| 676-682 | HMGIY | | | |
| 678-684 | CDXA | | | |
| 681-693 | NKX61 | | | |
| 681-696 | NFY | | | |
| 683-692 | GATA6 | | | |
| 683-689 | SRY | | | |
| 684-692 | PBX1 | | | |
| 685-691 | GATA | | | |
| 685-696 | CAAT | | | |

| | | | | |
|---|---|---|---|---|
| 685-691 | HOXA7 | | | |
| 702-707 | HNF4 | | | |
| 712-719 | STAT3 | | | |
| 727-734 | MZF1 | | | |
| 727-732 | USF2 | | | |
| 729-734 | CHCH | | | |
| 737-742 | HNF4 | | | |
| 742-755 | CDX2 | | | |
| 742-754 | TCF11 | | | |
| 743-757 | IPF1 | | | |
| 712-719 | STAT3 | 717-723 | CDXA | |
| 727-734 | MZF1 | | | |
| 727-732 | USF2 | | | |
| 729-734 | CHCH | | | |
| 737-742 | HNF4 | | | |
| 742-755 | CDX2 | | | |
| 742-754 | TCF11 | | | |
| 743-757 | IPF1 | | | |
| 746-753 | CAP | | | |
| 749-759 | AP1 | | | |
| | | 752-759 | STAT1, STAT4, STAT5A | |
| 750-758 | AP1 | 921-928 | STAT4, STAT6 | |
| 750-757 | AP1 | | | |
| 750-755 | CREB | | | |
| 757-764 | AP3 | | | |
| 760-767 | CAP | | | |
| 778-784 | CDXA | | | |
| 779-785 | TBP | | | |
| 795-801 | MOVOB | | | |
| 797-804 | CAP | | | |
| 799-806 | CACD | | | |
| 837-844 | GR | | | |
| 842-849 | CETS168 | | | |
| 855-863 | AP2ALPHA, | | | |
| 856-864 | HOXA3 | | | |
| 862-867 | AP2ALPHA, | | | |
| 864-873 | AP2GAMMA | | | |
| 866-873 | TEF1 | | | |
| 867-875 | PR | | | |
| 867-874 | STAT1, STAT3, | | | |
| 870-877 | STAT5A | | | |
| 870-879 | AR | | | |
| 871-877 | GR | | | |
| 871-876 | TFE | | | |
| 882-888 | TITF1 | | | |
| 883-892 | NKX25B | | | |
| 891-900 | NKX25, USF2 | | | |
| 892-900 | NKX25L | | | |
| 893-902 | HSF2 | | | |
| 900-911 | MYB | | | |
| 902-908 | EVI1 | | | |
| 904-913 | RUSH1A | | | |
| 910-923 | SRY | | | |
| 912-923 | SRY | | | |
| 914-921 | RUSH1A | | | |
| 915-920 | ETS2 | | | |
| | ETS | | | |
| | CAP | | | |
| | TEF1 | | | |
| 916-923 | GEN_INI3, | 979-986 | STAT4 | STAT4 |
| 919-924 | GEN_INI_B | | | |

| | | | |
|---|---|---|---|
| 924-936 | HNF4 | | |
| 925-932 | CEBP | | |
| 929-936 | STAT6 | | |
| 942-948 | AP3 | | |
| 944-951 | SREBP1 | | |
| 946-953 | GEN_INI3_B | | |
| 949-959 | CAP | | |
| 949-960 | E2F | | |
| 950-958 | E2F | | |
| 959-966 | E2F | | |
| 976-983 | STAT3 | | |
| 959-966 | STAT5A | | |
| 964-975 | STAT6 | | |
| 972-978 | CAAT | | |
| | CDXA | | |
| 973-985 | HFH8 | 997-1008 | FOXD3 |
| 974-981 | LPOLYA | | |
| 974-984 | PAX4 | | |
| 991-997 | TBP | | |
| 975-983 | PAX2 | | |
| 975-982 | TBP | | |
| 975-981 | CDXA | | |
| 976-982 | CDXA | | |
| 978-992 | ZBRK1 | | |
| 981-991 | AP1FJ | | |
| 981-992 | MEIS1 | | |
| 982-992 | TGIF | | |
| 983-994 | SREBP | | |
| 985-990 | CREB | | |
| 987-995 | LYF1 | | |
| 990-997 | AP3 | 1047-1053 | NKX25 |
| 991-1000 | GATA1, GATA2 | 1088-1097 | P53 |
| 991-999 | GATA3 | 1106-1113 | TBP |
| 991-1000 | GATA6 | 1107-1113 | CDXA |
| 996-1004 | LYF1 | | |
| 997-1008 | TTF1 | | |
| 997-1010 | USF | | |
| 997-1003 | MOVOB | | |
| 997-1010 | USF | | |
| 999-1009 | MYOGENIN, E2A | | |
| 999-1006 | ΔEF1 | | |
| 1000-1009 | EBOX | | |
| 1000-1007 | CAP | | |
| 1000-1006 | MYC | | |
| 1001-1006 | USF2 | | |
| 1002-1014 | GATA1 | | |
| 1002-1009 | TBX5B | | |
| 1003-1012 | GATA1, GATA2,GATA2, | | |
| 1004-1012 | GATA3, GATA6 | | |
| 1005-1011 | GATA3 | | |
| 1006-1013 | GATA | | |
| 1016-1022 | STAT1 | | |
| 1032-1043 | AP2REP | | |
| 1039-1056 | CEBP | | |
| 1045-1055 | HNF3 | | |
| 1047-1056 | LEF1TCF1 | | |
| 1047-1054 | LEF1 | | |
| 1048-1054 | LPOLYA, BRCA | | |
| 1048-1057 | LEF1B | | |
| 1049-1054 | RUSH1A | | |
| 1057-1063 | LEF1 | | |

| | |
|---|---|
| 1059-1067 | SRY |
| 1060-1067 | PAX2 |
| 1061-1067 | STAT4 |
| 1075-1082 | CDXA |
| 1076-1083 | STAT1 |
| 1077-1082 | CAP |
| 1085-1092 | TEF1 |
| 1088-1094 | CAP |
| | EN1 |

**Table A2.3.** Estimated positions of TFBS gained or lost exclusively by snakes in Island I sequence conserved fragments.

| **Position** *P. bivitatus* | **Snakes** | **Position** *A. mississippiensis* | **'Reptiles'-snakes** | **All-snakes** |
|---|---|---|---|---|
| 596-604 | PAX2 | 568-574 | NKX25 | NKX25 |
| | | 584-594 | ΔEF1 | ΔEF1 |
| | | 584-595 | LMO2COM | LMO2COM |
| | | 585-592 | MYOGENIN | - |
| | | 587-594 | E2A | E2A |
| | | 587-592 | USF2 | USF2 |
| 644-651 | STAT3, STAT4, | 611-618 | CAP | - |
| | STAT5A | 619-626 | CAP | - |
| 647-653 | STAT6 | 622-629 | CAP | - |
| | CDXA | 625-634 | GATA6 | - |
| | | 625-631 | SRY | - |
| | | 626-634 | PBX1 | - |
| | | 631-639 | PAX2 | - |
| | | 632-639 | STAT1 | - |
| | | 632-638 | HGMIY | - |
| | | 633-639 | CDXA | - |
| | | 652-661 | GATA6 | GATA6 |
| | | 654-661 | CAP | CAP |
| | | 655-663 | GATA3 | GATA3 |
| | | 657-664 | CAP | CAP |
| | | 660-673 | ERR1 | ERR1 |
| | | 660-671 | TFIIA | - |
| | | 663-672 | PR | - |
| | | 665-673 | T3R | T3R |
| | | 666-671 | HNF4 | HNF4 |
| | | 666-673 | GR | - |
| 666-677 | PXR | | | |
| 668-674 | EN1 | | | |
| 669-676 | STAT5A | | | |
| 674-690 | MTATA | | | |
| 674-688 | TATA | | | |
| 680-687 | OTX | 685-693 | GATA3 | GATA3 |
| 699-706 | STAT6 | 685-694 | GATA6 | GATA6 |
| | | 686-692 | GATA | GATA |
| | | 701-707 | CDXA | - |
| | | 706-713 | CAP | - |
| 710-719 | TEL2 | 711-724 | OCT1 | - |
| 711-719 | SMAD | 712-719 | CAP | - |
| 713-727 | TST1 | 718-725 | CAP | CAP |
| 714-722 | MSX1 | 721-730 | GATA6 | GATA6 |
| 715-721 | CDXA | | | |
| 716-722 | EN1 | | | |
| 716-730 | TST1 | | | |
| 717-727 | HNF3ALPHA | | | |
| 721-731 | OCT1 | | | |
| 723-730 | STAT6 | | | |
| 727-734 | CAP | | | |
| 727-736 | TATA | | | |
| 728-734 | CDXA | | | |
| 732-742 | HELIOSA | | | |
| 732-743 | NFAT | | | |
| 745-754 | AP4 | 750-756 | CDXA | - |
| 756-769 | CEBPA | 750-761 | NKX62 | NKX62 |
| 766-774 | PAX2, AREB6 | 751-759 | FOXM1 | - |
| 769-775 | SRY | 752-759 | OTX, STAT1, STAT4 | OTX, STAT1, |
| 769-778 | P53 | 753-759 | CDXA | STAT4 |

| | | | | |
|---|---|---|---|---|
| | | 753-765 | CEBPGAMMA | CDXA |
| | | 753-760 | LPOLYA | CEBPGAMMA |
| | | 753-762 | POU1F1 | LPOLYA |
| | | 754-766 | CEBPGAMMA | POU1F1 |
| | | 754-760 | TBP | CEBPGAMMA |
| | | 754-762 | PBX1 | TBP |
| | | 755-767 | OCT1 | PBX1 |
| | | 756-762 | CDXA | - |
| | | 757-763 | CDXA | CDXA |
| | | 757-765 | PBX1 | CDXA |
| | | | | PBX1 |
| 882-887 | TEF1 | 875-882 | MYOGENIN | |
| 884-893 | KAISO | 876-882 | NKX25B, NKX25 | |
| 885-893 | AP2ALPHA | 876-883 | CAP | |
| 894-901 | STAT3, STAT4, | 876-882 | MYC | |
| | STAT5A, STAT6 | 876-885 | NKX25 | |
| 895-902 | LFA1 | 877-882 | USF2 | |
| 897-903 | SRY | | | |
| 907-915 | TFIII | | | |
| 908-914 | MOVOB | | | |
| 909-914 | CHCH | | | |
| 909-915 | ETF | | | |
| 912-921 | GATA2 | | | |
| 912-919 | ETS | | | |
| 913-922 | HSF2 | | | |
| 913-919 | PEA3 | | | |
| 917-922 | HNF4 | | | |
| 918-926 | CIZ | | | |
| 921-928 | STAT1, STAT3, | | | |
| | STAT4, | | | |
| 921-927 | STAT5A, STAT6 | | | |
| 922-928 | HMGIY | | | |
| 926-933 | CDXA | | | |
| 926-936 | CAP | | | |
| 927-932 | HELIOSA | | | |
| 929-938 | TEF1 | | | |
| 942-953 | RUSH1A | | | |
| | NKX62 | | | |
| 942-949 | STAT6 | 922-930 | MSX1 | - |
| 943-952 | LHX3 | 925-938 | CEBPA | - |
| 944-955 | IPF1 | 926-937 | CEBP□ | - |
| 944-955 | NKX62 | 939-946 | CAP, ETS | - |
| 944-951 | STAT4 | 939-945 | PEA3 | PEA3 |
| 944-959 | S8 | 939-953 | TST1 | - |
| 945-958 | CHX10 | 942-949 | AP3 | AP3 |
| 945-955 | PAX4 | 943-949 | NKX25, CDXA | NKX25, CDXA |
| 945-951 | CDXA | 945-952 | STAT4, STAT6 | STAT4, STAT6 |
| 945-951 | NKX25L | 959-965 | CDXA | CDXA |
| 946-952 | CDXA | | | |
| 946-952 | NKX25L | | | |
| 946-954 | PBX1 | | | |
| 947-956 | LHX3 | | | |
| 948-955 | NKX25 | | | |
| 949-955 | NKX25L | | | |
| 949-960 | OCT1 | | | |
| 949-956 | OTX | | | |
| 1003-1018 | E47 | | | |
| 1005-1016 | MYOD, LMO2COM | | | |
| 1005-1012 | BRCA | | | |
| 1005-1015 | E12 | | | |
| 1007-1015 | AP4 | | | |
| 1010-1017 | GR | 997-1005 | AP1 | AP1 |

| | | | | |
|---|---|---|---|---|
| 1011-1018 | STAT5A | 997-1004 | CREBP1, CREB | CREBP1, |
| 1011-1020 | PR | 998-1005 | AP1 | CREB |
| | | 1000-1005 | CREB | AP1 |
| | | | | CREB |
| 1042-1049 | CAP | 1021-1037 | HSF1 | |
| 1043-1049 | NKX25L, CDXA | 1022-1030 | PAX2 | |
| 1043-1050 | STAT1, STAT3, | 1022-1028 | CDXA, LEF1B | |
| | STAT4, STAT5A, | 1022-1031 | RUSH1A | |
| | STAT5A | 1023-1028 | LEF1 | |
| | | 1025-1037 | CETS1P54 | |
| 1046-1052 | CDXA | 1026-1035 | CETS1P54 | - |
| 1051-1057 | NKX25L, CDXA | 1027-1038 | CEBPΔ | - |
| 1051-1058 | STAT5A | 1032-1039 | STAT3 | STAT3 |
| | | 1033-1038 | TEF1 | |
| 1064-1072 | PAX2 | 1074-1082 | FOXM1 | FOXM1 |
| 1067-1073 | TBP | 1082-1093 | CEBP | |
| 1077-1088 | CEBP | | | |
| 1082-1088 | CDXA | | | |
| 1090-1099 | VMYB | | | |
| 1091-1098 | CAP | | | |
| 1092-1100 | MYB | | | |
| 1092-1102 | MYB | | | |
| 1094-1101 | AP3 | | | |
| 1094-1102 | FOXM1 | | | |
| 1095-1101 | NKX25L | | | |
| 1096-1106 | GATA | | | |
| 1098-1107 | CDPCR3HD | | | |
| 1098-1104 | CDXA | | | |
| 1099-1108 | GATA2 | 1115-1124 | LEF1 | LEF1 |
| 1100-1106 | TBP, GATA | 1116-1122 | CDXA | CDXA |
| 1104-1109 | CHCH | 1116-1122 | LEF1B | LEF1B |
| 1104-1111 | MZF1 | 1116-1125 | RUSH1A | RUSH1A |
| 1105-1110 | CHCH | 1117-1122 | LEF1 | LEF1 |
| 1105-1111 | MOVOB | 1120-1135 | S8 | S8 |
| 1106-1115 | RUSH1A | 1121-1130 | NCX | NCX |
| 1107-1114 | STAT3, STAT5A, | 1123-1134 | CEBP | CEBP |
| 1109-1116 | STAT6 | 1123-1130 | HOXA4 | HOXA4 |
| 1123-1129 | PU1 | 1124-1136 | CEBP | CEBP |
| 1123-1130 | SREBP1 | 1125-1136 | NFAT | NFAT |
| 1125-1132 | BRCA | 1127-1136 | NFAT | NFAT |
| 1126-1132 | CACD | 1128-1135 | STAT1 | STAT1 |
| 1127-1134 | MOVOB | 1128-1135 | STAT4 | STAT4 |
| 1128-1133 | CAP | 1128-1135 | STAT6 | STAT6 |
| 1135-1144 | CHCH | 1139-1149 | ΔEF1 | - |
| 1138-1145 | FOXO1 | 1140-1151 | TBX5 | TBX5 |
| 1154-1165 | BRCA | 1140-1149 | TBX5 | TBX5 |
| 1163-1174 | HNF6 | 1142-1149 | BRCA | - |
| 1170-1176 | NKX3A | 1 | RFX | - |
| 1181-1186 | CDXA | | CEBP | CEBP |
| 1194-1203 | CHCH | 144-1152 | CEBP | CEBP |
| 1220-1226 | RUSH1A | 1144-1156 | OSF2, AML1 | OSF2, AML1 |
| | CDXA | 1149-1156 | | |
| | | 1149-1154 | | |
| 1226-1233 | STAT3, STAT5A, | 1154-1161 | HOXA4 | - |
| 1237-1244 | STAT6 | 1158-1169 | AREB6 | - |
| 1238-1249 | STAT4 | 1160-1171 | TBX5 | TBX5 |
| 1241-1248 | CEBP | 1160-1167 | CAP | - |
| | CAP | 1162-1169 | TBX5 | TBX5 |
| | | 1172-1177 | AML1 | - |
| | | 1173-1179 | EN1 | - |

# *CHAPTER III*

Evolution of limb development in marsupial mammals: noncoding RNAs associated with the regulation of *HOX* gene expression

## 3.1 ABSTRACT

Limb development in the tammar wallaby *Macropus eugenii* is characterized by accelerated forelimb development before birth, which allows the newborn to reach the pouch, while hindlimbs initially remain at a very early stage of differentiation and are effectively nonfunctional. Aditionally, kangaroos' feet lack digit I, develop syndactyl digits II and III and an elongated digit IV. *HOXD* genes are essential for limb patterning during tetrapod development, and its complex regulatory system might contribute to the origin of phenotypic novelties. Recent literature recognizes noncoding RNAs (ncRNAs) as a crucial parcel of this regulatory system, and investigation of their conservation and function can provide insights into mechanisms of tetrapod limb evolution. Here, I have identified and compared ncRNAs transcribed in tammar limbs at days 23 (d23) and d25 (d25) of pregnancy, using RNA-seq analyses, and real-time quantitative PCR (qPCR) to validate comparisons; performed conservation analyses using orthologue genomic sequences from opossum, Tasmanian devil, mouse and human; and characterized their spatiotemporal expression patterns in the developing autopodia by whole mount *in situ* hybridisation (WISH). I have also, for the first time, shown the expression patterns of *HOXD10* at d25, and *HOXD11-D12* at d23 and d25 in tammar wallaby autopodia, using WISH, and compared Hox coding and ncRNA expression patterns. I identified five lncRNAs (long noncoding RNAs) located in intergenic and intronic positions among terminal *HOXD* coding genes that are conserved among mammals, and three that are located downstream to *HOXD9* that are exclusive to marsupials. Transcriptional profiles in the limbs of the lncRNAs *XLOC46-49* and the putative pre-miRNA *XLOC53* partially resemble those of *HOXD10-13* genes. The lncRNA *XLOC46* shows spatial expression patterns in the autopodia that resemble those of mouse and tammar *HOXD11-12* genes, whereas *XLOC52* and *XLOC53* are expressed in patterns identical to those in the tammar, except for low *XLOC53* expression in the hind autopodium at d25. These findings suggest that the lncRNAs located intergenically and intronically to the terminal *HOXD* genes (*XLOC46-49*) may regulate expression of these genes during autopodium development in mammals. *XLOC52* and *XLOC53* likely regulate *HOXD* expression in marsupials, and may have contributed to the evolution of specific morphological traits in this lineage. Finally, *XLOC53* may represent a stage-specific pre-miRNA, contributing to the rapid development of the forelimbs compared to that of the hindlimbs in kangaroos. Knowledge on lncRNAs evolution and functions may reveal elements that contributed to morphological evolution in marsupials.

## 3.2 INTRODUCTION

Macropodid marsupials, as the tammar wallaby *Macropus eugenii* (Fig. 3.1), exhibit highly modified hindlimbs in comparison with other marsupials and with eutherians, and heterochrony between development of fore- and hindlimbs (Chew et al. 2012; Chew et al. 2014). When the tammar mother delivers her altricial young, it possesses relatively well-developed forelimbs, used to climb independently, using swimming motions, from the birth canal opening to one of the four teats in the pouch (Shaw & Renfree 2006; Schneider et al. 2009). At this stage, hindlimbs are still fetal and cannot help the young during its journey to the pouch, but after birth the hindlimbs develop very fast and become bigger than the forelimbs, allowing the animal to hop, an efficient way of locomotion for wallabies and kangaroos (Sears 2009; Keyte & Smith 2010; Chew et al. 2014). This specialized appendage is highly modified in comparison with the typical vertebrate autopodium that very often have five separate digits (Weisbecker & Nilsson 2008; Chew et al. 2012). In the tammar foot, digit I is absent even in the fetus, digits II and III are syndactyl (reduced, fused and bound by skin), and digit IV is elongated (Fig. 3.2; Chew et al., 2012; Jones, 1925; Weisbecker and Nilsson, 2008).

**Figure 3.1.** Tammar wallaby *Macropus eugenii* at the marsupial breeding colony from the University of Melbourne. Image credit: Prof. Geoff Shaw.

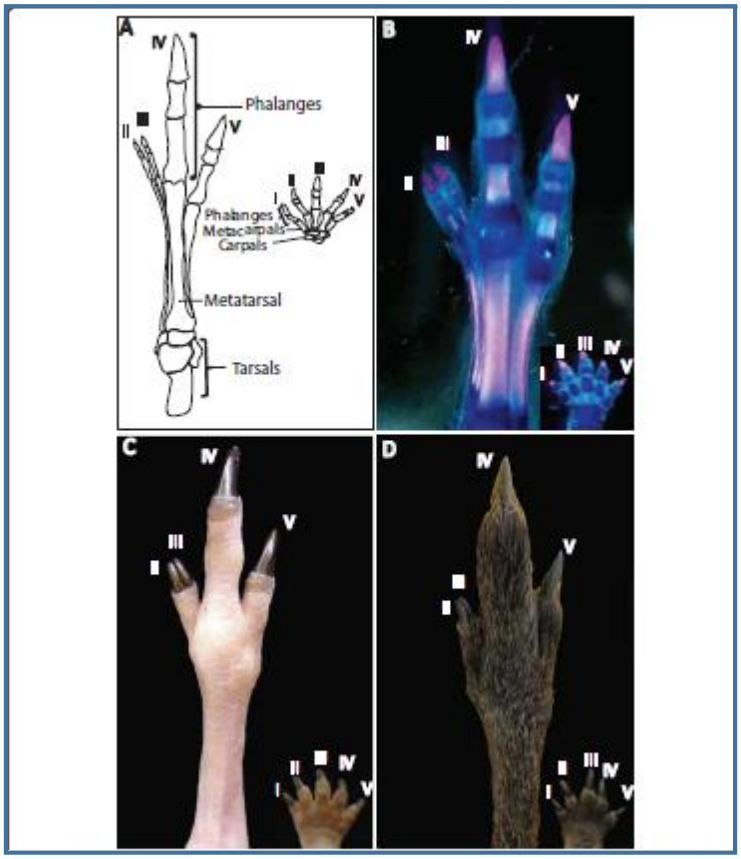


**Figure 3.2.** Morphology of the tammar wallaby autopodium. (A) Diagram of the macropodid fore- and hind autopodial bones. (B) Fore autopodium at day 150 post-partum (pp) and hind autopodium stained for bone and cartilage with alcian blue and counter-stained with alizarin red. (C) External morphology of fore- and hind autopodia at day 120 pp. (D) Adult tammar fore- and hind autopodia. In the fore autopodium, all five digits are present, whereas hind autipodium lacks digit I. Digits are numbered with Roman numerals. (Images copied from Chew et al., 2012).

The syndactyl morphology is a classic example of homoplasy, as it is identified in the foot of all species from the marsupial Orders Diprotodontia (koalas, wombats, possums, kangaroos, and allies) and Peramelemorpha (bilbies and bandicoots); such phenotype is characterized by ossification similarity and strong integration between the syndactyl pedal digits (Fig. 3.3; Weisbecker & Nilsson 2008). Because of the unlikelihood of a complex trait such as syndactyly evolving multiple times, this character was first used to classify marsupials in a "former" clade, the Syndactyla (Hall 1987; Luckett 1994; Marshall 1973; Szalay 1994), but such theory was refuted by a new classification based on dentition characteristics and

molecular systematics (Amrine-Madsen et al. 2003; Kirsch 1977; Nilsson et al. 2004; Phillips et al. 2006). Nowadays, Diprotodontia and Peramelemorpha are widely accepted as non-sister groups [(Kirsch 1977; Nilsson et al. 2004; Phillips et al. 2006), but see (Szalay 1994; Baker et al. 2004)]. The marsupial syndactyly common to both orders is considered non-adaptative because it apparently does not favor locomotion, so the morphological similarities identified argue for a parallelism settled on equivalent developmental pathways (Weisbecker & Nilsson 2008). A malformation, which was not deleterious because the marsupial syndactyl digits remain functional (differently from human syndactyly), apparently explains the origin of this morphology. It seems to have been fixed when the ancestral population went through a bottleneck event (Hall 1987), culminating in an ontogenetic constraint (Weisbecker & Nilsson 2008). The processes regulating morphogenesis of such differentiated appendages and the developmental heterochrony between the macropodid fore- and hindlimb remain however obscure (Chew et al. 2012; Deakin 2012).

Investigation of developmental processes underlying evolution of morphological diversity in marsupial limbs recalls *Hox* genes as main target candidates (Zákány & Duboule 1999; Zakany & Duboule 2007). Among these, *HoxD* genes are particularly relevant because they are intimately connected with limb development in all tetrapods, and their specific expression patterns are frequently associated with the origin of new limb morphologies (Schneider & Shubin 2013). *HoxD* genes are pleitropic, as their expression is also essential to pattern several other body structures during development, such as the anterior-posterior body axis (Deschamps & van Nes 2005; Mallo et al. 2010). The pleiotropic nature of *HoxD* genes can be illustrated by *HOXD13* mutation in human causing limb and genital malformations (Bruneau et al. 2001; Amiel et al. 2004; Sivanantharajah & Percival-Smith 2015). Phenotypic evolution involving changes in gene expression domains, particularly the pleiotropic ones, is often associated with the evolution of regulatory noncoding sequences (Spitz & Duboule 2008;

Spitz 2010b; Montavon et al. 2011). Despite the prominent morphological and developmental peculiarities of the tammar limbs and the fact that *HOXD13* gene expression patterns in the tammar limb is detected at earlier stages of embryonic development in the forelimb than in the hindlimb (which differs from its mouse orthologue), the *HoxD13* coding sequence is highly conserved between mouse and tammar (Chew et al. 2012). In scenarios where coding regions are too conserved, morphological differences may be explained by variation in the way that genes are expressed, which is modulated by regulatory elements. One important parcel of *HOX* gene expression regulatory elements is comprised by noncoding RNAs (ncRNAs) (Petruk et al. 2006; Kumar et al. 2015), including the micro- (miRNAs) (Yekta et al. 2008) and long noncoding RNAs (lncRNAs) (Rinn et al. 2007; Chew et al. 2012). These elements provide great specificity in gene expression during digit development (Montavon et al. 2011), which have been thoroughly investigated in traditional model systems.

Noncoding RNAs (ncRNAs) comprise the vast majority of genome transcripts (Ponting et al. 2009; Djebali et al. 2012), and are likely involved in evolution of specific phenotypes. Transcription of hundreds of ncRNAs from Hox loci have been observed in fly, mouse and human (Bae et al. 2002; Noordermeer & Duboule 2013; Dasen 2013). Amongst those are the miRNAs, comprising a class of short noncoding RNA molecules that regulate expression of their mRNA targets (Mansfield et al. 2004), some of which have been shown to regulate *HoxD* genes during limb development. In mice, limb bud development relies on three miRNAs, *miR-let-7c* and *miR-let-7e* and *miR-10b*, all involved in cell fate determination (Mansfield et al. 2004). In the tammar wallaby, both conserved (*mir-10a, mir-10b, mir-196a, mir-196a2, and mir-196b)* and one novel miRNA located in Hox clusters have been identified (Yu et al. 2012).

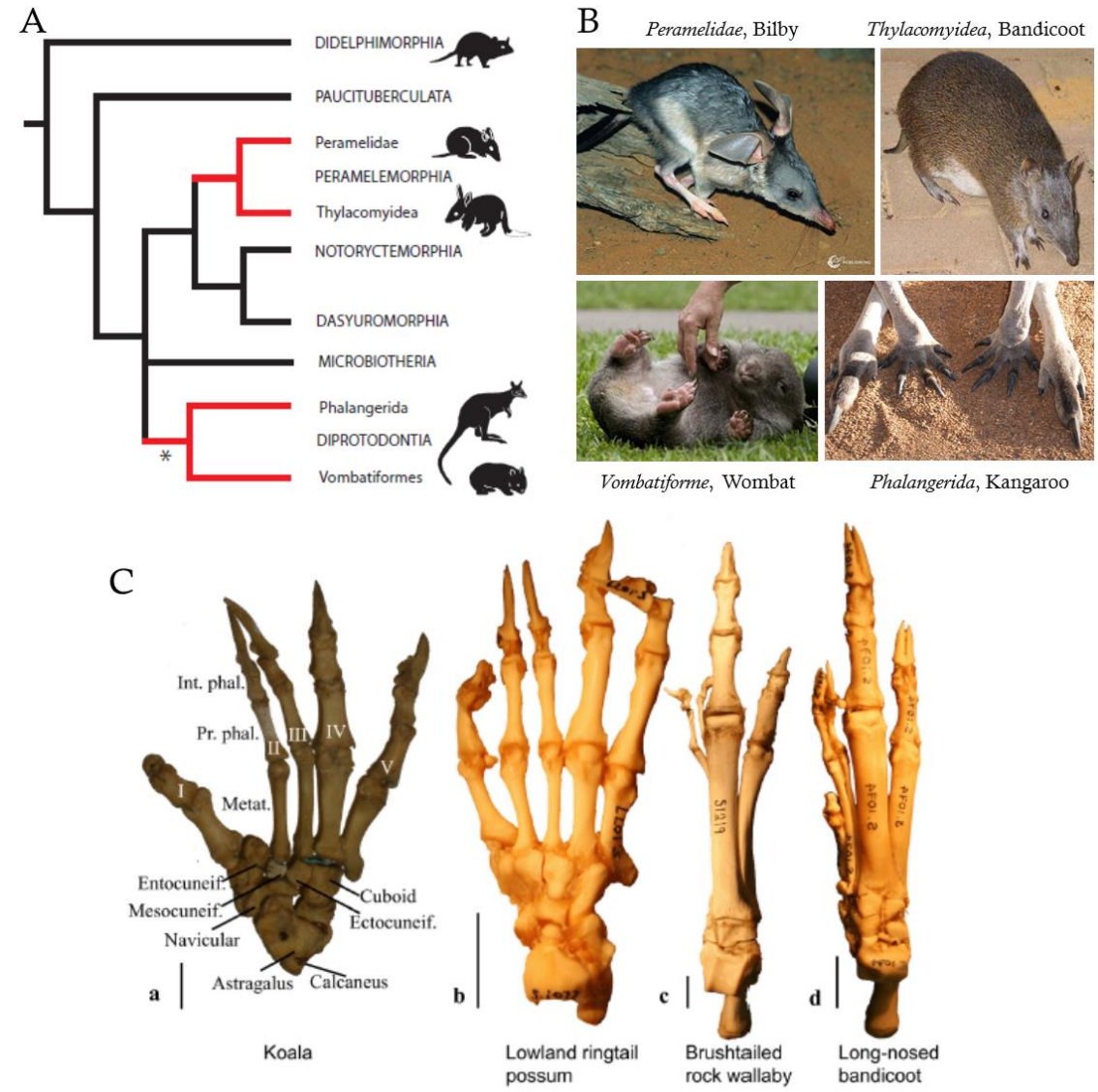


**Figure 3.3**. Syndactyl feet occur in both Peramelemorpha (bilbies and bandicoots, top B) and Diprotodontia (wombats and kangaroos, bottom B), which are not sister groups according to the phylogenies proposed by Weisbecker and Nilsson (2008) and Amrine-Madsen et al., (2003) (A). C) Right pedal skeleton of four syndacylous marsupials, showing tarsal bones and separated digits of a koala (a), a lowland ringtail possum *Pseudochirulus canescens* (b), a brush-tail rock wallaby (c), and a long-nosed bandicoot *Perameles nasuta* (d). Figures *a-c* are diprotodontian marsupials, while *d* is a peramelemorph. Int. phal., Intermediate phalanx; Metat., metatarsal; Prox. Phal, proximal phalanx. Roman numerals correspond to digit numbers. Vertical scale bar = 1 cm. Adapted from Figs. 1 and 3 from Weisbecker and Nilsson (2008).

Long non-coding RNAs, on the other hand, constitute an extremely variable class of transcripts that are longer than 200 base pairs (Saxena & Carninci 2011), which participate in

target gene silencing or activation, working as key *cis-* or *trans-* acting regulators of transcriptional and translational output (Ponting et al. 2009; Saxena & Carninci 2011; Dinger et al. 2008). These transcripts are fascinating for engaging several different genetic and epigenetic mechanisms involved in regulation of gene expression in specific cellular contexts and stages of differentiation (Rinn et al. 2007; Ponting et al. 2009; Blignaut et al. 2012; Dasen 2013), reason for these elements to be defined as 'fine-tuners' of cell fate (Fatica & Bozzoni 2014). The Hox lncRNAs *HOXA11AS*, *HOTAIRM1*, and *HOTAIR* regulate site-specific *Hox* gene expression in eutherians that have also been identified in the tammar wallaby (Yu et al. 2012). Among these, *HOTAIR* is known to play a role in the regulation of *HoxD* genes during limb development, although it originates from the HoxC cluster located on another chromosome, representing the first *trans*-acting lncRNAs ever described (Rinn et al. 2007; Saxena & Carninci 2011). In the tammar, it is expressed at the early head-fold stage of the embryo before limb buds develop (Chew et al. 2012; Yu et al. 2012). In human fibroblasts, *HOTAIR* knockdown decreases activity of *HOXD* genes repressor complexes, increasing their expression across 40 kilobases (Rinn et al. 2007); in mouse, it causes homeotic transformation of spinal vertebrae and limb defects (Dasen 2013; Fatica & Bozzoni 2014). Another lncRNA that influences limb development in the mouse, is "*HOXA* Transcript at the distal TIP" (*HOTTIP*), which is transcribed from 5′ end of the human HOXA *locus* upstream of *HOXA13* and coordinates the activation of several *HOXA* genes in vivo. *HOTTIP* knockdown in mouse limb buds decreases the expression of *HOXA13*, *HOXA11* and *HOXA10* and affects limb morphology in chick and mouse embryos, resembling some of the defects observed in mice lacking *HOXA11* and *HOXA13* (Wang et al. 2011; Fromental-Ramain et al. 1996; Small & Potter 1993; Davis et al. 1995; Fatica & Bozzoni 2014). Specific functional roles of these conserved lncRNAs were not yet investigated in marsupials.

NcRNAs are recently-discovered regulators, important for addressing questions about the evolution of lineage-specific transcriptional regulation (Yu et al. 2012). Investigating the lncRNAs of tammar wallaby HoxD cluster that are transcribed in developing limbs is particularly interesting due to its unique characteristics such as their digital formula, locomotion by hopping, and limb developmental heterochrony (Lindsay et al. 2012). The divergence time between eutherian and marsupials, estimated for about 160 Mya (Luo 2007), is a sufficiently long period for allowing regulatory complexes to evolve concomitantly with new traits, which turns mammals an interesting target for investigating ncRNA evolution (Yu et al. 2012). In this context, I have investigated conservation and spatiotemporal expression of *HOXD* ncRNAs transcribed in *M. eugenii* developing autopodia, in order to identify regulatory elements that may have contributed to the evolution of marsupial developmental characteristics and unique digit phenotypes. I predict that some of those ncRNAs will be conserved amongst mammals, whereas other ones might have evolved exclusively in marsupials' genomes, representing important candidates for modulating morphological evolution. I also hypothesize that some of the ncRNAs exhibit transcription profiles and expression patterns in the fore- and hind autopodia at distinct developmental stages that resemble those of *HOXD* coding genes, indicating possible regulatory roles of these ncRNAs during autopodial development.

