DOCTOR OF PHILOSOPHY

The Genetic Basis of Venom Variation in the Genus Echis: Causes, Correlates and Consequences

Casewell, Nicholas

Award date:
2010

Awarding institution:
Bangor University

PRIFYSGOL
BANGOR
UNIVERSITY

1884

The Genetic Basis of Venom Variation in the Genus *Echis*: Causes, Correlates and Consequences

Nicholas Robert Casewell

Supervisors: Wolfgang Wüster and Robert Harrison

Thesis to be submitted for the degree of
Doctor of Philosophy
at Bangor University

Molecular Ecology and Evolution of Reptiles Unit
School of Biological Sciences
Bangor University

March 31, 2010

SIGNED DECLARATION FORM

NOT TO BE INCLUDED IN THE DIGITISED THESIS

A member of the saw-scaled vipers - genus *Echis* (photograph by Wolfgang Wüster)

ABSTRACT

Variation in venom components is inherent to multiple taxonomical levels of the Serpentes and can impact significantly upon the symptomatology of envenoming and the efficacy of antivenoms. Snake venom composition is thought to be subject to strong natural selection as a result of adaptations to specific diets, although no direct link at the molecular level has elucidated the evolutionary adaptations responsible for driving the optimisation of venom components to specific prey items. Venom gland cDNA libraries were constructed for three species of the genus *Echis* (*E. pyramidum leakeyi*, *E. coloratus* and *E. carinatus sochureki*) to complement the existing *E. ocellatus* transcriptome. Generated expressed sequence tags were clustered with a modified CLOBB algorithm, which was demonstrated to confer increases in the integrity of cluster formation and membership over the standard CLOBB2 algorithm. Comparative analyses of multiple *Echis* venom gland transcriptomes revealed the presence of snake venom metalloproteinases (SVMP), C-type lectins, phopholipases A_2, serine proteases (SP), L-amino oxidases and growth factors throughout the genus. Putative novel venom proteins exhibiting similarity to lysosomal acid lipase/cholesteryl ester hydrolase and the metallopeptidases dipeptidyl peptidase III and neprilysin were also identified in the venom glands of individual species. Phylogenetic and gene tree parsimony analyses provide the first evidence of the genomic basis of snake venom adaptations as a response to alterations in diet, with SVMP and SP toxin families exhibiting diet-associated gene events that correlate strongly with a dietary shift to vertebrate feeding in *E. coloratus*. The diversification and retention of these coagulopathic and haemorrhagic toxins in *E. coloratus* correlates with significant differences in venom function in the form of *in vivo* haemorrhage, providing genetic and functional evidence of coevolution between diet and venom components. Selective evolutionary pressures were also determined to be capable of confounding the derivation of species relationships from toxin data, suggesting venom components should not be used as primary species identifiers. Finally, the *E. ocellatus* antivenom EchiTabG® was demonstrated to effectively neutralise the venom of African members of the genus *Echis* in spite of considerable intra-generic variation in venom components. These results strongly advocate the geographical expansion of EchiTabG® to treat *Echis* envenomations throughout the African continent.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisor Wolfgang Wüster for all of his help and guidance over the past three years. His enthusiasm, assistance and engagement of my ideas have been integral for my progression as a scientist and the resulting thesis you are reading today; for this I will always be grateful. I must also thank Wolfgang for the design of the project, which has provided a fantastic framework for the exploration of snake venom evolution and other aspects of biology that have fascinated me.

I would also like to thank my co-supervisors at the Alistair Reid Venom Research Unit at the Liverpool School of Tropical Medicine. Rob Harrison has provided invaluable advice on all aspects of this project and has had a significant impact upon my personal development as a scientist. I thank him also for encouraging me to delve into the immunological aspect of venoms; the skills I have developed subsequently are the result of his guidance. Simon Wagstaff has been hugely supportive throughout the past three years and has successfully guided me through the murky world of cDNA and sequence bioinformatics. He has always been there to provide guidance, especially with the (occasionally stupid) questions I need answering.

I must also thank MicroPharm Ltd and their staff for their sponsorship as a NERC PhD CASE partner. I thank John Landon and Ibrahim Al-Abdulla particularly for taking their time to show me how antivenoms are made, as well as raising antibodies and providing antivenom for my own research studies.

General thanks go to Paul Rowley for his expert herpetological assistance, Damien Egan and Paul Vercammen (Breeding Centre for Endangered Arabian Wildlife, United Arab Emirates) for providing specimens of *E. c. sochureki*, Jean-François Trape and Youssouph Mané (Institut de Recherche pour le Développement, Dakar) for fieldwork assistance, Ann Hedley and Mark Blaxter (NERC Molecular Genetics Facility, University of Edinburgh) for providing sequencing and bioinformatic advice regarding the PartiGene pipeline, Tim Booth, Bela Tiwari and Jorge Soares (NERC Environmental Bioinformatics Centre, Oxford) for general bioinformatic advice, Michael Berenbrink (School of Biological Sciences, University of Liverpool)

for assistance with SigmaPlot and Wayne Maddison (University of British Columbia, Canada) for help with Mesquite.

Big thanks go to Cath, Axel, Yvonne, Darren, Rachel and Camila for being great office mates, making me laugh and helping me through the more stressful times!

My family have always encouraged me to follow a career path that I am passionate about and they have provided every support possible during the past twenty-five years. I am eternally grateful for everything they have done for me and without them I would not be in this position today. I look forward to many more conversations where I attempt to explain what I have been doing for the past three years! Finally, I thank my wife Lisa, who has been by my side supporting my decisions for a lot longer than the past three years. I am grateful for the sacrifices she made in her career so that I may undertake this PhD and for always being there through the good and bad times. I can only apologise to her for my numerous stresses about work-related issues, but she has always responded with humour and a positive attitude that makes things better. Thanks for the tea and biscuits and thank you so much for everything else.

PREFACE

Chapter 1 describes introductory information on the nature of venoms, their evolution and the genus *Echis*, whilst Chapter 2 details the methods utilised to generate DNA sequence information from venom glands. The experimental chapters (Chapters 3-7) are presented in the form of publication papers and therefore contain detailed methodological sections outlining the specific methods utilised for each chapter of experimental work. All experimental work has been undertaken by myself except where otherwise specified at the end of an experimental chapter. Chapter 8 discusses and summarises the conclusions drawn from the experimental chapters.

Chapter 3 outlines a comparative study between differing bioinformatic algorithms that cluster expressed sequence tags (ESTs). The results strongly support the use of a modified CLOBB algorithm as the optimal method for clustering snake venom gland derived ESTs.

Chapter 4 presents comparative results of four sequenced *Echis* venom gland cDNA libraries in the form of transcriptomic profiles. Substantial intra-generic variation in the representation of toxin components was observed and three novel putative venom components are described. This chapter has been published in the journal BMC Genomics – the published manuscript is presented in Appendix VI.

Chapter 5 investigates the selective influence of diet on the evolution of venom components in the genus *Echis*. Gene tree parsimony analyses provide evidence of multiple toxin families exhibiting diet-associated gene events that correlate with a reversion to vertebrate-feeding. This chapter has been submitted for publication to the journal Proceedings of the National Academy of Sciences, USA and is pending reviewer and editorial decisions.

Chapter 6 assesses the value of venom-derived toxin family gene trees as species tree predictors. Gene tree parsimony of multiple toxin trees largely failed to infer species trees congruent with each other or robustly supported phylogenies. This chapter has been invited for resubmission for publication in the journal Molecular Biology and Evolution pending further reviewer and editorial decisions.

Chapter 7 describes immunological comparisons of *Echis* venoms with homologous and non-homologous antivenoms, alongside assessments of their neutralisation with the existing antivenom EchiTabG®. Successful non-homologous venom neutralisation of African *Echis* species highlights the potential for the geographical expansion of EchiTabG®. This chapter has been submitted for publication to the journal PLoS Neglected Tropical Diseases and is pending reviewer and editorial decisions.

CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 The origin of venom

Venom has evolved a number of times throughout the animal kingdom, including in the Gastropoda, Cephalopoda, Hymenoptera, Arachnida, Mammalia and Reptilia (Olivera *et al.* 1990; de Oliveira *et al.* 2006; Escoubas *et al.* 2006; Fry *et al.* 2006, 2009; Whittington *et al.* 2009). Reptilian venoms are a complex mixture of components which have a diverse array of actions on both natural prey items and humans (Chippaux *et al.* 1991). The components themselves are a mixture of proteins, peptides, carbohydrates, lipids, metal ions and organic compounds, with proteins and peptides accounting for the vast majority (Aird, 2002). These proteins and peptides show a high level of biological activity (Aird, 2002); their primary function is to kill or immobilize prey and/or to assist in the digestion of prey items (Karlsson, 1979; Hayes, 1991; Chippaux *et al.* 1991), rather than for use as a defensive mechanism (Li *et al.* 2005).

The origin of venom in reptiles appears to have arisen at a single point at the base of the iguanians approximately 200 million years ago (Fry *et al.* 2006). Snakes, iguanians and anguiomorph lizards share a number of basal toxin families (enzymatic and non-enzymatic toxins) that have been recruited into the venom gland prior to the separation of these lineages and form a clade termed the Toxicofera (Fry *et al.* 2006, 2008) (Figure 1.1). The presence of venom secreting glands corresponds to the presence of toxin families throughout these lineages. The iguanians demonstrate an ancestral form of venom secreting glands with presence of both maxillary (upper) and mandibular (lower) glands, whilst the more derived venom systems found within the anguimorphs and snakes are characterised by the loss of either maxillary or mandibular glands (Fry *et al.* 2006). Despite the atrophy of a venom secreting gland within these lineages, the venom delivery system has increased in efficiency; the anguimorphs produce venom from a gland in the lower jaw where ducts lead onto grooved teeth along the length of the mandible, whilst the

snakes produce venom in specialized glands in the upper jaw and use a mixture of delivery mechanisms including highly specialized fangs (Fry *et al.* 2006, 2008; Vonk *et al.* 2008). Furthermore, the complexity of a venom gland appears to be directly linked to the quantity of additional toxin recruitment events, providing a correlation between gland complexity and venom toxicity (Fry *et al.* 2006).

Despite evidence supporting the presence of venom secretion as a basal characteristic in the Serpentes (Fry *et al.* 2006, 2008), it is thought that only approximately 450 species are medically relevant to humans (Jackson, 2003). These medically relevant species are all members of the advanced snakes (superfamily Caenophidia) and include three monophyletic clades of independently evolved front-fanged snakes (Atractaspididae, Elapidae and Viperidae) (Vidal *et al.* 2007; Fry *et al.* 2008; Vonk *et al.* 2008). The evolution of a front-fanged delivery system is strongly associated with the recruitment of new venom toxin types or substantial diversification in existing toxin types which are presumably responsible for the increases in venom toxicity towards humans (Fry *et al.* 2008). Although the majority of medically important species contain hollow, high pressure, front-fanged delivery systems, a small number of non-front fanged snakes are also considered to be dangerous to humans (Harris and Goonetilleke, 2004; Fry *et al.* 2008).

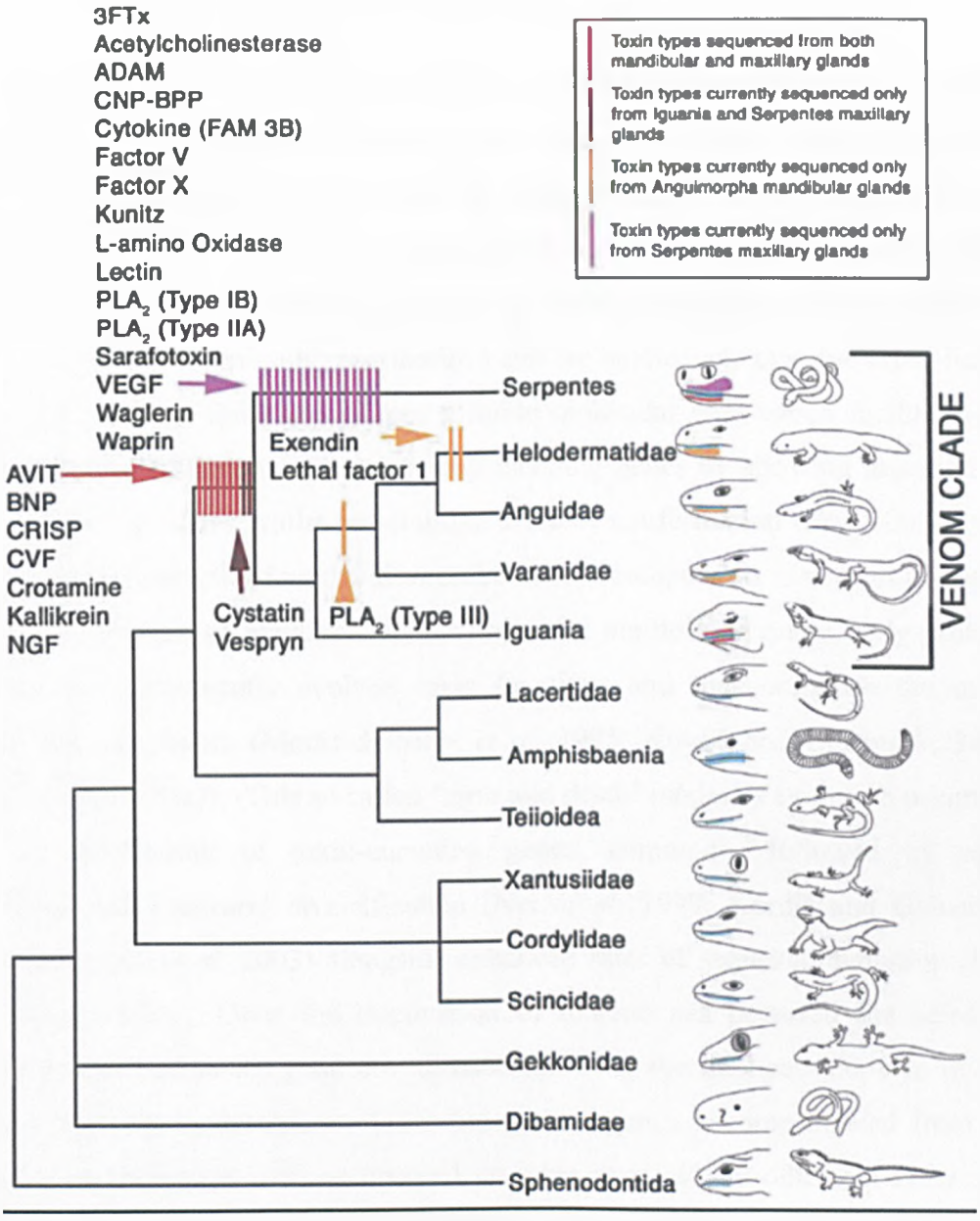

Figure 1.1. Relative glandular development and timing of toxin recruitment events mapped over the squamate reptile phylogeny (from Fry *et al.* 2006). Mucus-secreting glands are coloured blue; the ancestral form of the protein-secreting gland (serial, lobular and non-compound) red; the complex, derived form of the upper snake-venom gland (compound, encapsulated and with a lumen) fuchsia, and the complex, derived form of the anguimorph mandibular venom gland (compound, encapsulated and with a lumen) orange. Toxin family key: 3FTx, three-finger toxins; ADAM, a disintegrin and metalloproteinase; CNP-BPP, C-type natriuretic peptide-bradykinin-potentiating peptide; CVF, cobra venom factor; NGF, nerve growth factor; VEGF, vascular endothelial growth factor.