## 3.3 METHODS

*Tissue assessment: collecting tammar wallaby and programming gestation*

All sampling techniques and tissue collection procedures conformed to the Australian National Health and Medical Research Council guidelines (2013) and were approved by the University of Melbourne Animal Experimentation and Ethics Committees. The tammar wallaby *Macropus eugenii* were regularly collected at the Kangaroo Island, SA, and maintained in the marsupial colony of Prof. Marilyn Renfree's group in Melbourne, VIC. The expeditions were composed by four people who travelled to the Kangaroo Island and spent 7 days performing nocturnal collection of tammar individuals (approximately one hundred females and a couple of males). Researchers chased animals belonging to an overpopulation that inhabits local farm´s paddocks. Individuals collected over the night were screened the next day for reproductive stage based on pouch examination, and classified as: 1) pouch empty, dirty and no plug (absence of a young tammar in the pouch, which was dirty because the female was not close to give birth, no lactating mammary glands, and no evidence of sperm in the genital, which would indicate recent copulation); 2) pouch empty, dirty and plug; 3) pouch empty and clean (indicating proximity of birth); 4) pouch young (presence of a newborn fetus or fury young in the pouch; see Fig. 3.4); 5) pouch empty with lactating gland (indicating that the female was breast-feeding). Animals were also checked for injures, treated and ear-tagged, receiving a number that identifies each individual for future screening and experimental use. Collected animals were transported to the marsupial colony and distributed in several yards where they fed on green grass and received lucerne cubes (Fig. 3.5) as a supplement three times per week. In the colony, they were regularly screened for reproductive stage and rotated between yards during reproduction management (see Fig. 3.6 for collecting method).

**Figure 3.4.** Examining tammar wallaby pouches. The females are carryng pouch youngs, at an early stage on the left, and a late stage on the right.

**Figure 3.5.** Tammar wallaby feeding on lucern cubes at the marsupial colony. Image credit: Prof. Geoff Shaw.

**Figure 3.6.** Students catching a wallaby specimen with the help of a sack and a net, at the marsupial colony.

Most embryos used in this project were collected in January of 2014 and 2015 at days 23 (d23), 24 (d24) and 25 (d25) of pregnancy, as the tammar wallaby reproduction is seasonal, occurring from January to June. The tammar wallaby mate at a post-partum oestrus, but the blastocyst entries diapause controlled by both lactation and photoperiod (Renfree & Tyndale-Biscoe 1973; Browner et al. 2004), synchronizing births to occur around end of January and the first half of February (Andrewartha & Barker 1969; Berger 1966; Renfree & Tyndale-Biscoe 1973; Sadleir & Tyndale-Biscoe 1977). These reproductive characteristics allow programming females to give birth by removing the young from the mother's pouch and injecting the female with bromocriptine, a dopamine agonist that suppresses the peripartum pulse of plasma prolactin (Fletcher et al. 1990), inducing immediate blastocyst reactivation and allowing control of embryonic developmental stage. Five females were programmed and sacrificed on the d23 of pregnancy for obtaining additional embryos. Embryos' whole bodies or individualized limbs destined to RNA extraction were immediately frozen using liquid nitrogen and maintained at -80°C. The tissues used for whole-mount *in situ* hybridisation were dissected and immediately fixed at 4% paraformaldehyde (PFA)/PBS overnight at 4°C, then washed next day with PBS and 25, 50, 75 and 100% methanol (MeOH)/PBS for 20 min each step on ice, then stored at -20°C.

*Using RNA-seq data to access transcripts expressed in tammar wallaby developing limbs*

Total RNA extracted from forelimbs and hindlimbs from d23 and d25 (N=3, pooled together) was used to perform RNA-seq, in collaboration with Professor Asao Fujiyama and his team at National Institute of Genetics (Japan). The RNA-seq was analyzed through VLSCI Supercomputer Facility (https://www.vlsci.org.au) implemented to allow data processing. The protocol described in Trapnell et al. (2012) was used to identify transcripts of interest and to compare their expression levels between tissues and embryonic stages.

Transcription reads were mapped to *Macropus eugenii* genome (Renfree et al. 2011) using TopHat (Trapnell et al. 2009), a tool that also infers read spans when segments align to the genome far apart from one another, and estimates location of junction's splice sites, working independently from gene or splice site annotations (Trapnell et al. 2012). The resulting alignment files were inserted on to Cufflinks package (Trapnell et al. 2010) in order to generate a transcriptome assembly for each specific tissue and developmental stage. Using Cufflinks, the splicing structure of each gene was inferred, the expression level of each transfrag in the sample was quantified, and background or artefactual transfrags were filtered out. Subsequently, Cuffmerge was implemented to merge these assemblies together, generating a uniform database for calculating gene and transcript expression in each condition. Once each sample was assembled and all samples were merged, genes and transcripts that are differentially expressed or regulated between samples were identified from the final assembly. Both reads and merged assembly were inserted in Cuffdiff to calculate transcriptional levels and to test for statistical significance of observed changes. Text files reporting results were displayed in the software CummeRbund (http://compbio.mit.edu/cummeRbund/), which was used to create expression plots by converting Cufflinks output files into statistical objects that were then analyzed in R package version 2.2 and accessed through the Bioconductor website (http://www.bioconductor.org/).

Reads aligned to HoxA and HoxD clusters, in intergenic and intronic positions, were selected as target transcripts for investigation. As transcripts identified outside annotated genes can correspond either to new protein-coding genes or to noncoding RNA, I predicted coding potential of the sequences using online software ORF Finder (Rombel et al. 2002) and Coding Potential Assessment Tool (Wang et al., 2013) prior to investigating transcripts expression profiles and associated roles.

*Investigating conservation of HoxD cluster transcripts among mammals*

The transcripts from HoxD cluster identified by RNA-seq analyses were mapped into genomes of the marsupials opossum (*Monodelphis domestica*) and Tasmanian devil (*Sarcophilus harrisii*), and the eutherians mouse (*Mus musculus*) and human (*Homo sapiens*) using the BLAT tool on Ensemble Genome Browser database. Matches were individually evaluated, and those located in the HoxD cluster of the referred species were used to extract conserved regions. The genomic sequences corresponding to transcriptional fragments were used in the online software mVista to perform custom alignments. Annotation information, where present, was also valuable to infer known functions of these trancripts in other organisms.

*Confirming ncRNAs transcription by reverse transcriptase polymerase chain reaction, cloning and sequencing*

The construction of transcriptome assemblies allow for inferences of novel genes and transcripts, which must be validated by traditional cloning and PCR-based techniques (Trapnell et al. 2012). In order to validate transcription of *HOXA* and *HOXD* intergenic noncoding RNAs identified in the transcriptome, I first performed reverse transcriptase polymerase chain reaction (PCR). Total RNA was extracted from fore and hind autopodia dissected at d23 and d25 that have been instantly frozen, using the GenElute™ Mammalian Total RNA Miniprep Kit (Sigma-Aldrich, AU) and following the manufacturer's protocol. To assure that there was no genomic DNA in the samples, RNA was decontaminated using DNA-*free*™ kit (Ambion), following the manufacturer's protocol. DNA-decontaminated RNA was used as a template for a PCR using tammar *HOX* intronic primers (kind gift of Dr. Hongshi Yu) in order to verify that no amplification occurred, which would indicate genomic contamination. In case there was a smear band, decontamination reaction and PCR were repeated. RNA was finally stored at –80 °C.

Subsequently, cDNA was synthesized from RNA from fore- and hindlimb of d23 and d25 using SuperScript™ III First Strand Synthesis System for RT-PCR (Invitrogen), according to the manufacturer's protocol. The cDNA synthesis product was stored at -20°C and then used as template for PCR. For confirming transcription of intergenic noncoding RNAs identified by the transcriptome in the HoxA and HoxD clusters, primers were designed (Tables 3.1 and 3.2) using Primer 3 online software (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus. cgi/, Untergasser et al. (2007).

**Table 3.1.** Primers used for amplifying transcripts mapped to the tammar wallaby HoxD cluster; sequences are presented together with melting temperatures used in the PCRs.

| Transcript name/loci | Forward (F) and Reverse (R) Primers (3'-5') | PCR melting temperature |
|---|---|---|
| XLOC_033946 ***(XLOC46)*** | **F:**CCCTGTCACAAACCCTCACT<br>**R:**ATTCGACAAGAGCGAGCAAT | 58°C |
| XLOC_033947 ***(XLOC47)*** | **F:**AGCTCCCTGAATTATCCTTCAAC<br>**R:**AGGAGTTGCCCTTAAACAAGAAC | 60°C |
| XLOC_033948 ***(XLOC48)*** | **F:** CTTTGTTTGTGACTTGGGAAAAG<br>**R:**ATGAGAAACTGTTGTGGCAACTT | 60°C |
| XLOC_033949 ***(XLOC49)*** | **F:** TGGTTAGTAGCGAGGGGAAA<br>**R:** CCCTGTGTGGTTTCATAGAGAA | 55°C |
| XLOC_033950 ***(XLOC50)*** | **F:** CCGGTGCGGTAAATGTAAAC<br>**R:** GCGGGTTCACAAAGTAGGAG | 58°C |
| XLOC_033951 ***(XLOC51)*** | **F:** TTTTCCCTTCCTTGCCTTTT<br>**R:** ATGAGGGAAACACCTATATTGC | 52°C |
| XLOC_033952 ***(XLOC52)*** | **F:**TGTTAAAGTCCTTGCAAATTATGG<br>**R:** AGCGGCTGTTACACAAGCAT | 55°C |
| XLOC_033953 ***(XLOC53)*** | **F:** CAGGTGCTAGGTGGCATCTTA<br>**R:** GGGTTGAAGGAGGAGTGGTC | 60°C |
| XLOC_033954 ***(XLOC54)*** | **F:** CTACCCTCATCCCCCACTTT<br>**R:** AGGGCATTTCTCCCATACTG | 60°C |
| XLOC_033955 ***(XLOC55)*** | **F:** TTTGTTCTTCCTTGGGGTGA<br>**R:** GGGGAAGCTCTCTGTAGCTG | 52°C |

**Table 3.2.** List of primers used for amplifying transcripts mapped to the tammar wallaby HoxA cluster; sequences are presented together with melting temperatures used in the PCRs.

| Transcript name/loci | Forward (F) and Reverse (R) Primers (3'-5') | PCR melting temperature |
|---|---|---|
| XLOC_053797 ***(XLOC97)*-isoform 1** | **F:** CTCTGGTGAATGGGACCTGT **R:** AGCAAGCCAGACTTCAAAGC | 62°C |
| XLOC_053797 ***(XLOC97)*-isoform 2** | **F:** GACCGCAGTGGTTTTTCTGT **R:** AGCAAGCCAGACTTCAAAGC | 62°C |
| XLOC_053798 ***(XLOC98)*** | **F:** CTTGCCGAAGTCTCCTTGAC **R:** CAGCCTTCCTCTCAGAATGG | 55°C |
| (XLOC_053799) ***(XLOC99)*-isoform 1, 2 and 3** | **F:** GTGCTCAGCAGTGAGATGGA **R:** CCATGTTCTTGCCAGTCCTT | 55°C |
| XLOC_053799 ***(XLOC99)*-isoform 2** | **F:** GGGAACAACCCCTGTTCATA **R:** CCACTCTCTGCCACTTCTCC | - |
| XLOC_053799 ***(XLOC99)*-isoform 4** | **F:** GGTGAAGGTGCAGTCAGGAT **R:** GATGCCTGTATGTGGCCTCT | - |
| XLOC_053800 ***(XLOC800)*- isoform 1** | **F:** CCAACCCAGCAAACTTTCAG **R:** AGGTGCCAACAGAAGGAAAA | 55°C |
| XLOC_053800 ***(XLOC800)*- isoform 2** | **F:** GAGAGTCCCAGCTCTGGTCA **R:** AGGTGCCAACAGAAGGAAAA | - |
| XLOC_053800 ***(XLOC800)*- isoform 3** | **F:** GCCTGCTCAGTCACAATGAA **R:** CACACACCAGGGAAACTCCT | 55°C |

The efficiency of PCRs (d25 µl) were verified by electrophorese on 1% agarose gel. After amplification conditions were optimized, reactions were repeated using double volume (50 µl), and ran at 2% agarose gel, where DNA from the bands having suitable sizes were purified using *QIAquick Gel Extraction Kit* (Qiagen), following the manufacturer's protocol. The purified DNA was ligated to pGEM-t easy vector (*Promega*), according to the manufacturer's protocol. Ligation reactions (5 µL) were transformed into thermo-competent bacteria (JM109, Promega) through thermic shock at 42°C for 45 seconds. Bacteria were grown for one hour in 800 µL LB broth, centrifuged and plated on solid agar containing x-gal and ampicillin for double selection, incubated overnight at 37°C. Three white colonies of each plate were scrubbed using a sterile pipette tip into a PCR tube reaction containing specific primers of the putative insert; the pipette tips were incubated overnight in LB broth at 37°C. Clones exhibiting band having the expected sizes after PCR were purified using Wizard® Plus SV

Minipreps DNA Purification (Promega), following the manufacturer's protocol. Before purifying, 500 µL of bacteria in LB broth were added to 500 µL 50% glycerol and stored at -80°C for stock. Plasmid DNA samples were sent to sequencing for insert confirmation.

*Quantifying transcription using real time quantitative polymerase chain reaction (qPCR)*

Primers for qPCR were designed following the protocol described in Thornton and Basu (2011), and the oligonucleotides used are provided in Tables 3.3 and 3.4. The cDNAs were designed from RNA extracted of a minimum of 6 individuals for each condition compared (fore and hind autopodia of embryos with 23 and 25 days of development). The RNA extracted as previously described was quantified using Quibit equipment, and the same amount of RNA was used to synthesize every cDNA sample used. Standard curves were performed for every gene using cDNA serial dilutions. Pure DNA was diluted in 1 volume of nuclease free water, corresponding to dilution 1. Subsequent dilutions were 5, 25, 125, 625, 1265 or 10, 100, 1000, 2000, 10000. Optimal dilutions fell within the same range, and all cDNA pools were ultimately diluted 50 times for every differential expression experiment. Reactions were set up according the protocol from SYBR Green PCR Master mix (Thermo Fisher Scientific). Standard curves were managed on Max Pro software, where threshold values were manually adjusted in order to obtain cycle threshold (Ct) values based on primer efficiency (PE) and Pearson Correlation Coefficient (RSq) as closest as possible to 100% and 1, respectively (see individual values on Table 3.3).

**Table 3.3.** List of primers used to quantify tammar wallaby ncRNA from HoxD cluster by RT-qPCR. TN (transcript name); PC (primer concentration); PCR MT (melting temperature); TV (threshold value); RSq (R squared, Pearson Correlation Coefficient); PE (primer efficiency).

| TN | Forward (qF) and Reverse (qR) Primers (3'-5') | PCR MT (°C) | PC(M) | TV | RSq | PE (%) |
|---|---|---|---|---|---|---|
| TBP | **qF:**GGACAAACTGAAGCAAAGGGACC<br>**qR:**AGGGCATCATTGGGCTAAAGATAG | 58-60 | 0.5 | 1000 | 0.997 | 96.6 |
| GAPDH | **qF:**TCCCAATGTATCTGTTGTGGATCTG<br>**qR:**AACCATACTCATTGTCATACCAAGAAAT | 60 | 0.4 | 1000 | 0.999 | 91 |
| *XLOC46* | **qF:**AATCCCACAAATCCCTGACCC<br>**qR:**ACACAAAGAAAGAAACCCAAACCA | 64 | 1 | 1362.7 | 0.995 | 96.2 |
| *XLOC47* | **qF:**ATCGTTTGTGTCTTTCTTGCCC<br>**qR:**GCGTCCTGTCTCTCTGCC | 66 | 1 | 6565.8 | 0.998 | 100 |
| *XLOC48* | **qF:**GACAGAGGAGATTTATTTGGGAAGGG<br>**qR:**CTGGACTTGGATTTCAGGCGG | 60 | 1 | 1000 | 0.898 | 98.7 |
| *XLOC49* | **qF:**AAATATCTTCCCGTGTGTGGC<br>**qR:**AAATTCTCATCTACATTCTTGGGC | 55 | 0.5 | 546.08 | 0.997 | 99.8 |
| *XLOC53* | ***F:**CAGGTGCTAGGTGGCATCTTA<br>***R:**GGGTTGAAGGAGGAGTGGTC | 60 | 0.4 | 1000 | 0.992 | 97.2 |
| *XLOC54* | **qF:**AAAGTGAAGGTTTGATTCTATGCC<br>**qR:**CTCTCCTACATTCCATAAGCCC | 60 | 0.5 | 1368.5 | 0.99 | 105.5 |
| *HOXD10* | **qF:**CTACTTGCTCCTTCACTGCC<br>**qR:**CGGAAGAAATTAGTTTGGTTCGC | 60 | 0.5 | 8353.3 | 0.999 | 101.9 |
| *HOXD11* | **qF:**TGAATAATAGTTTCCGAGACTGGG<br>**qR:**ACTCAGACTTATGCCAATCCG | 60 | 0.5 | 3763.4 | 0.998 | 100 |
| *HOXD12* | **qF:**TACAAGGAGAAGAAGAAGTTAGCG<br>**qR:**ATTCTAATTGAGGTGGTAGGAGGG | 60 | 0.2 | 527.019 | 0.995 | 99.5 |
| *HOXD13* | **qF:**ATCTTCCTTTCCAGGGGATG<br>**qR:**ATGGCGTATTCGTTCTCCAG | 60 | 1 | 1528 | 0.998 | 100.7 |

*Primers designed following qPCR specific protocol did not successfully amplify this gene, so primers designed for cloning were used.

Differential expression experiments were set up according to Bustin et al. (2009), and all samples were compared in the same plate for each gene of interest. The plate model used was exemplified in Fig. 3.7. Real time-qPCR results were adjusted on Max pro software

according to individual threshold values, and analyzed both on REST (Pfaffl et al. 2002) statistical software and manually on Microsoft Excel using a normalization formula.

***RT-qPCR plate***

Date: _____

**Experiment: Hoxd13 differential expression**_____

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | Negative Control-Hoxd13 | Negative Control-Hoxd13 | Negative Control-Hoxd13 | Negative Control-Hoxd13 | Negative Control-Hoxd13 | Negative Control-Hoxd13 | Negative Control-TBP | Negative Control-TBP | Negative Control-TBP | Negative Control-GAPDH | Negative Control-GAPDH | Negative Control-GAPDH |
| B | Calibration-template (1:125 dilution) | Calibration-template (1:125 dilution) | Calibration-template (1:125 dilution) | Calibration TBP | Calibration TBP | Calibration TBP | Calibration GAPDH | Calibration GAPDH | Calibration GAPDH | Empty well | Empty well | Empty well |
| C | FL-d23 Replicate 1 | FL-d23 Replicate 1 | FL-d23 Replicate 1 | HL-d23 Replicate 1 | HL-d23 Replicate 1 | HL-d23 Replicate 1 | FL-d25 Replicate 1 | FL-d25 Replicate 1 | FL-d25 Replicate 1 | HL-d25 Replicate 1 | HL-d25 Replicate 1 | HL-d25 Replicate 1 |
| D | FL-d23 Replicate 2 | FL-d23 Replicate 2 | FL-d23 Replicate 2 | HL-d23 Replicate 2 | HL-d23 Replicate 2 | HL-d23 Replicate 2 | FL-d25 Replicate 2 | FL-d25 Replicate 2 | FL-d25 Replicate 2 | HL-d25 Replicate 2 | HL-d25 Replicate 2 | HL-d25 Replicate 2 |
| E | FL-d23 Replicate 3 | FL-d23 Replicate 3 | FL-d23 Replicate 3 | HL-d23 Replicate 3 | HL-d23 Replicate 3 | HL-d23 Replicate 3 | FL-d25 Replicate 3 | FL-d25 Replicate 3 | FL-d25 Replicate 3 | HL-d25 Replicate 3 | HL-d25 Replicate 3 | HL-d25 Replicate 3 |
| F | FL-d23 Replicate 4 | FL-d23 Replicate 4 | FL-d23 Replicate 4 | HL-d23 Replicate 4 | HL-d23 Replicate 4 | HL-d23 Replicate 4 | FL-d25 Replicate 4 | FL-d25 Replicate 4 | FL-d25 Replicate 4 | HL-d25 Replicate 4 | HL-d25 Replicate 4 | HL-d25 Replicate 4 |
| G | FL-d23 Replicate 5 | FL-d23 Replicate 5 | FL-d23 Replicate 5 | HL-d23 Replicate 5 | HL-d23 Replicate 5 | HL-d23 Replicate 5 | FL-d25 Replicate 5 | FL-d25 Replicate 5 | FL-d25 Replicate 5 | HL-d25 Replicate 5 | HL-d25 Replicate 5 | HL-d25 Replicate 5 |
| H | FL-d23 Replicate 6 | FL-d23 Replicate 6 | FL-d23 Replicate 6 | HL-d23 Replicate 6 | HL-d23 Replicate 6 | HL-d23 Replicate 6 | FL-d25 Replicate 6 | FL-d25 Replicate 6 | FL-d25 Replicate 6 | HL-d25 Replicate 6 | HL-d25 Replicate 6 | HL-d25 Replicate 6 |

**Figure 3.7.** Example of plate set up applied to all genes differential expression qPCR experiments performed to validate RNA-seq data.

*Spatial characterization of noncoding RNA expression profiles using whole-mount in situ hybridisation (WISH)*

Based on the expression profiles acquired by RNA-seq and qPCR data, I chose the lncRNAs *XLOC46*, *XLOC52* and *XLOC53* as target genes to perform expression spatial characterization using WISH. These genes have been chosen because they had higher transcriptional levels and were adequate representatives of distinct transcriptional profiles (see Results section). The coding genes *HOXD10*, *HOXD11* and *HOXD12* were chosen as reference to interpret lncRNAs expression profiles. I used lncRNAs plasmid DNA produced as previously described to design probes, and used the aforementioned methods to clone *HOX* genes coding sequences (Table 3.4).

**Table 3.4.** List of primers used to amplify tammar wallaby *HOXD* genes, together with melting temperatures used in the PCR reactions.

| Transcript name/loci | Forward (F) and Reverse (R) Primers (3'-5') | PCR melting temperature |
|---|---|---|
| ***HOXD9*** | **F:** GCTGTCCCTACACAAAATACCAG<br>**R:** CTGGCACCTAAAGCAAACTAAGA | 55°C |
| ***HOXD10*** | **F:** CCACAATTACACAGGGAATGTTT<br>**R:** TGAGGGACTTTAGCTGAGTCTTG | 58°C |
| ***HOXD11*** | **F:** TTGTCTGAGTTCCATACCGAGAT<br>**R:** TAGGGGAAAGTCATTTGACAAGA | 55°C |
| ***HOXD12*** | **F:** GCCTTTGAGTTCCCTTGTTACTT<br>**R:** ACCTACTGGGAGAGAAGTCATCC | 60°C |

In order to design RNA probes (riboprobes) necessary for performing WISH, plasmid DNA was first linearized and the orientation of inserts in the vectors was identified by PCRs. The following set of primers was used in four distinct reactions *per* target gene: 1) Gene specific forward primer + M13 forward primer (see pGEM-t easy vector protocol); 2) Gene specific reverse primer + M13 reverse primer (see pGEM-t easy protocol); 3) Gene specific forward primer + M13 reverse primer; 4) Gene specific reverse primer + M13 forward primer. As a result of these four reactions, I obtained one pair of equal bands presenting the approximate insert size, representing the predominant orientation of respective plasmid inserts. After the correct pair of primers was identified, PCRs using 50-100 µL were repeated in order to obtain a high number of insert copies. Samples were run in 2% agarose gel, and DNA from gel bands was purified as previously described. Purified DNA was used to perform *in vitro* transcription reactions according to DIG RNA Labeling Mix (Roche) according to the manufacturer's protocol.

The technique WISH was performed on limb buds preserved in 100% methanol and stored at -20°C as previously described. On the first day of WISH protocol application, pre-hybridisation was performed starting with rehydrating limbs through MeOH/PBTX series in reverse (75%, 50%, 25% and 0% MeOH) on ice. Limbs were washed three times with PBTX

for 5 minutes each time at room temperature on a rocker; then, they were bleached with 6% hydrogen peroxide in PBTX for 15 minutes at room temperature, washed three times in PBTX for 5 minutes each, and finally treated with 10µg/ml Proteinase K in PBTX at room temperature. Time for the Proteinase K treatment ranged between 1 and 2 hours, depending on tissue and developmental stage, representing the most critical step of this protocol.

Limbs were washed twice in PBTX for 5 minutes, re-fixed carefully with 0.2% glutaraldehyde/4% PFA in PBTX for 20 minutes at room temperature on rocker, washed twice in PBTX for 10 minutes, immersed in 500µl of pre-hybridisation solution (Table 3.5) and then incubated overnight at 65ºC. After this step, embryos can be stored for several days at -20ºC before following up with the experiment. In the second day, limbs were hybridised by adding 15-20 µl of probes *per* sample pre-warmed in 500µl pre-hybridisation solution, and were incubate overnight at 65ºC. On the third day, pre-hybridisation solution was removed from limbs and stored at -20 ºC, because they can be used again. Samples were washed for 5 mins with solutions pre-warmed in heating block to 65ºC, as follows: 1) 100% Solution 1 (Table 3.5); 2) 75% Solution 1/25% 2xSSC; 3) 50% Solution 1/50% 2xSSC and 4) 25% Solution 1/75% 2xSSC. Then, they were washed with 2xSSC, 0.1% CHAPS (Table 3.5) and 0.2xSSC, 0.1% CHAPS once for 30 mins each, at 65ºC on rocker. Subsequently, they were washed twice with TBTX 1X (Table 3.5) for 10 mins at room temperature rocker and preblocked with 10% sheep serum, 2% BSA in TBTX for 2-3 hours at room temperature rocker. Limbs were then removed, immersed in fresh pre-block solution with DIG antibody (1:1000) and rocked overnight at 4ºC. In the fourth day, limbs were washed five times with 1xTBTX containing 0.1% BSA for 30-60 minute. Either in the same or the next day, color reaction was started by washing embryos twice with 1xTBTX for 30 mins each on rocker at room temperature, then washed three times with freshly made NTMT (Table 3.5) for 10 mins each on rocker, at room temperature. They were incubated with 20 µL NBT/BCIP (Roche)

per 1 ml of NTMT, rocking for 20 min in the dark. Then, incubating samples were kept on the bench, protected from light, and checked periodically until color was developed (usually taking 2-6 hours). All reaction pairs (sense and antisense probes) were stopped at the same time, for allowing valid comparisons. To continue color reaction on the next day, limbs were washed in NTMT for 10mins, then TBTX for 10min and left in TBTX overnight at 4ºC in foil. To stop color reaction, they were washed in NTMT for 10mins and in PBTX for at least 20 min (but they were washed for several hours if background was intense), then washed in Ethanol/PBTX series of (25%, 50%, 75%, 100%, 75%, 50%, 25%) for 20-30 min each, and PBTX for 20-30 min. Finally, limbs were fixed in 4%PFA/PBS for 20 min and stored in 50% glycerol at 4 ºC. Photographs were taken under stereomicroscope coupled to a camera, with stained limb buds positioned in plastic plates filled with 50% glycerol, sometimes over an agarose substrate, under several dark and light backgrounds, in order to adjust images individually.

**Table 3.5.** Whole-mount *in situ* hybridisation reagents.

| Solution | Reagents |
|---|---|
| 5% CHAPS | 10g CHAPS; 200ml $MQH_20$ |
| 20X SSC | 350.6g NaCl; 176.4g Sodium citrate; 2lLM$QH_20$ (pH = 7.0) |
| 10x TBTX | 125ml 1M Tris.Cl pH7.5; 75ml 5M NaCl; 25ml 10% Triton X-100; 25ml $MQH_20$ |
| PBTX | 1L PBS; 1ml Triton X-100 |
| Hybridisation Solution | 25ml100% Formamide; 12.5 ml 20x SSC; 25 µl 100mg/ml Heparin; 500 µl 10% Triton X-100; 5ml 5% CHAPS; 50 mg Yeast RNA; 500 µl 500mM EDTA; 1g blocking powder; $MQH_20$ to 50 ml |
| Solution 1 | 25ml 100% Formamide; 12.5 ml 20XSSC; 500 µl 10% Triton-X; 5 ml CHAPS $MQH_2O$ to 50 ml |
| NMTM | 1 ml NaCl; 5 ml Tris pH9.5; 2.5 ml $MgCl_2$; 50 µl Twen 20; MQHO to 50 ml |

Because ncRNAs both sense and antisense probes induced strong staining, I checked transcription orientation to investigate if they were uni or bidirectionally transcribed. cDNA was synthesized from RNA extracted of limbs using the protocol previously described, except that primers specific to each ncRNA studied were used instead of oligo(DT) primers, using either forward or reverse primes in distinct reactions. cDNA from each reaction was used as template for PCRs using primers specifically designed (Table 3.6) to amplify a region internal to the first pair of primers.

**Table 3.6.** Relation of primers designed for amplifying tammar wallaby transcripts in order to check transcription orientation from cDNA synthesized using unidirectional specific primers (Table 3.1).

| Primers name | Forward (F) and Reverse (R) Primers (3'-5') | PCR melting temperature |
|---|---|---|
| **oriXLOC46** | **F:** GGCACAAAAATACCGCATCT | 55-60°C |
| | **R:** CTCAGGTCCTCACAGCAACA | |
| **oriXLOC52** | **F:** TATTCACGCCGAGTCAGTTG | 55-60°C |
| | **R:** TACCAGCCAACACAAGTGGA | |
| **qXLOC53** | **F:** GCTGGAGCTATGCAGCG | 60°C |
| | **R:** AAGGAGGAGTGGTCGGG | |

## 3.4 RESULTS

*RNA-seq transcripts identification and confirmation by reverse transcriptase PCR*

Based on analyses of transcriptome from limbs of embryos having 23 and 25 days of development, 10 and 4 transcripts were inferred respectively within HoxD and HoxA clusters. Curiously, transcripts mapped to tammar wallaby HoxD cluster presented only one isoform (see Table 3.1), whereas those in HoxA presented several ones (see Table 3.2). All transcripts from HoxD cluster, and also some isoforms of the transcripts encountered in HoxA cluster, were confirmed by reverse transcriptase PCR. However, primers for amplifying several isoforms of *HOXA* transcripts were not always specific (Fig. 3.8).

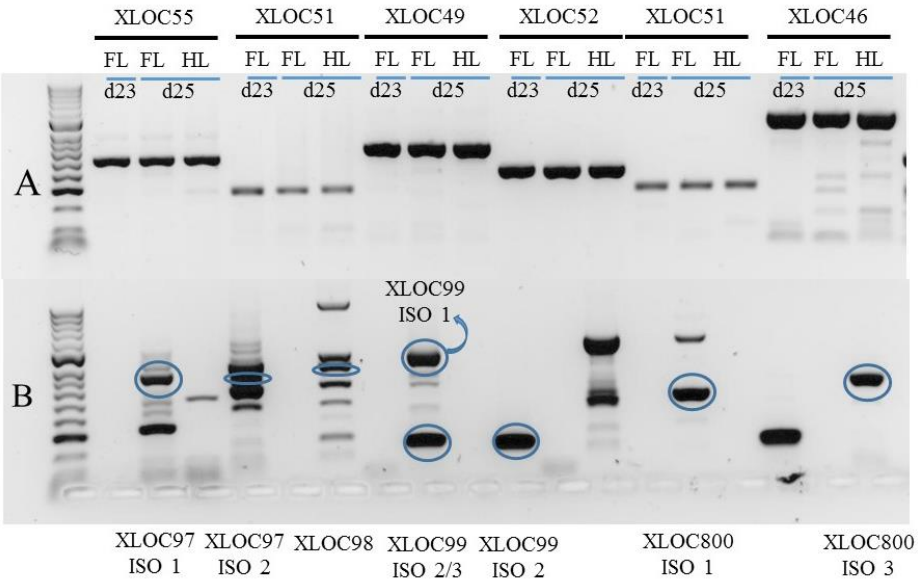


**Figure 3.8.** Image of agarose gel containing PCR amplification of HoxD (A) in forelimb (FL), d23, and hindlimb (HL), d23/d25; and HoxA (B) cluster transcripts from HL, d25. Most primers used for amplifying *HOXA* genes were not specific enough to isolate all transcripts isoforms.

Investigation of transcripts from HoxD cluster was the major focus of this study (Fig. 3.9), not only because it would be more viable to work with single isoforms, but also because

HoxD cluster regulatory elements have been more extensively studied in the literature, therefore representing a priority issue for investigation.

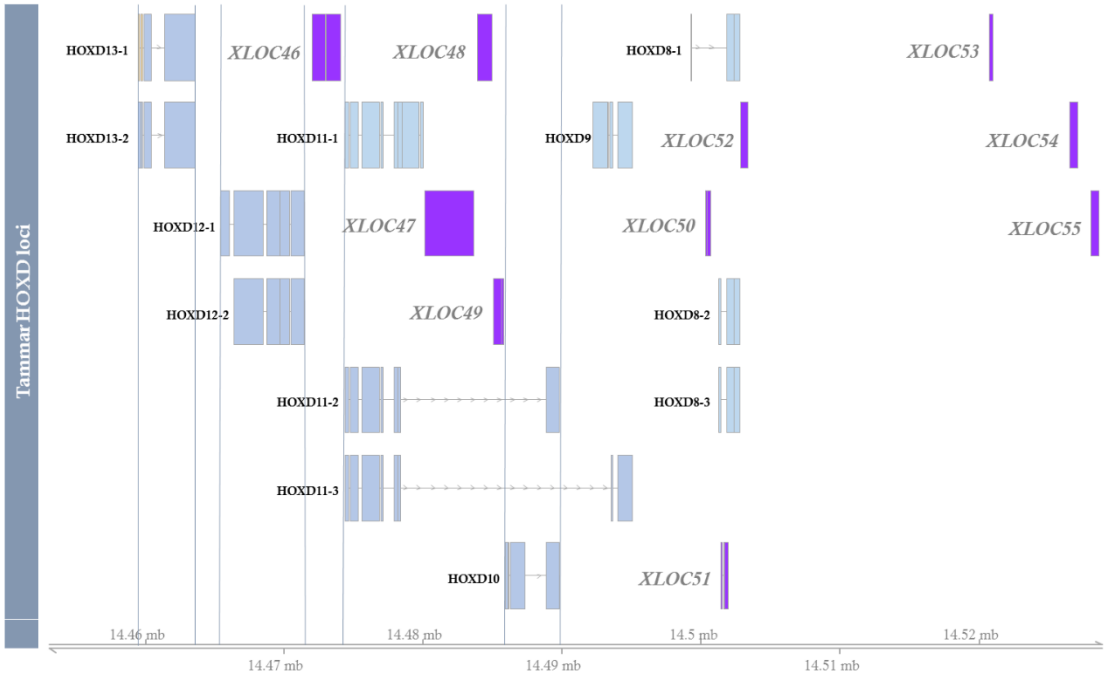


**Figure 3.9.** Tammar wallaby HoxD loci map illustrating location of the transcripts studied. Note that *HOXD10-13* genes appear in more than one isoform (e.g. *HOXD13-1* and *HOXD13-2*)

*Prediction of transcripts' coding potential*

Prediction of transcripts' coding potential using ORF finder showed that most transcripts had very low similarity (< 30%) with any coding sequence annotated in genomes from multiple species available on NCBI databases. However, one ORF was predicted in *XLOC49* that presents 59% similarity with a predicted forkhead protein J2-like from the fish *Pundamilia nyererei* (E-value=0.96, GenBank accession number XP005732030.1). Moreover, one ORF was predicted in *XLOC55* that presented 82% similarity with a hypothetical protein I79_004128 from the Chinese hamster *Cricetulus griseus* (e-value=$2e^{-20}$, GenBank accession number EGW02672.1; Xu et al., *unpublished data*). In order to reinforce coding potential predictions, I also evaluated sequences using CPAT online software (Wang et al. 2013). All

the ORFs predicted had very low probability of holding a coding potential, usually lower than 0.1, except for one ORF predicted in *XLOC50* (=0.172; see Table 3.7).

**Table 3.7.** Calculation of *HOX* intergenic transcripts coding potential.

| Sequence Name | RNA size | ORF size | Ficket Score | Hexamer Score | Coding Probability | Coding Label |
|---|---|---|---|---|---|---|
| **XLOC46** | 2125 | 147 | 0.6342 | -0.1065 | 0.02900 | no |
| **XLOC47** | 3592 | 174 | 0.7716 | 0.0076 | 0.06d259 | no |
| **XLOC48** | 1140 | 1140 | 0.6713 | 0.1852 | 0.03437 | no |
| **XLOC49** | 697 | 54 | 0.9044 | 0.1469 | 0.02229 | no |
| **XLOC50** | 551 | 24 | 1.0292 | 0.4485 | 0.17237 | no |
| **XLOC51** | 314 | 60 | 0.7389 | 0.7033 | 0.002d25 | no |
| **XLOC52** | 478 | 96 | 0.5834 | 0.1159 | 0.01817 | no |
| **XLOC53** | 167 | 96 | 0.9296 | 0.0167 | 0.05654 | no |
| **XLOC54** | 521 | 288 | 0.472 | 0.3607 | 0.02408 | no |
| **XLOC55** | 561 | 78 | 1.1732 | 0.3852 | 0.01896 | no |

*Conservation of HoxD cluster transcripts among mammals based on genomic data*

Investigation of HoxD cluster transcripts conservation among mammals was performed based on data available for genomic sequences deposited in Ensemble Genome Browser. All transcripts comprised conserved regions that were mapped onto opossum and Tasmanian devil HoxD clusters in similar regions to those where they matched sequences in tammar HoxD cluster. Several transcripts were also mapped to mouse and/or human HoxD clusters (Fig. 3.10).

Transcript XLOC_033946 (*XLOC46*), located at *locus* TCONS_00040045 and situated in between transcripts of *HOXD12* and *HOXD11* genes in the tammar HoxD cluster, was identified as fairly conserved among the marsupials studied and does not correspond to any

annotated region in the genome of these organisms. Only one fragment located approximately between base pairs at positions 1100 and 1200 of *XLOC46* sequence seems to be conserved also in mouse and human, a region that in these eutherians overlaps with *HOXD11* gene. The transcripts XLOC_033947 (*XLOC47*), XLOC_033948 (*XLOC48*), XLOC_033949 (*XLOC49*), respectively at *loci* TCONS_00040046, TCONS_00040047 and TCONS_00040048, were identified between *HOXD11-1* and *HOXD10* transcripts in tammar HoxD cluster, but also overlapping with *HOXD11-2* and *HOXD11-3* transcripts, possibly representing intronic lncRNAs. The transcript *XLOC47* exhibited a high level of conservation with opossum and Tasmanian devil genome sequences located in positions equivalent to its position in tammar HoxD cluster, and was not mapped onto any annotated gene in these species. In contrast, *XLOC47* was not identified as conserved in relation to mouse sequences, and only a low-conserved fragment located between positions 2900 and 3200 in the tammar sequence was mapped with human *HOXD10* gene. The transcript *XLOC48* seemed highly conserved among opossum, Tasmanian devil and mouse, and two small fragments (located between positions 300-400 and 600-700 in the tammar's transcript sequence) were identified as conserved in human. The transcript *XLOC48* overlapped human *HOXD10* gene and the noncoding processed mouse transcript Gm28309 (Acc. MGI5579015), while *XLOC49* was identified to be highly conserved among all the species studied; in the eutherians this conserved fragment corresponds to the region up to the first approximate 400 base pairs of the tammar´s sequence. This transcript, as well as *XLOC48*, also overlaps with human *HOXD10* gene and with mouse Gm28309 processed transcript.

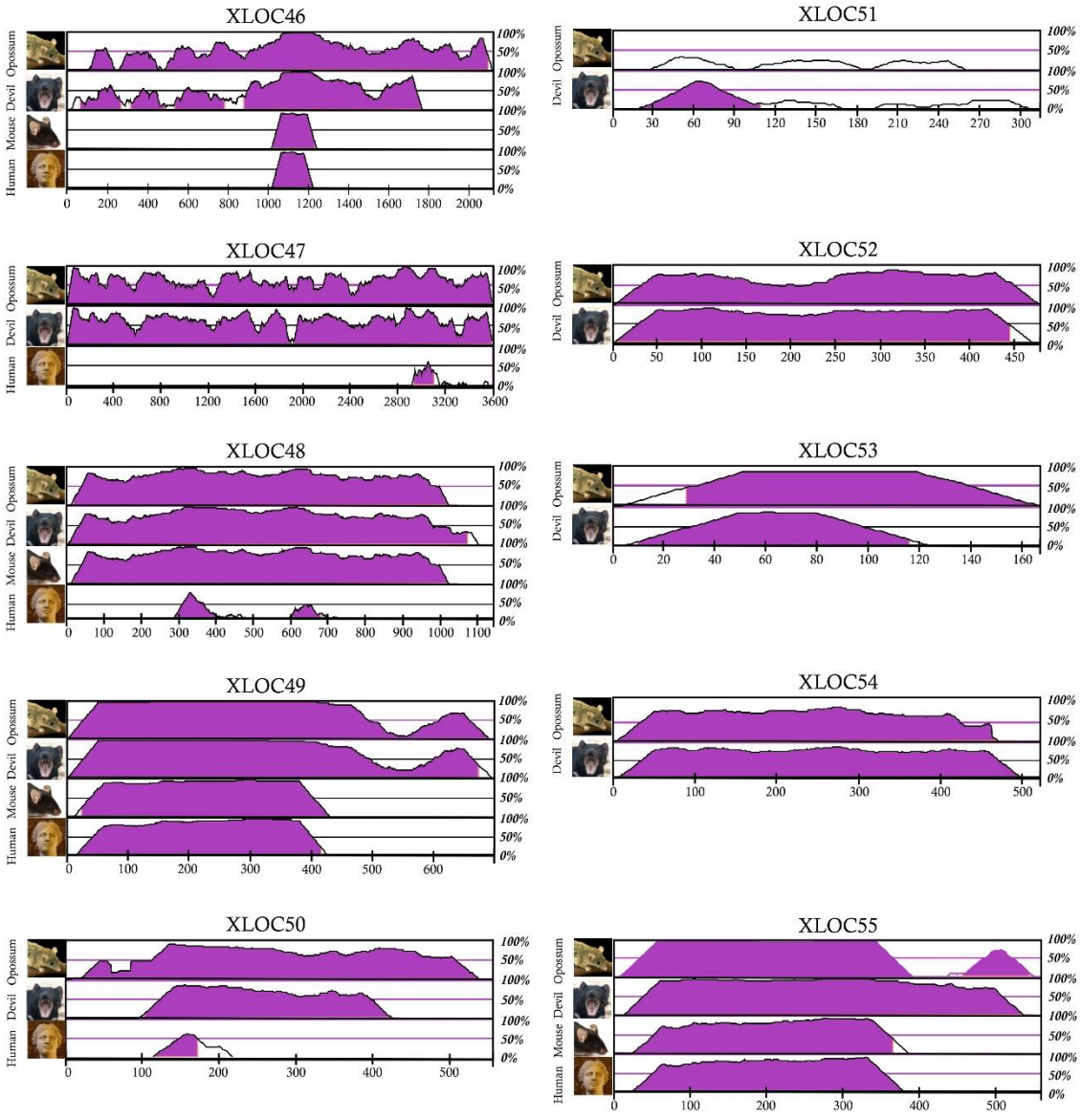


**Figure 3.10.** *HOX* ncRNAs conservation analyses based on genomic DNA sequences. Tammar transcripts were blasted on Ensemble Genome Browser onto genomes of opossum (*Monodelphis domestica*), Tasmanian devil (*Sarcophilus harrisii*), mouse (*Mus musculus*) and human (*Homo sapiens*). Conserved fragments were aligned using mVista online software. The transcripts *XLOX46*, *XLOC47*, *XLOC48*, *XLOC49*, *XLOC50* and *XLOC55* are conserved amongst mammals, whereas *XLOC51*, *XLOC52*, *XLOC53* and *XLOC54* seem to be marsupial elements.

Transcript XLOC_033950 (*XLOC50*), situated at *locus* TCONS_033949 and positioned between *HOXD9* and *HOXD8-2*/*HOXD8-3* transcripts in the tammar transcriptome as well as within *HOXD8-1* transcript, possibly represents another intronic lncRNA. The transcript *XLOC50* mapped to *HOXD8* genes of human, opossum and Tasmanian devil, and to human

HOXD-AS2, a noncoding transcript named HoxD cluster antisense RNA2, but had no match with mouse genome. This transcript seems highly conserved among the marsupial species studied, but only 100 base pairs (between positions 110 and 210 in tammar's transcript sequence) segment matched the human genome. The transcript XLOC_033951 (*XLOC51*), located at *locus* TCONS_033950, was identified as the least-conserved element of all studied transcripts, exhibiting low conservation degree even when compared only with marsupial species. It overlapped with tammar *HOXD8* transcripts, and also with opossum and Tasmanian devil *HOXD8* genes, but it did not match any region in the eutherians genome. This transcript apparently represents an intronic lncRNAs exclusive to marsupials. In contrast, XLOC_033952 (*XLOC52*), positioned at *locus* TCONS_033951, seems highly conserved among the marsupials studied and, although located downstream to *HOXD8* transcripts in the tammar HoxD cluster, it overlapped with opossum´s and Tasmanian devil *HOXD8* genes. This transcript also did not match to any region in the eutherians genome, possibly representing an intergenic lncRNAs exclusive to marsupials. Transcript XLOC_033953 (*XLOC53*), located at *locus* TCONS_033952, is positioned downstream to *HOXD8* and did not overlap with any annotated genes in the marsupials' genomes, although mapped in equivalent regions of their HoxD clusters. However, no matches for this transcript occur when it is searched in the eutherian genomes. The transcript XLOC_033954 (*XLOC54*) is positioned at *locus* TCONS_033953 and also exhibits some degree of conservation among marsupials, overlapping with opossum *HOXD4* gene and with a novel homeodomain-like protein in the Tasmanian devil (ensemble access code ENSSHAT00000019110), although it did not map the genome of eutherian species. Finally, XLOC_033955 (*XLOC55*), positioned at *locus* TCONS_033954, also overlapped with the same homeodomain-like protein from the Tasmanian devil, as well as with mouse *Hoxd3* gene and a transcript for mice Gm28230 protein (Havana gene 0TTMUSG000000045590, Uniprot AOAOAOMG91, MGI5578935).

*XLOC55* also matched human *HOXD3* gene and AC009336.19 transcript (Havana gene OTTHUMG0000013d2517, Uniprot AOAO87WSZ3).

*Comparison of HOX gene expression profiles between fore- and hindlimbs from d23 and d25*

Differential expression analyses showed consistently that terminal coding *HOXD* genes (*HOXD13-10;* see Fig. 3.11) exhibit overall similar expression levels when compared between fore- and hindlimbs from d23 and d25 of embryonic development. Based on results provided both by CummeRbund analyses (n=3) and by relative-qPCR (n=6), our data indicate that transcription of these genes were higher in hindlimbs when compared with forelimbs, and higher in forelimbs at d23 than those at d25. However, there were some differences between both data sources regarding comparisons of transcriptional levels in hindlimbs at d23 and d25. Based on qPCRs, *HOXD10* was upregulated in hindlimbs at day d25 [p(H1)=0.0d25], corroborating the transcriptome data. Transcriptional levels of *HOXD11* [p(H1)=0.131] and *HOXD13* [p(H1)=0.605] were not statistically different in hindlimbs at d23 and d25, although the mean tends to be higher at d25 in both cases. However, in the transcriptomic sample average, both *HOXD11* and *HOXD13* exhibited more copies of transcripts at d23. The gene *HOXD12* seems upregulated at d23 [p(H1)=0.000].

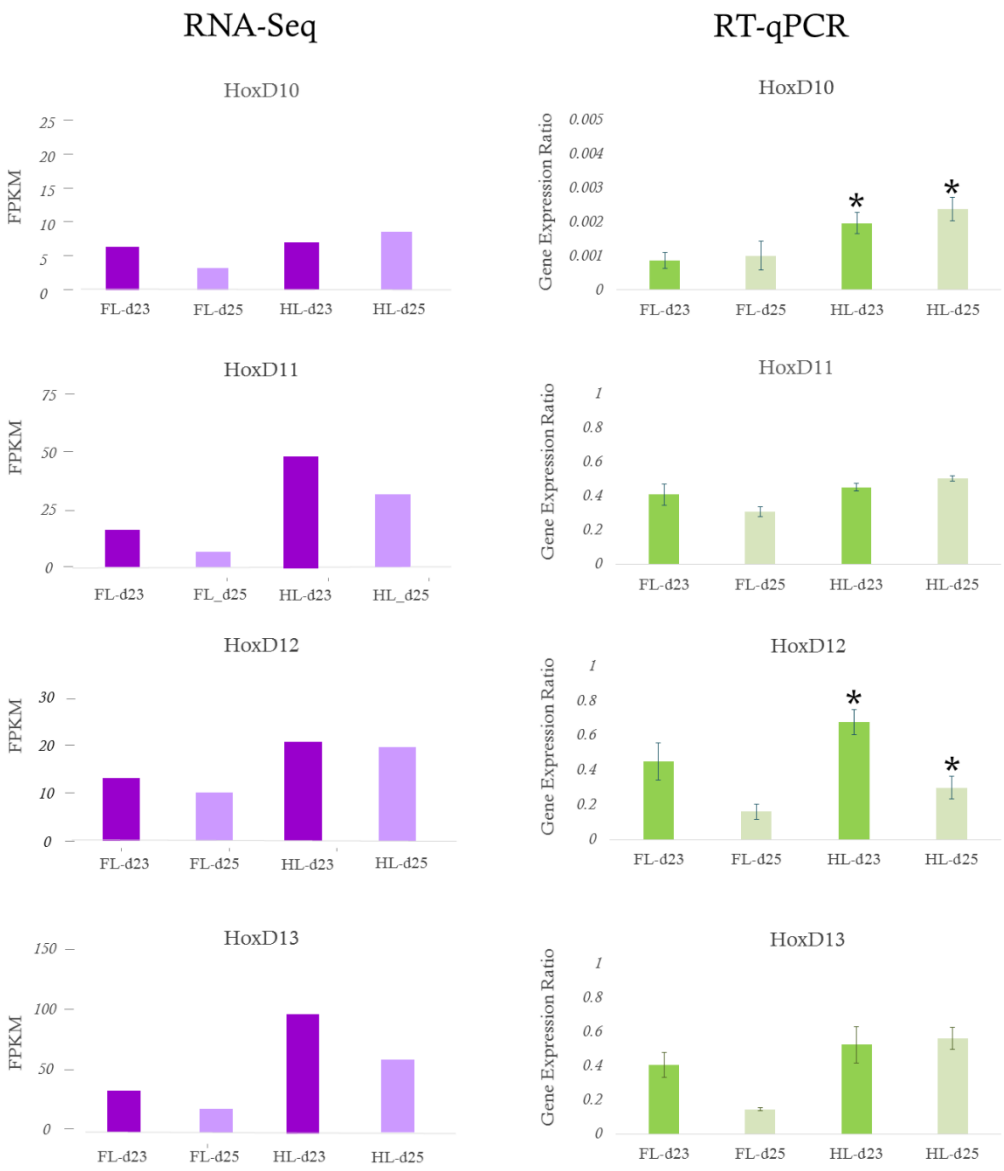


**Figure 3.11.** Differential expression of *HOXD10-13* coding transcripts among fore- (FL) and hindlimbs (HL) at d23 and d25 of pregnancy, calculated using CummeRbund in R statistical environment (RNA-seq, N=3) and REST/Excel (qPCR, N=6). FPKM= Fragments Per Kilobase of transcript per Million mapped reads. * highlights statistically significant differences that are relevant to my discussion.

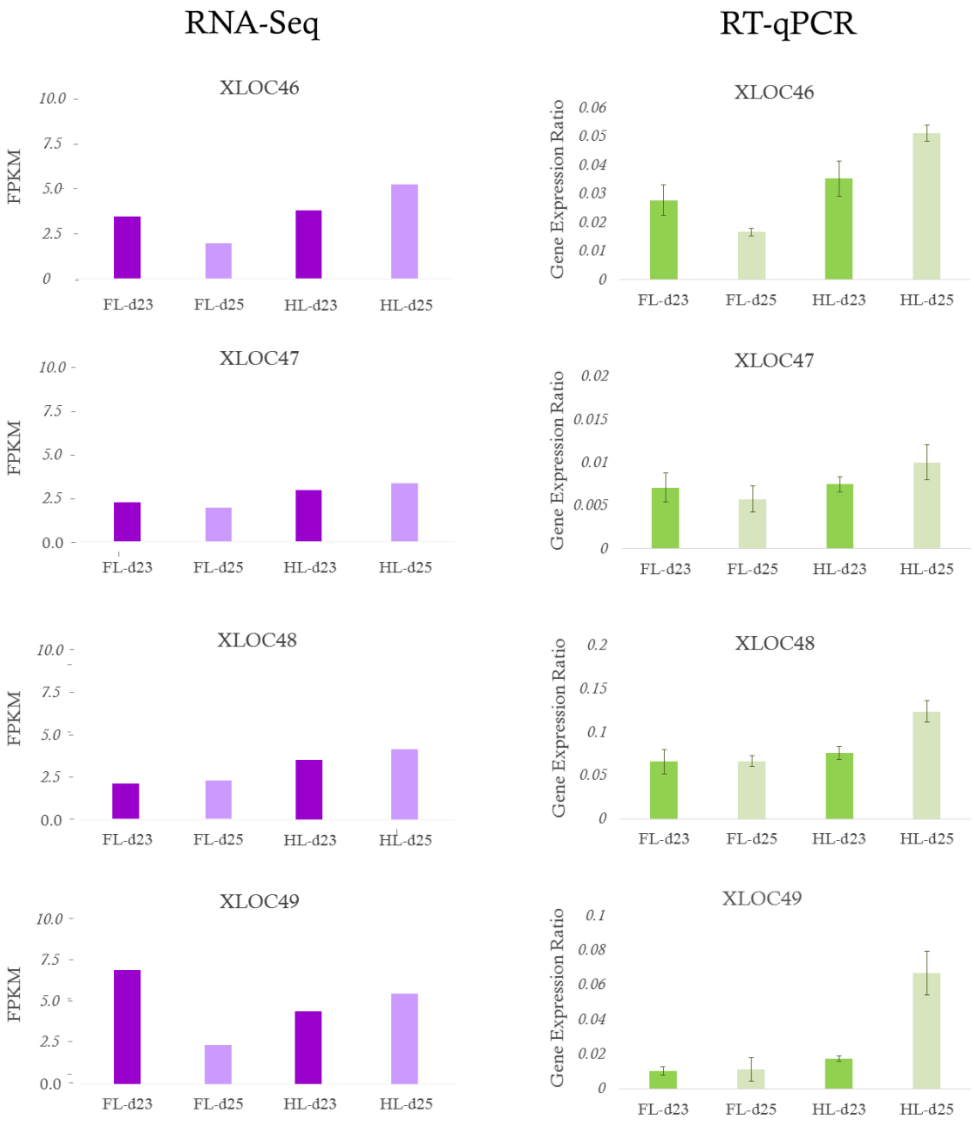


**Figure 3.12.** Differential expression of intergenic ncRNAs transcripts of HoxD cluster between fore- (FL) and hindlimbs (FL) at d23 and d25, using CummeRbund in R statistical environment (RNA-seq, N=3) and REST/Excel (qPCR, N=6).

The lncRNAs located in intergenic and intronic positions among terminal coding genes (*XLOC46*, *XLOC47*, *XLOC48* and *XLO49*; Fig. 3.14.) exhibited approximately the same range of transcriptional levels, and varied according to tissue and developmental stage within a range from 2.5 to 7.5 FPKM (Fragments Per Kilobase of transcript per Million mapped reads). They showed similar transcription profiles, which were quite consistent between both approaches used (RNA-seq and qPCR), resembling overall *HOXD10*, *HOXD11* and *HOXD13* transcription profiles as established by qPCR: transcription was usually higher in hindlimbs, tending to be

magnified at d23 in the forelimbs and at d25 in the hindlimbs. The transcript *XLOC49* was the only exception where transcriptome data differed from qPCR data, and exhibited higher transcription level in the forelimb at d23 (Fig. 3.12, bottom left).

The ncRNAs that surround *HOXD8* transcripts (*XLOC50*, *XLOC51* and *XLOC52*), which seem exclusive to marsupials, presented variable expression levels between limb tissues and developmental stages compared (Fig. 3.13B), and were not validated by qPCR. The transcripts *XLOC50* and *XLOC52* seem to be more intensely transcribed in forelimbs at d25, although their FPKM was in general low. In contrast, *XLOC51* seems approximately twenty times more transcribed in the forelimb at d23 than in the other conditions. The noncoding elements situated upstream to *HOXD4*, *XLOC53* and *XLOC54* exhibited consistent transcription profiles among each other, as established by qPCR but not corroborated by RNA-seq (Fig. 3.13A). These profiles also partially resembled the coding *HOXD* genes transcriptional profiles, as presented higher levels in both limb tissues at d23, and in the hindlimbs the transcription levels were higher than in the forelimbs. Curiously, according to RNA-seq, both ncRNAs were transcribed at higher levels in the forelimbs at d23, where *XLOC53* presented an extraordinarily elevated FPKM (~150). Lastly, the *XLOC55* element, which is possibly a coding gene still not annotated in the tammar wallaby genome, exhibited low FPKM that was higher in the forelimb at d23, a result not validated by qPCR.

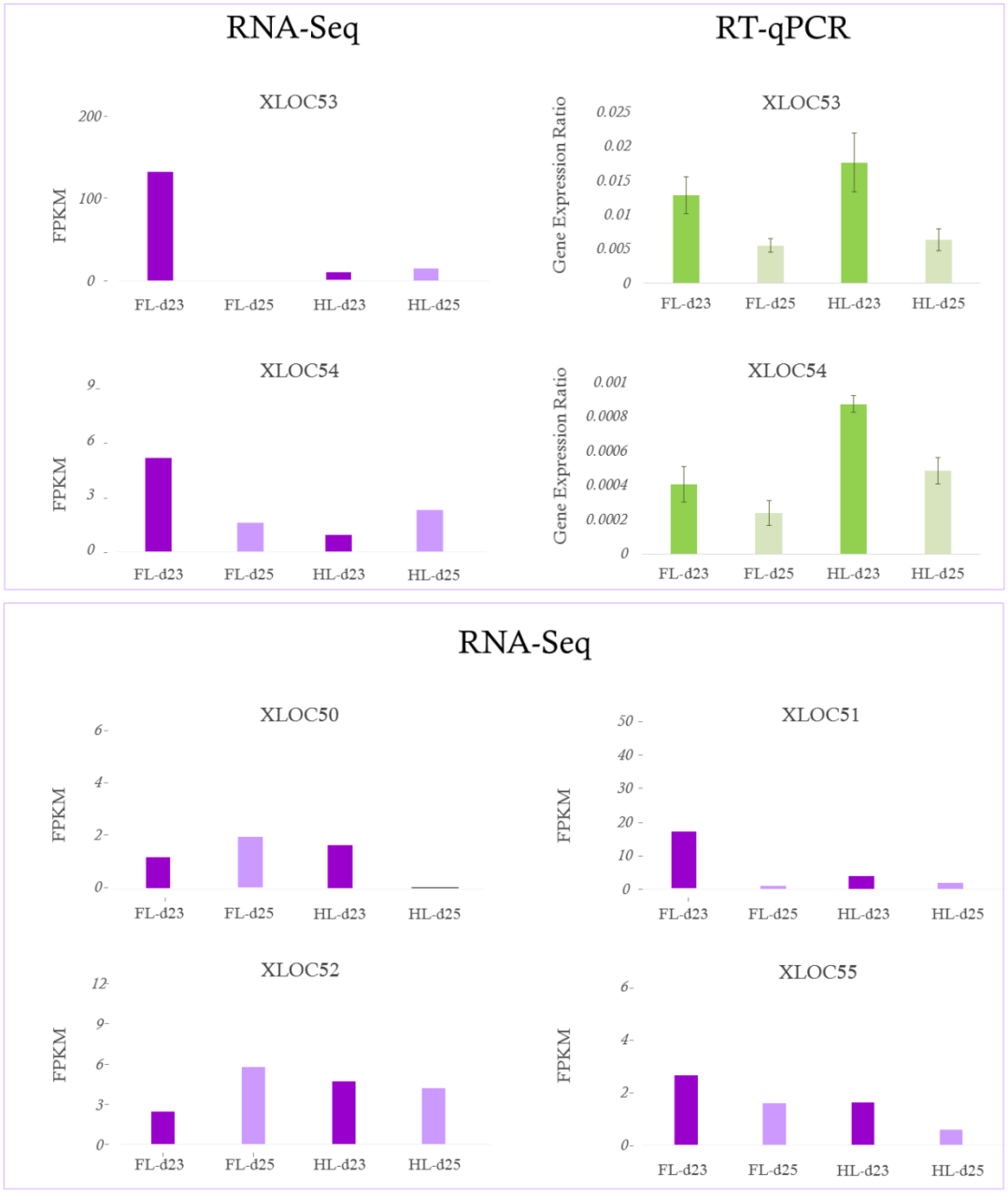


**Figure 3.13.** Differential expression of intergenic ncRNAs transcripts of HoxD cluster between fore- (FL) and hindlimbs (HL) at d23 and d25, using CummeRbund in R statistical environment (RNA-seq, N=3; A and B) and REST/Excel (qPCR, N=6; A).

*Spatial characterization of HOX genes expression profiles in autopodia at d23 and d25*

The expression patterns of *HOXD10*, *HOXD11* and *HOXD12* genes in *M. eugenii* hindlimb and fore autopodia were characterized by WISH. These experiments did not include *HOXD13* gene as such characterization has been previously performed by Chew et al. (2012).

Spatial characterization of these three *HoxD* genes established a basis for comparison with the expression patterns of the three ncRNAs which expression patterns were investigated: the lncRNA *XLOC46*, transcribed from a region between *HOXD11* and *HOXD12* genes, the lncRNA *XLOC52* and the pre-miRNA *XLOC53*, situaded upstream to *HOXD4* gene.

Expression patterns of *HOXD10-12* genes in fore- and hind autopodia during embryonic development were characterized when WISH was performed using DIG-labeled antisense probes specificaly designed for staining sense transcripts of *HOXD10-12* genes (Fig. 3.14A, C, E, G, I and K, Fig. 3.15A-F, and Fig. 3.16A-G, see also appendix for additional replicates). When sense probes for the same *HOX* transcripts were used instead, lack of staining was observed in most samples, indicating that the *HOX* genes studied are transcribed in sense direction (Fig. 3.14B,D, F, H, J and L, and Fig. 3.16H).

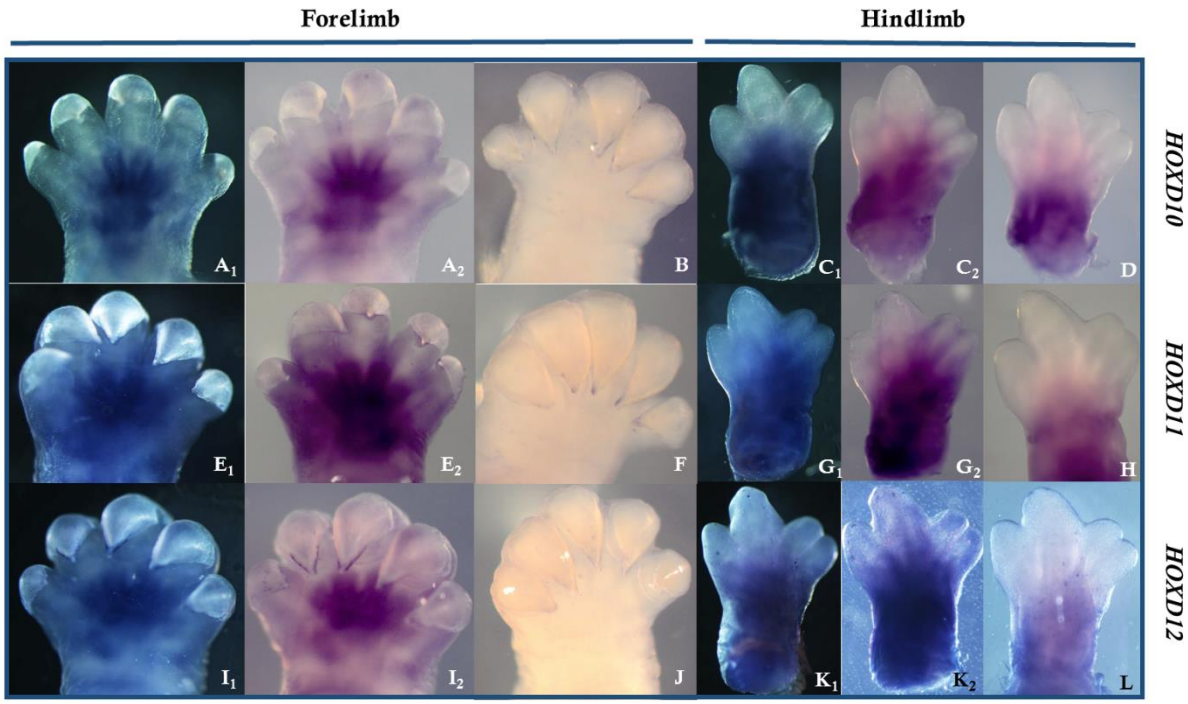


**Figure 3.14.** Whole-mount *in situ* hybridisation representing the expression patterns of *HOXD10* (A-D, n=4), *HOXD11* (E-H, n=4) and *HOXD12* (I-M, n=3) genes in fore (A, B, E, F, I, J) and hind (C, D, G, H, K, L) autopodia at d25 in tammar wallaby embryos. Pictures at A, C, E, G and I correspond to samples hybridized with antisense probes, which images are provided in both black ($A_1$, $C_1$, $E_1$, $G_1$, $I_1$ and $K_1$) and white ($A_2$, $C_2$, $E_2$, $G_2$, $I_2$, $K_2$) backgrounds. Pictures B, D, F, H, J and L correspond to hybridisation performed using sense probes.

Expression patterns of genes *HOX11-12* in fore autopodium at d23 were spread from the beggining of digital condensations, up on the interdigital tissues and surrounding digital pre-bone cartilage above the digital tips (Fig.3.15A, D). The staining appeared to migrate centripetally during development on d24 (Fig. 3.15B, E) and d25 (Fig. 3.14A, E, I and Fig.3.15C, F), turning remarckably concentrated in the central area of the paddle at the start of the digital condensations and the proximal boundary of interdigital tissues, away from the tips of autopodium by d25. The transcripts *XLOC52* and *XLOC53* exhibited similar expression patterns in fore autopodia in every developmental stage studied (Fig. 3.15G-K), but expression of *XLOC53* was not observed on most fore autopodia at d25 (5/6) (Fig 3.15L).

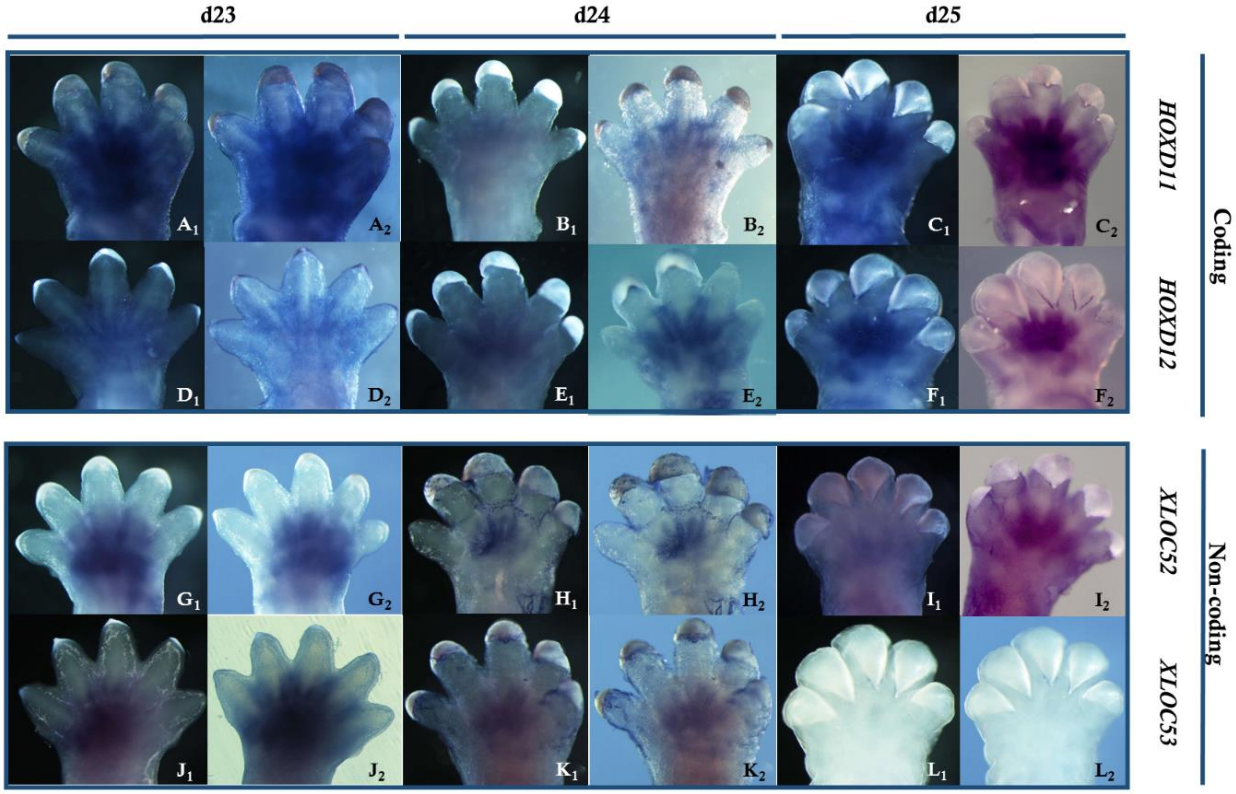


**Figure 3.15.** Whole-mount *in situ* hybridisation representing the expression patterns of *HOXD11* [(A) d23, n=2; (B) d24, n=3; (C) d25, n=4)], *HOXD12* [(D) d23, n=4; (E) d24, n=6; (F) day d25, n=3)], *XLOC52* [(G) d23, n=3; (H) d24, n=2; (I) day d25, n=1)] and *XLOC53* [(J) d23, n=5; (K) d24, n=3; (L) day d25, n=6)] antisense probes during fore autopodial development on d23-d25 tammar wallaby embryos. Images of every sample were provided both in dark and light backgrounds, in order to register respectively morphology and staining. Note that the transcription profiles are in general very similar

among coding and noncoding *HOX* in each developmental stage; however, *XLOC53* staining on d25 is abrogated, suggesting low transcription level at this stage.