1.2 Recruitment and evolution of venom components

The majority of toxin families found in the venom of the advanced snakes are closely related to secretory proteins, implying their recruitment from body tissues (Fry, 2005). There appears to be no specific location that toxin-encoding genes are recruited from, with evidence of recruitment from tissues as diverse as the brain, liver and the salivary glands (Fry, 2005). However, the proteins that are recruited typically originate from multigene families and are extensively cysteine cross-linked (Fry, 2005). Cross-linking promotes a stable molecular core which facilitates the functional diversification of these protein-encoding genes by allowing mutations to non-structural residues whilst maintaining a stable conformation (Fry, 2005). The large degree of diversity found within snake venom components is a result of a gene duplicating method of evolution demonstrated by multiple recruited body proteins which have subsequently evolved toxic functions and thus comprise the major constituents of venom (Moura-da-Silva *et al.* 1995; Kordiš and Gubenšek, 2000; Župunski *et al.* 2003). This so called 'birth and death' model of evolution occurs by frequent duplication of toxin-encoding genes, commonly followed by rapid functional and structural diversification (Nei *et al.* 1997; Kordiš and Gubenšek, 2000; Župunski *et al.* 2003) alongside enhanced rates of sequence evolution (Kini and Chan, 1999). Once the duplication of a gene has occurred the selection pressures attached to the gene are released allowing the duplicate copy to evolve without functional constraints. Over time, some genes become deleted from the genome via processes such as unequal crossing-over, whilst other genes become redundant and degenerate into pseudogenes (Li *et al.* 2005). However, some genes diversify into new functional proteins and it is common within venom to find a range of toxins with different actions that are encoded by multigene families (Fry *et al.* 2003a; Fox and Serrano, 2005; Lynch, 2007). The existence in venom of functionally diverse isoforms of the same protein family reflects accelerated Darwinian evolution (e.g. Moura da Silva *et al.* 1996; Ohno *et al.* 2003). Furthermore, the high rate of non-synonymous to synonymous substitutions described from a number of toxin multigene families, indicates that natural selection is acting to diversify coding sequences at an accelerated rate (e.g. Nakashima *et al.* 1995; Kordiš and Gubenšek, 2000; Lynch, 2007). Consequently, toxin functions within venom are thought to be progressive leading to neofunctionalizations within

specific toxin families (Lynch, 2007), as demonstrated by the phospholipase A_2 type II myotoxins, where a novel non-hydrolytic mechanism to induce membrane damage has arisen following an amino acid substitution of aspartate to lysine at residue 49 (Diaz *et al.* 1991; Rufini *et al.* 1992; van den Bergh *et al.* 1998).

1.3 Toxic components of snake venom

The numerous highly biologically active protein and peptide components of snake venoms (Aird, 2002) have traditionally been classified into two groups; enzymes which are limited by the time of the enzymatic reaction to cause a toxicological effect, and toxins which typically exhibit a dose-dependent mechanism of action (Chippaux, 2006). More recently it has become typical for toxinologists to describe all of the pathological components present in venom as toxins. A number of toxins present in snake venoms are basal to the Toxicofera (Figure 1.1), whilst a substantial number have been recruited into the venom gland at the base of the advanced snake radiation (Fry *et al.* 2006, 2008) (Figure 1.2). Following the divergence of lineages within the Caenophidia, further recruitments of toxin families have occurred within each of the medically important lineages, with the recruitment of novel toxins occurring at least once in each of the lineages which have independently evolved a front-fanged venom delivery system (Vidal *et al.* 2007; Fry *et al.* 2008) (Figure 1.2). A large number of toxin types have been characterised from snake venoms, including three finger toxins, dendrotoxins, lectins, phospholipases, metalloproteinases and serine proteases (e.g. Harvey and Karlsson, 1980; Ogilve and Gartner, 1984; Machado *et al.* 1993, Serrano *et al.* 1993, Gutiérrez *et al.* 1995, Harvey, 2001; Fry *et al.* 2003b; Braga *et al.* 2006). For the majority of toxin types, the basal bioactivities of the toxins have been determined (Table 1.1), thus implying a potential role in envenoming. It is typical that the venom from any one species will contain multiple toxin isoforms that represent multiple toxin families (e.g. Junqueira-de-Azevedo and Ho, 2002; Juárez *et al.* 2004; Bazaa *et al.* 2005; Wagstaff and Harrison, 2006; Calvete *et al.* 2007; Wagstaff *et al.* 2009); this multitude of venom components exhibiting differing bioactivities presents a complex picture when attempting to determine symptomatology in both humans and prey. Venom components function primarily to immobilise and kill prey through a complex

network of disparate molecular targets but synergistic pathways (Chippaux, 1991). For example, within a number of viper venoms the presence of multiple metalloproteinases, serine proteases and C-type lectins work in combination to consume blood clotting factors, increase the permeability of the vascular vessels and inhibit platelet aggregation leading to a compromised vascular system; characterized by presenting haemorrhage and coagulopathy (Morita, 2005; Kini, 2006; Yamazaki and Morita, 2007). Moreover, within specific toxin families, there can be substantial diversification as a result of gene duplications, with a number of different gene products being expressed in the venom for each toxin family, with each gene product likely acting upon different molecular targets and causing a myriad of effects (Fry *et al.* 2003a; Harrison *et al.* 2003; Wagstaff and Harrison, 2006; Wagstaff *et al.* 2009).

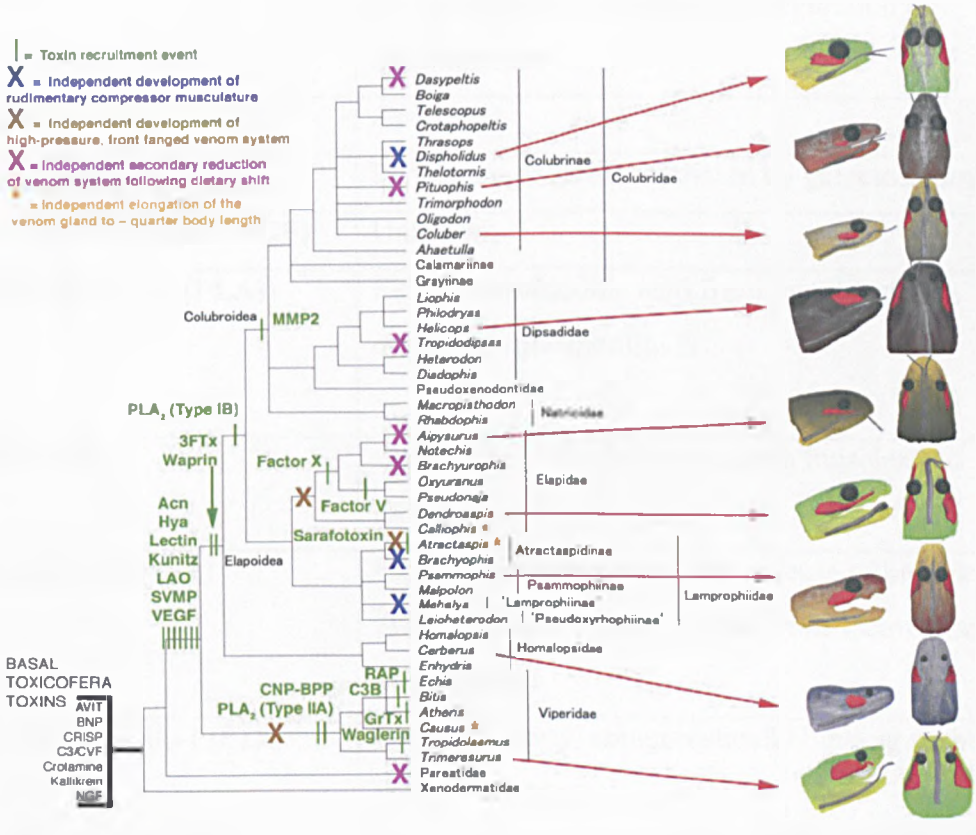

Figure 1.2. Cladogram of the evolutionary relationships of advanced snakes showing the relative timing of toxin recruitment events and derivations of the venom system (from Fry *et al.* 2008). MRI images are shown for representatives. Toxin key: Acn, Acetylcholine esterase; LAO, L-amino oxidase; C3B, FAMC3B cytokine; CNP-BPP, C-type natriuretic peptide-bradykinin-potentiating peptide; GrTx, glycine-rich toxin; Hya, hyaluronidase; RAP, renin-like aspartic protease; VEGF, vascular endothelial growth factor.

Toxin type	Basal toxic activities
Cysteine-rich secretory proteins (CRISP)	Paralysis of peripheral smooth muscle and induction of hypothermia
Disintegrin/metalloproteinase (ADAM)	Tissue necrosis, fibrinolytic and haemorrhagic activity, inhibition of platelet aggregation
Factor V	Combines with toxic form of factor X to covert prothrombin to thrombin
Factor X	Conversion of prothrombin to thrombin in the presence of factor V, calcium and phospholipids
Kallikrein	Increase of vascular permeability and production of hypotension in addition to stimulation of inflammation
L-amino oxidase (LAO)	Apoptosis
C-type lectins (CTL)	Platelet aggregation mediated by galactose binding
Nerve growth factor (NGF)	Unknown
Phospholipase A_2 (PLA_2)	Release arachidonic acid from the plasma membrane phospholipids
Prokinecticin 2	Constriction of intestinal smooth muscles and induction of hyperalgesia
Serine protease (SP)	Fibrin(ogen)olytic activities, release of bradykinin, inducing hypotension, activation of factor V and plasminogen
Three finger toxins (3FTx)	α-neurotoxicity, antagonistically binding to the nicotinic acetycholine receptor
Vascular endothelial growth factor (VEGF)	Increase in the permeability of the vascular bed and binding of heparin, results in hypotension and shock

Table 1.1. Variation in the basal bioactivities of major toxin types. Those commonly found in Viperidae venoms are highlighted in bold. Adapted from Fry (2005).

More recently, complementary DNA (cDNA) methods have been implemented to assess the whole venom gland composition of a species, rather than to focus on the isolation of specific toxins (Junqueira-de-Azevedo and Ho, 2002). This transcriptomic technique has proven to be particularly powerful as it generates an overview of the diversity and expression levels of toxin family secretion in the venom gland, whilst also allowing the discovery of novel toxin families (e.g., Junqueira-de-Azevedo and Ho, 2002, Fry *et al.* 2006, 2008). This method has been implemented on venom glands from a number of lineages within the Toxicofera, but the most comprehensive sequenced venom gland cDNA libraries to date come from the Viperidae (Junqueira-de-Azevedo and Ho, 2002; Francischetti *et al.* 2004; Kashima *et al.* 2004; Cidade *et al.* 2006; Junqueira-de-Azevedo *et al.* 2006; Wagstaff and Harrison, 2006; Zhang *et al.* 2006; Pahari *et al.* 2007; Casewell *et al.* 2009; Neiva *et al.* 2009). Although there are considerable differences between the relative expression levels of the toxin families found within different Viperidae venom gland transcriptomes, the presence of the toxin families themselves is typically consistent, with representation from: snake venom metalloproteinases (SVMP), phospholipases A_2 (PLA_2), serine proteases (SP), disintegrins (DIS), C-type lectins (CTL), L-amino oxidases (LAO), cysteine rich secretory proteins (CRISP) and growth factors (GF) (Junqueira-de-Azevedo and Ho, 2002; Francischetti *et al.* 2004; Kashima *et al.* 2004; Cidade *et al.* 2006; Junqueira-de-Azevedo *et al.* 2006; Wagstaff and Harrison, 2006; Zhang *et al.* 2006; Pahari *et al.* 2007; Casewell *et al.* 2009; Neiva *et al.* 2009). Consequently, as a result of the vast sequence data generated from these studies, novel putative Viperidae toxins have been described, including a multi-Kunitz protease inhibitor from *Bitis arietans* and renin-like aspartic proteases from *Echis ocellatus* (Francischetti *et al.* 2004; Wagstaff and Harrison, 2006). In addition to venom gland transcriptomes, a number of proteomic surveys have been undertaken to analyse the toxin composition of crude venom (e.g. Juárez *et al.* 2004; Bazaa *et al.* 2005; Sanz *et al.* 2006; Calvete *et al.* 2007; Angulo *et al.* 2008). These studies have confirmed the presence of the majority of toxin families identified from venom gland transcriptomes, whilst good accordance of toxin composition has been determined between the techniques (Wagstaff *et al.* 2009). However, the wealth of DNA sequence data produced by a transcriptomic approach is particularly advantageous for evolutionary assessments of variation in toxin components.