The expression patterns of *HOXD* in hind autopodia at d23-d25 were overally similar to the expression of these genes on fore autopodium, exhibiting a pattern that migrated centripetally from interdigital areas towards the center of the paddle and proximal boundary of interdigital tissues. At d22, expression was concentrated in the distal anterior half of the bud (see Fig. 3.16A2, from left to right), located in regions where future digits V and IV develop. The second stained field observed beneath the limb bud corresponds to a zeugopodium domain. At d23, expression started to be defined in the interdigital region, surrounding pre-bone cartilage around digit V and also between digits III and II. A stained field developed in the center of the paddle, where the proximal boundary of carpal pre-bone cartilage form (see Fig. 3.16B and E). At d24, the staining was intensified in the interdigital areas described, turning strongly concentrated in the shape of a line that bifurcates at distal extremities. Additional fields also developed on the outer boundaries of digits V and II (see Fig. 3.16$C_2$ and $F_2$). At day d25, the expression pattern migrated towars digits III and II, being more concentrated around them, and the signal became weak in other areas (see Fig. 3.14C and G, Fig. 3.16D and G). Similarly to fore autopodia, the expression patterns shown by *XLOC52* (Fig. 3.16J-M) and *XLOC53* (Fig. 3.16L-O) were equivalent to the *HOX* genes pattens. Expression of *XLOC53* on hind autopodium at d25 (Fig. 3.16O) was strong, contrasting to its expression in the forelimbs at this developmental stage (see also appendix for additional replicates).

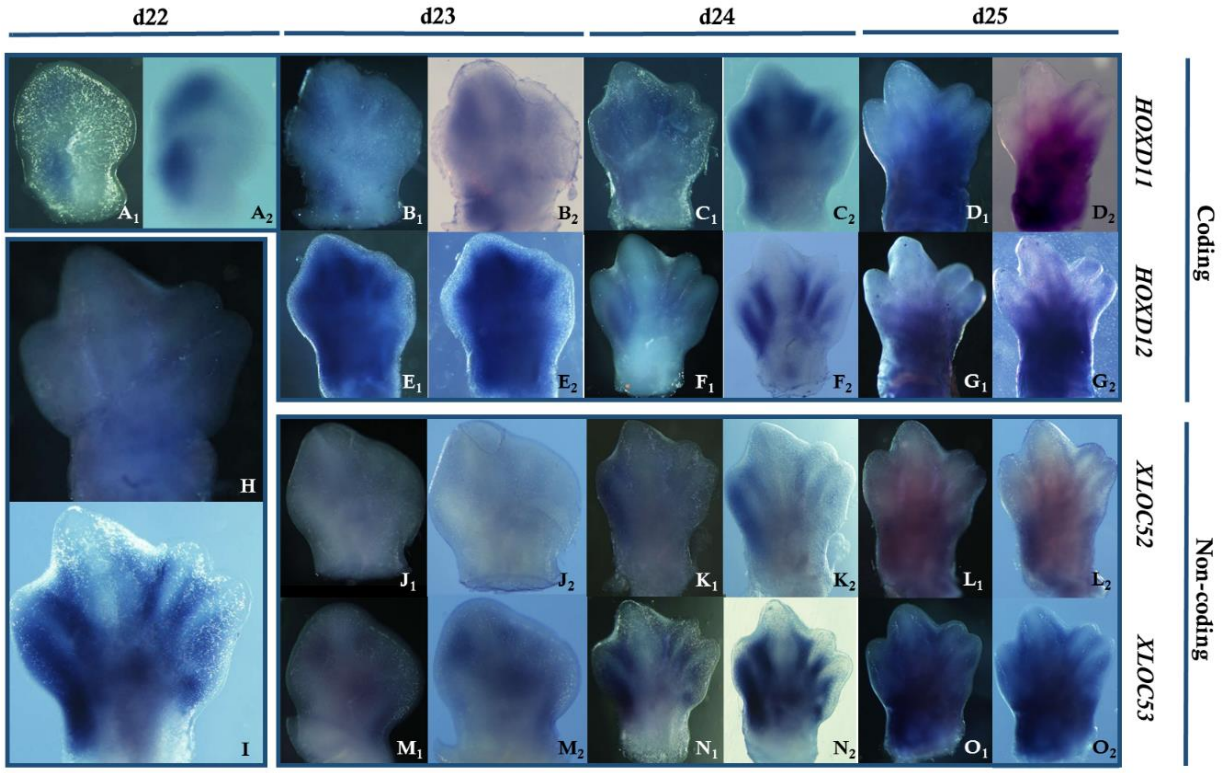


**Figure 3.16.** Whole-mount *in situ* hybridisation representing the expression patterns of *HOXD11* [(A) d22, n=1; (B) d23, n=4; (C) d24, n=3, (D) d25, n=3)], *HOXD12* [(E) d23, n=1; (F) d24, n=6; (G) d25, n=3)], *XLOC52* [(J) d23, n=2; (K) d24, n=2; (L) d25, n=2)] and *XLOC53* [(M) d23, n=1; (N) d24, n=5; (O) d25, n=3)] antisense probes during hind autopodial development on d22-d25 tammar wallaby embryos. Images of each sample are provided both in dark and light backgrounds, in order to register respectively morphology and staining. Note that the expression patterns were very similar among coding and noncoding *HOX* in all developmental stages. Hybridisation performed using sense probes are illustrated in H for coding RNA (*HOXD12*) and I for ncRNA (*XLOC53*); the ncRNA transcripts studied are bidirectional, therefore sense and antisense probes show similar staining profiles, whereas coding RNAs are transcribed unidirectionally.

Expression patterns of the lncRNA *XLOC46* were also similar to that of *HOX* genes in fore autopodia at d23-d25 (Fig. 3.17A-D) and hind autopodia at d25 (Fig. 3.17G-H), although it was expressed in different fields on hind autopodia at d23 and d24 (Fig. 3.17E-F). Expression of this transcript at these stages, rather than well-defined in strong lines in interdigital regions, seemed more concentrated on the outer boundaries of digital tips, staining the whole distal area of the paddle. Expression was also identified between digits and in the

center of the paddle, but exhibiting a more outspread and "blurred" manner. Interestingly, expression seems to migrate to a similar area where all the other genes described are expressed at d25, by the proximal boundary of interdigital tissues (Fig. 3.17G-H, see also appendix).

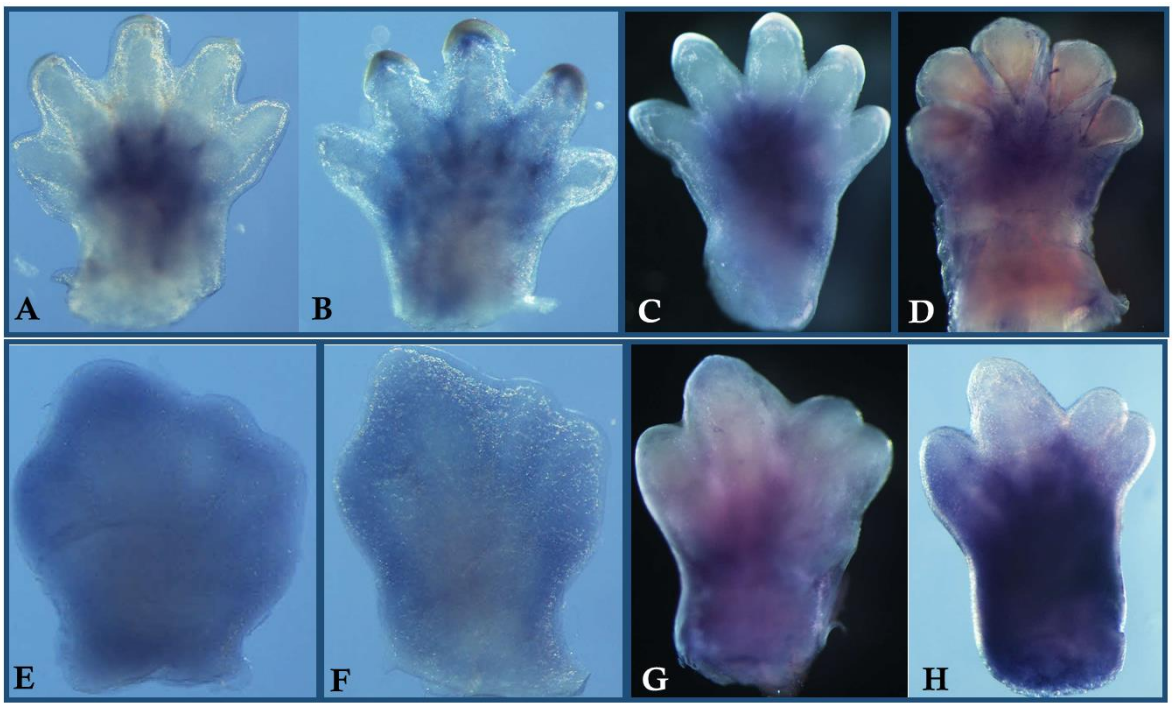


**Figure 3.17.** Whole-mount *in situ* hybridisation representing the expression patterns of *XLOC46* sense and antisense probes during fore (A-D) and hind (E-H) autopodial development on d23-d25 tammar wallaby embryos. (A) antisense probe, d23, n=3; (B) sense probe, d23, n=3; (C) sense probe, d24, n=1; (D) sense probe, d25, n=2; (E) antisense probe, d23, n=3; (F) antisense probe, d24, n=3; (G) antisense probe, d25, n=1; (H) sense probe, d25, n=2. The transcript *XLOC46* is bidirectional, and its sense and antisense probes provide similar staining patterns. Expression patterns of *XLOC46* differ from those of *HOX*, provided in Fig. 3.19 in autopodia at d23 (E) and d24 (F), where staining seems to surround digital tips.

Differently from coding genes, the ncRNAs are transcribed in a bidirectional manner (sense and antisense transcripts), as sense and antisense probes result in the same expression patterns (see Fig. 3.16I and Fig. 3.17B-D, H, see also appendix). This information was confirmed by synthesizing strand-specific cDNA for these 3 ncRNAs, and amplyfing the same fragments from both sense and antisense strands, which were confirmed by DNA sequencing.

## 3.5 DISCUSSION

The present study explored conservation and spatiotemporal expression of ncRNAs during limb development in a group of marsupials characterized by a peculiar autopodial morphology (syndactyl feet) that is established heterochronically in relation to forelimbs during embryo development. In this study, at least eigth ncRNAs were identified in intergenic and intronic positions of the HoxD cluster that are transcribed in the forelimbs and hindlimbs of the tammar wallaby *M. eugenii* during embryonic development. Our analyses based on genomic DNA sequences suggest that three among these noncoding elements may be exclusive to the Order Marsupialia, whereas five others are conserved among therian mammals. The spatiotemporal expression patterns of *HOX* noncoding and *HOX12-10* coding genes have been characterized in developing tammar wallaby limbs for the first time. The similarity in expression patterns and transcriptional profiles between fore- and hind autopodia at different embryonic stages suggests that some of these ncRNAs regulate terminal *HOXD* gene expression during autopodial development.

*Terminal HOXD gene expression patterns and transcriptional profiles in the tammar wallaby limbs– characterization of coding transcripts*

To my knowledge, this is the first characterization of *HOXD10-12* transcriptional profiles and expression patterns in any marsupial. The transcriptional profiles of terminal *HOXD* genes reinforce the idea that these genes are important for stablishing heterochrony between fore- and hindlimbs (Chew et al. 2012; Chew et al. 2014). Differences between fore- and hindlimbs at d23 and d25 in transcriptional levels of the genes studied consistently indicate that levels of coding *HOXD10-13* genes transcription are higher in hindlimbs when compared to forelimbs. Additionally, transcription is more intense in the forelimbs at d23 than

at d25. On hindlimbs, however, there is variation when comparing qPCR and RNA-seq data, which may be due to variability in the number of replicates (N=6, compared to N=3 for RNA-seq), limb tissues (whole limbs were used for RNA-seq but only autopodium was used for qPCRs) and developmental stages (developmental processes range significantly during one day) used. Therefore, qPCR data, which correspond to the most accurate data set of this investigation because of replicate numbers and the use of RNA extracted from autopodia, indicates that transcription of terminal *HOXD* genes is more accentuated at d25 than d23 (but see in Fig. 3.11. that the opposite trend was revealed by the transcriptome, d23 > d25). These results are likely explained by limb developmental heterochrony driving accelerated forelimb development, approximately two days ahead of the hindlimbs (Chew et al. 2012). Expression of *HOXD13* in tammar forelimbs starts at d18, but it does not begin in the hindlimbs until d21.5 (Chew et al. 2012). The forelimbs at d25, one day before birth, are nearly fully developed, explaining a decrease in *HOXD* gene transcription when compared with forelimbs at d23. Hindlimbs, however, still exhibit an early developmental stage at d25, when levels of terminal *HOXD* gene expression would be accentuated (Chew et al. 2012).

Alike the transcriptional profiles, *HOXD* gene expression patterns suggest a role for these genes on regulating heterochrony. The patterns exhibited by these genes resemble those of *HOXD13,* which have been characterized in the pioneering study by Chew et al. (2012) (see Fig 3.18). The study by Chew et al. (2012), however, did not detect *HOXD13* expression in the limbs at d25, when strong *HOXD10-12* expression is still detected. The lack of staining at d25 in the experiments by Chew et al. (2012) was likely caused by poor probe penetration, as the specimens are larger and less permeable to the probes at the day before birth. Similarly to *HOXD13*, expression patterns of *HOXD10-12* in the tammar fore autopodia resemble expression of these genes in mouse, although gene expression in hind autopodia differs considerably between the two species. As suggested by Chew et al. (2012), variation of *HOX*

domains may contribute to the heterochrony and phenotypic differences induced through development that result in marsupial syndactyl digits. The gene structure of *HOXD13* and predicted proteins seem highly conserved with their eutherian orthologues despite the morphological differences in hindlimb patterning, except for a difference in the polyserine tract of *HOXD13* (Chew et al. 2012). These findings indicate that phenotypic differences in limb development most likely result from evolution of regulatory modules, including regulatory ncRNAs, which can influence *HOX* gene expression domains (Woltering et al. 2014). Despite the differences observed in *HOXD* gene expression patterns in tammar hind autopodia, the stripe-like staining identified at d24 of gestation seem to recapitulate in wallabies the vertebrate ancestral Turing-type mechanism of digital patterning, in which the dose of distal *HOX* genes modulates the digit period or wavelength (Sheth et al. 2012).

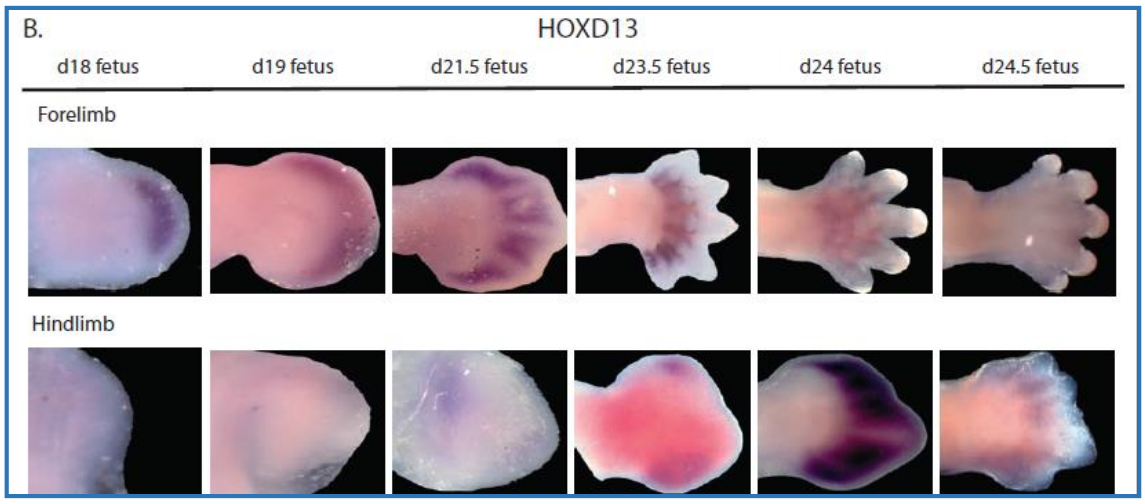


**Figure 3.18.** mRNA wholemount *in situ* hybridisation for expression of *HOXD13* (B) in the tammar forelimb and hindlimb from d18 to d24.5 of pregnancy. Top images are forelimbs, with the corresponding hindlimb in the lower panel. Expression was first visible at d8.5 in the forelimb. The gene *HOXD13* in the forelimb was expressed distally from day 18 and became interdigital in the forelimb by d21.5. In the hindlimb, *HOXD13* mRNA was not detected at d18 and d19, but it was weakly expressed in the distal region at d21.5. By d23.5, the gene was expressed in the anterior and posterior points of the limb and after d24 encompassed the distal end of the hindlimb and became interdigital by d24.5. All images are dorsal and orientated to point distally (n = 3). (Figure copied from Fig. 6B of Chew et al., 2012).

*The promoter-positioned lncRNAs of HoxD cluster: XLOC46-49 may represent regulatory elements of terminal HOXD gene expression during limb development conserved among mammals*

Mapping the ncRNAs identified from the marsupial limb transcriptome to genomic sequences of multiple mammalian species indicated that the intergenic *XLOC46* and the intragenic (or intronic) *XLOC47*, *XLOC48*, *XLOC49* seem to have originated in a common ancestor of marsupials and eutherians, as they matched sequences in their respective equivalent positions of opossum, Tasmanian devil, mouse and human HoxD cluster. Conservation analyses reveal that these transcripts present characteristics consistent with the lncRNAs described in the literature: they are generally poorly conserved but retain conservation of some fragments such as promoter, exon regions, and position in the genome, relative to their target gene (Wang et al. 2004; Pang et al. 2006; Li & Chang 2014). According to Li and Chang (2014), lncRNA genomic position is conserved across organisms, which indicates that syntenic lncRNAs may exhibit conserved functions and that functional structures can be maintained under weaker selective constraints during evolution (Rinn et al. 2007). For example, the lncRNA *HOTAIR* is highly conserved in synteny and structure among vertebrates (Rinn et al. 2007), including marsupials (Yu et al. 2012), and performs similar roles in human and mouse (Ulitsky et al. 2011; Li et al. 2013). However, although the conservation of specific lncRNAs sequences are believed to provide the necessary structural basis for its function, recent studies (Ralston 2013; Schorderet & Duboule 2011; Fatica & Bozzoni 2014) evidence that mouse long intergenic noncoding RNAs (lincRNA) do not always reflect the same functions performed in humans. Under this rational, even those lncRNAs conserved among mammals could have evolved specific functions in different clades, having the potential to contribute for evolution of morphological peculiarities in marsupial limbs.

The results from conservation analyses discussed above, together with spatiotemporal transcriptional profiles and expression patterns, suggest that *XLOC46-49* may represent

mammalian elements that regulate expression of terminal *HOXD* genes during limb development. The qPCRs and RNA-seq results for lncRNAs were very consistent. Comparative transcriptional levels of the lncRNAs *XLOC46-49*, situated in intergenic and intronic positions among the terminal *HOXD* genes, estimated by both approaches used, seem equivalent to the transcriptional profile of coding *HOX* genes as reported here by qPCR. Transcription was always higher in the hindlimbs than in the forelimbs. When examining only forelimbs, transcription tended to be higher at d23. In de hindlimbs, oppositely, transcription was magnified at d25. As aforementioned, the accentuated expression of *HOXD* coding and noncoding transcripts on hindlimbs (and the lower expression of these transcripts in the forelimbs at d25) by the final stages of pregnancy may be explained by the fact that, at this stage, hindlimbs are still nonfunctional and d25 is the day when hind digits are being defined (Chew et al. 2012). At the same time, at d25, forelimbs are already functional, exhibiting well-developed digits and claws, almost ready to climb from the mother's birth canal to the pouch (Chew et al. 2012). The overlapping expression domains of coding and noncoding transcriptional profiles suggest that the intergenic and intronic lncRNAs among terminal *HOXD* gene may regulate *HOXD* gene expression during autopodial development. Several lncRNAs that are intergenic to key developmental genes have been shown to have promoter or enhancer roles, regulating their expression (Rinn et al. 2007).

Our conservation findings were also reinforced by WISH of the lncRNA *XLOC46* in autopodia, showing an expression pattern that resembles tammar *HOX* coding genes in the fore autopodia, but resembles more the mouse *HOXD13* expression pattern in the hind autopodia. At d23 and d24 of pregnancy, instead of being strongly concentrated in the interdigital tissues like *HOXD11-12* genes, lncRNA *XLOC46* expression spreads to the distal boundary of the paddle, surrounding digital pre-condensation. This pattern is observed in mouse hind autopodia at developmental stages E12 and E12.5 (Fig. 3.19, Chew et al. 2012).

Because the lncRNAs *XLOC46-49* are conserved among marsupials and eutherian mammals, it is possible that they constitute a regulatory element of the HoxD cluster system, regulating terminal *HOXD* expression during autopodium development in mammals, and possibly other tetrapods.

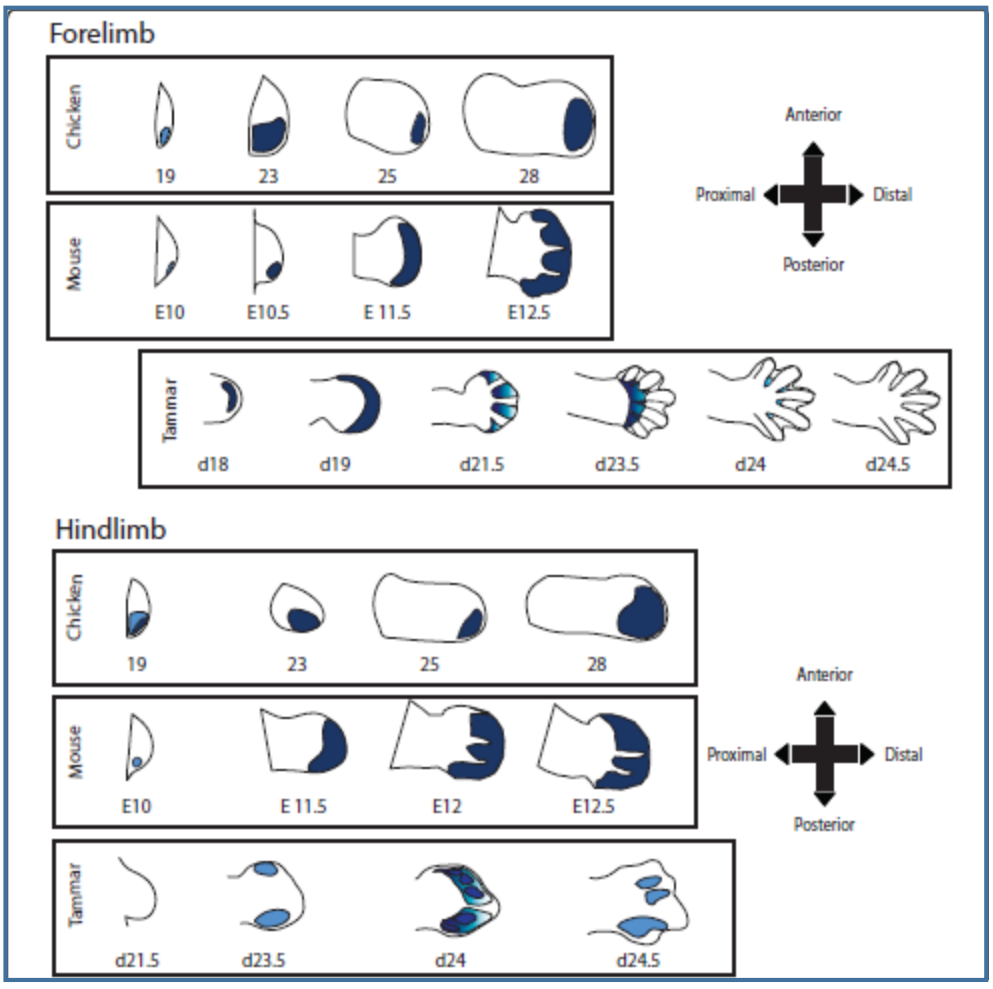


**Figure 3.19.** Comparisons of *HOXD13* expression patterns among mouse, chicken and tammar. Staining patterns were traced from published mRNA staining for mouse and chicken. [Mouse: (Keyte & Smith 2012; C. Chen et al. 2005); chicken:(Nelson et al. 1996)]. Dotted lines indicate digital rays. (Image retrieved from Fig. 8 of Chew et al. 2012).

Roles for the mammalian transcripts *XLOC50* and *XLOC55* remain unknown. The putative lncRNA *XLOC50* had expression profiles that differed from *HOXD* genes and other putative lncRNAs, but were not validated by qPCR, therefore there is no evidence that this

element is a *HOXD* gene regulatory elements. The transcript *XLOC55*, on the other hand, apparently corresponds to a coding region, orthologue of a recently identified protein in Tasmanian devil (*Sarcophilus harrissii;* Ensemble access code ENSSHAT00000019110) and of a Gm28230 mouse protein.

*The ncRNAs of HoxD cluster downstream to HOXD9: XLOC51-53 may represent marsupial-exclusive regulatory elements of terminal HOXD gene expression during limb development*

Among the transcripts that did not match any sequences of eutherian genomes, *XLOC51*, *XLOC52* possibly represent intronic lncRNAs, whereas *XLOC53* likely represents a microRNA precursor (pre-miRNA), as its sequence is shorter than 200 base pairs, the minimum length of lncRNA sequences (Saxena & Carninci 2011). *XLOC54,* however, alike *XLOC55*, possibly corresponds to a recently identified protein in Tasmanian devil (*Sarcophilus harrissii;* Ensemble access code ENSSHAT00000019110). The existence of some ncRNAs that evolved exclusively in the marsupial lineage agrees with the characterization of the transcriptional landscape of the human Hox clusters, where most ncRNAs are conserved among vertebrates, but a few represent novel elements. Human Hox cluster was characterized in 11 anatomic sites, where 231 ncRNAs occurred in known transcribed regions of more than 30 kilobases, 15% of which are exclusive to human (Rinn et al. 2007).

The transcripts *XLOC53* and *XLOC54* are located downstream to *HOXD9* and also exhibit transcription profiles similar to those of terminal *HOXD* genes; its transcription in the hindlimbs is higher than in the forelimbs, and transcription at d23 is higher than at d25 in both fore and hind autopodia, as suggested by qPCRs. Results from RNA-seq, however, suggest that transcription in the forelimb at d23 is extremely accentuated, overtaking even transcription levels in the hindlimbs. Remarkably, *XLOC53* transcription reaches a value of nearly 150 FPKM at d23 in the forelimb, while transcription in the forelimb at d25 is null.

Despite such differences, both data sources confirm that expression in the fore autopodium at d23 is at a much higher level than at d25, a result confirmed by WISH. Expression patterns of *XLOC52* and *XLOC53* perfectly match those of *HOX11-12* genes both in fore and hind atupodia. Comparison of ncRNAs expressions patterns in humans with those of neighboring *HOX* genes shows that 90% are coordinately induced in association with protein coding gene expression (Rinn et al. 2007; Fatica & Bozzoni 2014). Some of the ncRNAs that are expressed in a similar spatiotemporal profile as neighboring protein-coding genes had their function in the regulation of *HOX* genes investigated (Fatica & Bozzoni 2014), and they either activate (Sanchez-Elsner 2006; Schmitt et al. 2005; Ringrose & Paro 2007) or silence transcription (Petruk et al. 2006; Rinn et al. 2007). The similarity of expression patterns between *XLOC52*, *XLOC53* and the coding *HOXD* genes, together with the evidence that they may be elements exclusive to marsupials, suggest that these elements modulate *HOXD* gene expression, possibly contributing to establish the phenotypic peculiarities that characterize the kangaroo family. However, there was a difference between expression patterns of terminal *HOXD* gene and *XLOC53*: there was no staining of this putative pre-miRNA in the fore autopodium at d25, corroborating our data showing that transcription of this element in the forelimbs at d25 was very low. At this stage, the penetration of probe in the forelimbs is more challenging and the lack of staining was likely due to a very low level of *XLOC53* expression at d25. The reduction of *XLOC53* expression at d25 suggests an important stage-specific role likely associated to the heterochrony in development between fore and hindlimbs (Chew et al. 2014), therefore *XLOC53* may help accelerating forelimb development before birth.

## 3.6 CONCLUSIONS

Conservation of HoxD ncRNAs regulatory elements expressed in tammar wallaby limbs was evaluated in this study, and five lncRNAs located in intergenic and intronic positions among terminal *HOXD* coding genes that are conserved among mammals were identified. These lncRNAs present similar transcriptional profiles to those of tammar coding terminal *HOXD* genes, and *XLOC46* exhibits expression patterns that resemble those of mouse and tammar terminal *HOXD* genes. These findings evidence that these lncRNAs may regulate *HOXD* expression during autopodium development in mammals. Three ncRNAs that are exclusive to marsupials were also identified. The transcripts *XLOC52* and *XLOC53* exhibit expression patterns that reproduce exactly those of tammar *HOXD11-12* coding genes, therefore they may regulate *HOXD* expression and likely have contributed to the evolution of the unique kangaroo family limb morphology. Further, *XLOC53* is only expressed at low levels in the hind autopodia at d25, which could indicate a stage-specific role of this pre-miRNA to accelerate forelimb development before birth, thus contributing to their developmental heterochrony. Investigating the spatiotemporal transcriptional profiles of Hox ncRNAs in the tammar limb represents a first step towards understanding the specific regulatory mechanisms that contribute to establish autopodial peculiarities from such highly conserved gene clusters and the mechanisms of evolution of unique limb morphologies in therian mammals.

## 3.7 APPENDIX

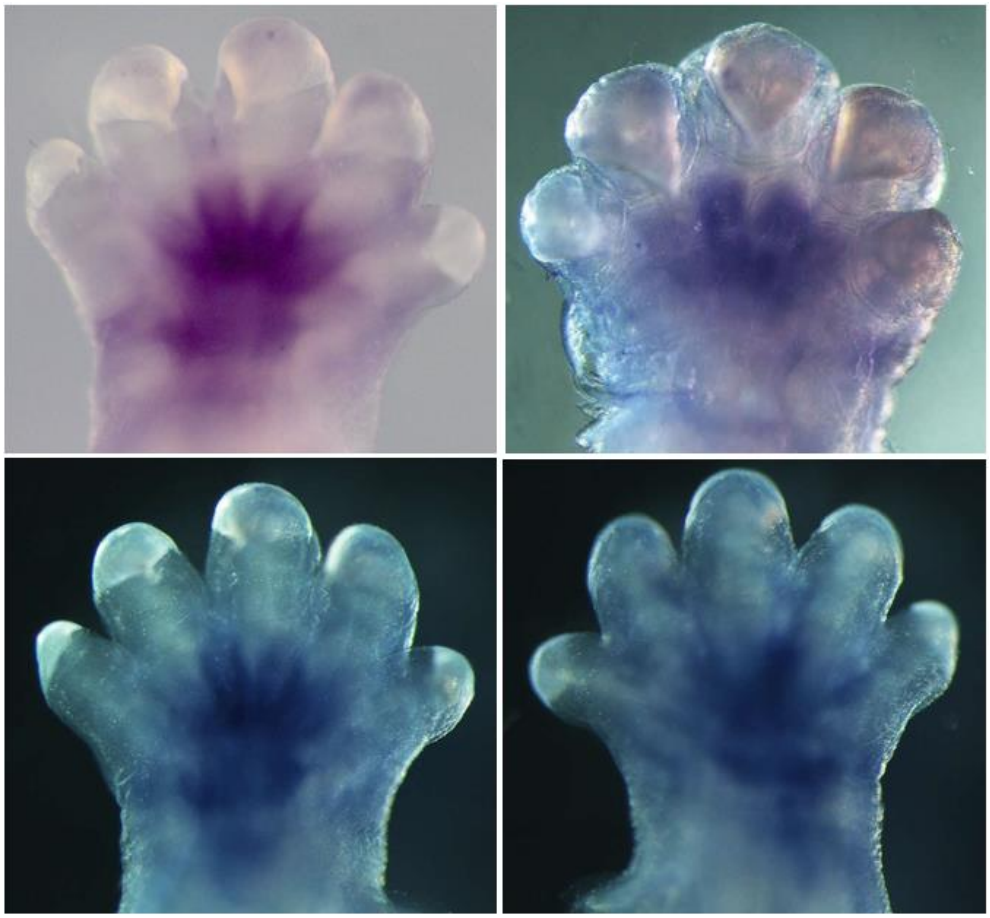

**Figure A3.1.** Expression of *HOXD10* in fore autopodium at d25 - antisense probe

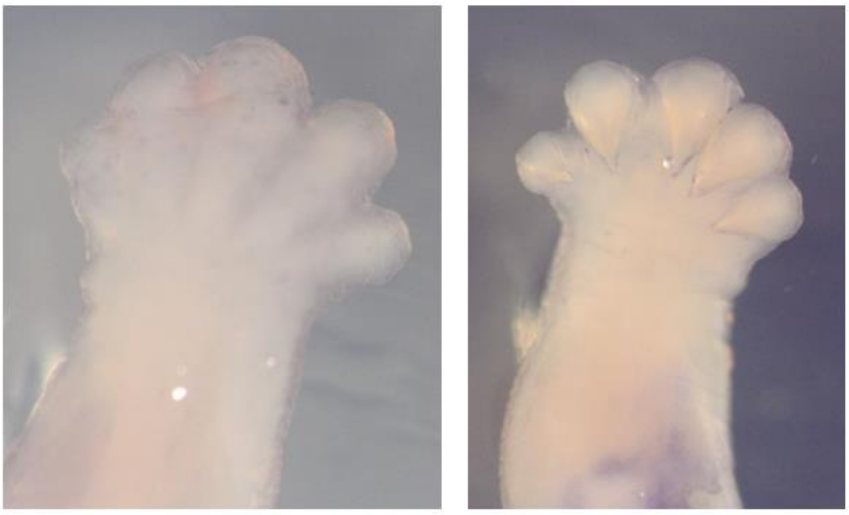

**Figure A3.2.** Expression of *HOXD10* in fore autopodium at d25– sense probe.

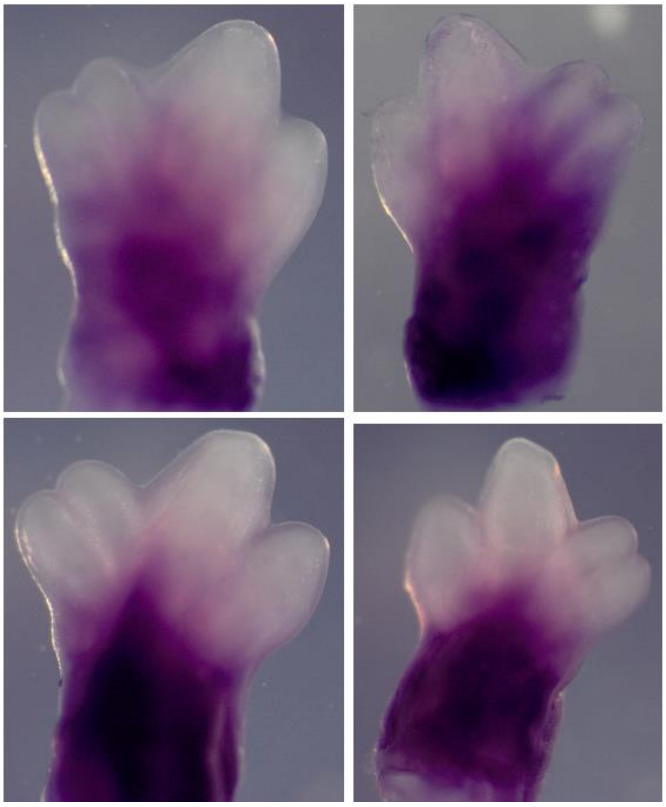

**Figure A3.3.** Expression of *HOXD10* in hind autopodium at d25 – antisense probe.

**Figure A3.4.** Expression of *HOXD10* in hind autopodium at d25– sense probe.

**Figure A3.5.** Expression of *HOXD11* in fore autopodium at d23 – antisense probe.

**Figure A3.6.** Expression of *HOXD11* in fore autopodium at d23 – sense probe.

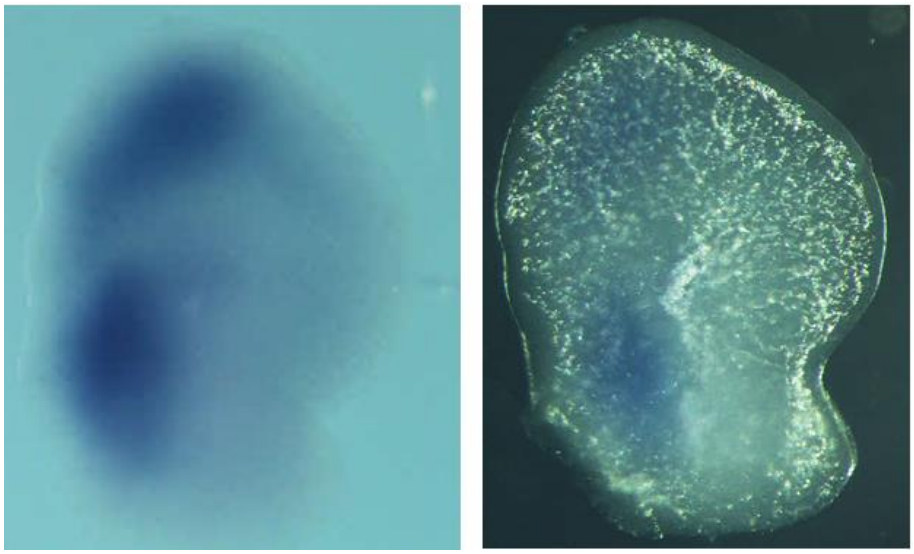

**Figure A3.7.** Expression of *HOXD11* in hind autopodium at d22 – antisense probe.

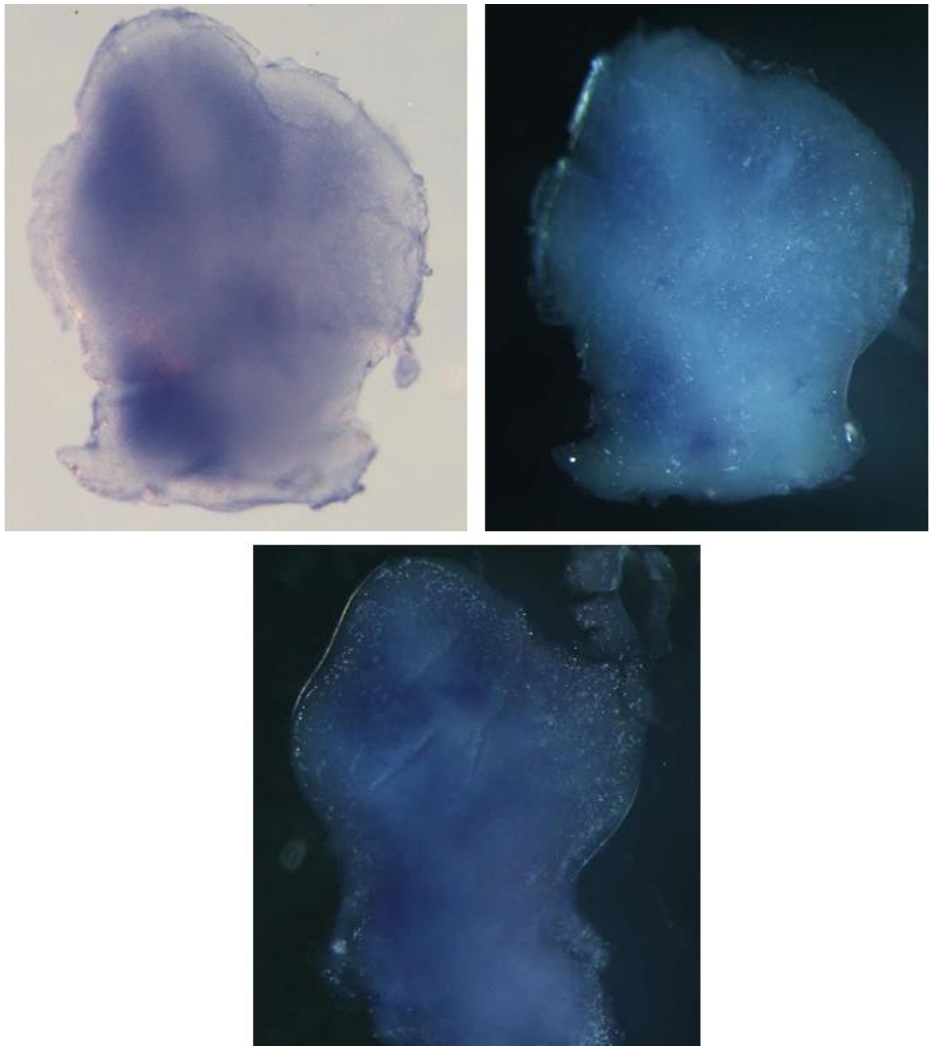

**Figure A3.8.** Expression of *HOXD11* in hind autopodium at d23 – antisense probe.

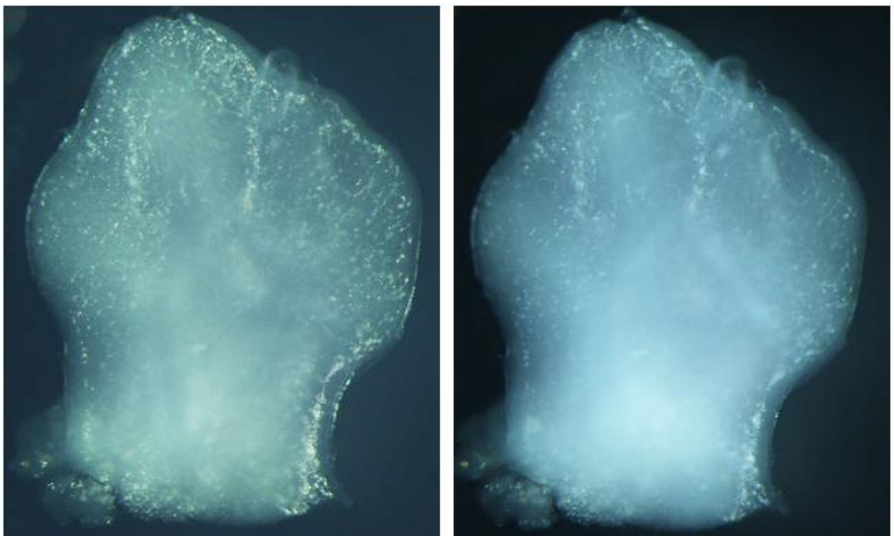

**Figure A3.9.** Expression of *HOXD11* in hind autopodium at d23 – antisense probe.

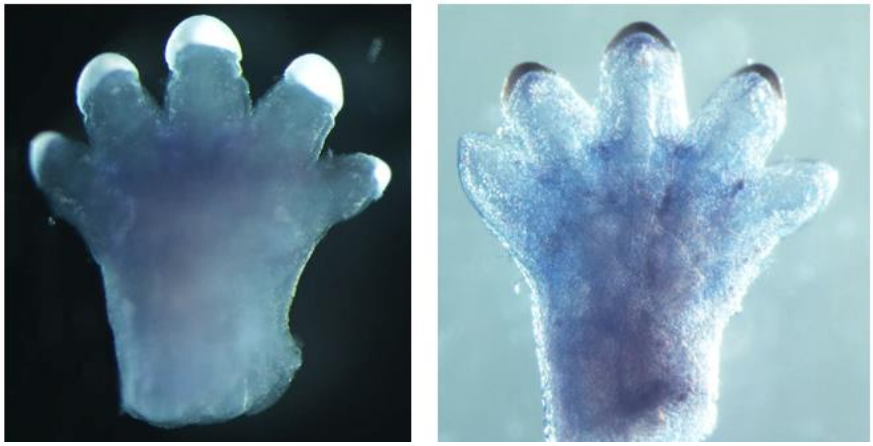

**Figure A3.10.** Expression of *HOXD11* in fore autopodium at d24 – antisense probe.

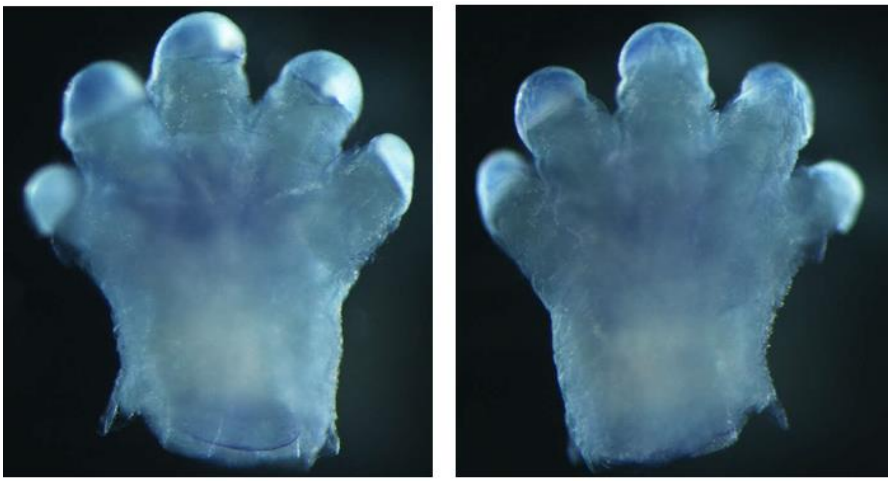

**Figure A3.11.** Expression of *HOXD11* in fore autopodium at d24 – sense probe.

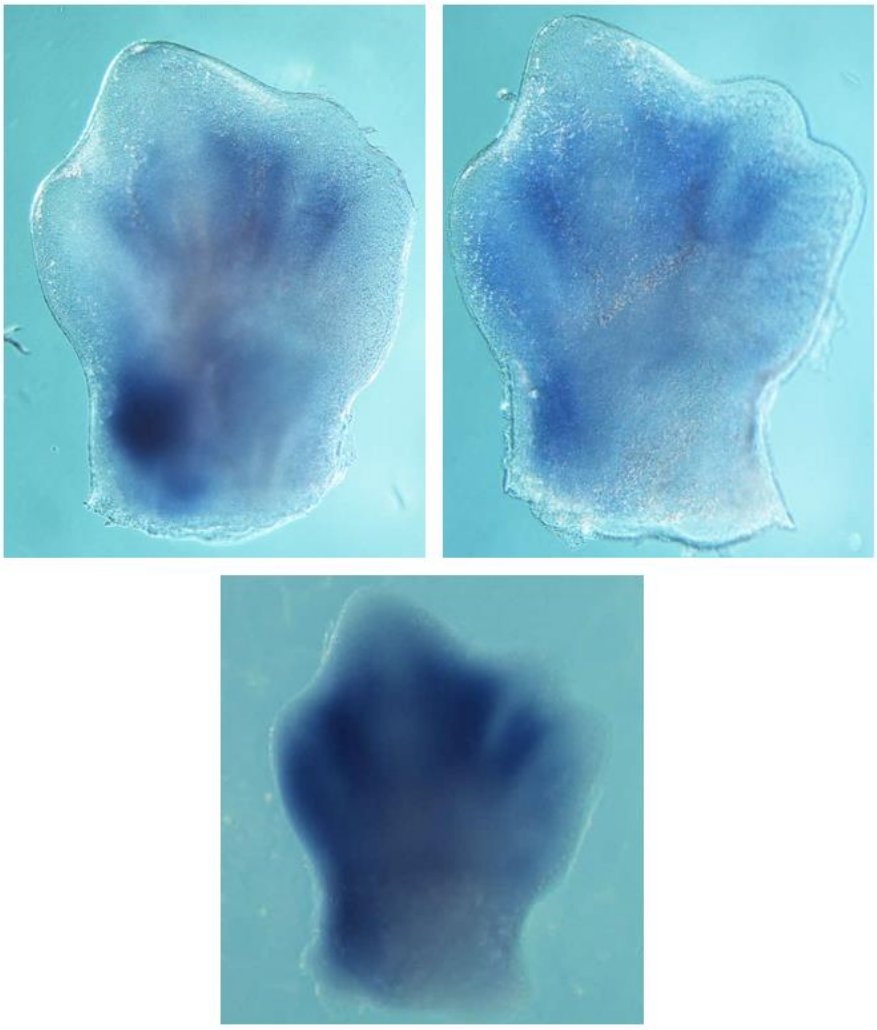

**Figure A3.12.** Expression of *HOXD11* in hind autopodium at d24 – antisense probe.

**Figure A3.13.** Expression of *HOXD11* in hind autopodium at d24 – sense probe.

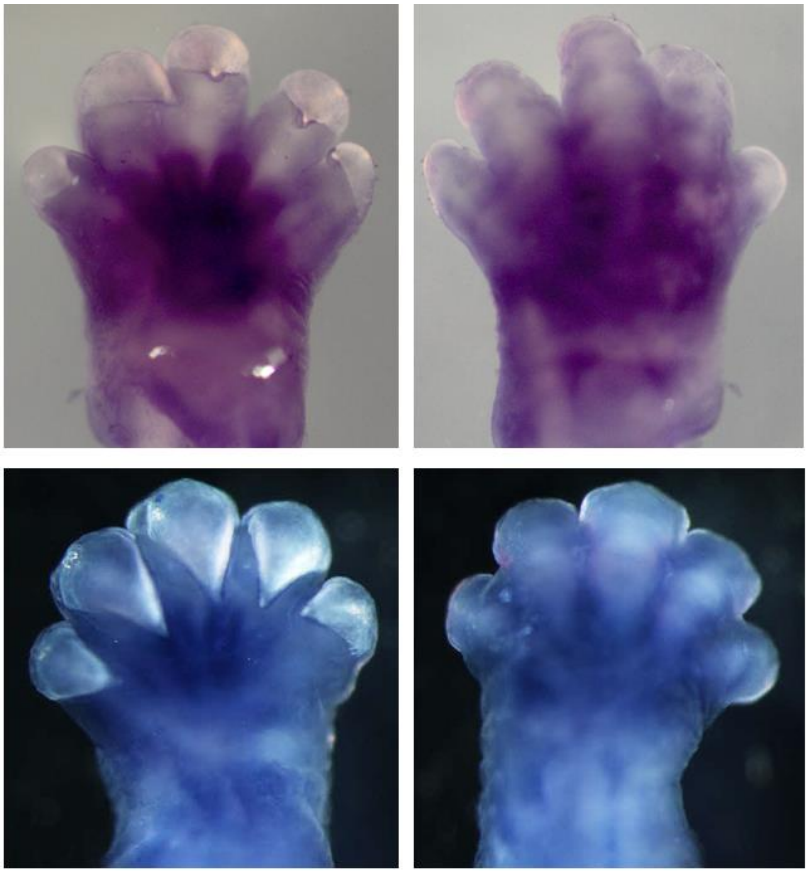

**Figure A3.14.** Expression of *HOXD11* in fore autopodium at d25 – antisense probe.

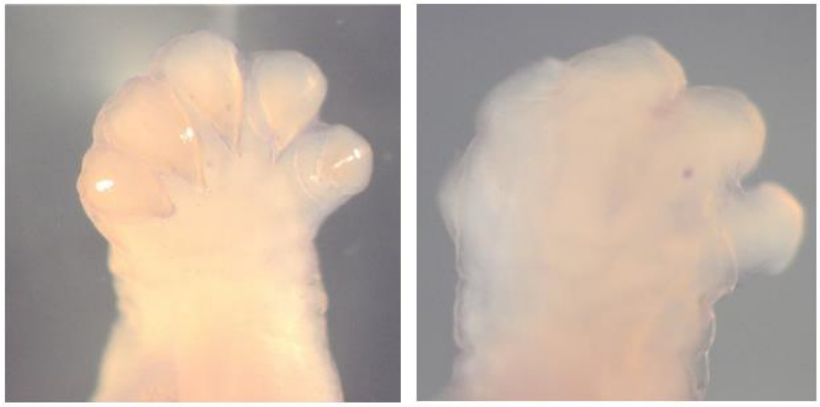

**Figure A3.15.** Expression of *HOXD11* in fore autopodium at d25 – sense probe.

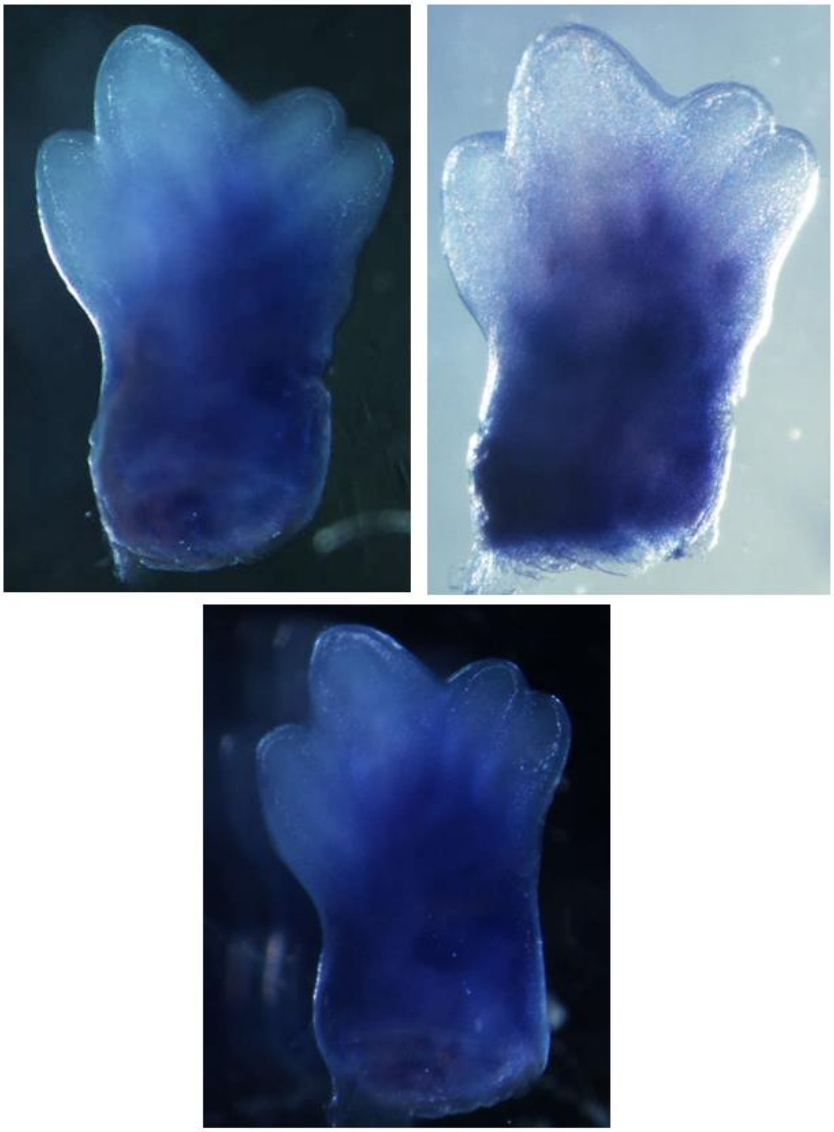

**Figure A3.16.** Expression of *HOXD11* in hind autopodium at d25 – antisense probe.

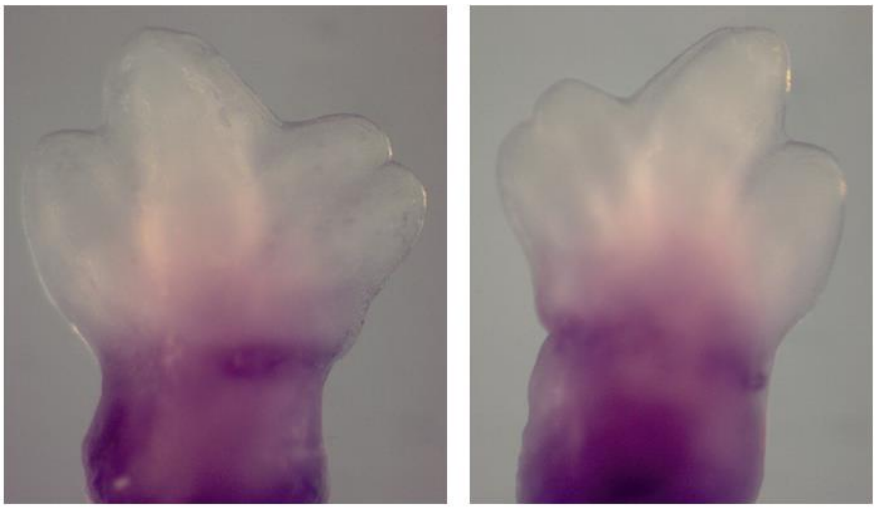

**Figure A3.17.** Expression of *HOXD11* in fore autopodium at d25 – sense probe.

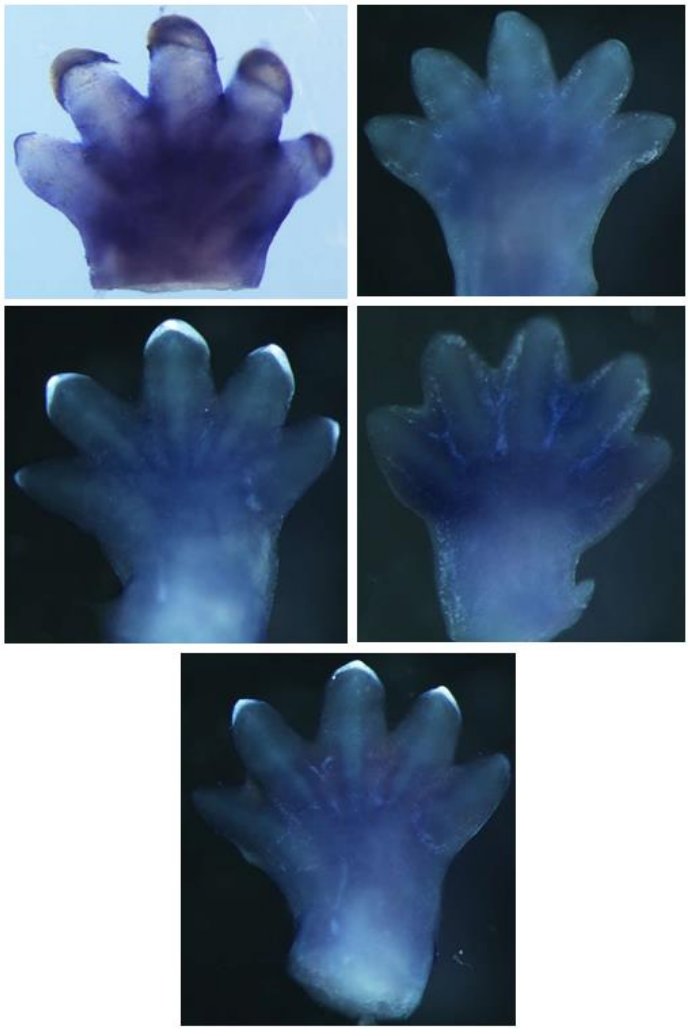

**Figure A3.18.** Expression of *HOXD12* in fore autopodium at d23 – antisense probe.

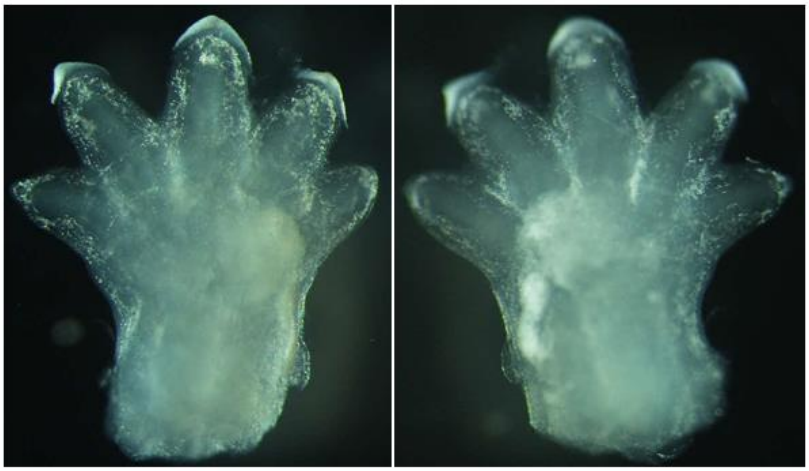

**Figure A3.19.** Expression of *HOXD12* in fore autopodium at d23 – sense probe.

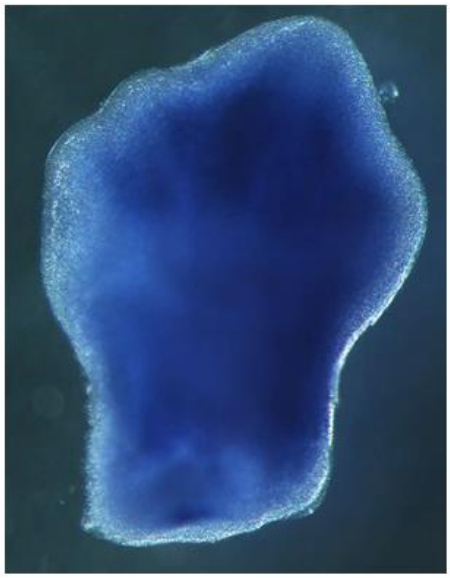

**Figure A3.20.** Expression of *HOXD12* in hind autopodium at d23 – antisense probe.

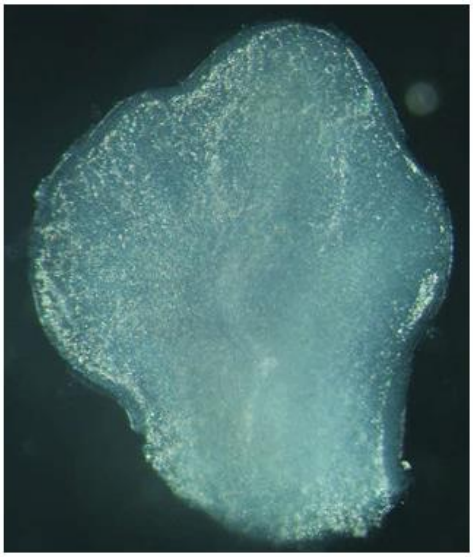

**Figure A3.21.** Expression of *HOXD12* in hind autopodium at d23 – sense probe.

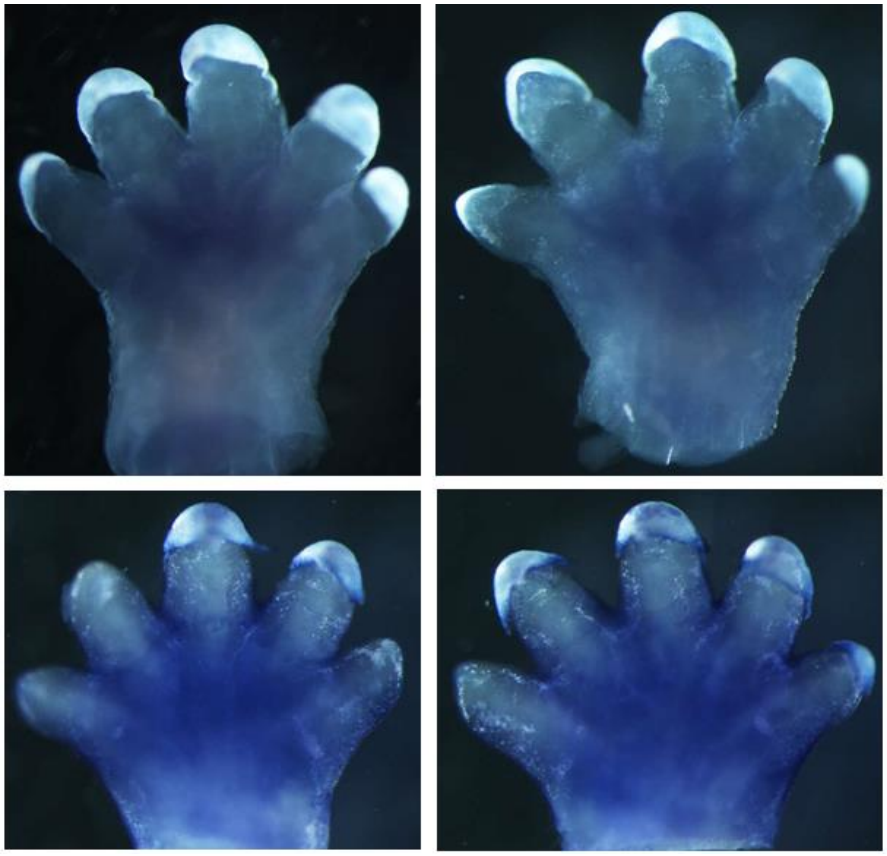

**Figure A3.22.** Expression of *HOXD12* in fore autopodium at d24 – antisense probe.

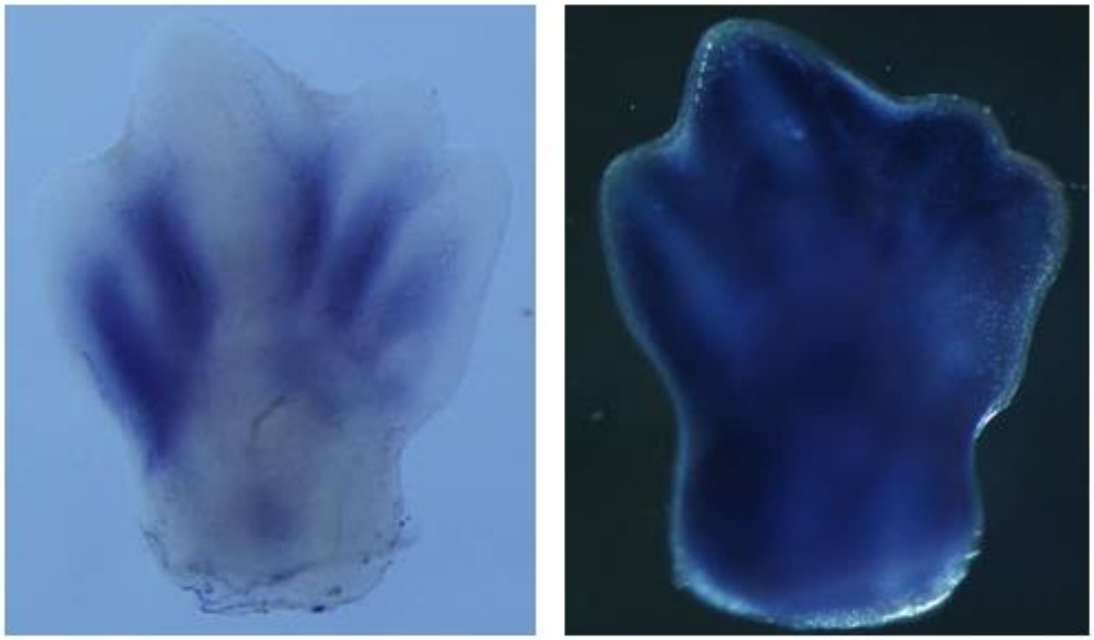

**Figure A3.23.** Expression of *HOXD12* in hind autopodium at d24 – antisense probe.

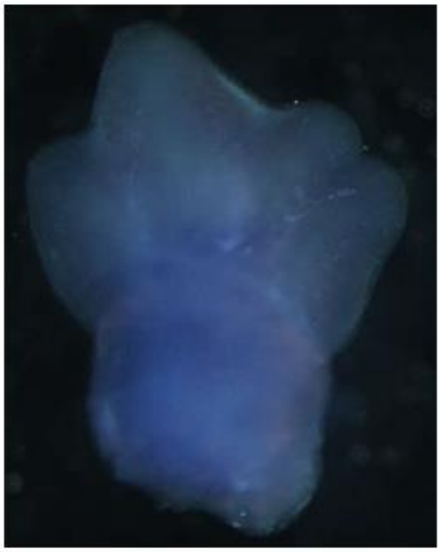

**Figure A3.24.** Expression of *HOXD12* in hind autopodium at d24 – sense probe.

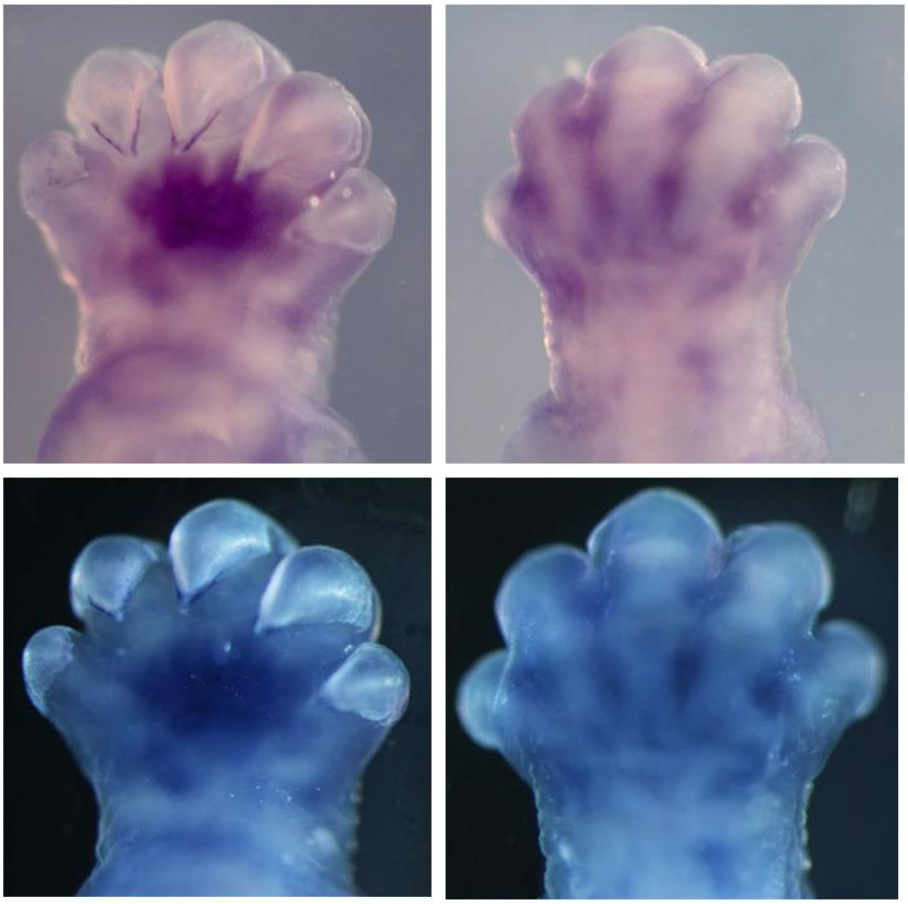

**Figure A3.d25.** Expression of *HOXD12* in fore autopodium at d25 – antisense probe.

**Figure A3.26.** Expression of *HOXD12* in fore autopodium at d25 – sense probe.

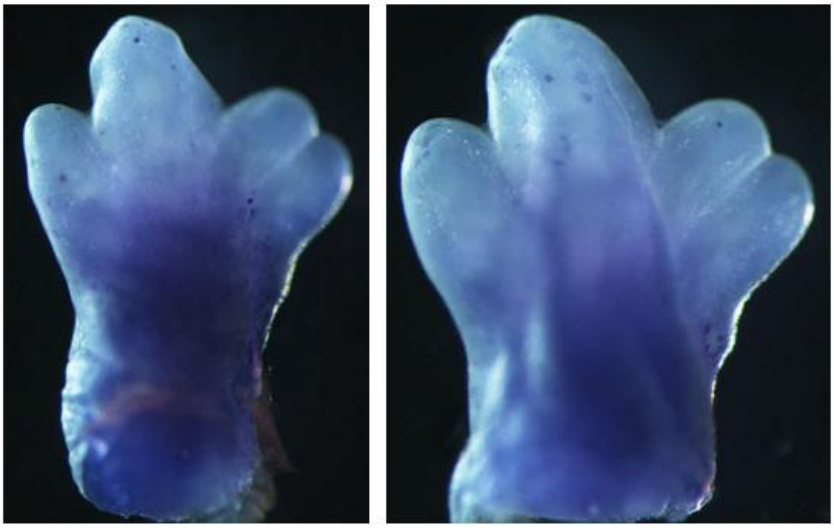

**Figure A3.27.** Expression of *HOXD12* in hind autopodium at d25 – antisense probe.

**Figure A3.28.** Expression of *HOXD12* hind autopodium at d25, HL– sense probe.

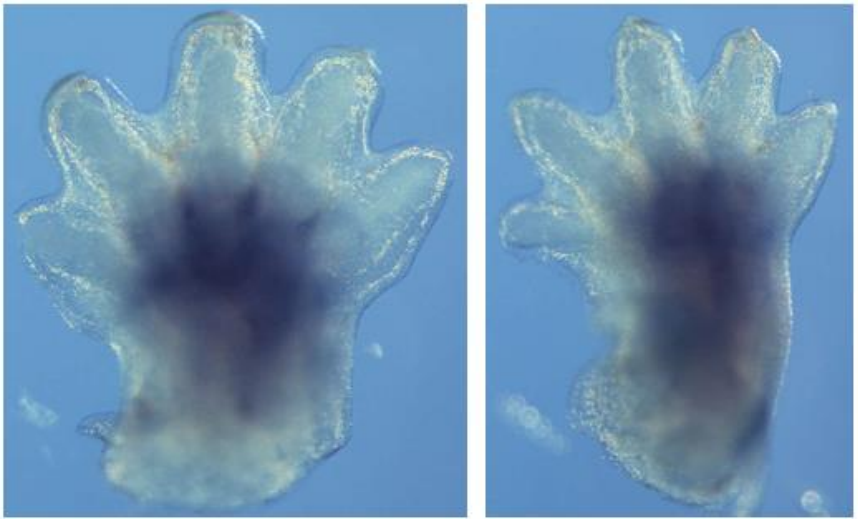

**Figure A3.29.** Expression of *XLOC46* in fore autopodium at d23 – antisense probe.

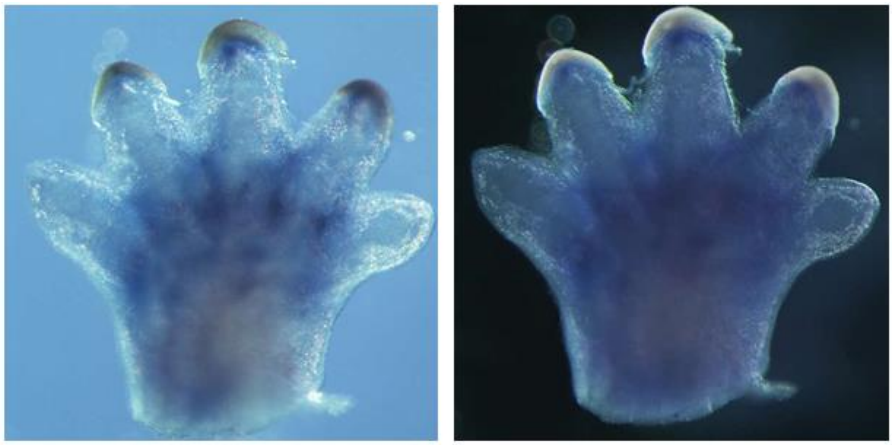

**Figure A3.30.** Expression of *XLOC46* in fore autopodium at d23 – sense probe.

**Figure A3.31.** Expression of *XLOC46* in hind autopodium at d23 – antisense probe.

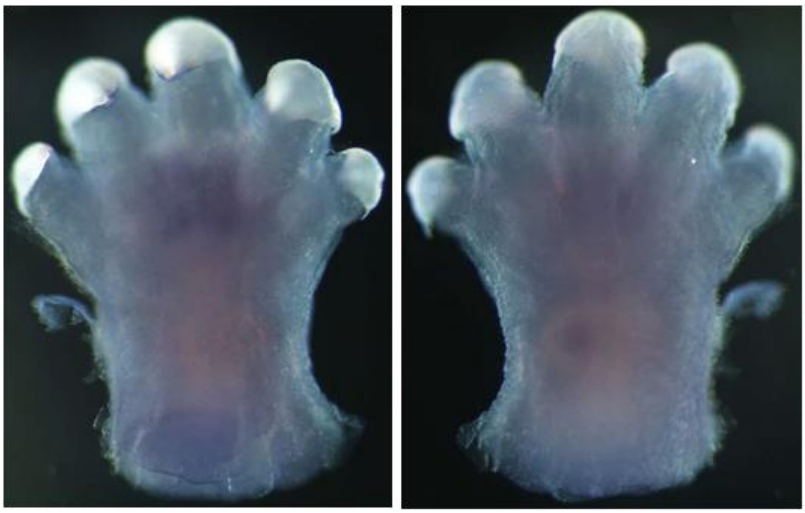

**Figure A3.32.** Expression of *XLOC46* in fore autopodium at d24 – antisense probe.

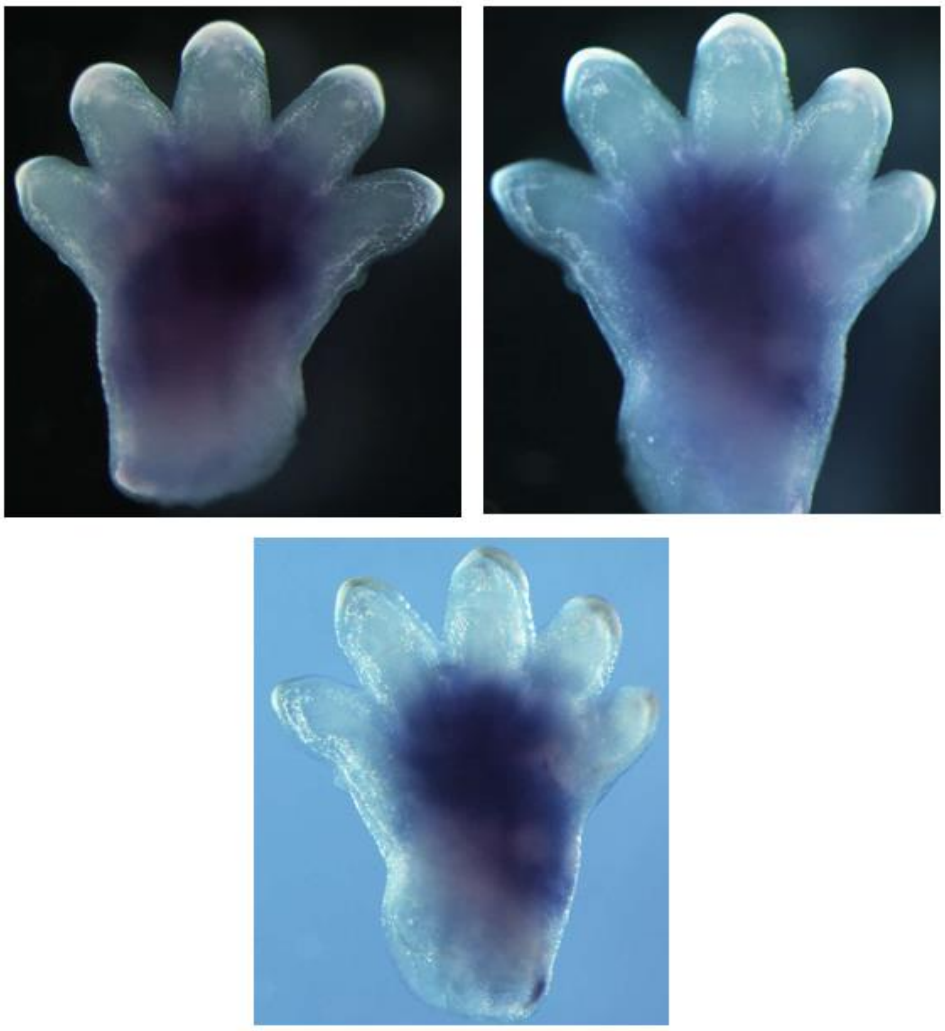

**Figure A3.33.** Expression of *XLOC46* in fore autopodium at d24 – sense probe.

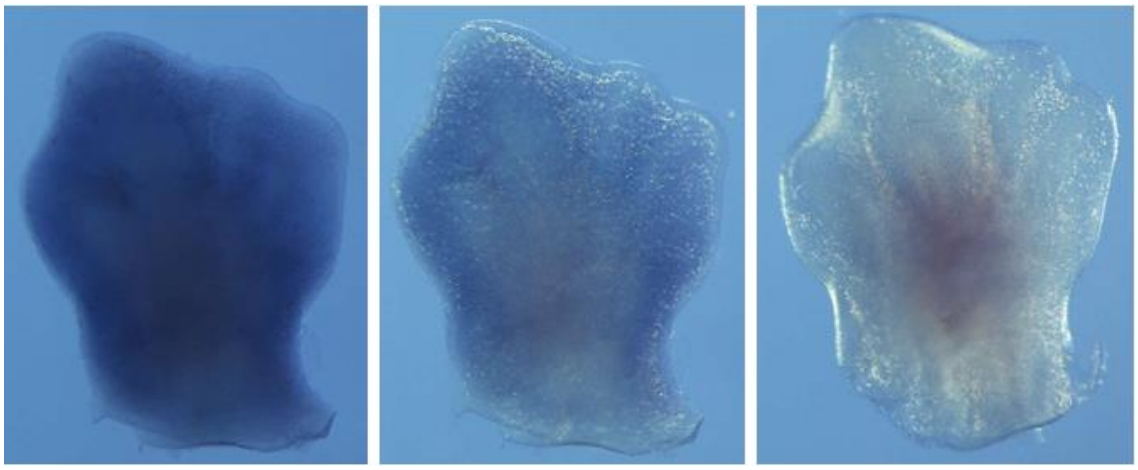

**Figure A3.34.** Expression of *XLOC46* in hind autopodium at d24 – antisense probe.

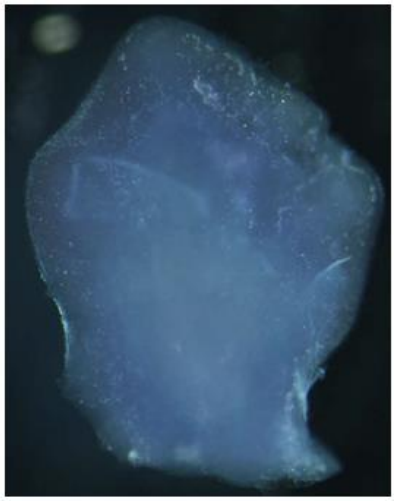

**Figure A3.35.** Expression of *XLOC46* in hind autopodium at d24 – sense probe.

**Figure A3.36.** Expression of *XLOC46* in fore autopodium at d25 – antisense probe.

**Figure A3.37.** Expression of *XLOC46* in fore autopodium at d25 – sense probe.

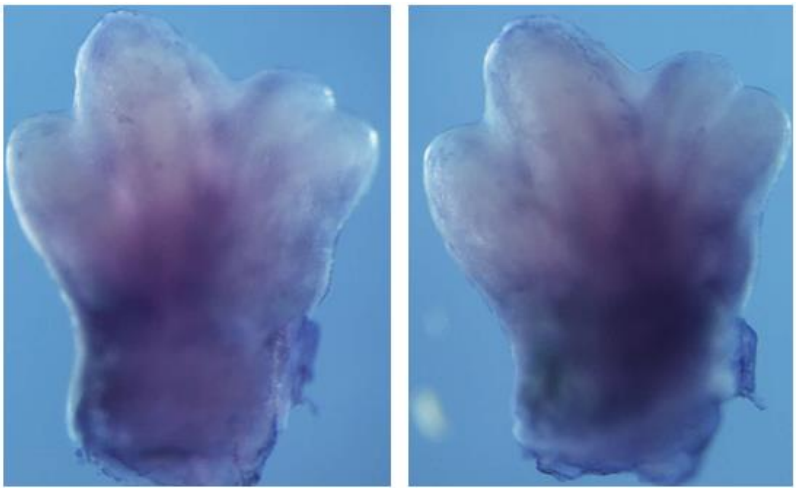

**Figure A3.38.** Expression of *XLOC46* in hind autopodium at d25 – antisense probe.

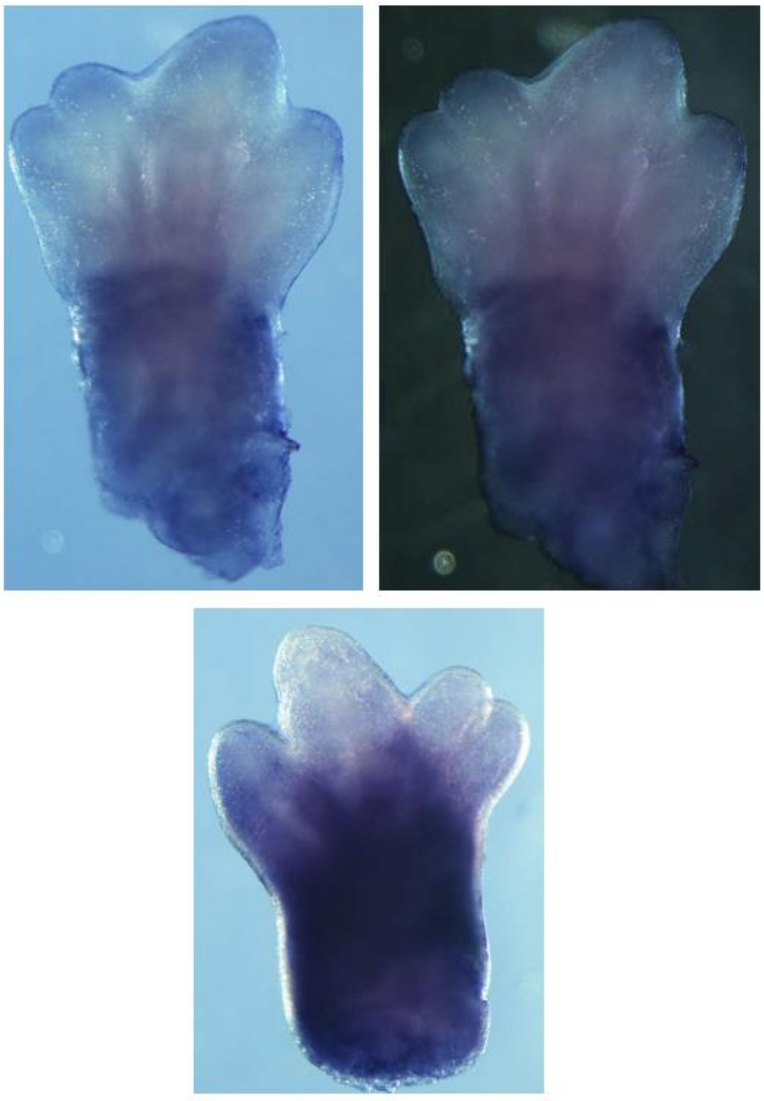

**Figure A3.39.** Expression of *XLOC46* in hind autopodium at d25 – sense probe.

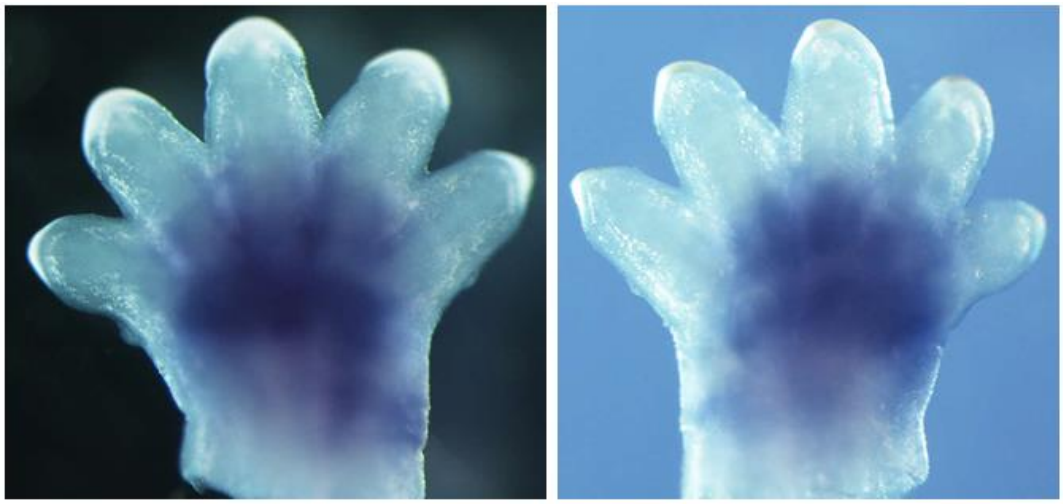

**Figure A3.40.** Expression of *XLOC52* in fore autopodium at d23 – antisense probe.

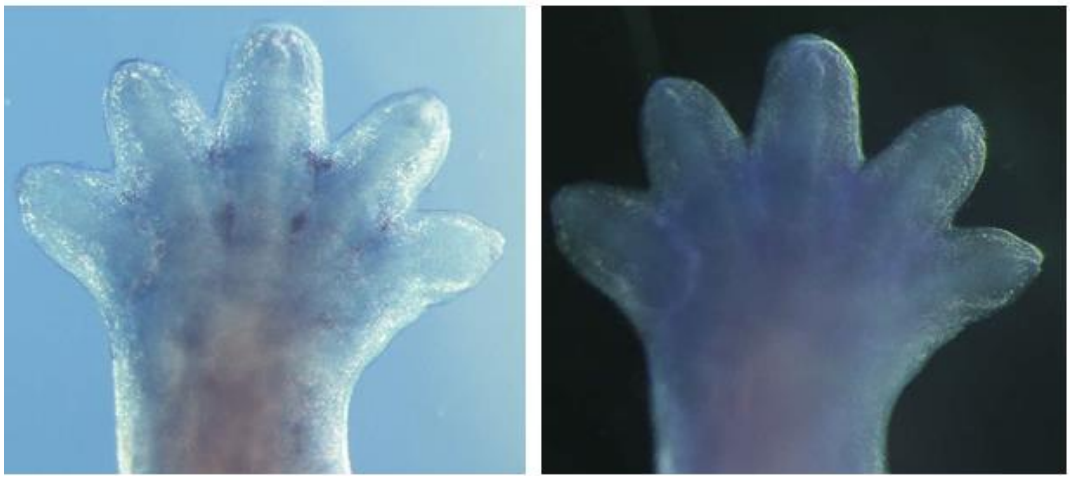

**Figure A3.41.** Expression of *XLOC52* in fore autopodium at d23 – sense probe.

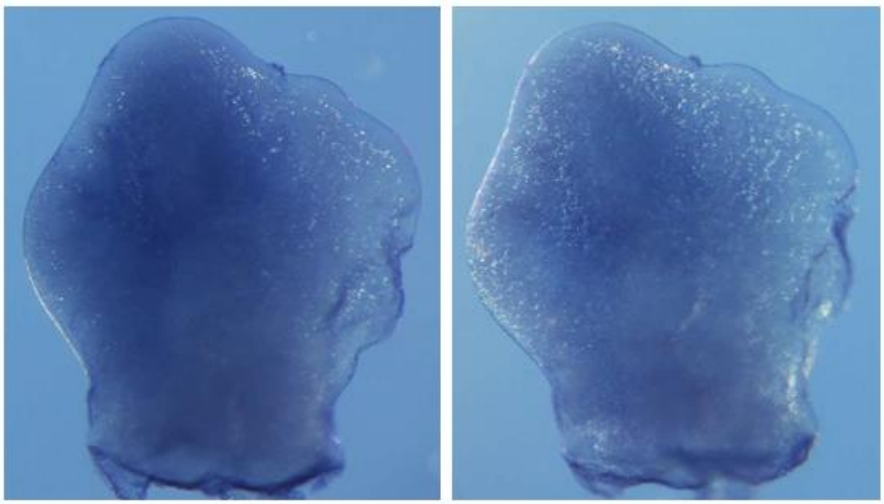

**Figure A3.42.** Expression of *XLOC52* in hind autopodium at d23 – antisense probe.

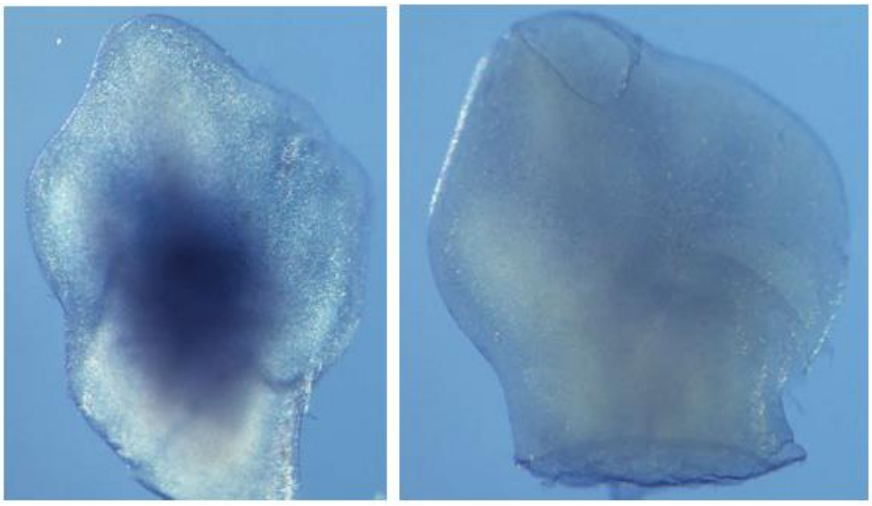

**Figure A3.43.** Expression of *XLOC52* in hind autopodium at d23 – sense probe.

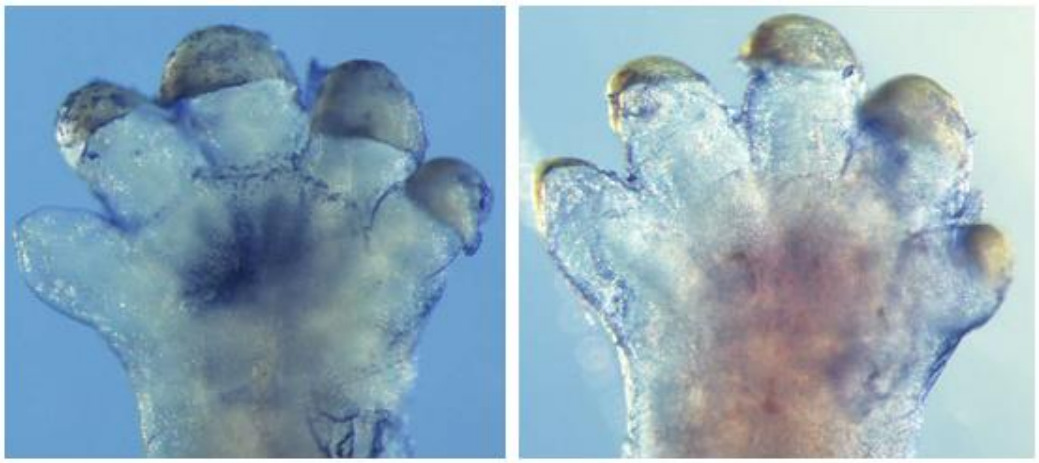

**Figure A3.44.** Expression of *XLOC52* in fore autopodium at d24 – antisense probe.

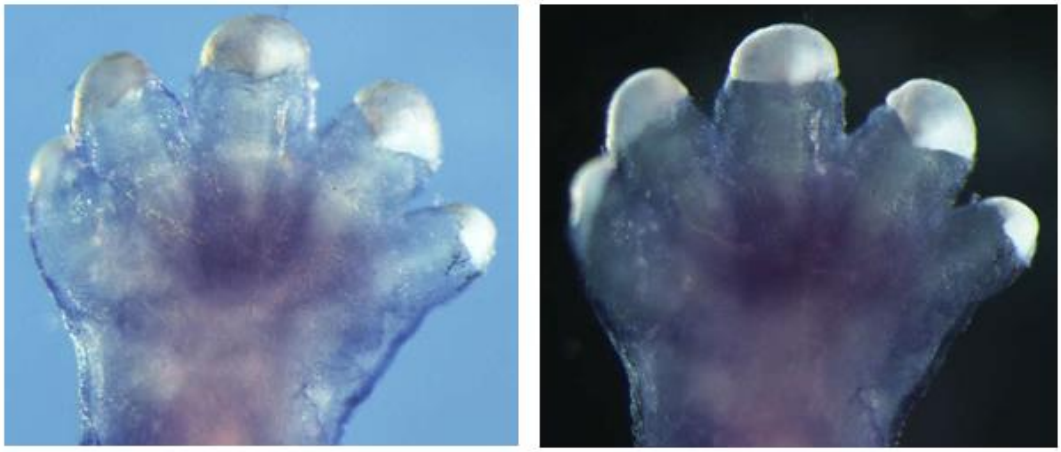

**Figure A3.45.** Expression of *XLOC52* in fore autopodium at d24 – sense probe.

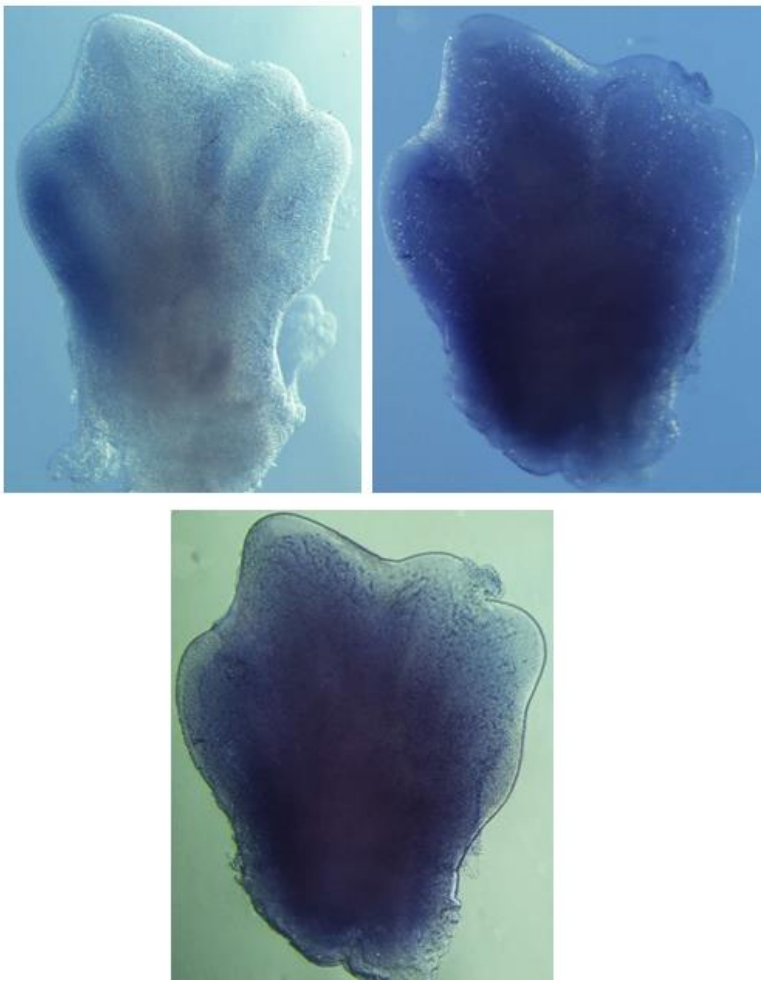

**Figure A3.46.** Expression of *XLOC52* in hind autopodium at d24 – antisense probe.

**Figure A3.47.** Expression of *XLOC52* in fore autopodium at d25 – antisense probe.

**Figure A3.48.** Expression of *XLOC52* in fore autopodium at d25 – sense probe.

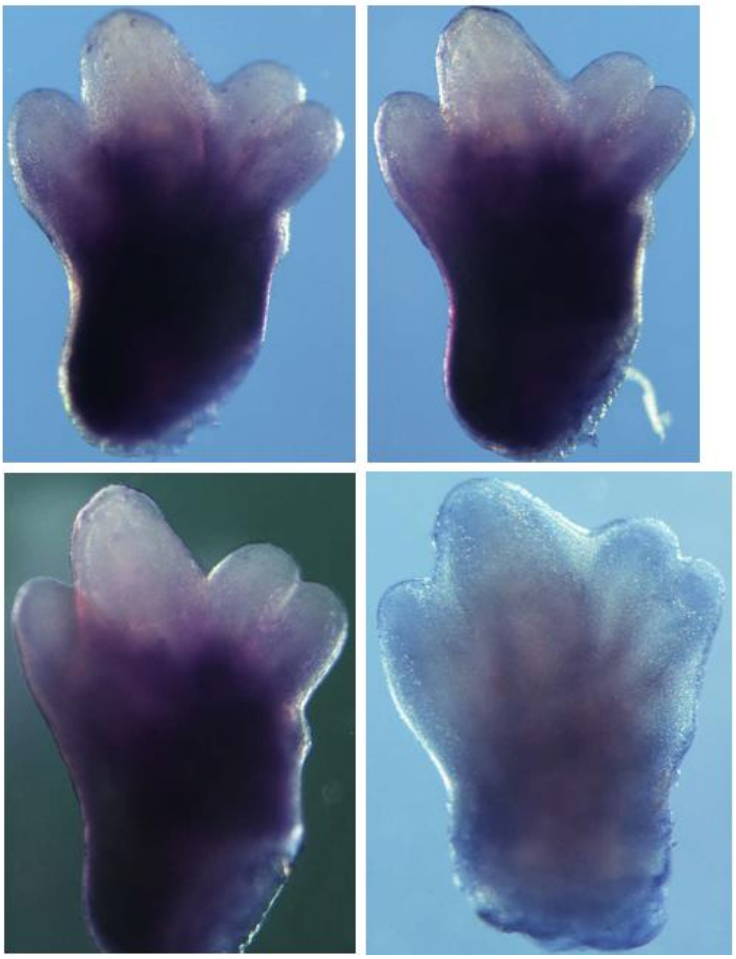

**Figure A3.49.** Expression of *XLOC52* in hind autopodium at d25 – antisense probe.

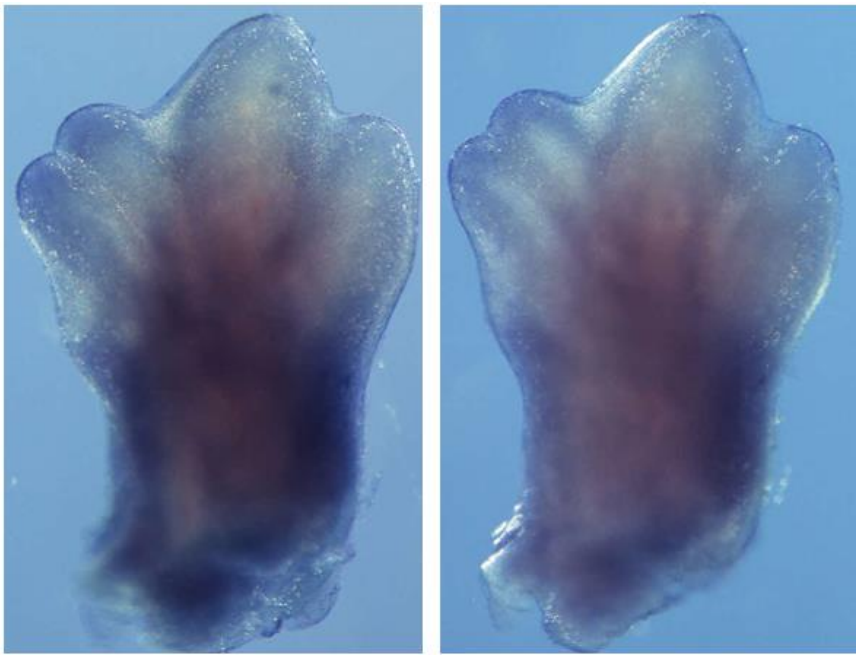

**Figure A3.50.** Expression of *XLOC52* in hind autopodium at d25 – sense probe.

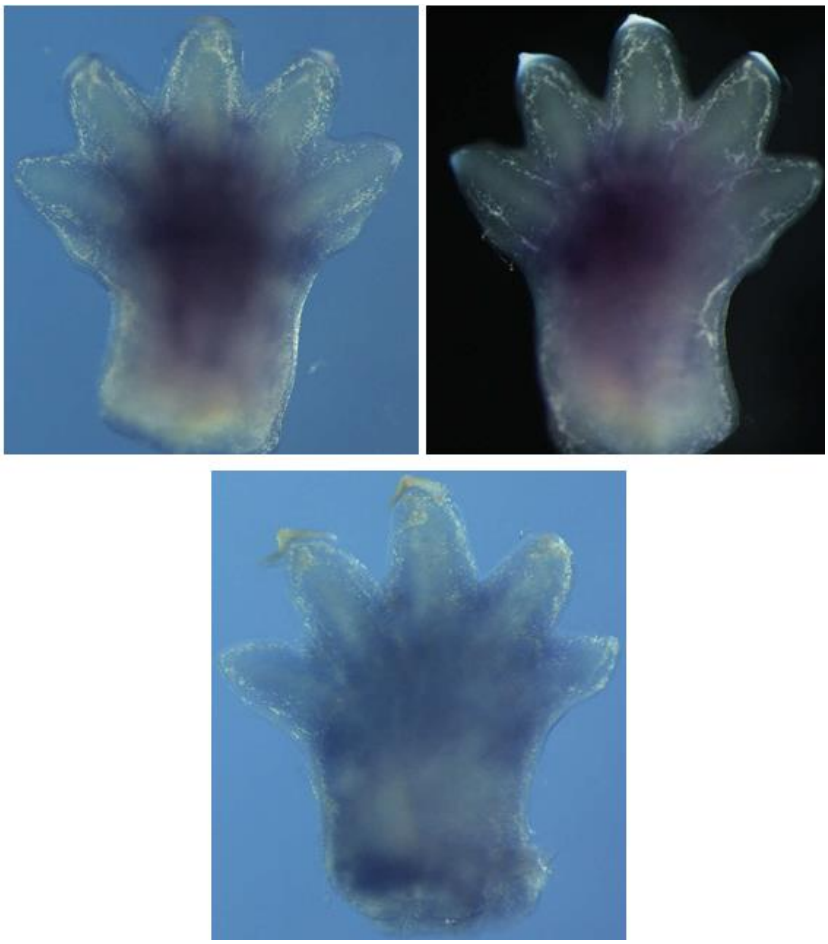

**Figure A3.51.** Expression of *XLOC53* in fore autopodium at d23 – antisense probe.

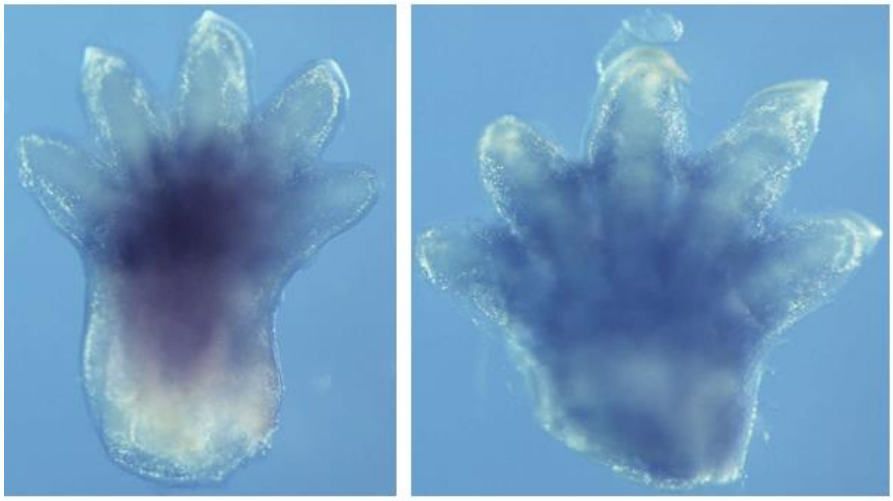

**Figure A3.52.** Expression of *XLOC53* in fore autopodium at d23 – sense probe.

**Figure A3.53.** Expression of *XLOC53* in hind autopodium at d23 – antisense probe.

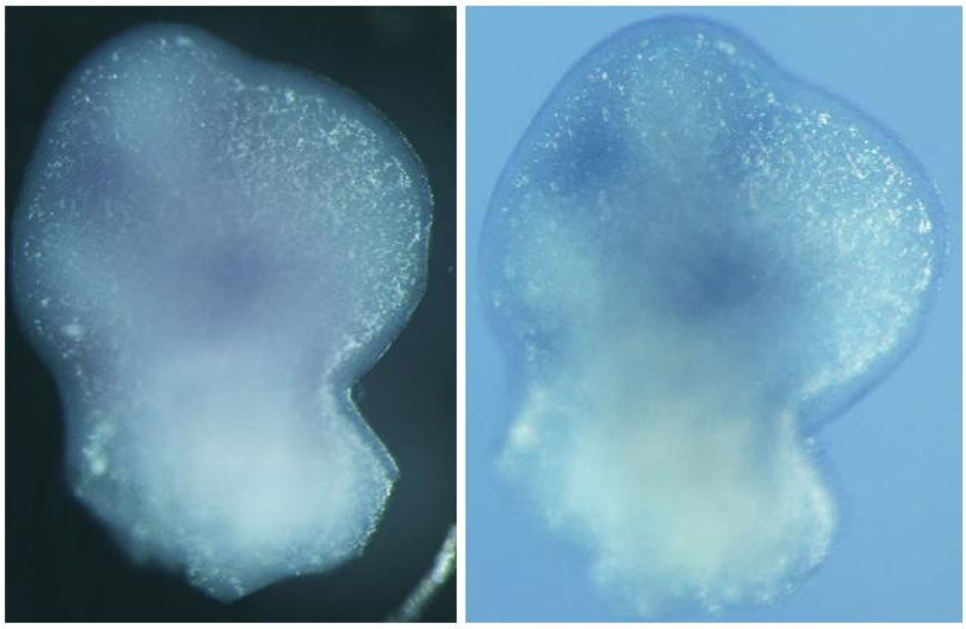

**Figure A3.54.** Expression of *XLOC53* in hind autopodium at d23 – sense probe.

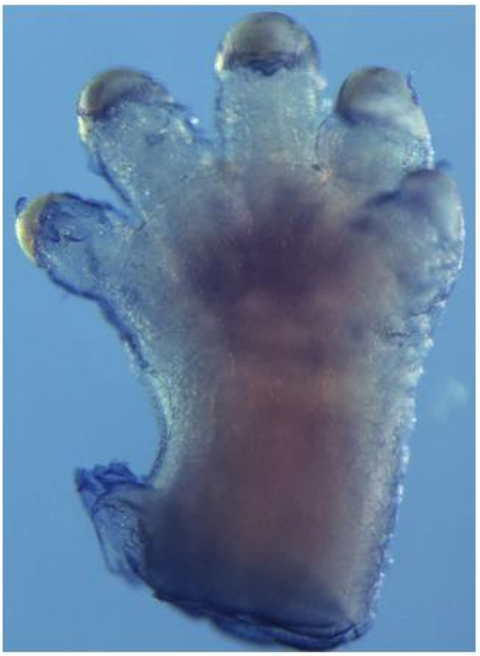

**Figure A3.55.** Expression of *XLOC53* in fore autopodium at d24 – antisense probe.

**Figure A3.56.** Expression of *XLOC53* in fore autopodium at d24 – sense probe.

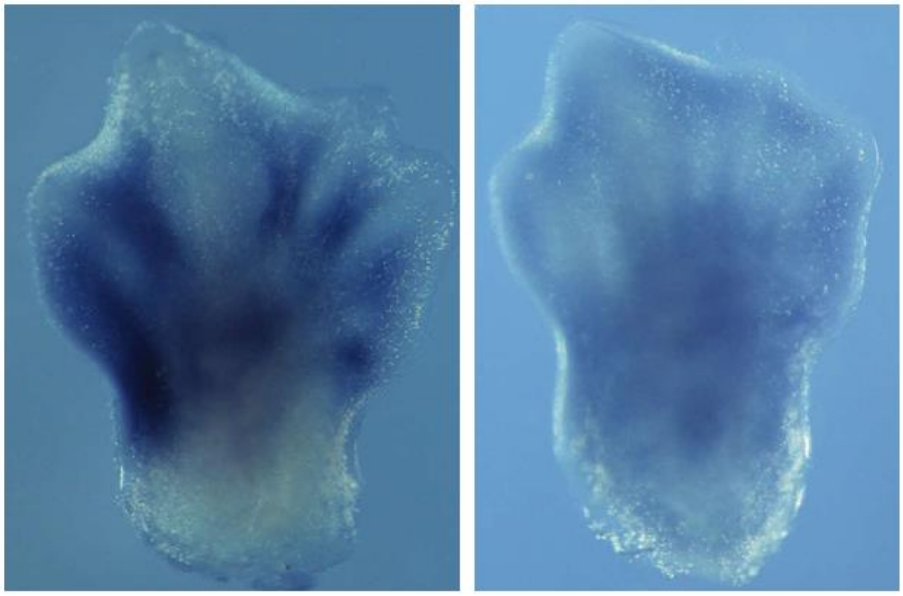

**Figure A3.57.** Expression of *XLOC53* in hind autopodium at d24 – antisense probe.

**Figure A3.58.** Expression of *XLOC53* in hind autopodium at d24 – sense probe.

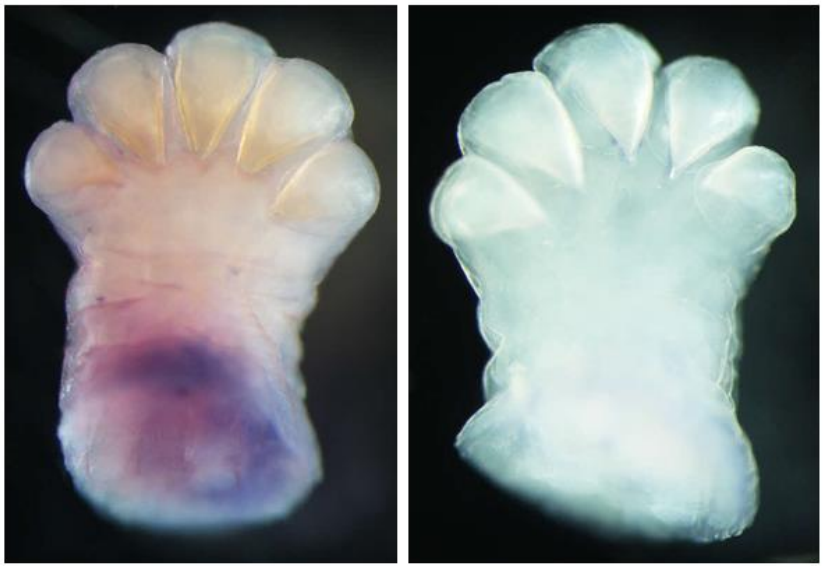

**Figure A3.59.** Expression of *XLOC53* in fore autopodium at d25 – antisense probe.

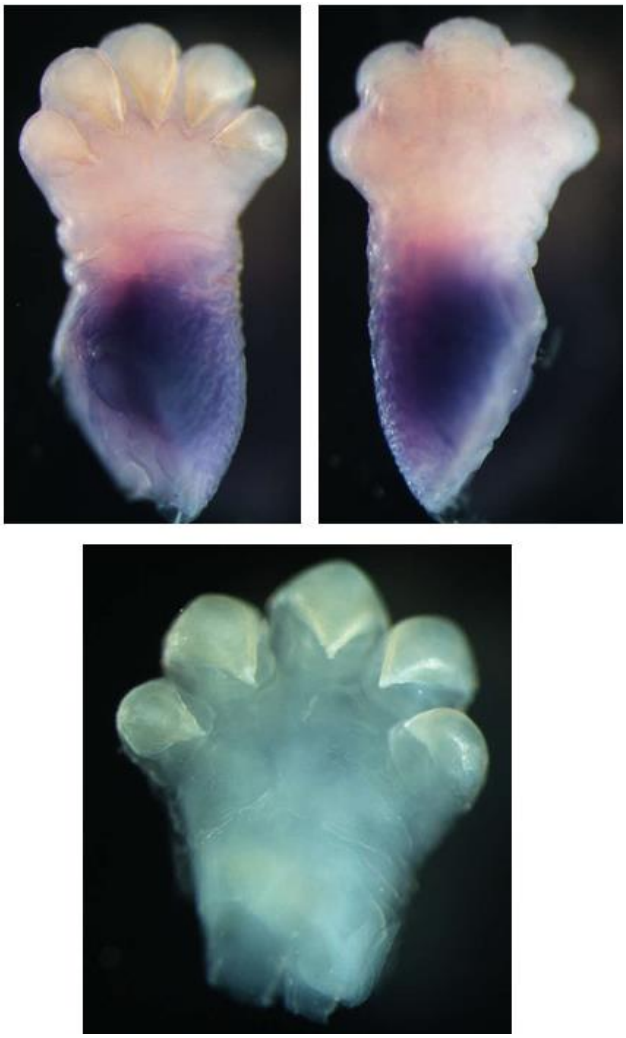

**Figure A3.60.** Expression of *XLOC53* in fore autopodium at d25 – sense probe.

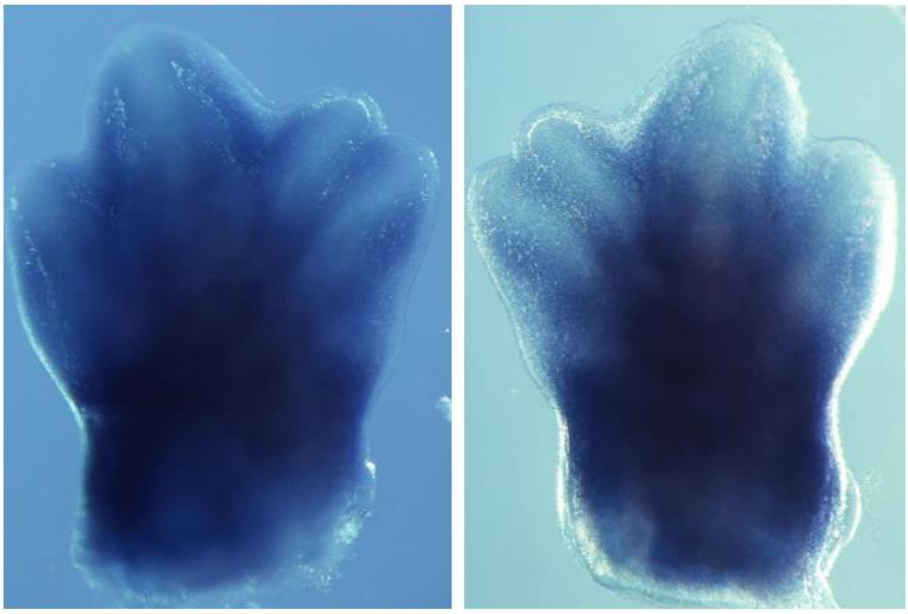

**Figure A3.61.** Expression of *XLOC53* in hind autopodium at d25 – antisense probe.

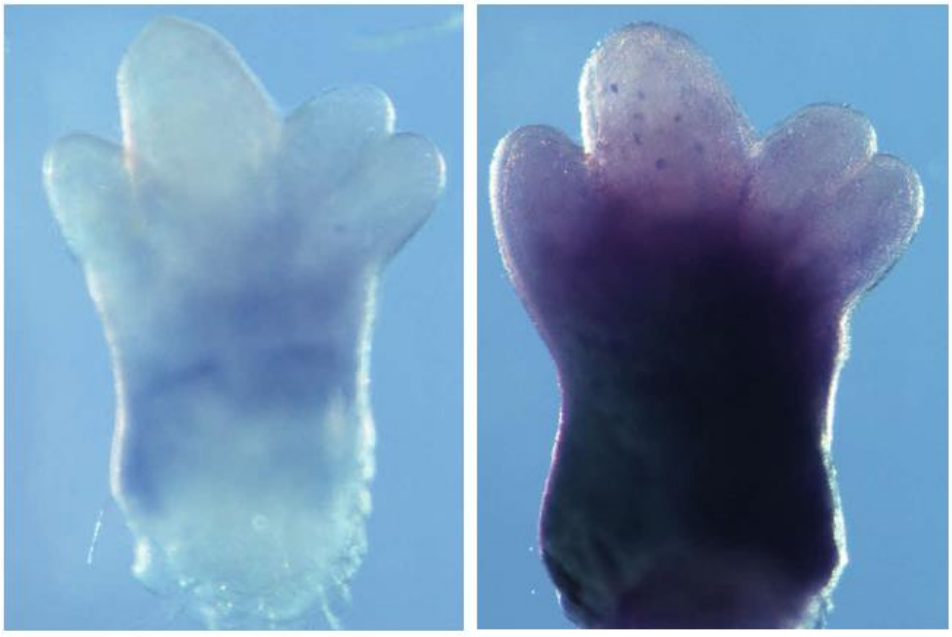

**Figure A3.62.** Expression of *XLOC53* in hind autopodium at d25 – sense probe.

# *CONCLUDING REMARKS AND FUTURE PERSPECTIVES*

The idea of morphological evolution *happening together with* the evolution of regulatory elements of developmental genes is the framework connecting the three chapters of this PhD Thesis. By saying "*happening together with*", I try to group different scenarios, such as 'being driven by', 'leaving footprints on' or 'having its paths opened by', permeating causal and consequential correlations of evolution both at phenotypic and genetic levels. The regulatory network focused was the HoxD cluster because it constitutes an intriguing and complex system involved in the development of several body structures. This study brought important clues into developmental aspects of scarcely explored evolutionary contexts, which may guide future work into unravelling evolutionary developmental mechanisms involved in morphological divergence among related lineages.

Chapter I explored whether independent evolution of snakelike morphologies in Serpentes and Amphisbaenia registered regulatory signatures that might argue for common developmental mechanisms underlying recurrent origin of such phenotypes among Squamata. One footprint identified in *CsB*, a *Hoxd* centromeric regulatory region, seems remarkable because it was predicted in all of the 14 species analyzed from each snakelike group, encompassing loss of TFBS involved in limb developmental pathways and gain of a TFBS essential for AP axis development in tetrapods. This signature, together with a functional dissection of *CsB*, provides evidence for a possible ancient role of *CR2-CsB* fragment during axial elongation. This study showed how strong and consistent predictions based solely on bioinformatics can be, as long as a series of steps is embraced to ascertain accuracy of predictions derived from robust sampling. This approach is important because validation of the predicted convergent signature would represent a dual groundbreaking insight to the field of Evolutionary Developmental Biology. The confirmation of these prediction would reinforce the still relegated role of *CsB* regulatory element for regulating *Hoxd* genes during