1.4 Venom variation

Transcriptomic and proteomic overviews reveal the presence of numerous toxin components in the venom and venom gland of a particular species (e.g. Wagstaff and Harrison, 2006; Wagstaff *et al.* 2009), highlighting the complexity of snake venom composition. Despite the common origin of many venom components at the base of the Toxicofera, the divergence of the Caenophidia has resulted in the separate evolution of a number of venom components (Fry *et al.* 2006, 2008) (Figure 1.2) and the consequential observation of venom variation between species (reviewed in Chippaux, 1991). Notably, variation in venom has been observed at all taxonomic levels: inter-family, inter-genus, inter-species and intra-species (Chippaux, 1991). Early studies by Lamb (1902, 1904) revealed variation in snake venom by testing the cross-reactivity of venoms from a number of species against antivenom raised against *Naja naja* venom (Elapidae); cross reactivity was only found in one species. Subsequently, the examination of electrophoretic patterns of crotalids (Viperidae) and elapids showed substantial variation between the two families. Of 119 distinct bands from all species only 22 were shown to be found in two or more species and only three of these were present in both the elapids and crotalids (Bertke *et al.* 1966). This not only implied substantial variation in venom composition between these two families but also within each family, with only a small proportion of bands found in any two different species. However, some studies have shown that venom cross reactivity between sub-families can occur. For example, commercial polyvalent antivenom was found to be effective at neutralising the effects of two Viperinae and four Crotalidae venoms (both family Viperidae), whereas monovalent antivenom was ineffective (Kornalik and Táborská, 1989). Studies of the genus *Bothrops* found that venom was either: markedly coagulant with little fibrinolytic activity, exhibited both activities at a high level, had low coagulant with a high fibrinolytic activity or was weak in both (Rosenfeld *et al.* 1959). This highlights the degree of inter-specific venom variation within this particular genus and indicated that no specific functional activity is occurring within the genus *Bothrops*. Other studies on crotalids have also highlighted the lack of a genus-specific activity; it was found that the variation in venom activity was no greater between three genera than it was between the species representing them (Githens, 1935; Minton, 1956). Subsequently, inter-species variation has been observed in the Asian pit vipers

(genus *Trimeresurus*) (Tan *et al.* 1989); the bite of *T. malabaricus* was shown to cause extensive local tissue damage compared to other *Trimeresurus* species, which exhibit less or no local tissue necrosis (Tan *et al.* 1989; Gowda *et al.* 2006a). Differences were also found between the lethality of these species; *T. malabaricus* was found to have non-lethal venom in the majority of cases compared with other species of the same genus (Gowda *et al.* 2006a). Notably, venom variation also occurs at the intra-specific level. Variability was found in the electrophoretic profiles from the venom of eight midget faded rattlesnakes (*Crotalus viridis concolor*) (Glenn and Straight, 1977) and functional intra-specific venom variation has also been noted in studies on the yellow and white venom variants produced by *Vipera ammodytes*, *Daboia russelii* and *Crotalus helleri*. Although a number of venom activities observed from *V. ammodytes* were comparable, the yellow venoms were determined to contain a greater quantity of L-amino acid oxidase (Kornalik and Master, 1964; Master and Kornalik, 1965). A similar result was found in the venom of *D. russelii*, with the necrotising action of the yellow venom stronger than the white (Kornalik and Master, 1964). In the case of *C. helleri*, the white venom showed greater fibrinolytic and proteolytic activity whilst the yellow venom was more toxic and haemorrhagic (Galán *et al.* 2004). Individual venom variability has also been found between parents and siblings (Táborská and Kornalik, 1985; Kornalik and Táborská, 1988), with one study demonstrating as much variability in the venom of related snakes as those that were unrelated (Kornalik and Táborská, 1988).

Geographical variation may play an integral part in the processes involved in venom variation (Chippaux *et al.* 1991). Early studies on the action of *Crotalus terrificus terrificus* (now *Crotalus durissus terrificus*) venom identified geographical variation as a potential factor in the intra-specific variation found in the venom (Barrio and Brazil, 1951). Two distinct responses of the venom were found: one characterised by seizures and paralysis from Argentina, Paraguay and Bolivia, whilst those characterised by muscle flaccidity were found in areas of Brazil (Barrio and Brazil, 1951). Distinct geographical delineations were also found in Costa Rica, where biochemical variation in the venom of *Bothrops nummifera* (now *Atropoides mexicanus*) was associated with either Pacific or Atlantic zone origins due to

reproductive isolation of populations (Jimenez-Porras, 1964). A similar situation was found in the venom composition of *Bothrops asper* which also exhibited Atlantic and Pacific zone variants (Aragon-Ortiz and Gubenšek, 1981). Geographical variation was also observed in the venom of *Echis* species (Schaeffer, 1987); however, because of the nature of this species complex (morphologically indistinguishable different species), variation at the inter-species and subspecies level cannot be discounted as the true reason for the differences in venom composition (Chippaux *et al.* 1991). True geographical variation was observed in close populations of the Mojave rattlesnake (*Crotalus scutulatus scutulatus*) following the discovery of a venom type found in the north-eastern part of their range which exhibited consistently higher LD_{50} values (Glenn and Straight, 1978). The situation was further elucidated by the description of two divergent populations with no significant external morphological differences, but that differed in the presence or absence of a phospholipase A_2 toxin known as Mojave toxin (Glenn *et al.* 1983; Rael *et al.* 1984). Although evidence suggests that the two venom populations were historically isolated, no barrier to interbreeding between the populations was found, highlighted by the fact that an intergrade population was subsequently described (Glenn and Straight, 1989). Geographical venom variation can also be found in isolated populations of morphologically indistinguishable snakes; as in the case of the black tiger snake (*Notechis ater niger* – now *Notechis scutatus*) on island populations off the south coast of Australia (Williams and White, 1987) and the Habu pit viper (*Trimeresurus flavoviridus* – now *Protobothrops flavoviridis*) on island populations from the Okinawa Islands (Sadahiro and Omori-Satoh, 1980).

A number of studies have suggested that significant differences in venom compositional activity could have implications for the classification of a species (e.g. Jimenez-Porras, 1967; Bernadsky *et al.* 1986; Tan *et al.* 1989). However, the high level of intra-specific venom variation found in these studies and in other cases (Boche *et al.* 1981; Daltry *et al.* 1996b) question the validity of such theories without supporting morphological and taxonomical data. Nevertheless, Detrait and Saint Girons (1979) found results that supported the classification of Elapidae and Viperidae when comparing antigens of venoms from both families; thus showing

that a good correlation can be found between immunological venom data alongside morphological observations (Detrait and Saint Girons, 1979; Saint Girons and Detrait, 1980). More recently, proteomic venom profiles have been advocated for use as taxonomic markers in the genera *Bitis* and *Atropoides* (Calvete *et al.* 2007; Angulo *et al.* 2008); similarity coefficients of *Bitis* venom proteins were interpreted to be informative for the reconstruction of the evolutionary history of congeneric taxa, whilst disparate venom profiles were obtained from *A. nummifer* and *A. picadoi* despite minimal morphological variation. Furthermore, venom protein sequences from two toxin families (phospholipases A_2 and short neurotoxins) were demonstrated to be successful in reconciling a species tree derived from members of the Elapidae (Slowinski *et al.* 1997), again suggesting that venom proteins may provide taxonomically informative data. However, the lack of node support values for the generated species trees produced by this study prevents any assessment of the uncertainty inherent to the derived species relationships (Page and Cotton, 2000; Sanderson and McMahon, 2007). More rigorous assessments of the validity of venom components use as taxonomic markers are required, particularly considering evidence that other factors, such as diet and geography, may also strongly influence venom composition (e.g., Jimenez-Porras, 1964; Daltry *et al.* 1996a). Consequently, if venom components suffer selective evolutionary pressures independent to neutral phylogenetic processes, the evolutionary history of these components may not correspond to the true species relationship.

1.5 Evolutionary basis of venom variation

The evolution of multicomponent, multifunctional venom containing a diverse array of enzymes and proteins is thought to provide an advantage to the snake in prey acquisition and digestion (Mebs, 1999). However, a number of theories exist as to whether specific selection pressures are independently driving the accelerated evolution and subsequent diversification of venom components, thereby causing observed cases of venom variation. One theory suggests the ongoing evolution of venom can be driven by predator-prey interactions (Poran *et al.* 1987; Biardi *et al.* 2000). Poran *et al.* (1987) demonstrated resistance to Northern Pacific rattlesnake (*Crotalus oreganus*) venom in natural prey species. Resistance to venom was found

amongst populations of California ground squirrels (*Spermophilus beecheyi*) in varying localities; the level of resistance depended directly upon the density of *C. oreganus* in each locality. Further work on the same species determined that blood sera from ground squirrels in rattlesnake abundant areas inhibited *C. oreganus* venom more effectively than venom from two allopatric rattlesnake species, with particular neutralisation of venom metalloprotease and haemolytic activity, thus indicating evolutionary specialisation (Biardi *et al.* 2000, 2006). The fact that the inhibition of venom proteases has been found in a preferred prey item was said to provide a model for an evolutionary arms race, whereby prey resistance induces corresponding changes in venom toxins in order to maintain their effectiveness (Biardi *et al.* 2000, 2006). It is hypothesised that this coevolutionary process may drive structural rearrangements in venom toxins and resistance proteins in a pattern that will vary across populations and species (Biardi *et al.* 2000). Prey resistance was also demonstrated in eels when subjected to the venom of two different sea snakes, *Aipysurus laevis* and *Laticauda colubrina* (Heatwole and Poran, 1995). Two species of eels tested were syntopic and therefore probable prey to the sea snakes; these were found to be highly resistant to the venom. The eels that were sympatric but unlikely to be preyed upon and an allopatric species were highly susceptible to the venom (Heatwole and Poran, 1995). This is another case of specific venom resistance, indicating an origin via coevolution. Resistance to the venom was found to be greater in the specialized eel feeder, *L. colubrina*, than in the more generalist feeder, *A. laevis*. It was therefore hypothesized that *L. colubrina* exerts a greater selection pressure for resistance by feeding continuously on specific species of eels, or that *A. laevis* may have a broader spectrum of venom toxins which have been generated to be effective against a larger range of prey (Heatwole and Poran, 1995). In this case it appears that the basis for selection of resistance has arisen as a defense against specific predators rather than as a general hardiness based on phylogenetic position (Zimmerman *et al.* 1992; Heatwole and Powell, 1998). However, Mebs (1999) questioned the influence of a predator-prey co-evolutionary relationship by suggesting that it is not evident, particularly in the viperids, that more powerful venom is evolving to counteract prey resistance. However, evidence from Southern Pacific rattlesnakes (*Crotalus helleri*) demonstrated that venoms from the same locality were capable of inducing significant differences in functional activities and were neutralized to different extents by the sera of prey items (Galán *et al.* 2004). It

has been suggested that snakes feeding on a wide diversity of prey items will require a multiplicity of toxin types in order to counteract the variety of prey defense systems and physiological targets (Fry *et al.* 2003b). Evidence from venomous marine gastropods supports this theory; venom duct transcriptomes revealed that a specialist diet correlated with a reduction in the number of venom components compared to the diversity found in species with broad dietary width (Remigio and Duda, 2008).

A number of authors have proposed an 'overkill' hypothesis of venom evolution suggesting, due to the high levels of apparent toxicity of many snake venoms and the correspondingly large doses injected, that the type of prey item is irrelevant because the loss of a particular venom component may easily be compensated by other lethal factors (Sasa, 1999a, 1999b; Mebs, 2001). Therefore variation in venom composition is unlikely to be subject to natural selection for lethality to prey, but rather results from neutral evolutionary processes (Sasa, 1999a, 1999b; Mebs, 2001). However, the overkill hypothesis overlooks the influence of venom resistance in natural prey items, whereby substantial increases in venom may be required to subdue a syntopic prey item (Heatwole and Poran, 1995; Biardi *et al.* 2000, 2006). Furthermore, venom production has been demonstrated to be metabolically costly, and evidence from some snakes suggest an ability for a species to 'meter' the amount of venom injected into a prey item based upon prey size (Hayes *et al.* 1995; McCue, 2006). Furthermore, studies on the genus *Echis* (Viperidae) suggest that despite some species demonstrating a higher lethality towards natural prey items, the speed with which prey was incapacitated was not associated with venom lethality, implying that venom toxicity may be adaptive in terms of metabolic saving, by reducing venom expenditure (Barlow *et al.* 2009). Combined, this data contradicts the assumption that snakes inject venom in amounts far greater than the lethal dose required, but rather that a trade-off exists between the metabolic cost of venom synthesis and foraging efficiency alongside complex predator-prey interactions.

A number of authors hypothesise that snake venom composition is subject to strong natural selection and that venom diversity results from adaptation to specific diets

(Daltry *et al.* 1996a, Wüster *et al.* 1999, Kordiš and Gubenšek, 2000). Multivariate analysis of isoelectrically focused *Calloselasma rhodostoma* venoms revealed a close association between venom composition and the diet of populations. Geographical distance and phylogenetic relationships between populations were rejected as correlates due to insignificant results (Daltry *et al.* 1996a). It was suggested that natural selection has allowed different *C. rhodostoma* populations to produce venoms appropriate for subduing and digesting the local diet. Therefore, the susceptibility and availability of prey items are likely to play an important role in the evolution of venom components; venom composition may directly reflect the prey animals and hence the feeding habits of the snake. (Daltry *et al.* 1996a). Mebs (1999) questions the findings of Daltry *et al.* (1996a), stating that electrophoretic patterns cannot provide clues for biological activities, such as lethality for a certain type of prey, or high or low enzymatic activity. However, a number of other studies, using varied techniques, have produced correlations between venom variation and diet. Creer *et al.* (2003) used matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI TOF-MS) isoelectric focusing to analyse the variation in phospholipase A_2 toxins from geographically diverse populations of *Trimeresurus stejnegeri* and demonstrated a correlation between venom variation and selection for regional diets. Proteomic analysis of four *Sistrurus catenatus* subspecies found a correlation between the complexity of venom components and the proportion of mammals found in the snakes diet (Sanz *et al.* 2006). Li *et al.* (2005) found further support for diet as a driving factor in the evolution of venom components by analysing molecular toxin data from the marbled sea snake (*Aipysurus eydouxii*). A dinucleotide deletion in the only three finger toxin expressed in *A. eydouxii* venom was found to result in a truncated, inactive form of the toxin (Li *et al.* 2005) corresponding to a reduction in *A. eydouxii* venom toxicity compared to other members of the genus (Tu, 1974). The inactivity of this three finger toxin appears to be a secondary result of the adaptation of *A. eydouxii* to the new dietary habit of feeding exclusively on fish eggs, rendering venom unnecessary for prey capture (Li *et al.* 2005).