development of vertebrates AP axis. Moreover, it would consent a scenario of parallel snakelike evolution in snakes and amphisbaenians, involving similar changes in nucleotide sequence of the same regulatory element that likely account for establishment of limbless-elongated phenotypes. This study nurtures very interesting questions for future projects, especially after availability of snake genomes: 1) Are there additional signatures of independent snakelike evolution in the third conserved peak of *CsB*, which is displaced about 2 kb downstream in comparison with other vertebrates? 2) Does the complete snake *CsB* retain its regulatory potential on limb buds? 3) If we replace the snake sequence fragment where the snakelike regulatory signature was identified by its mouse orthologous region, does the *CsB* display its full regulatory capacity?

Chapter II described that snake evolution impaired regulatory ability of two limb-related regulatory elements, *CNS65* (*HoxD* telomeric landscape) and *Island I* (*HoxD* centromeric landscape). Functional evidence of their regulatory capacities degeneration was complemented with comparative analyses of predicted TFBS, which identified one putative stilopod/zeugopod-specific segment in *CNS65* and three autopod-specific candidate segments in *Island I.* Such regions are likely comprised in limb-specific regulatory modules within these enhancers that potentially evolved under relaxed selection together with or after limb loss. This chapter inspires future studies towards answering whether substitutions of the regions where the signatures were predicted in the sequences of snakes by their mouse orthologue segments would recover their regulatory activities in transgenic mice limb buds. Moreover, one might ask if mouse sequences of these elements would still elicit a similar regulatory activity if the same fragments of snakes occupied the regions of their mouse orthologues? Both complementary approaches would evaluate if the loss of regulatory function in the snake enhancers resides in degeneration of the candidate segments proposed here, also testing the relevance of these motifs for limb development. This chapter demonstrates that studies on the

molecular evolution of regulatory sequences contribute for identifying specific elements dedicated to unique developmental processes affected by particular evolutionary events. Specifically, I showed that by focusing on functional loss of regulatory sequences it is possible to select fragments within sequences that are possibly module-specific, using simple TFBS prediction (not as strict as predictions performed in Chapter I, because they are based on functional evidence), an approach that paves the way for discovering relevant elements within CREs that are dedicated to development of specific morphological structures in vertebrates.

Chapter III explored transcripts of noncoding sequences in the HoxD cluster of a marsupial family that exhibits several limb peculiarities when compared to other limbed vertebrates. This unity is connected with Chapter I through the conceptual framework focusing on recurrent evolution of specific phenotypes: evolution of particular traits in the tammar wallaby (i.e. the syndactylous feet) is also observed in a distantly related marsupial group (Peramelemorpha). The scenario is slightly different, however, as marsupial syndactyly is interpreted as a non-adaptive trait because its relevance for locomotion remains unknown; Diprotodontia (the family to which the kangaroos belong), for instance, is the most diversified marsupial order regarding locomotion and ecology (Weisbecker & Nilsson 2008). The most accepted current hypothesis states that marsupial syndactyly represents a case of parallelism established through an ontogenetic constraint (Weisbecker & Nilsson 2008). Although the topic has not been directly approached in my Thesis, this study represents an important step towards understanding the enigmatic genetic system that builds autopodial morphology in marsupials, as ncRNAs have been disclosed in the past decade as important regulatory elements. I explored transcription profiles and expression patterns of HoxD ncRNAs identified in embryo limbs, among which five lncRNAs were conserved in mammals and three of them might represent marsupial elements. These putative marsupial elements may constitute important regulatory elements involved in morphological evolution, therefore

representing interesting targets for functional investigation. As most ncRNAs examined showed similar transcriptional profiles and overlapping expression domains with their putative target *HOXD* coding genes, they may regulate *HOXD* expression during autopodial development. Because expression of the putative pre-miRNA *XLOC53* decreased in forelimb at the day before birth, it may also be involved in defining heterochrony between forelimbs and hindlimbs, a peculiarity of marsupials. This work provides the first foundation for exploring function of ncRNAs from cluster HoxD in the scenario of limb development and the evolutionary mechanisms linked to autopodial diversification in marsupials. Next steps of investigation may involve performing functional assays to access specific ncRNA roles during development, such as interference RNA and overexpression assays using tammar wallaby cell cultures where both coding and noncoding HoxD regions here studied are transcribed. As future questions addressed, I exemplify: 1) does inhibition or overexpression of these ncRNAs affect expression profiles of their putative target terminal *HOXD* genes? 2) do these ncRNAs act in *cis* or *trans*-regulatory mechanisms?

Overall, the present PhD Thesis validates the general hypothesis that regulatory elements of HoxD cluster carry histories of morphological divergence both in mammals and reptiles, placing the foundations for understanding developmental mechanisms that originated the diversity of extant vertebrate forms standing in our Planet.

## LITERATURE CITED

Abouheif, E., 2008. Parallelism as the Pattern and Process of Mesodermevolution. *Evolution and Development*, 10(1), pp.3–5.

Agarwal, P. et al., 2003. Tbx5 is essential for forelimb bud initiation following patterning of the limb field in the mouse embryo. *Development (Cambridge, England)*, 130(3), pp.623–633.

Akhtar, W. & Veenstra, G.J.C., 2011. TBP-related factors: a paradigm of diversity in transcription initiation. *Cell & bioscience*, 1(1), p.23.

Akker, E. Van Den & Forlani, S., 2002. Cdx1 and Cdx2 have overlapping functions in anteroposterior patterning and posterior axis elongation. *Development*, 129, pp.2181–2193.

Alexandrovich, A. et al., 2008. A role for GATA-6 in vertebrate chondrogenesis. *Developmental Biology*, 314(2), pp.457–470.

Amemiya, C.T. et al., 2013. The African coelacanth genome provides insights into tetrapod evolution. *Nature*, 496(7445), pp.311–6.

Amiel, J. et al., 2004. Polyalanine expansions in human. *Human Molecular Genetics*, 13(2), pp.235–243.

Amrine-Madsen, H. et al., 2003. Nuclear gene sequences provide evidence for the monophyly of australidelphian marsupials. *Molecular phylogenetics and evolution*, 28, pp.186–196.

Andrewartha, H. & Barker, S., 1969. Introduction to a Study of the Ecology of the Kangaroo Island Wallaby Protemnodon-eugenii within Flinders Chase Kangaroo Island South-Australia. *Transactions of the Royal Society of South Australia*, 93, pp.127–132.

Andrey, G. et al., 2013. A Switch Between Topological Domains Underlies HoxD Genes Collinearity in Mouse Limbs. *Science*, 340, p.1195.

Apesteguía, S. & Zaher, H., 2006. A Cretaceous terrestrial snake with robust hindlimbs and a sacrum. *Nature*, 440, pp.1037–1040.

Archambeault, S., Taylor, J.A. & Crow, K.D., 2014. HoxA and HoxD expression in a variety of vertebrate body plan features reveals an ancient origin for the distal Hox program. *EvoDevo*, 5(44), pp.1–10.

Aulehla, A. et al., 2003. Wnt3a plays a major role in the segmentation clock controlling somitogenesis. *Developmental Cell*, 4, pp.395–406.

Averof, M. & Akam, M., 1995. Hox genes and the diversification of insect and crustacean body plans. *Nature*, 376, pp.420–423.

Bae, E. et al., 2002. Characterization of the intergenic RNA profile at abdominal-A and Abdominal-B in the Drosophila bithorax complex. *Proc Natl Acad Sci U S A*, 99(26), pp.16847–16852.

Baker, M. et al., 2004. Relationships among the families and orders of marsupials and the major mammalian lineages based on recombination activating gene-1. *Journal of Mammalian Evolution*, 11, pp.11–16.

Balhoff, J.P. & Wray, G.A., 2005. Evolutionary analysis of the well characterized endo16 promoter reveals substantial variation within functional sites. *Proceedings of the National Academy of Sciences of the United States of America*, 102(24), pp.8591–6.

Bejder, L. & Hall, B.K., 2002. Limbs in whales and limblessness in other vertebrates: Mechanisms of evolutionary and developmental transformation and loss. *Evolution and Development*, 4(6), pp.445–458.

Bellairs, B.Y.A.D.A. & Underwood, G., 1950. The origin of snakes.

Bellon, E., Luyten, F.P. & Tylzanowski, P., 2009. Delta-EF1 is a negative regulator of Ihh in the developing growth plate. *Journal of Cell Biology*, 187(5), pp.685–699.

Bel-Vialar, S., Itasaki, N. & Krumlauf, R., 2002. Initiating Hox gene expression: in the early chick neural tube differential sensitivity to FGF and RA signaling subdivides the HoxB genes in two distinct groups. *Development*, 129(22), pp.5103–5115.

Benton, M.J., 1990. Phylogeny of the major tetrapod groups: morphological data and divergence dates. *Journal of molecular evolution*, 30(5), pp.409–24.

Berge, D. ten et al., 1998. Prx1 and Prx2 in skeletogenesis: roles in the craniofacial region, inner ear and limbs. *Development (Cambridge, England)*, 125(19), pp.3831–3842.

Berger, P., 1966. No TitleEleven-month “Embryonic Diapause” Marsupial. *Letters to Nature*, 211, pp.435–436.

Blignaut, M., Morris, K. V & Scripps, T., 2012. regulation of gene expression A book edited by Kevin Morris. , 7(6), pp.664–666.

Bonnelye, E. et al., 1997. The ERR-1 orphan receptor is a transcriptional activator expressed during bone development. *Molecular endocrinology*, 11(7), pp.905–916.

Brandley, M.C., Huelsenbeck, J.P. & Wiens, J.J., 2008. Rates and patterns in the evolution of snake-like body form in squamate reptiles: evidence for repeated re-evolution of lost digits and long-term persistence of intermediate body forms. *Evolution*, 62(8), pp.2042–64.

Brand-Saberi, B., 2005. Genetic and epigenetic control of skeletal muscle development. *Annals of Anatomy*, 187(3), pp.199–207.

Browner, N.C., Sellak, H. & Lincoln, T.M., 2004. Downregulation of cGMP-dependent protein kinase expression by inflammatory cytokines in vascular smooth muscle cells. *American journal of physiology. Cell physiology*, 287(1), pp.C88–96.

Bruneau, S. et al., 2001. The Mouse Hoxd13spdh Mutation, a Polyalanine Expansion Similar to Human Type II Synpolydactyly (SPD), Disrupts the Function but Not the Expression of Other Hoxd Genes. *Developmental Biology*, 237(2), pp.345–353.

Burel, A. et al., 2006. Role of HOXA7 to HOXA13 and PBX1 genes in various forms of MRKH syndrome (congenital absence of uterus and vagina). *Journal of negative results in biomedicine*, 5, p.4.

Burke, A. et al., 1995. Hox genes and the evolution of vertebrate axial morphology. *Development*, 121(2), pp.333–346.

Büscher, D. et al., 1997. Evidence for genetic control of Sonic hedgehog by Gli3 in mouse limb development. *Mechanisms of development*, 62(2), pp.175–82.

Bustin, S. a. et al., 2009. The MIQE guidelines:Minimum Information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, 55(4), pp.611–622.

Caldwell, M., 2003. " Without a leg to stand on": on the evolution and development of axial elongation and limblessness in tetrapods. *Canadian Journal of Earth Sciences*, 40, pp.573–588.

Caldwell, M.W. & Lee, M.S.Y., 1997. A snake with legs from the marine Cretaceous of the Middle East. *Letters to Nature*, 386, pp.705–709.

Capdevila, J. & Belmonte, J., 2001. Patterning mechanisms controlling vertebrate limb development. *Annual review of cell and developmental biology*, 17, pp.87–132.

Capdevila, J. & Izpisúa-Belmonte, J.-C., 2001. Patterning Mechanisms Controlling Vertebrate Limb Development. *Annual review of cell and developmental biology*, 17, pp.87–132.

Capellini, T.D., Handschuh, K., et al., 2011. Control of pelvic girdle development by genes of the Pbx family and Emx2. *Developmental dynamics : an official publication of the American Association of Anatomists*, 240(5), pp.1173–89.

Capellini, T.D. et al., 2008. Pbx1/Pbx2 govern axial skeletal development by controlling Polycomb and Hox in mesoderm and Pax1/Pax9 in sclerotome. *Developmental biology*, 321(2), pp.500–14.

Capellini, T.D. et al., 2006. Pbx1/Pbx2 requirement for distal limb patterning is mediated by the hierarchical control of Hox gene spatial distribution and Shh expression. *Development (Cambridge, England)*, 133(11), pp.2263–73.

Capellini, T.D. et al., 2010. Scapula development is governed by genetic interactions of PBX1 with its family members and with Emx2 via their cooperative control of Alx1. *Development*, 137, pp.2559–2569.

Capellini, T.D., Zappavigna, V. & Selleri, L., 2011. Pbx homeodomain proteins: TALEnted regulators of limb patterning and outgrowth. *Developmental dynamics: an official publication of the American Association of Anatomists*, 240(5), pp.1063–86.

Carapuço, M. et al., 2005. Hox genes specify vertebral types in the presomitic mesoderm. *Genes & development*, 19(18), pp.2116–21.

Carroll, S.B., 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*, 134(1), pp.25–36.

Carroll, S.B., 2005. Evolution at two levels: on genes and form. *PLoS biology*, 3(7), pp.e245–e245.

Carroll, S.B., 1995. Homeotic genes and the evolution of arthropods and chordates. *Nature*, 376, pp.479–485.

Catala, M., Teillet, M.A. & Le Douarin, N.M., 1995. Organization and development of the tail bud analyzed with the quail-chick chimaera system. *Mechanisms of development*, 51(1), pp.51–65.

Chen, A., Ginty, D. & Fan, C.-M., 2005. Protein kinase A signalling via CREB controls myogenesis induced by Wnt proteins. *Nature*, 433(7023), pp.317–322.

Chen, C. et al., 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic acids research*, 33(20), p.e179.

Chen, F. & Capecchi, M.R., 1999. Paralogous mouse Hox genes, Hoxa9, Hoxb9, and Hoxd9, function together to control development of the mammary gland in response to pregnancy. *Proceedings of the National Academy of Sciences of the United States of America*, 96, pp.541–546.

Chew, K.Y. et al., 2014. Heterochrony in the regulation of the developing marsupial limb. *Developmental dynamics : an official publication of the American Association of Anatomists*, 243(2), pp.324–338.

Chew, K.Y. et al., 2012. HOXA13 and HOXD13 expression during development of the syndactylous digits in the marsupial Macropus eugenii. *BMC developmental biology*, 12(2), pp.1–15.

Chimal-Monroy, J. et al., 2003. Analysis of the molecular cascade responsible for mesodermal limb chondrogenesis: Sox genes and BMP signaling. *Developmental Biology*, 257(2), pp.292–301.

Cho, E. & Dressler, G.R., 1998. TCF-4 binds β -catenin and is expressed in distinct regions of the embryonic brain and limbs. *Mechanisms of Development*, 77(1), pp.9–18.

Chotteau-Lelievre, A. et al., 2001. Expression patterns of the Ets transcription factors from the PEA3 group during early stages of mouse development. *Mechanisms of Development*, 108(1-2), pp.191–195.

Coates, M. & Ruta, M., 2000. Nice snake, shame about the legs. *Trends in Ecology & Evolution*, 15(12), pp.503–507.

Cohn, M. & Tickle, C., 1999. Developmental basis of limblessness and axial patterning in snakes. *Nature*, 399, pp.474–479.

Colosimo, P.F., 2005. Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. *Science*, 307(5717), pp.1928–1933.

Conrad, J.L., 2008. Phylogeny And Systematics Of Squamata (Reptilia) Based On Morphology. *Bulletin of the American Museum of Natural History*, 310(310), pp.1–182.

Crow, K.D. & Wagner, G.P., 2006. Proceedings of the SMBE Tri-National Young Investigators' Workshop 2005. What is the role of genome duplication in the evolution of complexity and diversity? *Molecular biology and evolution*, 23(5), pp.887–92.

Darriba, D. et al., 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*, 9(8), p.772.

Dasen, J.S., 2013. Long noncoding RNAs in development: solidifying the Lncs to Hox gene regulation. *Cell reports*, 5(1), pp.1–2.

Davis, a P. et al., 1995. Absence of radius and ulna in mice lacking hoxa-11 and hoxd-11. *Nature*, 375(6534), pp.791–795.

Davis, A. & Capecchi, M., 1994. Axial homeosis and appendicular skeleton defects in mice with a targeted disruption of hoxd-11. *Development (Cambridge, England)*, 120(8), pp.2187–2198.

Deakin, J.E., 2012. Marsupial genome sequences: providing insight into evolution and disease. *Scientifica*, 2012, pp.1–22.

Delpretti, S. et al., 2013. Multiple Enhancers Regulate Hoxd Genes and the Hotdog LncRNA during Cecum Budding. *Cell Reports*, 5(1), pp.137–150.

Denans, N., Iimura, T. & Pourquié, O., 2015. Hox genes control vertebrate body elongation by collinear Wnt repression. *eLife*, 4, pp.1–33.

Dequéant, M.-L. & Pourquié, O., 2008. Segmental patterning of the vertebrate embryonic axis. *Nature reviews. Genetics*, 9(5), pp.370–82.

Deschamps, J., 2007. Ancestral and recently recruited global control of the Hox genes in development. *Current opinion in genetics & development*, 17(5), pp.422–7.

Deschamps, J. & van Nes, J., 2005. Developmental regulation of the Hox genes during axial morphogenesis in the mouse. *Development*, 132(13), pp.2931–42.

Deutsch, U., Dressier, R. & Gruss, P., 1988. Pax 7, A Member of a Paired Box Homologous Murine Gene Family, Is Expressed in Segmented Structures during Development. *Cell*, 53, pp.617–625.

Dinger, M.E. et al., 2008. Differentiating protein-coding and noncoding RNA: challenges and ambiguities. *PLoS computational biology*, 4(11), p.e1000176.

Di-Poï, N. et al., 2010. Changes in Hox genes' structure and function during the evolution of the squamate body plan. *Nature*, 464(7285), pp.99–103.

Di-Poï, N., Montoya-Burgos, J.I. & Duboule, D., 2009. Atypical relaxation of structural constraints in Hox gene clusters of the green anole lizard. *Genome research*, 19(4), pp.602–10.

Dixon, J. et al., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485, pp.376–380.

Djebali, S. et al., 2012. Landscape of transcription in human cells. *Nature*, 489(7414), pp.101–8.

Dollé, P. et al., 1989. Coordinate expression of the murine Hox-5 complex homeobox-containig genes during limb pattern formation. *Nature*, 342, pp.767–72.

Duboule, D., 2007. The rise and fall of Hox gene clusters. *Development*, 134(14), pp.2549–2560.

Duboule, D., 1992. The vertebrate limb: a model system to study the Hox/HOM gene network during development and evolution. *BioEssays*, 14(6), pp.375–384.