The functional significance of adaptations to specific prey have been tested by measuring the effects of venom on natural prey items for a number of different snake

species, including coral snakes (*Micrurus sp.*) and Eurasian vipers (*Vipera sp.*) (Jorge-da-Silva and Aird, 2001; Starkov *et al.* 2007). In both cases venom was demonstrated to be most toxic to natural prey species rather than non-prey species. Although these results suggest adaptation, these correlations do not rule out the possibility of phylogenetic constraint, whereby similarity in venom characteristics and diet may be the result of common ancestry rather than selection (Barlow *et al.* 2009). Subsequently, similar studies have been undertaken in a more robust manner by interpreting the results within a phylogenetic framework (Barlow *et al.* 2009; Gibbs and Mackessy, 2009). Barlow *et al.* (2009) demonstrated that venom toxicity and diet have co-evolved within the genus *Echis* in respect to arthropod prey items, whilst venom toxicity to mice in the genus *Sistrurus* correlated to the proportion of mammals found in the snakes diet and appears to be a major axis for evolution within this genus (Gibbs and Mackessy, 2009). This combination of results reinforces the apparent strong relationship between the evolution of venom composition and feeding adaptations in snakes. However, as yet no direct link at the molecular level has elucidated the evolutionary adaptations driving venom composition optimisation to specific prey items.

1.6 The symptomatology of envenoming

Envenoming by venomous snakes is estimated to cause as many as 94,000-125,000 deaths per year worldwide (Chippaux, 1998; Kasturiratne *et al.* 2008). Aside from mortality, bites by venomous snakes can cause substantial long term morbidity, particularly in cases where significant necrosis occurs (Chippaux, 2006). Current estimates suggest that up to 5.5-6 million people are subject to snake bites each year, with between 0.4-2.6 million people exhibiting clinical problems of varying severity as a result of envenomation (Chippaux, 1998; Kasturiratne *et al.* 2008). Due to the large variation in venom composition, a variety of symptoms can arise after envenomation, such as bleeding, shock or necrosis (Mebs, 1999). A number of clinically significant effects leading to potential morbidity and mortality include: flaccid paralysis, systemic myolysis, coagulopathy and haemorrhage, renal damage and failure, cardiotoxicity and local tissue injury (White, 2005). These symptoms are caused by the action of venom toxins which have varying molecular targets and

enzymatic activities (Lee, 1979; Mebs, 1999). In particular, it is common for snakes to possess either a markedly neurotoxic (e.g. Harvey *et al.* 1994; Ramasamy *et al.* 2005) or proteolytic/haemorrhagic (e.g. Bjarnason and Fox, 1994, 1995; Gowda *et al.* 2006b) venom. Typically a viperid bite will cause predominately local effects, such as swelling, and in severe cases, necrosis, at the site of the bite (e.g. Annobil, 1993; Tan and Ponnudurai, 1996). The systemic effects of Viperidae bites are far more complex as venom components, such as SVMPs and SPs are often haemorrhagic, and can be procoagulant, anticoagulant and/or fibrinolytic in form (Siigur and Siigur, 1992; Morita, 2005). In serious cases severe haemorrhaging, consumption coagulopathy and renal failure can occur, leading to death (Than-Than *et al.* 1988; Soe *et al.* 1993). In contrast, members of the family Elapidae tend to provoke systemic responses that are typically neurotoxic and non-haemorrhagic (Shelke *et al.* 2002). The neurotoxins found in snake venoms are widely assumed to be a mix of presynaptic and/or postsynaptic toxins (Shelke *et al.* 2002). Presynaptic neurotoxins act by bind to the presynaptic membrane, causing the inhibition of neurotransmitter release (Montecucco and Rossetto, 2000), whilst postsynaptic neurotoxins bind to acetylcholine receptors and inhibit impulse formation (Charpentier *et al.* 1990; Gawade, 2004). Typical indications of neurotoxic envenomation include: ptosis, ophthalmoplegia, dysphoria, ataxia and general weakness, leading to paralysis and respiratory failure in severe cases (Goonetilleke and Harris, 2002). However, there are several documented cases where local tissue damage and severe coagulopathy has been caused by Elapidae bites (Warrell *et al.* 1976, White, 2005) and where neurotoxicity has been exhibited following Viperidae bites (Kularatne and Ratnatunga, 1999; Shelke *et al.* 2002). Furthermore, the identification of typically Viperidae toxins such as SVMPs in Elapids and neurotoxic proteins from viperid venom glands indicates the complexity of defining symptomatology based upon snake lineages and the subsequent effect that these assumptions can have in cases of severe envenomation (Jan *et al.* 2002; Junqueira-de-Azevedo *et al.* 2006; Fry *et al.* 2008).

The observation that venom variation is an extremely complex yet common occurrence within the advanced snakes and can be influenced by a number of factors relating to the life history of a species (Chippaux, 1991), highlights the importance

of characterising venom variation in respect to pathology. Factors such as phylogenetic position, geographical location, prey selection and predator-prey interactions can combine to radically alter the venom composition of closely related species or populations of snakes. It is therefore unsurprising that such alterations in venom composition impact upon the varied clinical manifestations observed following envenoming and subsequent antivenom therapy. For example, venom from the Russell's viper (*Daboia russelii*) exhibits procoagulant activities in the west, south and north of India, whilst in the east, the venom was found to be procoaglant at low concentrations and anticoagulant at high concentrations (Prasad *et al.* 1999). Additionally, the venom of the spectacled cobra (*Naja naja*) was observed as neurotoxic and procoagulant in the east of India, whilst myotoxic and procoagulant in western regions (Shashidharamurthy *et al.* 2002). Characterizing the venom variability of closely related species and populations has major implications for the treatment of snake bite; knowledge of venom variation allows for increases in efficacy, and in some cases, such as those above, medical personnel may have to choose appropriate antivenom depending on the geographical locality of the bite (Chippaux, 1991). The production of effective antivenom is therefore fundamentally dependent upon the knowledge of the variability of venoms within and between specific localities of medically important snakes (Barrio and Brazil, 1951; Warrell, 1985; Warrell *et al.* 1989; Theakston *et al.* 1989; Galán *et al.* 2004). It is clear that variation in venom may have an impact on both primary venom research and the management of snakebite, including the selection of antivenoms and most importantly, the selection of specimens for antivenom production (Chippaux, 1991).

1.7 The genus *Echis*

The genus *Echis* (Schneider, 1801) contains a group of small Viperidae snakes, from the sub-family Viperinae, known as the saw scaled vipers (Spawls *et al.* 2004). Saw scaled vipers inhabit a wide geographical range, stretching from India and Sri Lanka in the east, across the Arabian peninsula to Mauritania and Senegal in west Africa (Cherlin, 1990; Whitaker and Captain, 2004; Spawls *et al.* 2004; Trape and Mané, 2006; Arnold *et al.* 2009; Pook *et al.* 2009). This genus can also be found in northern Africa up to the Mediterranean Sea and as far south as northern parts of

Kenya (Figure 1.3) (Cherlin, 1990; Spawls *et al.* 2004; Arnold *et al.* 2009; Pook *et al.* 2009). Members of this genus are typically small with an average length for adult specimens ranging between 400 and 600mm, up to a maximum of 800/900 mm (Whitaker and Captain, 2004; Spawls *et al.* 2004). Despite being predominately cryptic species, the triangular head commonly contains a marking, such as crosses or arrows, which can help in identifying a species (Cherlin, 1983). The saw-scaled vipers display varying colour variation, ranging from sand coloured through to dark brown or grey (Spawls *et al.* 2004). They exhibit vertical eye pupils and can be either oviparous or viviparous (Whitaker and Captain, 2004; Spawls *et al.* 2004). *Echis* species have short and thin tails and strongly keeled scales which contain a saw-tooth ridge for which the species is named. When these scales are rubbed together in a characteristic defensive position (Figure 1.4) they produce a loud warning 'rasping' noise (Whitaker and Captain, 2004; Spawls *et al.* 2004). These snakes are terrestrial and predominately nocturnal; their primary habitat is dry savannah (Spawls *et al.* 2004). A comprehensive description of scale measurements and other descriptive factors is given by Cherlin (1990).

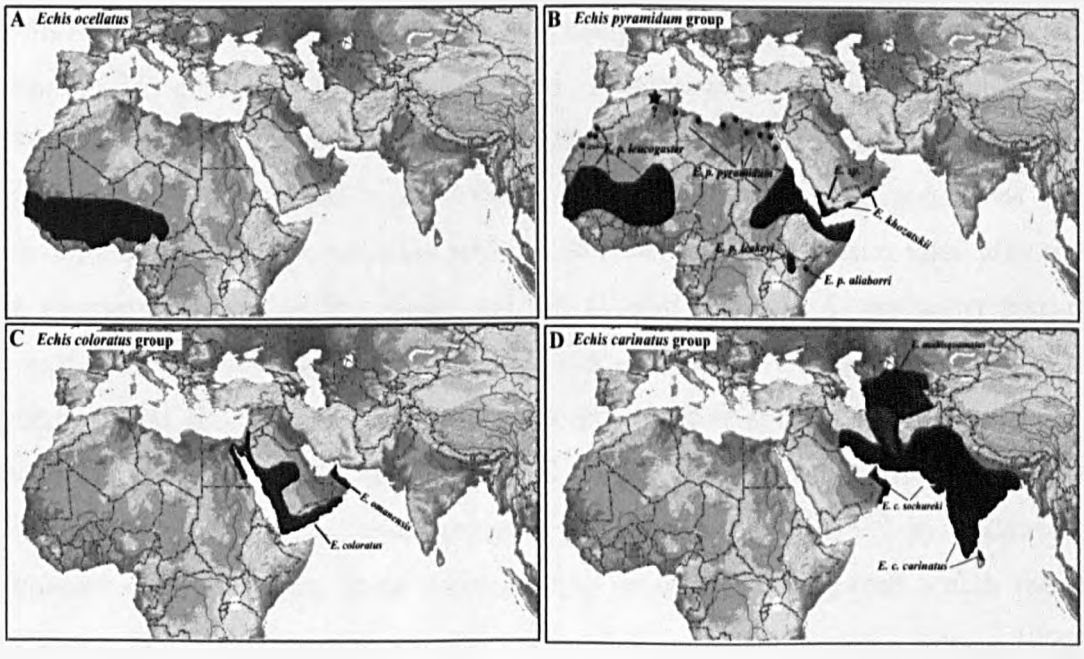

Figure 1.3. A distribution map showing the range of the four main species groups of the genus *Echis* (from Arnold *et al.* 2009).

Figure 1.4. Photographs of *Echis pyramidum leakeyi* and *Echis coloratus*. Note the variation in colour and the characteristic coiled defensive position. Whilst in this position the snake is able to rub its scales against each other to produce a rasping saw like sound. Photographs by Wolfgang Wüster.

The taxonomy of the genus *Echis* has been in a state of flux for some time; as many as twelve species and seven sub-species have been described (Cherlin, 1990), but there exists little consensus on the real number of species in the complex (Wüster *et al.* 1997; David and Ineich, 1999). More recently, the taxonomy has been partially resolved, with strong support for the monophyly of four species groups; the *E. carinatus*, *E. ocellatus, E. pyramidum* and *E. coloratus* complexes (Arnold *et al.* 2009; Barlow *et al.* 2009; Pook *et al.* 2009). The most comprehensive study, by Pook *et al.* (2009), used over 4000bp of mitochondrial gene sequences and determined that the *E. pyramidum* and *E. coloratus* groups are sister taxa, although the interrelationships of this clade and the *E. ocellatus* and *E. carinatus* species groups were unresolved (Figure 1.5) (Pook *et al.* 2009). Using a combination of mitochondrial and nuclear genes and one representative species from each species group, Barlow *et al.* (2009) recovered the *E. carinatus* group as the sister group of all other *Echis*, and the *E. ocellatus* group as the sister group of the *E. pyramidum*/*E. coloratus* clade. Despite three species being previously recognized within the *E. carinatus* group (*E. carinatus*, *E. sochureki* and *E. multisquamatus*) (Cherlin, 1990), Pook *et al.* (2009) determined little divergence between these species implying the presence of one species, *E. carinatus*. Further sampling is required to exclude the possibility of sub-species status, particularly given the clinal variations exhibited by *E. sochureki* and *E. multisquamatus* when compared to *E. carinatus* (Auffenberg and

Rehman, 1991; Pook *et al.* 2009). The *E. ocellatus* species group contains two species, *E. ocellatus*, from the majority of western Africa and *E. jogeri* from south-east Senegal and Mali (Pook *et al.* 2009). The *E. coloratus* species group also contains two species; *E. coloratus* from the Middle East and Egypt and *E. omanensis*, a closely related form from the eastern corner of the Arabian Peninsula (Figure 1.3) (Pook *et al.* 2009). The remaining species group, the *E. pyramidum* species complex, is less clear, although Pook *et al.* (2009) determined the presence of at least four species, *E. pyramidum*, *E. leucogaster*, *E. borkini* and *E. khozatskii*, with an undetermined number of members which may yet be classified as separate species following further sampling.

The saw scaled vipers feed on a variety of prey, including both vertebrates and invertebrates (Barlow *et al.* 2009). Notably, there appears to be a substantial shift in feeding habits between monophyletic species groups; stomach content samples indicate that the *E. carinatus, E. pyramidum* and *E. ocellatus* species groups feed on both vertebrates and invertebrates, with scorpions making up a significant proportion of the invertebrates (Barlow *et al.* 2009). Conversely, the *E. coloratus* species group appears to feed almost exclusively on vertebrates (Barlow *et al.* 2009) (Figure 1.6). Furthermore, the toxicity of *Echis* venom was demonstrated to have co-evolved alongside a shift in dietary preference, with an increase in the proportion of arthropods contributing to diet correlating with an increase in venom toxicity to scorpions (Barlow *et al.* 2009) (Figure 1.6). These results were interpreted within a phylogenetic framework and determined that the co-evolution of these two factors had occurred twice within the genus *Echis*, with a basal shift towards feeding on arthropod prey items and corresponding high venom toxicity towards these prey, followed by a secondary shift in diet within the *E. coloratus* species group leading to a reduction of venom toxicity (Barlow *et al.* 2009) (Figure 1.7). The reason for these shifts in target prey is unclear, although it is possible to speculate that the invertebrate feeding groups have adopted a more opportunistic form of feeding which has led to the incorporation of invertebrates into their diet, whilst prey availability may also be a factor (Wüster W, personal communication).

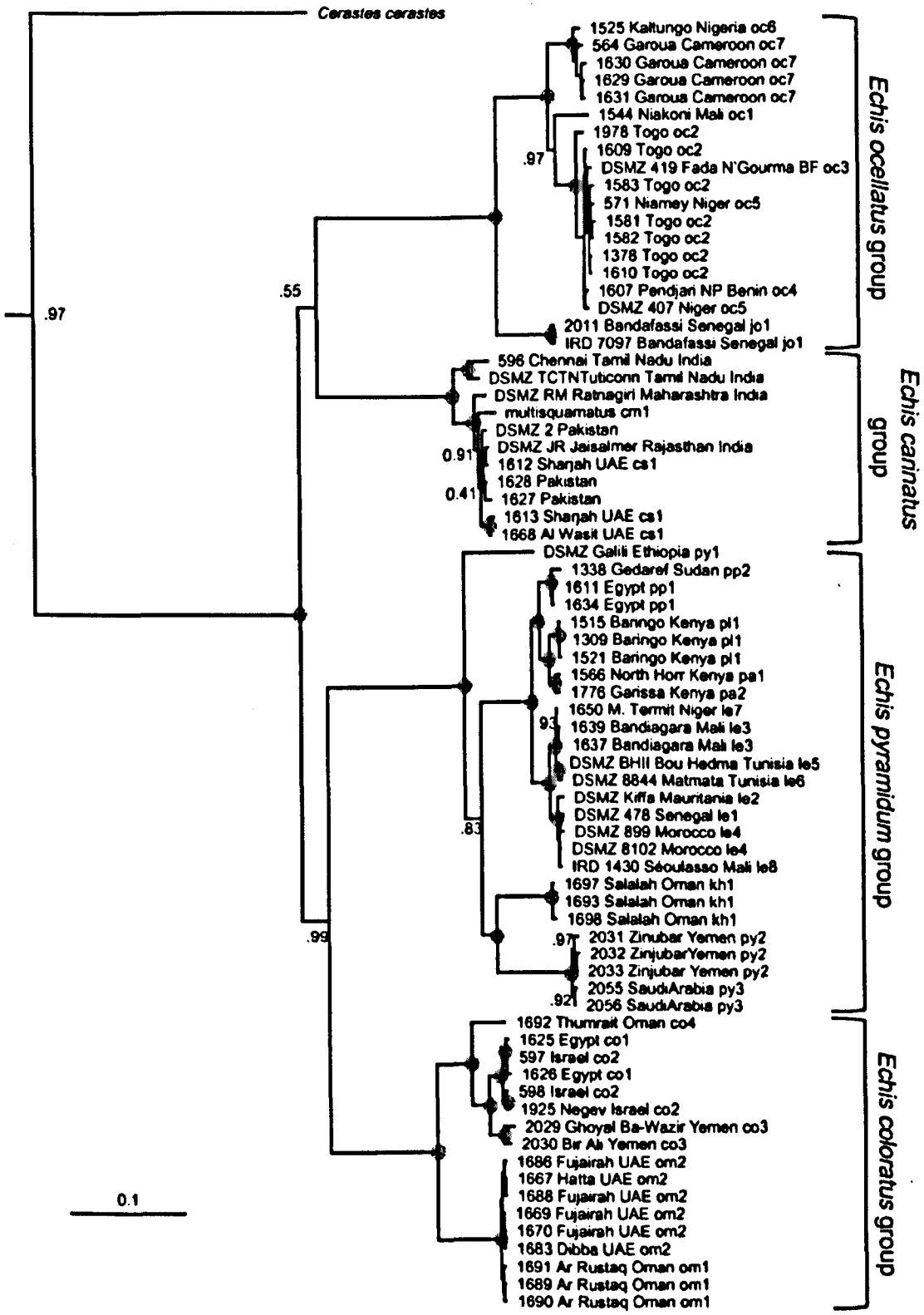

Figure 1.5. Bayseian Inference phylogeny of the genus *Echis* (from Pook *et al.* 2009). Outgroup taxon to the *Echis* clade is *Cerastes cerastes*. Nodes with grey circles represent a Bayesian posterior probability of 1.00.

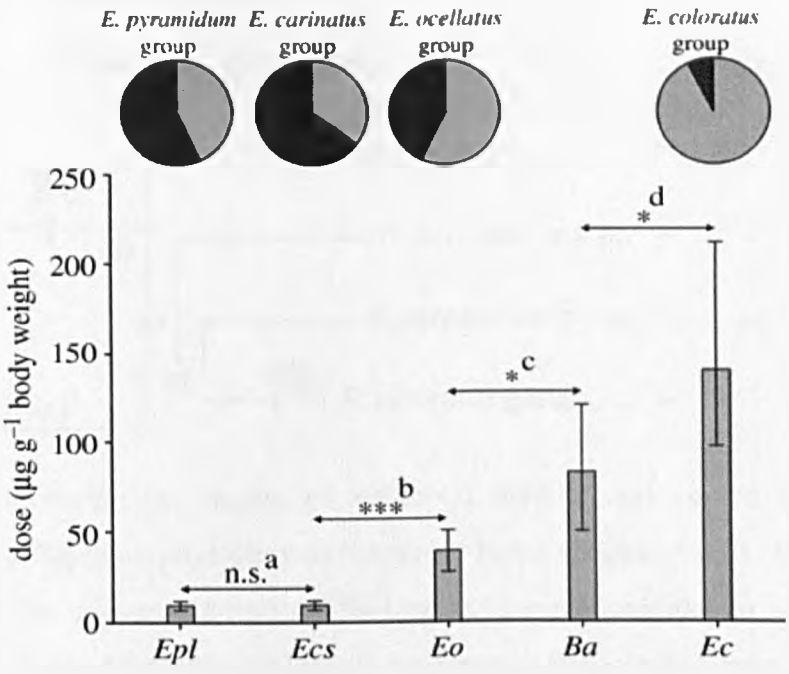

Figure 1.6. An increase in the proportion of arthropods in the diet of *Echis* species correlates with an increase in venom toxicity against scorpions (from Barlow *et al.* 2009). The pie-charts show the proportion of arthropods (black portion) and vertebrates (grey portion) consumed by each *Echis* species group based on stomach content analysis. Scorpion LD_{50} measurements for the venoms are represented by the bars with error bars showing 95% confidence intervals. Pair-wise statistical comparisons are shown by asterisks (* = $P<0.05$, *** = $P<0.001$, n.s. = not significant). The vertebrate feeding outgroup is represented by *Bitis arietans* (*Ba*). *Epl* = *E. pyramidum leakeyi*, *Ecs* = *E. carinatus sochureki*, *Eo* = *E. ocellatus* and *Ec* = *E. coloratus*.

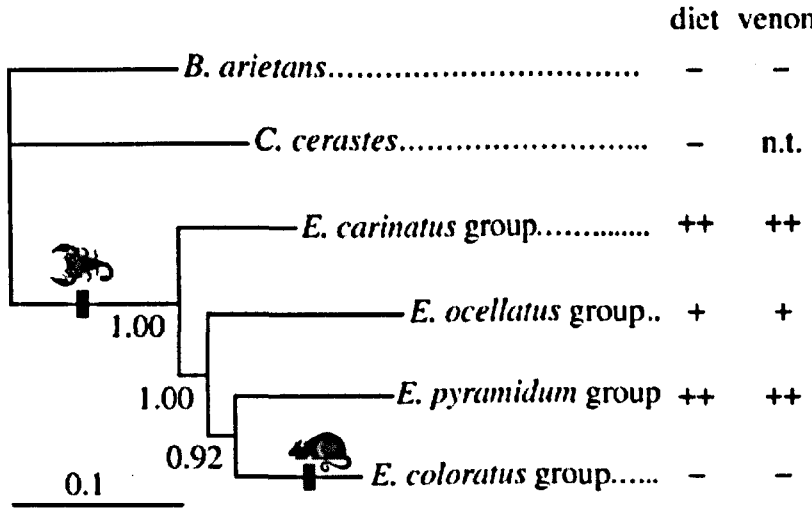

Figure 1.7. Mapping the degree of arthropod feeding and venom toxicities to scorpions to a Bayesian phylogeny of the major *Echis* species groups (from Barlow *et al.* 2009). The degree of arthropod feeding and venom toxicities to scorpions are shown to the right of the tree (++ high, + moderate, - low, n.t. not tested). Instances of dietary shifts in prey type accompanied by co-evolution of venom composition are indicated by bars along branches. The outgroups *Bitis arietans* and *Cerastes cerastes* were included to root the tree and infer the timing of dietary shifts.

Differences in lethality to prey items is likely to rely largely upon the variation in components that are present in the venom. A number of venom proteins with varying activities have previously been described from members of the genus *Echis*, including SVMP prothrombin activators (Nishida *et al.* 1995; Yamada *et al.* 1996), myotoxic PLA_2s (Jasti *et al.* 2004a; Zhou *et al.* 2008) and CTL and disintegrin inhibitors of platelet aggregation (Peng *et al.* 1993; Jasti *et al.* 2004b; Juárez *et al.* 2006a). More recently, a representative overview of the venom gland composition of one species, *E. ocellatus*, was determined by cDNA library construction (Wagstaff and Harrison, 2006). This venom gland transcriptome identified the snake venom metalloproteinases (SVMPs) as the major toxin components present, with ~60% of all toxin sequences encoding them (Figure 1.8) (Wagstaff and Harrison, 2006). Substantial diversity was established within this abundant expression of SVMPs, including representation of all four SVMP subclasses (PI-IV), suggesting that this toxin family may be fundamental for venom function by members of this genus (Wagstaff and Harrison, 2006; Wagstaff *et al.* 2009). Furthermore, a large number

of SVMP inhibitory transcripts (SVMPIs - previously termed bradykinin potentiating peptides in Wagstaff and Harrison, 2006) were discovered in the venom gland library and demonstrated to inhibit both SVMP activity and venom-induced haemorrhage in mice (Wagstaff *et al.* 2008). It was hypothesised that the presence of SVMPIs aids the inhibition of SVMPs during glandular storage; the relatively low abundance of SVMPIs determined from proteomic analysis of *E. ocellatus* venom supports this theory (Figure 1.8) (Wagstaff *et al.* 2008, 2009). A number of other toxin families were determined from the *E. ocellatus* venom gland transcriptome, including PLA_2s, CTLs, SPs, LAOs, growth factors and a putative new toxin family, termed the renin-like aspartic proteases; all were present in relatively low expression levels (1-10%) compared to the SVMPs (Figure 1.8) (Wagstaff and Harrison, 2006). Proteomic analysis of *E. ocellatus* venom revealed a number of consistencies with the transcriptomic expression (Figure 1.8), suggesting that venom gland transcriptomes may produce a partial representative reflection of proteomic venom expression (Wagstaff *et al.* 2009). The primary differences that occur, including representation of disintegrins and DC-fragments, likely reflect proteolytic processing of SVMP precursors (Wagstaff *et al.* 2009). The transcriptomic analysis of *E. ocellatus* has produced a comprehensive database which supplies substantial DNA sequence information on the numerous toxins present in the venom gland of this species. This sequence data has subsequently been utilized for other purposes, including studies aimed at increasing the efficacy of antivenoms (Wagstaff *et al.* 2006); the authors identified sequences encoding variable structural and immunogenic epitopes thought to be responsible for *E. ocellatus* induced haemorrhage. Subsequently, synthetic DNA immunogens were designed based upon these epitopes and demonstrated to successfully neutralize haemorrhage *in vivo* (Wagstaff *et al.* 2006).

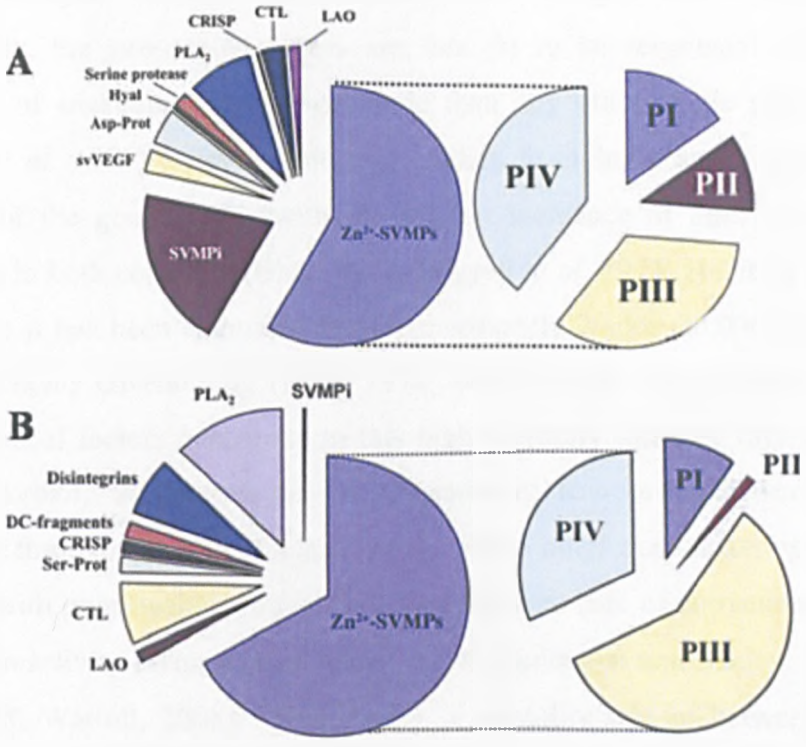

Figure 1.8. The composition of the *E. ocellatus* venom gland (A) transcriptome and (B) proteome (from Wagstaff *et al.* 2009). Key: DC-fragment, disintegrin/cysteine-rich fragment from PIII snake venom Zn2+-metalloproteinase (SVMPs); LAO, L-amino acid oxidase; PLA2, phospholipase A_2; CRISP, cysteine-rich secretory protein; CTL, C-type lectin-like protein; Ser-Prot, serine proteinase; Asp-Prot, aspartic proteinase; SVMPi, snake venom metalloproteinase inhibitors; Hyal, hyaluronidase. The relative abundances of the different classes of SVMPs (PI-PIV) predicted from the proteomic and transcriptomic analyses are highlighted.