Duboule, D., 1998. Vertebrate hox gene regulation: clustering and/or colinearity? *Current opinion in genetics & development*, 8(5), pp.514–518.

Duboule, D. & Morata, G., 1994. Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends in Genetics*, 10(10), pp.358–364.

Duprez, D. et al., 1996. Overexpression of BMP-2 and BMP-4 alters the size and shape of developing skeletal elements in the chick limb. *Mechanisms of Development*, 57(2), pp.145–157.

Economides, K.D., 2003. Hoxb13 is required for normal differentiation and secretory function of the ventral prostate. *Development*, 130(10), pp.2061–2069.

Fatica, A. & Bozzoni, I., 2014. Long non-coding RNAs: new players in cell differentiation and development. *Nature reviews. Genetics*, 15(1), pp.7–21.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, pp.783–791.

Felsenstein, J., 2005. Phylip (phylogeny inference package) version 3.6 (2004) Distributed by the author.

Fisher, S. et al., 2006. Conservation of RET Regulatory Function from Human to Zebrafish Without Sequence Similarity. *Science*, 312(5771), pp.276–279.

Fletcher, T., Shaw, G. & Renfree, M., 1990. Effects of bromocriptine at parturition in the tammar wallaby, Macropus eugenii. *Reproduction, Fertility and Development*, 2(1), pp.79–88.

Fromental-Ramain, C. et al., 1996. Hoxa-13 and Hoxd-13 play a crucial role in the patterning of the limb autopod. *Development*, 122(10), pp.2997–3011.

Fukuda, K. & Kikuchi, Y., 2005. Endoderm development in vertebrates: fate mapping, induction and regional specification. *Development, growth & differentiation*, 47, pp.343–55.

Gajovic, S., Kostovic-Knezevic, L. & Svajger, A., 1993. Morphological evidence for secondary formation of the tail gut in the rat embryo. *Anatomy and Embriology*, 187(3), pp.291–297.

Galceran, J. et al., 1999. Wnt-/- -like phenotype and limb deficiency in Lef1-/-Tcf1-/- mice. *Genes*, 13, pp.709–717.

Gehrke, A.R. et al., 2014. Deep conservation of wrist and digit enhancers in fish. *Proceedings of the National Academy of Sciences*, 112(3), pp.803–808.

Gompel, N. et al., 2005. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature*, 433, pp.481–487.

Gonzalez, F., Duboule, D. & Spitz, F., 2007. Transgenic analysis of Hoxd gene regulation during digit development. *Developmental biology*, 306(2), pp.847–859.

Goulding, M.D. et al., 1991. Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. *The EMBO journal*, 10(5), pp.1135–47.

Grant, P.R., 1999. *Ecology and evolution of Darwin's finches*, Princeton, NJ.: Princeton Univ. Press.

Griffith, C., Wiley, M. & Sanders, E., 1992. The vertebrate tail bud: three germ layers from one tissue. *Anatomy and Embriology*, 185, pp.101–113.

Guerreiro, I. et al., 2013. Role of a polymorphism in a Hox/Pax-responsive enhancer in the evolution of the vertebrate spine. *Proceedings of the National Academy of Sciences of the United States of America*, 110(26), pp.10682–6.

Gustavsson, P. et al., 2007. Increased expression of Grainyhead-like-3 rescues spina bifida in a folate-resistant mouse model. *Human molecular genetics*, 16(21), pp.2640–6.

Hall, B.K., 2012. Parallelism, deep homology, and evo-devo. *Evolution & development*, 14(1), pp.29–33.

Hall, L., 1987. Syndactyly in marsupials-problems and prophecies. In *Possums and Opossums*. pp. 245–255.

Head, J.J., 2015. Fossil calibration dates for molecular phylogenetic analysis of snakes 1 : *Paleontologica Electronica*, pp.1–17.

Head, J.J. & Polly, P.D., 2015. Evolution of the snake body form reveals homoplasy in amniote Hox gene function. *Nature*, 520, pp.86–89.

Higgins, D.G., Bleasby, A.J. & Fuchs, R., 1992. CLUSTAL V: improved software for multiple sequence alignment. *Computer applications in the biosciences*, 8(189-191).

Hillis, D. & Bull, J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic biology*, 42(2), pp.182–192.

Hirata, M. et al., 2009. C/EBPβ promotes transition from proliferation to hypertrophic differentiation of chondrocytes through transactivation of p57Kip2. *PLoS ONE*, 4(2), p.e4543.

Hogan, P.G. et al., 2003. Transcriptional regulation by calcium, calcineurin. *and NFAT. Genes Dev.*, 17(18), pp.2205–2232.

Hosaka, T. et al., 2004. Disruption of forkhead transcription factor (FOXO) family members in mice reveals their functional diversification. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9), pp.2975–2980.

James, R. & Kazenwadel, J., 1991. Homeobox gene expression in the intestinal epithelium of adult mice. *The Journal of biological chemistry*, 266(5), pp.3246–51.

Jeong, S., Rokas, A. & Carroll, S.B., 2006. Regulation of body pigmentation by the Abdominal-B Hox protein and its gain and loss in Drosophila evolution. *Cell*, 125(7), pp.1387–99.

Jones, W.F., 1925. *The mammals of south australia*, Adelaide, Australia: Government Printer.

Kanagae, Y. et al., 1998. Role of Rel/NF-kB transcription factors during the outgrowth of the vertebrate limb. *Nature*, 392, pp.611–614.

Karamboulas, K., Dranse, H.J. & Underhill, T.M., 2010. Regulation of BMP-dependent chondrogenesis in early limb mesenchyme by TGFbeta signals. *Journal of cell science*, 123, pp.2068–2076.

Kardon, G., Harfe, B.D. & Tabin, C.J., 2003. A Tcf4-positive mesodermal population provides a prepattern for vertebrate limb muscle patterning. *Developmental Cell*, 5(6), pp.937–944.

Kawakami, Y. et al., 2006. Wnt/beta-catenin signaling regulates vertebrate limb regeneration. *Genes & Development*, (858), pp.3232–3237.

Kel, A.E., 2003. MATCHTM: a tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Research*, 31(13), pp.3576–3579.

Kessel, M. & Gruss, P., 1991. Homeotic transformations of murine vertebrae and concomitant alteration of Hox codes induced by retinoic acid. *Cell*, 67(1), pp.89–104.

Keyte, A. & Smith, K.K., 2012. Heterochrony in somitogenesis rate in a model marsupial, Monodelphis domestica. *Evolution & development*, 14(1), pp.93–103.

Keyte, A.L. & Smith, K.K., 2010. Developmental origins of precocial forelimbs in marsupial neonates. *Development (Cambridge, England)*, 137(24), pp.4283–4294.

Kirsch, J., 1977. The comparative serology of Marsupialia, and a classification of marsupial. *Australian Journal of Zoology Supplementary Series*, 25(52), pp.1–152.

Kmita, M. et al., 2005. Early developmental arrest of mammalian limbs lacking HoxA/HoxD gene function. *Nature*, 435(7045), pp.1113–6.

Kmita, M. & Duboule, D., 2003. Organizing axes in time and space; 25 years of colinear tinkering. *Science (New York, N.Y.)*, 301(5631), pp.331–333.

Knezevic, V., De Santo, R. & Mackem, S., 1998. Continuing organizer function during chick tail development. *Development*, 125(10), pp.1791–801.

Kohlsdorf, T. et al., 2010. Data and Data Interpretation in the Study of Limb Evolution: a Reply To Galis Et Al. on the Reevolution of Digits in the Lizard Genus Bachia. *Evolution*, 64(8), pp.2477–2485.

Kohlsdorf, T. & Wagner, G., 2006. Evidence for the reversibility of digit loss: a phylogenetic study of limb evolution in Bachia (Gymnophthalmidae: Squamata). *Evolution*, 60(9), pp.1896–1912.

Kondrashov, N. et al., 2011. Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell*, 145(3), pp.383–97.

Kos, L., Chiang, C. & Mahon, K.A., 1998. Mediolateral patterning of somites: multiple axial signals, including Sonic hedgehog, regulate Nkx-3.1 expression. *Mechanisms of development*, 70(1-2), pp.25–34.

Kozhemyakina, E., Ionescu, A. & Lassar, A.B., 2014. GATA6 Is a Crucial Regulator of Shh in the Limb Bud. *PLoS Genetics*, 10(1), p.e1004072.

Krawchuk, D. et al., 2010. Twist1 activity thresholds define multiple functions in limb development. *Developmental biology*, 347(1), pp.133–46.

Kumar, B. et al., 2015. Analysis of dynamic changes in retinoid induced transcription and epigenetic profiles of murine Hox clusters in ES cells. *Genome research*, in press.

Lee, M.S.Y., 2005. Molecular evidence and marine snake origins. *Biology Letters*, 1(2), pp.227–230.

Lemons, D. & McGinnis, W., 2006. Genomic Evolution of Hox Gene Clusters. *Science*, 313(5795), pp.1918–1922.

Lemos, B. et al., 2004. Phylogenetic footprinting reveals extensive conservation of Sonic Hedgehog (SHH) regulatory elements. *Genomics*, 84(3), pp.511–523.

Lewis, E.B., 1978. A gene complex controlling segmentation in Drosophila. *Nature*, 276(5688), pp.565–570.

Li, L. et al., 2013. Targeted disruption of Hotair leads to homeotic transformation and gene derepression. *Cell reports*, 5(1), pp.3–12.

Li, L. & Chang, H.Y., 2014. Physiological roles of long noncoding RNAs: insight from knockout mice. *Trends in Cell Biology*, 24(10), pp.594–602.

Lindsay, J. et al., 2012. Unique small RNA signatures uncovered in the tammar wallaby genome. *BMC genomics*, 13, p.559.

Long, F. et al., 2001. The CREB family of activators is required for endochondral bone development. *Development*, 128(4), pp.541–550.

Lorda-Diez, C.I. et al., 2009. Transforming growth factors β coordinate cartilage and tendon differentiation in the developing limb mesenchyme. *Journal of Biological Chemistry*, 284(43), pp.29988–29996.

Losos, J.B., Hillis, D.M. & Greene, H.W., 2012. Who speaks with a forked tongue. *Science*, 338, pp.1428–1429.

Luckett, W., 1994. Suprafamilial relationships within Marsupialia: resolution and discordance from multidisciplinary data. *Journal of Mammalian Evolution*, 2, pp.225–283.

Luo, Z.-X., 2007. Transformation and diversification in early mammal evolution. *Nature*, 450, pp.1011–1019.

Mallo, M., Wellik, D.M. & Deschamps, J., 2010. Hox genes and regional patterning of the vertebrate body plan. *Developmental Biology*, 344(1), pp.7–15.

Mansfield, J.H. et al., 2004. MicroRNA-responsive "sensor" transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. *Nature genetics*, 36(10), pp.1079–1083.

Marigo, V. & Tabin, C.J., 1996. Regulation of patched by sonic hedgehog in the developing neural tube. *Proceedings of the National Academy of Sciences of the United States of America*, 93(18), pp.9346–51.

Marshall, C.R., Raff, E.C. & Raff, R.A., 1994. Dollo's law and the death and resurrection of genes. *Proceedings of the National Academy of Sciences of the United States of America*, 91(25), pp.12283–12287.

Marshall, L., 1973. Evolution of the peramelid tarsus. *Proceedings of the Royal Society of Victoria*, 85, pp.51–60.

McMahon, J.A. et al., 1998. Noggin-mediated antagonism of BMP signaling is required for growth and patterning of the neural tube and somite. *Genes & Development*, 12, pp.1438–1452.

Mercader, N. et al., 1999. Conserved regulation of proximodistal limb axis development by Meis1/Hth. *Nature*, 402(6760), pp.425–429.

Mercader, N., Tanaka, E.M. & Torres, M., 2005. Proximodistal identity during vertebrate limb regeneration is regulated by Meis homeodomain proteins. *Development*, 132(18), pp.4131–4142.

Moens, C.B. & Selleri, L., 2006. Hox cofactors in vertebrate development. *Developmental biology*, 291(2), pp.193–206.

Montavon, T. et al., 2011. A regulatory archipelago controls Hox genes transcription in digits. *Cell*, 147(5), pp.1132–1145.

Montavon, T. et al., 2008. Modeling Hox gene regulation in digits: reverse collinearity and the molecular origin of thumbness. *Genes & development*, 22(3), pp.346–359.

Montavon, T. & Soshnikova, N., 2014. Hox gene regulation and timing in embryogenesis. *Seminars in cell & developmental biology*, 34C, pp.76–84.

Mott, T. & Vieites, D.R., 2009. Molecular phylogenetics reveals extreme morphological homoplasy in Brazilian worm lizards challenging current taxonomy. *Molecular Phylogenetics and Evolution*, 51(2), pp.190–200.

Münsterberg, A. & Lassar, A., 1995. Combinatorial signals from the neural tube, floor plate and notochord induce myogenic bHLH gene expression in the somite. *Development*, 121, pp.651–660.

Nelson, C.E. et al., 1996. Analysis of Hox gene expression in the chick limb bud. *Development*, 122(5), pp.1449–1466.

Niederreither, K. et al., 2002. Embryonic retinoic acid synthesis is required for forelimb growth and anteroposterior patterning in the mouse. *Development*, 129(15), pp.3563–3574.

Nilsson, M. et al., 2004. Marsupial relationships and a timeline for marsupial radiation in South Gondwana. *Gene*, 340(2), pp.189–196.

Noordermeer, D. & Duboule, D., 2013. Chromatin Architectures and Hox Gene Collinearity. In *Current Topics in Developmental Biology*. pp. 113–148.

Nora, E. et al., 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398), pp.381–385.

O'Rourke, M.P. et al., 2002. Twist Plays an Essential Role in FGF and SHH Signal Transduction during Mouse Limb Development. *Developmental Biology*, 248(1), pp.143–156.

Ovcharenko, I. et al., 2005. Mulan: multiple-sequence local alignment and visualization for studying function and evolution. *Genome research*, 15(1), pp.184–194.

Packard, D.S.J. & Jacobson, A.G., 1976. The influence of axial structures on chick somite formation. *Developmental Biology*, 53(1), pp.36–48.

Palci, A. & Caldwell, M.W., 2007. Vestigial forelimbs and axial elongation in a 95 million-year-old non-snake squamate. *Journal of Vertebrate Paleontology*, 27(1), pp.1–7.

Pang, K.C., Frith, M.C. & Mattick, J.S., 2006. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends in genetics*, 22(1), pp.1–5.

Panopoulou, G. & Poustka, a, 2005. Timing and mechanism of ancient vertebrate genome duplications – the adventure of a hypothesis. *Trends in Genetics*, 21(10), pp.559–567.

Park, H.L. et al., 2000. Mouse Gli1 mutants are viable but have defects in SHH signaling in combination with a Gli2 mutation. *Development*, 127(8), pp.1593–1605.

Parker, H.J. et al., 2011. Ancient Pbx-Hox signatures define hundreds of vertebrate developmental enhancers. *BMC Genomics*, 12, p.637.

Parr, B.A. & Mcmahon, A.P., 1995. Dorsalizing signal Wnt-7a required for normal polarity of D–V and A–P axes of mouse limb. *Nature*, 374, pp.350–353.

Peer, Y. Van de, Maere, S. & Meyer, A., 2009. The evolutionary significance of ancient genome duplications. *Nature Reviews Genetics*, 10, pp.725–732.

Pellegrino, K. et al., 2001. A molecular perspective on the evolution of microteiid lizards (Squamata, Gymnophthalmidae), and a new classification for the family. *Biological Journal of the Linnean Society*, 74(3), pp.315–338.

Petruk, S. et al., 2006. Transcription of bxd Noncoding RNAs Promoted by Trithorax Represses Ubx in cis by Transcriptional Interference. *Cell*, 127(6), pp.1209–1221.

Pfaffl, M.W., Horgan, G.W. & Dempfle, L., 2002. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic acids research*, 30(9), p.e36.

Phillips, M.J. et al., 2006. Combined mitochondrial and nuclear DNA sequences resolve the interrelations of the major Australasian marsupial radiations. *Systematic biology*, 55(1), pp.122–137.

Ponting, C.P., Oliver, P.L. & Reik, W., 2009. Evolution and functions of long noncoding RNAs. *Cell*, 136(4), pp.629–641.

Pough, H., 1998. *Herpetology* 1st ed., Prentice Hall.

Price, M. et al., 1992. Regional expression of the homeobox gene Nkx-2.2 in the developing mammalian forebrain. *Neuron*, 8(2), pp.241–55.

Provot, S. & Schipani, E., 2005. Molecular mechanisms of endochondral bone development. *Biochemical and Biophysical Research Communications*, 328(3), pp.658–665.

Prud´homme, B. et al., 2006. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature*, 440(7087), pp.1050–3.

Pyron, R.A., Burbrink, F.T. & Wiens, J.J., 2013. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC evolutionary biology*, 13, p.93.

Rallis, C. et al., 2003. Tbx5 is required for forelimb bud formation and continued outgrowth. *Development*, 130(12), pp.2741–2751.

Ralston, G., 2013. Article HOTAIR : Can we make sense of this anti-sense RNA ? *Chemistry in New Zealand*, (October), pp.116–119.

Ranger, A.M. et al., 2000. The nuclear factor of activated T cells (NFAT) transcription factor NFATp (NFATc2) is a repressor of chondrogenesis. *The Journal of experimental medicine*, 191(1), pp.9–22.

Renfree, M.B. et al., 2011. Genome sequence of an Australian kangaroo, Macropus eugenii, provides insight into the evolution of mammalian reproduction and development. *Genome Biology*, 12(8), p.R81.

Renfree, M.B. & Tyndale-Biscoe, C.H., 1973. Intrauterine Development after Diapause in the Marsupial Macropus eugenii. *Developmental biology*, 32, pp.28–40.

Reynolds, R., Niemiller, M. & Revell, L., 2014. Toward a Tree-of-Life for the boas and pythons: Multilocus species-level phylogeny with unprecedented taxon sampling. *Molecular Phylogenetics and Evolution*, 71, pp.201–213.

Riddle, R.D. et al., 1993. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell*, 75(7), pp.1401–1416.

Rieppel, O. et al., 2003. The anatomy and relationships of Haasiophis terrasanctus, a fossil snake with well-developed hind limbs from the Mid-Cretaceous of the middle east. *Journal of Paleontology*, 77(3), pp.536–558.

Ringrose, L. & Paro, R., 2007. Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development*, 134(2), pp.223–232.

Rinn, J.L. et al., 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, 129(7), pp.1311–1323.

Roberts, D.J. et al., 1995. Sonic hedgehog is an endodermal signal inducing Bmp-4 and Hox genes during induction and regionalization of the chick hindgut. *Development*, 121(10), pp.3163–74.

Rombel, I.T. et al., 2002. ORF-FINDER: a vector for high-throughput gene identification. *Gene*, 282(1-2), pp.33–41.

Rudnicki, M. a & Jaenisch, R., 1995. The MyoD family of transcription factors and skeletal myogenesis. *BioEssays*, 17(3), pp.203–209.

Rundle, H.D. et al., 2000. Natural Selection and Parallel Speciation in Sympatric Sticklebacks. *Science*, 287, pp.306–308.

Sadleir, R.M.F.S. & Tyndale-Biscoe, C.H., 1977. Photoperiod and the Termination of Embryonis Diapause in the Marsupial eugenii. *Biology of reproduction*, 16, pp.605–608.

Sahni, M. et al., 1999. FGF signaling inhibits chondrocyte proliferation and regulates bone development through the STAT-1 pathway FGF signaling inhibits chondrocyte proliferation and regulates bone development through the STAT-1 pathway. *Genes & Development*, 13, pp.1361–1366.

Saleh, M. et al., 2000. Cell Signaling Switches HOX-PBX Complexes from Repressors to Activators of Transcription Mediated by Histone Deacetylases and Histone Acetyltransferases. *Molecular and Cellular Biology*, 20(22), pp.8623–8633.

Sanchez-Elsner, T., 2006. Noncoding RNAs of Trithorax Response Elements Recruit Drosophila Ash1 to Ultrabithorax. *Science*, 311(5764), pp.1118–1123.

Sanger, T.J. & Gibson- Brown, J.J., 2004. The developmental bases of limb reduction and body elongation in squamates. *Evolution*, 58(9), pp.2103–2106.

Santa Barbara, P. & Roberts, D., 2002. Tail gut endoderm and gut/genitourinary/tail development: a new tissue-specific role for Hoxa13. *Development*, 129(3), pp.551–61.

Saxena, A. & Carninci, P., 2011. Long non-coding RNA modifies chromatin: epigenetic silencing by long non-coding RNAs. *BioEssays*, 33(11), pp.830–839.

Scanlon, J.D. et al., 1999. The palaeocology of the primitive snake Pachyrhachis. *Historical Biology*, 13, pp.127–152.

Schmitt, S., Prestel, M. & Paro, R., 2005. Intergenic transcription through a Polycomb group response element counteracts silencing. *Genes and Development*, 19(6), pp.697–708.

Schneider, I. et al., 2011. Appendage expression driven by the Hoxd Global Control Region is an ancient gnathostome feature. *Proceedings of the National Academy of Sciences of the United States of America*, 108(31), pp.1–5.

Schneider, I. & Shubin, N.H., 2013. The origin of the tetrapod limb: from expeditions to enhancers. *Trends in genetics*, 29(7), pp.419–426.

Schneider, N.Y. et al., 2009. The olfactory system of the tammar wallaby is developed at birth and directs the neonate to its mother's pouch odours. *Reproduction*, 138(5), pp.849–857.

Schorderet, P. & Duboule, D., 2011. Structural and Functional Differences in the Long Non-Coding RNA Hotair in Mouse and Human. *PLoS Genetics*, 7(5), p.e1002071.

Scott, V., Morgan, E.A. & Stadler, H.S., 2005. Genitourinary functions of Hoxa13 and Hoxd13. *Journal of biochemistry*, 137(6), pp.671–6.

Sears, K.E., 2009. Differences in the timing of prechondrogenic limb development in mammals: the marsupial-placental dichotomy resolved. *Evolution; international journal of organic evolution*, 63(8), pp.2193–2200.

Sela-Donenfeld, D. & Kalcheim, C., 2000. Inhibition of noggin expression in the dorsal neural tube by somitogenesis: a mechanism for coordinating the timing of neural crest emigration. *Development*, 127, pp.4845–4854.

Selleri, L. et al., 2001. Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation. *Development*, 128(18), pp.3543–57.

Sensiate, L.A. et al., 2014. Dact gene expression profiles suggest a role for this gene family in integrating Wnt and TGF-β signaling pathways during chicken limb development. *Developmental Dynamics*, 243(3), pp.428–439.

Shahi, M.H., Rey, J.A. & Castresana, J.S., 2012. The sonic hedgehog-GLI1 signaling pathway in brain tumor development. *Expert Opinion on Therapeutic Targets*, 16(12), pp.1227–1238.

Shapiro, M.D. et al., 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, 428(6984), pp.717–723.

Shapiro, M.D., Bell, M.A. & Kingsley, D.M., 2006. Parallel genetic origins of pelvic reduction in vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, 103(37), pp.13753–13758.

Shapiro, M.D., Hanken, J. & Rosenthal, N., 2003. Developmental basis of evolutionary digit loss in the Australian lizard Hemiergis. *Journal of experimental zoology. Part B*, 297(1), pp.48–56.

Shaw, G. & Renfree, M.B., 2006. Parturition and perfect prematurity- birth in marsupials. *Australian Journal of Zoology*, 54, pp.139–149.

Sheth, R. et al., 2012. Hox genes regulate digit patterning by controlling the wavelength of a Turing-type mechanism. *Science (New York, N.Y.)*, 338(6113), pp.1476–80.

Singarete, M. et al., 2015. Molecular evolution of HoxA13 and the multiple origins of limbless morphologies in amphibians and reptiles. *Genetics and Molecular Biology*, 38(3), p.ISSN 1678–4685.

Sivanantharajah, L. & Percival-Smith, A., 2015. Differential pleiotropy and HOX functional organization. *Developmental Biology*, 398(1), pp.1–10.

Small, K.M. & Potter, S.S., 1993. Homeotic transformations and limb defects in Hox A11 mutant mice. *Genes & Development*, 7(12), pp.2318–2328.

Soshnikova, N., 2014. Hox genes regulation in vertebrates. *Developmental dynamics*, 243(1), pp.49–58.

Soshnikova, N. & Duboule, D., 2009. Epigenetic temporal control of mouse Hox genes in vivo. *Science*, 324(5932), pp.1320–3.

Spitz, F., 2010. Control of Vertebrate Hox Clusters by Remote and Global Cis -Acting Regulatory Sequences. In J. S. Deutsch, ed. *Hox Genes: Studies from the 20th to the 21st Century*. Heidelberg: Landes Bioscience and Springer+Business Media, pp. 63–78.

Spitz, F. et al., 2005. Inversion-induced disruption of the Hoxd cluster leads to the partition of regulatory landscapes. *Nature genetics*, 37(8), pp.889–93.

Spitz, F. et al., 2001. Large scale transgenic and cluster deletion analysis of the HoxD complex separate an ancestral regulatory module from evolutionary innovations. *Genes & development*, 15(17), pp.2209–2214.

Spitz, F. & Duboule, D., 2008. Global control regions and regulatory landscapes in vertebrate development and evolution. In *Advances in genetics*. pp. 176–205.

Spitz, F., Gonzalez, F. & Duboule, D., 2003. A Global Control Region Defines a Chromosomal Regulatory Landscape Containing the HoxD Cluster. *Cell*, 113, pp.405–417.

Stanfel, M.N. et al., 2005. Regulation of organ development by the NKX-homeodomain factors: an NKX code. *Cellular and molecular biology*, 51, pp.OL785–799.

Stricker, S. et al., 2002. Role of Runx genes in chondrocyte differentiation. *Developmental biology*, 245(1), pp.95–108.

Sucena, E. et al., 2003. Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature*, 424(6951), pp.935–938.

Suryamohan, K. & Halfon, M.S., 2015. Identifying transcriptional cis - regulatory modules in animal genomes. *WIREs Developmental Biology*, 4(4), pp.59–84.

Szalay, F., 1994. *Evolutionary history of the marsupials and an analysis of osteological characters*, New York: Cambridge University Press.

Takagi, T. et al., 1998. DeltaEF1, a zinc finger and homeodomain transcription factor, is required for skeleton patterning in multiple lineages. *Development*, 125(1), pp.21–31.

Takahashi, Y. et al., 2004. Expression profiles of 39 HOX genes in normal human adult organs and anaplastic thyroid cancer cell lines by quantitative real-time RT-PCR system. *Experimental Cell Research*, 293(1), pp.144–153.

Tarchini, B. & Duboule, D., 2006. Control of Hoxd genes' collinearity during early limb development. *Developmental cell*, 10(1), pp.93–103.

Taylor, H.S., Vanden Heuvel, G.B. & Igarashi, P., 1997. A conserved Hox axis in the mouse and human female reproductive system: late establishment and persistent adult expression of the Hoxa cluster genes. *Biology of reproduction*, 57(6), pp.1338–1345.

Tchernov, E. et al., 2000. A fossil snake with limbs. *Science*, 287, pp.2010–2012.

The UniProt Consortium, 2008. The universal protein resource (UniProt). *Nucleic acids research*, p.36; D190–D195.

Thompson, J.D., Higgins, D.G. & Gibson., T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*, 22(22), pp.4673–4680.

Thornton, B. & Basu, C., 2011. Real-time PCR (qPCR) primer design using free online software. *Biochemistry and Molecular Biology Education*, 39(2), pp.145–154.

Tickle, C., 2006. Making digit patterns in the vertebrate limb. *Nature reviews. Molecular cell biology*, 7(1), pp.45–53.

Trapnell, C. et al., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*, 7(3), pp.562–78.

Trapnell, C. et al., 2010. Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nature Biotechnology*, 28(5), pp.511–515.

Trapnell, C., Pachter, L. & Salzberg, S.L., 2009. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9), pp.1105–1111.

Tschopp, P. et al., 2014. A relative shift in cloacal location repositions external genitalia in amniote evolution. *Nature*, 516, pp.391–394.

Tschopp, P. et al., 2011. Reshuffling genomic landscapes to study the regulatory evolution of Hox gene clusters. *Proceedings of the National Academy of Sciences of the United States of America*, 108(26), pp.10632–10637.

Tschopp, P. & Duboule, D., 2011. A genetic approach to the transcriptional regulation of Hox gene clusters. *Annual review of genetics*, 45, pp.145–166.

Uetz, P., 2007. *The EMBL reptile database*, Heidelberg: European Molecular Biology Laboratory.

Ulitsky, I. et al., 2011. Conserved Function of lincRNAs in Vertebrate Embryonic Development despite Rapid Sequence Evolution. *Cell*, 147(7), pp.1537–1550.

Untergasser, A. et al., 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research*, 35, pp.71–74.

Vargas, A.O. et al., 2008. The evolution of HoxD-11 expression in the bird wing: insights from Alligator mississippiensis. *PloS one*, 3(10), p.e3325.

Van de Ven, C. et al., 2011. Concerted involvement of Cdx/Hox genes and Wnt signaling in morphogenesis of the caudal neural tube and cloacal derivatives from the posterior growth zone. *Development*, 138(16), pp.3451–3462.

Vidal, N. et al., 2010. Blindsnake evolutionary tree reveals long history on Gondwana. *Biology letters*, 6(4), pp.558–561.

Vidal, N. & Hedges, S.B., 2004. Molecular evidence for a terrestrial origin of snakes. *Proceedings of the Royal Society B: Biological Sciences*, 271, pp.S226–S229.

Vidal, N. & Hedges, S.B., 2009. The molecular evolutionary tree of lizards, snakes, and amphisbaenians. *Evolution*, 332, pp.129–139.

Vidal, N. & Hedges, S.B., 2005. The phylogeny of squamate reptiles (lizards, snakes, and amphisbaenians) inferred from nine nuclear protein-coding genes. *Comptes rendus biologies*, 328(10-11), pp.1000–1008.

Villarejo-Balcells, B. et al., 2011. Expression pattern of the FoxO1 gene during mouse embryonic development. *Gene Expression Patterns*, 11(5-6), pp.299–308.

Villavicencio, E.H. et al., 2002. Cooperative E-box regulation of human GLI1 by TWIST and USF. *Genesis*, 32(4), pp.247–258.

Wake, D.B., Wake, M.H. & Specht, C.D., 2011. Homoplasy: from detecting pattern to determining process and mechanism of evolution. *Science*, 331(6020), pp.1032–5.

Wang, J. et al., 2004. Mouse transcriptome: Neutral evolution of "non-coding" complementary DNAs. *Nature*, 431.

Wang, K.C. et al., 2011. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*, 472(7341), pp.120–124.

Wang, L. et al., 2013. CPAT: Coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Research*, 41(6), pp.1–7.

Wang, X. & Chamberlin, H.M., 2002. Multiple regulatory changes contribute to the evolution of the Caenorhabditis lin-48 ovo gene. *Genes & development*, 16(18), pp.2345–9.

Warot, X. et al., 1997. Gene dosage-dependent effects of the Hoxa-13 and Hoxd-13 mutations on morphogenesis of the terminal parts of the digestive and urogenital tracts. *Development*, 124(23), pp.4781–91.

Warren, R. et al., 1994. Evolution of homeotic gene regulation and function in flies and butterflies. *Nature*, 372, pp.458–461.

Weisbecker, V. & Nilsson, M., 2008. Integration, heterochrony, and adaptation in pedal digits of syndactylous marsupials. *BMC evolutionary biology*, 8(160), pp.1–14.

Wellik, D.M. & Capecchi, M.R., 2003. Hox10 and Hox11 Genes Are Required to Globally Pattern the Mammalian Skeleton. *Science*, 301, pp.363–367.

Welscher, P. te et al., 2002. Progression of vertebrate limb development through SHH-mediated counteraction of GLI3. *Science*, 298(5594), pp.827–830.

Wiens, J.J., Brandley, M.C. & Reeder, T.W., 2006. Why does a trait evolve multiple times within a clade? Repeated evolution of snakelike body form in squamate reptiles. *Evolution*, 60(1), pp.123–41.

Wiens, J.J. & Slingluff, J.L., 2001. How lizards turn into snakes : A phylogenetic analyses of body-form evolution in anguid lizards. *Evolution*, 55(11), pp.2303–2318.

Williams, J.G., 2000. STAT signalling in cell proliferation and in development. *Current Opinion in Genetics and Development*, 10(5), pp.503–507.

Wolf, C. et al., 1991. The M-twist gene of Mus is expressed in subsets of mesodermal cells and is closely related to the Xenopus X-twi and the Drosophila twist genes. *Developmental biology*, 143(2), pp.363–73.

Woltering, J.M. et al., 2009. Axial patterning in snakes and caecilians: evidence for an alternative interpretation of the Hox code. *Developmental biology*, 332(1), pp.82–89.

Woltering, J.M. et al., 2014. Conservation and divergence of regulatory strategies at Hox Loci and the origin of tetrapod digits. *PLoS biology*, 12(1), p.e1001773.

Woltering, J.M., 2012. From lizard to snake; behind the evolution of an extreme body plan. *Current genomics*, 13(4), pp.289–99.

Xue, S. et al., 2015. RNA regulons in Hox 5' UTRs confer ribosome specificity to gene regulation. *Nature*, 517(7532), pp.33–8.

Yasutake, J., Inohaya, K. & Kudo, A., 2004. Twist functions in vertebral column formation in medaka, Oryzias latipes. *Mechanisms of development*, 121(7-8), pp.883–94.

Yekta, S., Tabin, C.J. & Bartel, D.P., 2008. MicroRNAs in the Hox network: an apparent link to posterior prevalence. *Nature reviews. Genetics*, 9(10), pp.789–96.

Yokouchi, Y., Sakiyama, J. & Kuroiwa, A., 1995. Coordinated expression of Abd-B subfamily genes of the HoxA cluster in the developing digestive tract of chick embryo. *Developmental biology*, 169, pp.76–89.

Yoshida, C. a. et al., 2004. Runx2 and Runx3 are essential for chondrocyte maturation, and Runx2 regulates limb growth through induction of Indian hedgehog. *Genes and Development*, 18(8), pp.952–963.

Young, T. et al., 2009. Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. *Developmental cell*, 17(4), pp.516–26.

Yu, H. et al., 2012. Evolution of coding and non-coding genes in HOX clusters of a marsupial. *BMC Genomics*, 13(251), pp.1–15.

Zakany, J. & Duboule, D., 2007. The role of Hox genes during vertebrate limb development. *Current opinion in genetics & development*, 17(4), pp.359–366.

Zákány, J. & Duboule, D., 1999. Hox genes in digit development and evolution. *Cell and tissue research*, 296(1), pp.19–25.

Zákány, J., Kmita, M. & Duboule, D., 2004. A dual role for Hox genes in limb anterior-posterior asymmetry. *Science (New York, N.Y.)*, 304(5677), pp.1669–1672.

Zinzen, R.P. et al., 2006. Evolution of the ventral midline in insect embryos. *Developmental cell*, 11(6), pp.895–902.

Zuniga, A. et al., 2004. Mouse limb deformity mutations disrupt a global control region within the large regulatory landscape required for Gremlin expression. *Genes & development*, 18(13), pp.1553–64.

Zuniga, A. et al., 2002. Mouse Twist is required for fibroblast growth factor-mediated epithelial-mesenchymal signalling and cell survival during limb morphogenesis. *Mechanisms of development*, 114(1-2), pp.51–9.