Envenoming by members of the genus *Echis* typically induces systemic symptoms such as spontaneous bleeding, disseminated intravascular coagulation and haemolysis and local effects such as necrosis, swelling, blistering and oedema (Warrell *et al.* 1977; Porath *et al.* 1992; Benbassat and Shalev, 1993; Gillissen *et al.* 1994; Ali *et al.* 2004; Kochar *et al.* 2007). The venom of the saw-scaled vipers contain numerous anticoagulant and pro-coagulant factors (Chen and Tsai, 1996; Warrell, 1996) and can cause extensive bleeding by methods such as: disseminated intravascular coagulation due to the activation of factor V and factor X, the continuous activation of fibrinogen and the breakdown of the vascular endothelium

by haemorrhagins (Warrell and Arnett 1976; Chugh, 1989; Warrell, 1996). Additionally, the saw-scaled vipers are thought to be responsible for a greater proportion of snakebite deaths worldwide than any other single genus of snakes (Warrell *et al.* 1977). Epidemiological studies from India and Nigeria implicate members of the genus *Echis* with the highest incidence of bites and number of mortalities in both countries (Bhat, 1974; Warrell *et al.* 1977; Habib *et al.* 2001); in India alone it has been estimated that approximately 20,000-30,000 people die per year from *Echis* envenoming (Bhat, 1974; World Health Organisation, 1999). A combination of factors contribute to this high mortality rate: the high incidence of *Echis* snakebite, the possession of a markedly haemorrhagic venom, a high occurrence throughout parts of a large geographical range encompassing a number of countries with poor healthcare facilities, and a severe lack of antivenom availability and cross-reactivity (Warrell and Arnett, 1976; Benbassat and Shalev, 1993; Visser *et al.* 2008; Warrell, 2008). Historically, a mortality rate of between 10-20% is typical in cases of envenoming where antivenom is not administered (Warrell *et al.* 1977; Pugh and Theakston, 1980). However, a number of monospecific and polyspecific antivenoms are produced against the venom of *Echis* species and typically reduce mortality rates to between 2-8% (Warrell *et al.* 1977). Nevertheless there are increasing reports that antivenom availability and cross-reactivity are a problem (Warrell and Arnett, 1976; Visser *et al.* 2008; Warrell, 2008), as demonstrated by the ineffectiveness of *E. carinatus* antivenom to treat patients envenomed by *E. carinatus sochureki* and *E. ocellatus* (Kochar *et al.* 2007; Visser *et al.* 2008) and antivenom raised against west and east African species to treat bites from a Tunisian member of the *E. pyramidum* complex (Gillissen *et al.* 1994). As the production of effective antivenom is fundamentally dependent upon the knowledge of the variability of venoms within and between specific localities and species (e.g. Theakston *et al.* 1989; Galán *et al.* 2004), assessing the venom variation between these species is integral to increasing antivenom efficacy.

1.8 Aims

The primary aim of this project is to elucidate the genetic basis of venom variation within the genus *Echis* and to determine whether dietary selection pressures are

responsible for the evolution of venom components. Venom variation within the genus *Echis* has previously been inferred from lethality studies on invertebrates; lethality was correlated to the proportion of invertebrates comprising the diet of the species tested (Barlow *et al.* 2009). Barlow *et al.* (2009) mapped the coevolutionary position of the dietary shifts and venom toxicity to a strongly supported mitochondrial and nuclear phylogeny and determined that a shift to invertebrate feeding from vertebrate feeding likely occurred at the base of the *Echis* radiation, whilst a subsequent reversion to vertebrate feeding occurred within the *E. coloratus* species group (Figure 1.7). In order to identify the venom components that may be responsible for conferring increases in toxicity to invertebrate prey items, the venom composition of members of the genus *Echis* must first be elucidated. A venom gland transcriptomic approach will be adopted for three representatives (*E. coloratus*, *E. pyramidum leakeyi* and *E. carinatus sochureki*) of the four major species groups (Pook *et al.* 2009), to complement the previously constructed *E. ocellatus* transcriptome (Wagstaff and Harrison, 2006). The production of venom gland cDNA libraries coupled with the generation of ~1000 expressed sequence tags (ESTs) for each species will provide substantial DNA sequence information regarding the representation of toxins present in the venom glands. Comparisons of the toxin encoding profiles from the four members of the genus may reveal correlations with dietary composition, perhaps through the recruitment of novel venom toxins or increases in expression of specific components. However, in order to fully assess the nature of venom variation and the influence of diet, thorough phylogenetic analyses will be undertaken on the major toxin families in order to reveal patterns of gene duplication and loss. These analyses will be undertaken by mapping toxin gene trees generated by Bayesian Inference to the rigorously supported species trees of Barlow *et al.* (2009) and Pook *et al.* (2009). Patterns of gene duplication and loss will then be correlated with the phylogenetic position of the dietary shifts determined by Barlow *et al.* (2009) in order to infer whether dietary selection pressures are influencing the evolution of specific venom components in the genus *Echis*.

The production of representative toxin family gene trees, alongside rigorously supported species trees will also provide the opportunity to assess whether rapidly

evolving gene families, such as snake venom toxins, can be used as accurate predictors of species relationships. A number of studies have suggested that differences in venom composition and activity could have implications for the classification of a species (e.g. Jimenez-Porras, 1967; Bernadsky *et al.* 1986; Tan *et al.* 1989; Calvete *et al.* 2007; Angulo *et al.* 2008). However, these studies derived taxonomic inferences predominately through similarities and differences in venom profiles rather than through rigorous phylogenetic approaches; to date only a few studies have attempted to incorporate information from the evolution of toxins alongside that of the species (e.g. Slowinski *et al.* 1997; Fry *et al.* 2002). Slowinski *et al.* (1997) attempted to assess whether patterns of toxin sequence evolution are congruent with the evolutionary history of the species sampled (Slowinski *et al.* 1997), despite a number of toxinological studies simply assuming that a gene tree accurately represents the organismal phylogeny (e.g. Okuda *et al.* 2001; Tsai *et al.* 2004, 2007). Whilst Slowinski *et al.* (1997) successfully reconciled toxin gene trees to the species relationship of members of the Elapidae, they used a combination of gene trees to derive a reconciliation with the species tree, with each gene tree contributing partially to the species tree. Furthermore, the absence of node support values for the generated Elapidae species tree prevented any assessment of the uncertainty inherent to the derived species relationships (Page and Cotton, 2000; Sanderson and McMahon, 2007). The generation of comprehensive toxin EST gene sequences and corresponding gene trees derived by Bayesian Inference will provide the data necessary for rigorous assessments of species tree node support values by incorporating gene tree uncertainty present in entire Bayesian posterior distributions. The inclusion of posterior distributions, coupled with multiple heuristic tree searches and the subsequent derivation of a consensus tree (see Buckley *et al.* 2006; Oliver, 2008), will provide an accurate measure of support for the inferred species relationship and subsequent interpretation of the value of toxin families as taxonomical markers.

The final aim of this study is to determine the effect venom variation in the genus *Echis* may have upon antivenom therapy. The production of effective antivenom is fundamentally dependent upon the knowledge of venom variation within and between localities and species (e.g. Theakston *et al.* 1989; Galán *et al.* 2004). A

number of monospecific and polyspecific antivenoms are raised against the venom of different *Echis* species and have been effective at substantially reducing mortality rates (e.g. Bhat, 1974; Warrell *et al.* 1977). However, there are reports that antivenom cross-reactivity remains a problem within this genus (Gillissen *et al.* 1994; Kochar *et al.* 2007; Visser *et al.* 2008). In order to assess the influence venom variation has upon therapeutic outcomes, monospecific antibodies will be raised against the venom from the four *Echis* species used to construct the venom gland transcriptomes. Subsequently, lethality comparisons of the *Echis* venoms will be undertaken alongside immunological assessments of the cross-neutralisation of these venoms by the monospecific antivenoms. The *in vivo* neutralisation of the *Echis* venoms with the commercial monospecific *E. ocellatus* antivenom EchiTabG® (MicroPharm Ltd, UK) will be assessed alongside 'antivenomic' (e.g. Lomonte *et al.* 2008; Calvete *et al.* 2009; Gutiérrez *et al.* 2009) studies attempting to identify venom components that fail to bind to EchiTabG®.

CHAPTER 2

METHODS

Methods specific to the experimental chapters can be found in their respective chapters (3-7). Buffers and stock solutions used throughout the course of this experimental work can be found in Appendix I.

2.1 Venom gland cDNA library construction

Venom gland cDNA libraries were constructed from ten specimens each of three species of saw-scaled viper; *E. coloratus* (Egypt), *E. pyramidum leakeyi* (Kenya) and *E. carinatus sochureki* (United Arab Emirates). Snakes were confirmed as the identified species based on morphological and phylogenetic analyses in the form of scale counts and mitochondrial gene sequencing (Wüster, W., personal communication). The methodology outlined below was followed using identical procedures to the *E. ocellatus* (Nigeria) venom gland cDNA library construction described by Wagstaff and Harrison (2006). Briefly, RNA was extracted from the venom glands and messenger RNA (mRNA) purified by selection of RNA containing poly (A^+) tails. Complementary DNA (cDNA) was constructed by hybridizing a primer to the mRNA, reverse transcriptase of the DNA first strand followed by DNA polymerase of the second strand. An adapter was ligated to the open end of the cDNA followed by size fractionation by column chromatography. Recombination of cDNA into pDONR222 *E. coli* was undertaken using lambda integration facilitated by the *att*-containing primer and adapter capping the 5' and 3' end of the cDNA. Successful recombination of cDNA clones was determined by kanamycin selection following the transformation of recombinants into phage resistant cells. The finalised cDNA library was qualified by determining the library size and variation in inset sizes.

2.2 Dissection of venom glands

Snakes were sacrificed by decapitation under licensed procedures approved by the UK Home Office. The mandibular bone was cut towards the middle of the head on each side. The venom gland was identified on top of the muscle tissue beneath the skin. The surrounding muscle tissue was cut, allowing the gland to be separated and removed (Figure 2.1). The process was repeated for the other side of the head. Once removed, the glands were put on ice and weighed before being snap frozen in liquid nitrogen (Table 2.1). This process was repeated for ten specimens per species.

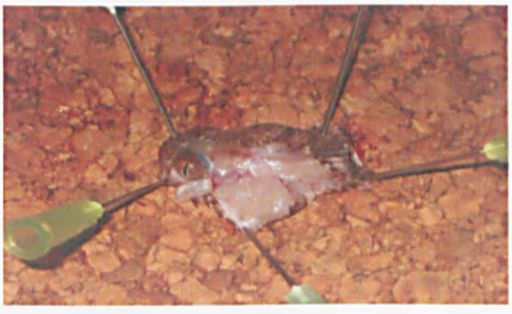

Figure 2.1. Dissection of venom glands demonstrating the separation of venom gland (below the eye) from muscle tissue.

2.3 RNA extraction

RNA extraction was carried out using a pestle and mortar partially submerged in liquid nitrogen. The ten venom gland samples (twenty glands) for each species were ground individually whilst submerged in liquid nitrogen (Figure 2.2). The pooled samples were then ground to a fine powder, collected and remaining liquid nitrogen was allowed to bubble off. The sample was weighed to determine percentage tissue recovery (Table 2.1). Using RNAase free equipment, trireagent was added at 1ml per 75mg of tissue recovered, followed by mixing and homogenisation of the tissue. The pooled homogenate was realiquotted and extracted according to the manufacturer's protocol for TriReagent (Sigma, UK). 0.2ml of chloroform was added to each sample and shaken vigorously. Centrifugation was carried out in a desktop centrifuge (Biofuge Fresco, Heraeus Centrifuges, UK) at 4543 x g for fifteen minutes at 4°C; the top aqueous layer was removed and stored on ice. 500µl of

isopropanol was added to each sample and mixed before further centrifugation for ten minutes at 4°C. The supernatant was removed from each tube leaving an RNA pellet. 1ml of 75% diethylpyrocarbonate (DEPC) treated ethanol was added to each sample before centrifugation for five minutes at 4°C, this step was then repeated to ensure the removal of salts and remaining isopropanol. RNA pellets were allowed to dry before resuspension in DEPC treated double-distilled water (ddH_2O) by pipetting. The samples were subsequently incubated at 60°C with occasional pipetting to aid resuspension. The quantity of RNA was determined using a LD1000-series nanodrop spectrophotometer (ThermoScientific, USA) (Table 2.1).

	E. coloratus	*E. p. leakeyi*	*E. c. sochureki*
Total pooled venom gland weight (mg)	594.1	507.0	309.4
Average venom gland weight (mg)	59.41	50.70	30.94
Pre-RNA extraction weight (mg)	833.0	436.7	335.6
RNA extraction % recovery	140%*	86%	108%*
Post-RNA extraction weight (µg)	732.56	1108.8	693.04
Post-mRNA purification weight (µg)	37.34	14.90	32.90
Post-mRNA purification weight assuming 50% purity (µg)	18.67	7.45	16.45
Pre-cDNA synthesis concentration (µg/µl)	3.37	4.68	2.54
Volume required to yield 10µg for cDNA synthesis (µl)	3	2.2	4
Pre-recombination weight (ng)	412.2	180.0	316.0
Final library size (number of clones)	5.54×10^{-07}	1.56×10^{-08}	5.04×10^{-07}

Table 2.1. Summary statistics for venom gland cDNA library construction of three members of the genus *Echis*. * The most likely explanation for greater than 100% recovery is superfrozen water collecting in the sample tube, although inaccurate measuring balance and human error cannot be excluded.

2.4 mRNA purification

mRNA was purified using 1X oligo-dT affinity chromatography according to the Illustra mRNA purification kit protocol (GE Healthcare (Amersham Biosciences), UK). The number of columns required was calculated by assuming that 1-2% of the total extracted RNA is mRNA combined with the maximum amount of mRNA specified for each column (1.25mg). Columns were prepared and the storage buffer drained followed by two 1ml washes with high-salt buffer. The sample was heated at 65°C for five minutes before cooling on ice for two minutes. 10µl of 1M Tris-Cl (pH 7.5), 2µl of 0.5M EDTA (ethylenediaminetetraacetic acid) and DEPC ddH_2O to a volume of 1ml were added to the sample in addition to 0.2ml of sample buffer. The sample was introduced to the column and spun in a RT6000D centrifuge (Sorvall Centrifuges, UK) at 1282.3 rpm (350 x g for a 190mm swing out rotor) for two minutes. The flow through was reserved before 0.25ml of high salt buffer was added to the column and centrifuged for two minutes. The high salt wash was repeated and followed by three 0.25ml low salt washes. Throughput was discarded and columns were placed in 15ml tubes for sample collection; elution was obtained by centrifugation using four 0.25ml additions of elution buffer. The quantity of mRNA purified was determined by nanodrop (Table 2.1). 100µl of ice cold sample buffer, 10µl of glycogen solution and 2.5ml of 100% ethanol was added for storage overnight. The sample was then placed at -20°C.

Figure 2.2. Pestle and mortar partially submerged in liquid nitrogen. The sample was ground in the mortar whilst fully submerged in liquid nitrogen.

2.5 cDNA synthesis

cDNA library construction was carried out according to the manufacturer's protocols for the CloneMiner cDNA library construction kit (Invitrogen, UK). A minimum of 5µg of mRNA was required for optimal cDNA library construction. The total mRNA previously extracted was assumed to be of 50% purity (Table 2.1), therefore 10µg of each species-specific mRNA sample was removed from -20°C storage and placed at -80°C for ten minutes. Each sample was subsequently separated into 1.5ml eppendorf tubes and centrifuged at 4543 x g for ten minutes at 4°C. The supernatant was removed and the pellet washed in 1ml of 75% DEPC ethanol and centrifuged twice for five minutes. The supernatant was removed again and the pellet was allowed to dry at room temperature for fifteen minutes before resuspension in 5µl DEPC ddH_2O. The sample was subsequently incubated at 45°C for three minutes to aid resuspension, before nanodropping to confirm the concentration and calculate the appropriate volume of mRNA required for cDNA library construction (Table 2.1). Remaining mRNA was placed at -80°C for long term storage.

2.5.1 First strand synthesis

The sample was made up to 9µl using DEPC ddH_2O, before 1µl of Biotin-attB2-Oligo(dT) primer (Invitrogen, UK) and 1µl of 10mM deoxyribonucleotide triphosphates (dNTPs) were added. The sample was mixed by pipetting and incubated at 65°C for five minutes and 45°C for two minutes. 4µl of 5X first strand buffer, 2µl of 0.1M dithiothreitol (DTT) and 1µl of DEPC ddH_2O were mixed, centrifuged and incubated at 45°C before addition to the sample and incubation at 45°C for two minutes. Superscript II Reverse Transcriptase (Invitrogen, UK) was added up to a volume of 20µl, and mixed by pippeting before incubation at 45°C for sixty minutes.

2.5.2 Second strand synthesis

The incubated sample was placed on ice to cool before the addition of 92µl of DEPC ddH_2O, 30µl 5X second strand buffer, 3µl 10mM dNTPs, 1µl *E. coli* DNA ligase

(Invitrogen, UK), 1µl *E. coli* DNA polymerase I (Invitrogen, UK) and 1µl *E. coli* RNase H (Invitrogen, UK). The sample was mixed by pipetting and centrifuged for two seconds before being incubated at 16°C for two hours. Subsequently, 2µl of T4 DNA polymerase (Invitrogen, UK) was added to the sample and incubated at 16°C for five minutes before 10µl of 0.5M EDTA was added to stop the synthesis reaction. The sample was transferred to a 0.5ml tube and 160µl of phenol:chloroform:isoamyl alcohol (25:24:1) was added. The sample was shaken vigorously for one minute and centrifuged at 4543 x g at room temperature for five minutes. The top aqueous layer was removed and 1µl of glycogen, 80µl of 7.5M NH_4OAc (ammonium acetate) and 600µl of 100% ethanol was added to the sample before storage at -80°C. After ten minutes the sample was centrifuged at 4543 x g at 4°C for twenty five minutes before phenol extraction and precipitation was undertaken using ice cold ethanol; the supernatant was removed and 150µl of 70% ethanol was added to the sample before further centrifugation for two minutes. The ethanol wash was repeated and the supernatant discarded. The pellet was allowed to dry at room temperature for ten minutes before resuspension in 18µl of DEPC ddH_2O and centrifuging for two seconds. The sample was then placed on ice prior to ligation of the attB1 adapter.

2.5.3 Ligating the attB1 adapter

10µl of 5X adapter buffer, 10µl of attB1 adapter (Invitrogen, UK), 7µl of 0.1M DTT and 5µl of T4 DNA ligase (Invitrogen, UK) was added to the sample on ice and mixed by pipetting. The sample was subsequently incubated at 16°C for 24 hours.

2.6 Size fractionation of cDNA

Size fractionation was carried out by column chromatography according to the CloneMiner cDNA library construction kit protocol (Invitrogen, UK). Following ligation of the attB1 adapter, the sample was incubated at 70°C for ten minutes to inactivate the DNA ligase and placed on ice. Columns were prepared and the flow rate and fraction sizes were measured to assess column integrity (flow rate=30-40 seconds/drop, drop size=25µl-35µl). The column was washed four times with 0.8ml

of TEN buffer and then left to drain until dry. 100µl of TEN buffer was added to the sample and mixed by pipetting before addition to the column and collection in tube number 1. This process was repeated and collected into tube number 2. Subsequently, further additions of 100µl of TEN buffer were added to the column and single drops collected in tube numbers 3-20. Collecting tubes were then placed on ice. Fraction sizes and cumulative volume were measured using a pipette, before the concentration of each sample was measured by nanodrop. The amount of cDNA in each fraction was calculated (Tables 2.2-2.4). Fractions with a minus concentration of cDNA were discarded, apart from the sample prior to the first positive reading which may contain high quality transcripts of undetected cDNA. Tubes were also discarded once the total volume reached 600µl in order to prevent contamination of the library with short, partial length 3'-end inserts and adapter sequences. Remaining fractions were pooled together to a quantity of 480ng, significantly more than the manufacturer's minimum requirement (60ng). Additional cDNA was pooled to remove potential size selection biases and to incorporate cDNA inserts as small as 250bp, so not to exclude small toxin encoding transcripts (Wagstaff and Harrison, 2006). 1µl of glycogen was added together with 0.5 volumes (of pooled cDNA) of 7.5M NH_4OAc and 2.5 volumes (of pooled cDNA and ammonium acetate) of 100% ethanol, before storage at -80°C.

2.7 Recombination reaction

The quantity of sample required to yield 87.5ng of cDNA for the recombination reaction (480ng – as determined above) was removed from -80°C and centrifuged at 4543 x g at 4°C for twenty five minutes. The supernatant was discarded before two 150µl 70% ethanol washes and centrifugation at 4°C for two minutes were undertaken. The pellet was allowed to dry at room temperature for ten minutes and subsequently resuspended in 5µl of TE buffer by pipetting. The sample was nanodropped in order to confirm the concentration (Table 2.1). The optimal quantity of cDNA for transformation (87.5ng) was retained before the addition of ddH_2O up to 4µl. 1µl of pDONR222 vector (Invitrogen, UK) and 2µl of 5X BP Clonase reaction buffer (Invitrogen, UK) was added to the sample, yielding a total volume of 7µl. BP Clonase enzyme mix (Invitrogen, UK) was removed from -80°C storage and

thawed on ice for two minutes prior to brief vortexing. 3µl of BP Clonase enzyme mix was added to the sample and mixed by pipetting; the sample was left to incubate at 25°C for 20 hours.

2.8 Transformation

Following incubation, the sample was centrifuged briefly before 2µl of Proteinase K (Invitrogen, UK) was added. The sample was then incubated at 37°C for fifteen minutes and 75°C for ten minutes before being placed on ice. 90µl of sterile H_2O, 1µl of glycogen, 50µl of NH_4OAc and 375µl of 100% ethanol was added. The sample was inverted and placed at -80°C for twenty five minutes before centrifugation at 4543 x g at 4°C for twenty five minutes. The supernatant was discarded and two ethanol washes were carried out using 150µl of 70% ethanol before further centrifuging for two minutes. The pellet was allowed to dry for ten minutes and resuspended in 9µl of TE buffer by pipetting. 1.5µl of the sample was transferred to six individual tubes before the addition of 50µl of Electromax DH10B T1 phage resistant cells (Invitrogen, UK) to each sample. The samples were then transferred into Gene Pulser 0.1cm cuvettes (Bio-Rad, UK) before MicroPulser electroporation at 2.00kV (Bio-Rad, UK). Subsequently, 1ml of SOC media was added to each sample before mixing in a shaking incubator at 225 rpm for seventy minutes at 37°C. Following incubation, the samples were pooled together producing a total volume of 6.3ml, an equal volume of freezing media (60% SOC medium:40% glycerol) was added and mixed by pipetting. The sample was transferred to -80°C for long term storage.

Fraction	Fraction volume (µl)	Cumulative volume (µl)	Concentration of cDNA (ng/µl)	Quantity of cDNA (ng)
1	167	167	-0.05	Discarded
2	80	247	-0.26	Discarded
3	43	290	-0.50	Discarded
4	42	332	-0.13	None detected
5	42.5	374.5	1.51	61.91
6	42	416.5	5.53	223.97
7	42.5	459	11.37	466.17
8	42	501	15.53	628.97
9	39	540	14.88	580.32
10	40	580	12.97	518.80
11	42	622	14.52	Discarded >600µl
12	41	663	22.17	Discarded >600µl
13	41	704	36.92	Discarded >600µl
14	43	747	56.72	Discarded >600µl
15	41	788	93.29	Discarded >600µl
16	40	828	108.33	Discarded >600µl
17	42	870	157.68	Discarded >600µl
18	41	911	160.84	Discarded >600µl
19	42	953	162.78	Discarded >600µl
20	40	993	155.61	Discarded >600µl

Table 2.2. Size fractionation statistics for *E. coloratus* venom gland cDNA library construction. Red text indicates the fractions which were completely or partially retained for recombination.

Fraction	Fraction volume (µl)	Cumulative volume (µl)	Concentration of cDNA (ng/µl)	Quantity of cDNA (ng)
1	133	133	-0.46	Discarded
2	124	257	-0.45	Discarded
3	39	296	-0.81	Discarded
4	39	335	-0.96	None detected
5	41	376	0.46	17.48
6	39	415	2.68	96.48
7	38	453	6.96	243.60
8	40	493	8.23	304.51
9	39	532	11.73	422.28
10	38	570	17.10	598.50
11	39	609	24.83	Discarded >600µl
12	40	649	52.75	Discarded >600µl
13	39	688	61.81	Discarded >600µl
14	39	727	107.56	Discarded >600µl
15	38	765	122.93	Discarded >600µl
16	39	804	157.58	Discarded >600µl
17	39	843	185.71	Discarded >600µl
18	39	882	174.11	Discarded >600µl
19	39	921	167.90	Discarded >600µl
20	40	961	144.70	Discarded >600µl
21	41	1002	94.53	Discarded >600µl

Table 2.3. Size fractionation statistics for *E. p. leakeyi* venom gland cDNA library construction. Red text indicates the fractions which were completely or partially retained for recombination.

Fraction	Fraction volume (µl)	Cumulative volume (µl)	Concentration of cDNA (ng/µl)	Quantity of cDNA (ng)
1	121	121	-0.28	Discarded
2	121	242	-0.30	Discarded
3	38	280	-0.18	Discarded
4	38	318	-0.30	None detected
5	39	357	0.08	2.8
6	39	396	2.44	85.4
7	39	435	4.46	156.1
8	39	474	5.68	198.8
9	39	513	6.20	217.0
10	39	552	7.92	277.2
11	40	592	9.95	358.2
12	40	632	14.61	Discarded >600µl
13	40	672	24.08	Discarded >600µl
14	40	712	41.17	Discarded >600µl
15	40	752	70.97	Discarded >600µl
16	40	792	119.48	Discarded >600µl
17	40	832	181.07	Discarded >600µl
18	40	872	175.69	Discarded >600µl
19	40	912	193.99	Discarded >600µl
20	40	952	191.45	Discarded >600µl

Table 2.4. Size fractionation statistics for *E. c. sochureki* venom gland cDNA library construction. Red text indicates the fractions which were completely or partially retained for recombination.

2.9 Qualifying the libraries

In order to quantify the size of the cDNA library, 100µl of the final sample was added to 900µl of SOC medium, before repeated dilutions were made up to a concentration of 10^{-4}. 100µl of each dilution was plated on two LB agar plates containing 50µg/ml kanamycin and incubated at 37°C overnight. Subsequently, colonies were counted on each plate and the size of each library was calculated based upon the dilution factors and the total stored cDNA library volume (Table 2.1). In order to assess the variation in cDNA library insert sizes, implying successful transformation, minipreps were carried out on thirty randomly selected colonies for each library using the Qiaprep miniprep kit (Qiagen, UK). 3ml of LB medium containing 50µg/ml of kanamycin was added to thirty 18ml tubes. 30 colonies were picked from a mixture of the plates used to assess library size using pipette tips which were ejected into the media and incubated overnight at 37°C. The samples were centrifuged for two minutes at 4543 x g and the supernatant removed. 250µl of Buffer P1 (Qiagen, UK) was added to each sample before resuspension by pipetting for one minute. Subsequently, 250µl of Buffer P2 (Qiagen, UK) was added and the tubes were inverted five times to mix. 350µl of Buffer N3 (Qiagen, UK) was then added immediately and mixed by inverting the tubes before further centrifugation for ten minutes. The supernatant was removed and added to a Qiaprep spin column (Qiagen, UK) and centrifuged for forty-five seconds. The flow through was discarded and 0.5ml of Buffer PB (Qiagen, UK) was added to the column and centrifuged for forty-five seconds. The flow through was discarded again and 0.75ml of Buffer PE (Qiagen, UK) was added and centrifuged for 45 seconds. The flow through was discarded before the columns were centrifuged for one minute. Collecting tubes were placed beneath the columns and 50µl of sterile water was added to each column and left to stand for one minute. The columns were centrifuged again for one minute and the eluted samples placed on ice. The samples were digested at 37°C overnight following the addition of 12.5µl of sterile water and restriction enzymes (0.5µl of BsrG1 and 2µl of NE Buffer 2 (New England Biolabs)). 5µl of 6X slow optical buffer was added to each sample, before electrophoresis on a 1% TAE buffer agarose gel at 100V for fifty minutes. Variation in the size of inserts, ranging from ~250bp to ~5000bp, was observed implying successful venom gland library construction (Figures 2.3-2.5). The insert sizes

observed were consistent with those obtained during construction of the *E. ocellatus* venom gland cDNA library (Wagstaff, S. C., personal communication).

2.10 Sequencing preparation

In order to prepare cDNA library clones for 96 well plate Sanger sequencing, colonies were first grown on LB agar plates containing 50µl/ml of kanamycin as described previously. Clearly defined individual colonies were picked using pipette tips and incubated in individual wells containing 150µl LB broth (containing 8% glycerol and 50µl/ml of kanamycin) for 10-15 minutes before the tips were discarded. Plates were covered and incubated at 37°C overnight, then split into duplicate and stored at -80°C prior to sequencing.

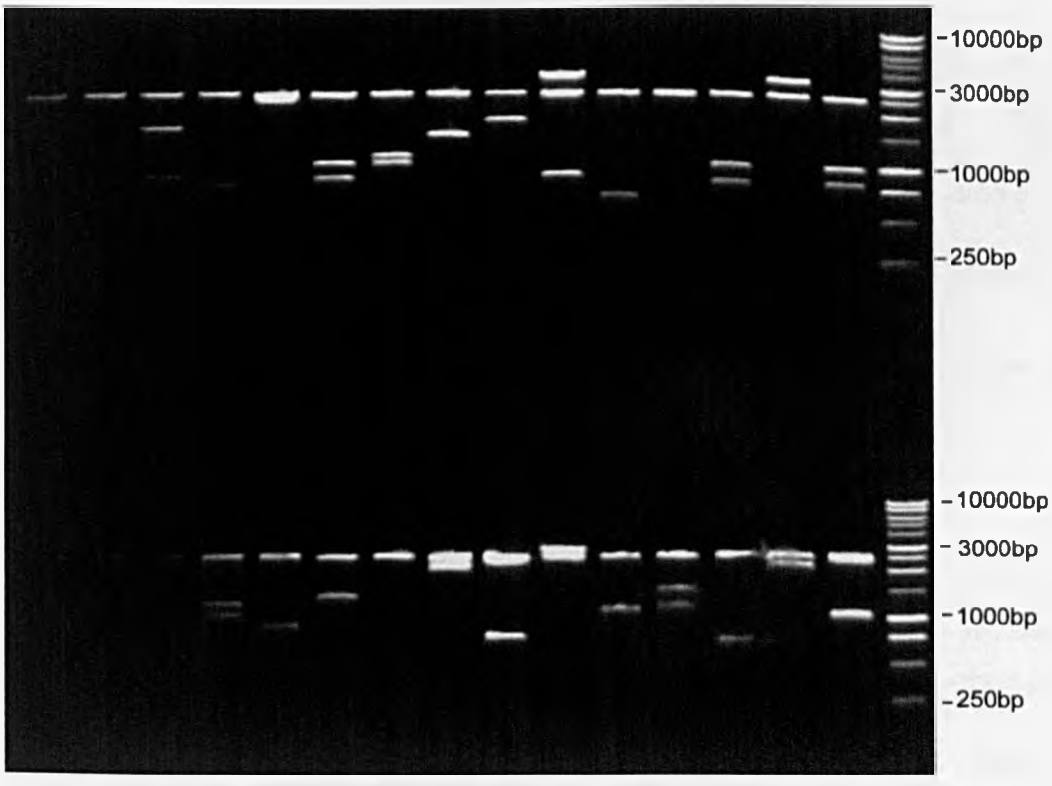

Figure 2.3. Quantification of the *E. coloratus* venom gland library by insert size. Inserts vary from ~250bp to ~4000bp.

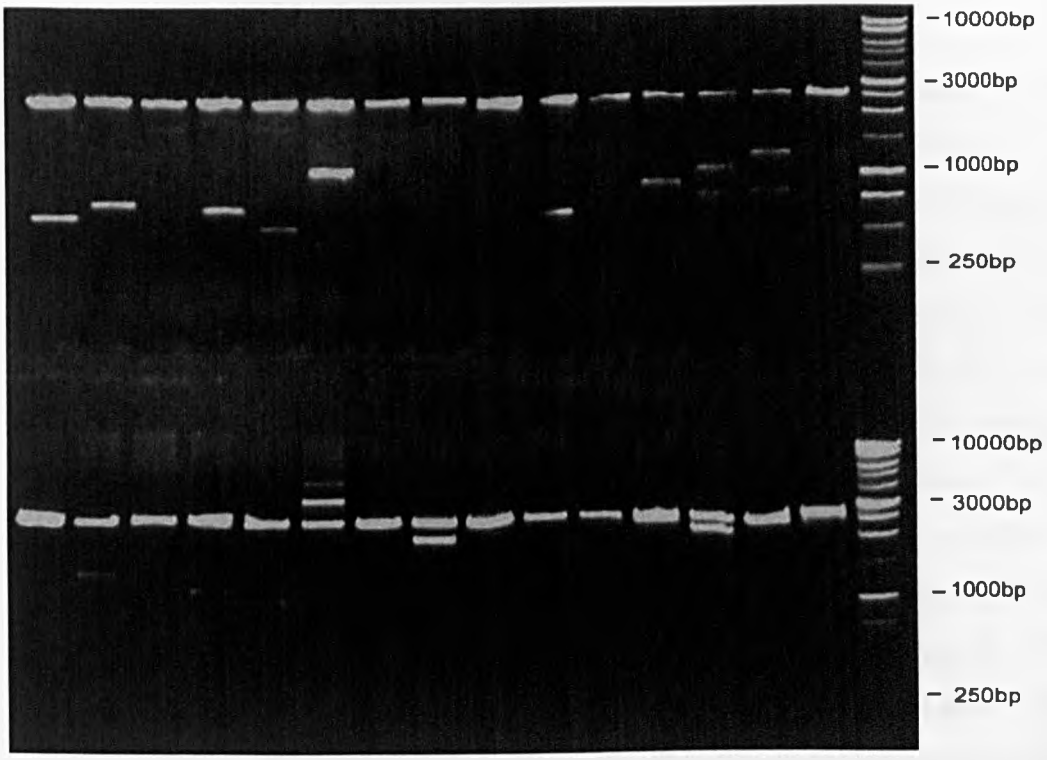

Figure 2.4. Quantification of the *E. p. leakeyi* venom gland library by insert size. Inserts vary from ~250bp to ~5000bp.

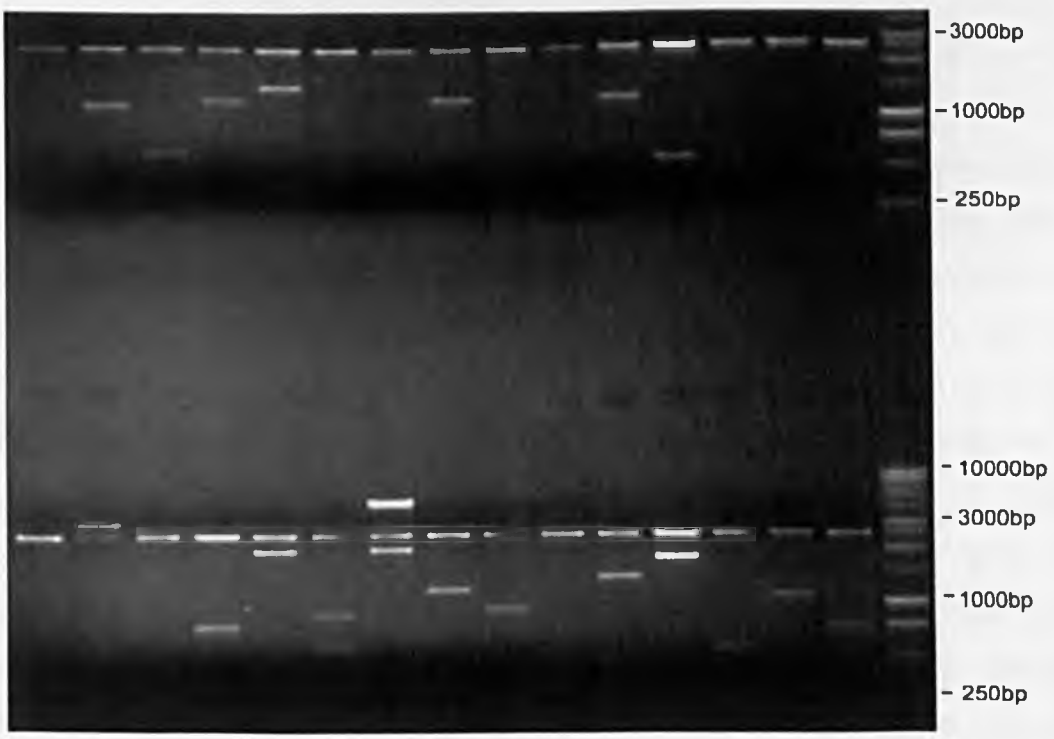

Figure 2.5. Quantification of the *E. c. sochureki* venom gland library by insert size. Inserts vary from ~250bp to ~5000bp.

2.11 Sequencing and bioinformatics

Sequencing of cDNA library clones was undertaken by Sanger sequencing (Natural Environmental Research Council (NERC) Molecular Genetics Facility – The GenePool, University of Edinburgh) using M13 primers and an ABI 3730 capillary sequencing instrument. EST processing and partial genome construction was undertaken on an Intel dual-core 2.8GHz workstation running the PartiGene pipeline on Bio-Linux 4.0 (http://envgen.ox.ac.uk) which is based on the Debian GNU/Linux distribution. The PartiGene pipeline (version 3.0) was preinstalled on Bio-Linux 4.0 alongside a number of programs that are freely available and essential for the functioning of PartiGene; DECODER (contact the authors, rgscerg@gsc.riken.go-.jp), ESTscan (http://www.isrec.isb-sib.ch/ftp-server/ESTScan/), postgreSQL (http-://www.postgresql.org), NCBI BLAST (http://www.ncbi.nlm.nih.gov/BLAST/), Bioperl (http://www.bioperl.org) and EMBOSS (http://www.hgmp.mrc.ac.uk/Soft-ware/EMBOSS). The remaining bioinformatic tools that the PartiGene pipeline is dependent on, phred, phrap and cross_match, were acquired by contacting the authors through the phrap website (http://www.phrap.org).

2.11.1 Trace2dbEST

Raw trace DNA sequence files were renamed into the NERC Environmental Genomics naming scheme for PartiGene pipeline processing. The naming scheme consisted of two letters representing a species identifier, followed by a maximum of five letters representing a library identifier and subsequently the plate and well number. The three identifiers are separated by underscores. The naming scheme for the *Echis coloratus* library was Ec_venom, where Ec represents the first letters of the taxonomical binomial and venom represents the tissue utilised for library construction. The *Echis pyramidum leakeyi* library used the binomial Ep and the *Echis carinatus sochureki* library utilised Es to avoid confusion with *Echis coloratus*. Renamed trace files were parsed through the PartiGene pipeline, beginning with Trace2dbEST (version 2.1.1). Trace2dbEST is an interactive script that processes raw sequencer trace data into quality submissable expressed sequence tags (ESTs) and formats this data into dbEST (EST database) submission files (Parkinson *et al.* 2004). dbEST submission pages were created to provide

information regarding the tissue type and construction methodology of the cDNA library, author contact details, future publication details and specific EST information. Submission pages were created prior to data processing because sequences are labelled with this information as they are processed for ease of future submission to sequence databases. Raw trace chromatograms were processed in Trace2dbEST using the advanced settings; sequences were processed in groups of 96 (representing a single 96 well sequencing plate) for subsequent tracking of cluster membership in PartiGene. Initially, the phred script was utilised and performed trace file base calling to a high accuracy and discrimination level (Ewing *et al.* 1998; Ewing and Green, 1998), facilitating the removal of poor quality sequences. The phred quality cut off was set at 150 high quality bases per sequence; ESTs with less than 150 high quality base pairs were automatically excluded from the dataset. Cross_match was implemented to screen and remove contaminating vector sequences; the vector sequence for the CloneMiner cDNA library vector pDONR222 (Invitrogen, UK) was provided for identification and exclusion. The remaining Trace2dbEST settings were set as default, apart from the trimming of poly(A) tails which was increased to 15 to aid clustering (Wagstaff, S. C., personal communication). The BLAST (basic local alignment search tool) annotation of processed DNA sequences in Trace2dbEST was declined and ESTs were withheld from submission to dbEST at this point.

2.11.2 PartiGene

Trace2dbEST output files were parsed into PartiGene and clustered sequentially into putative gene products (clusters) using a CLOBB (cluster on the basis of BLAST similarity) algorithm (Parkinson, 2002) modified to increase clustering stringencies to 95% (provided by S. C. Wagstaff). The use of this modified algorithm in preference to the standard PartiGene CLOBB algorithm was assessed using a test dataset; the results of this assessment advocating the use of modified CLOBB as the clustering algorithm of choice are described in Chapter 3. ESTs were clustered incrementally with modified CLOBB in order to track the addition of ESTs to clusters as the number of Trace2dbEST processed 96 well plates increase. The placement of ESTs in to clusters containing more than one EST can be used as an

assessment of sequencing coverage – the point where new ESTs are placed in existing clusters rather creating novel clusters implies a representative level of sequencing has been achieved (Wagstaff and Harrison, 2006). The clustered datasets were assembled to produce contiguous sequences derived from the ESTs that represent each cluster. BLAST annotations of contiguous sequences were undertaken against UniProt (v56.2) and TrEMBL (v39.2) protein databases, whilst nucleotide and protein annotations were derived from separate databases containing only Serpentes nucleotide and protein sequences derived from the same UniProt and TrEMBL release versions. Annotated ESTs generated from the four *Echis* species cDNA libraries (including *E. ocellatus* - Wagstaff and Harrison, 2006) were used to construct a postgreSQL database, generated in PartiGene, termed 'all_echis'. The EST sequences generated from the venom gland transcriptomes have been submitted into the dbEST division of the public database GenBank: *E. coloratus* [GenBank: GR947900-GR948969], *E. c. sochureki* [GenBank: GR948970-GR950126] and *E. p. leakeyi* [GenBank: GR950127-GR951204].

2.12 EST identification

Clusters exhibiting significant BLAST annotation (>$1e^{-05}$) with venom toxin families were identified from each venom gland transcriptome using annotation searches of the 'all_echis' postgreSQL database. An example of the SQL command used for these searches was:

SELECT clus_id FROM blast where description like '%phospholipase%';

Toxin specific statistics were subsequently generated for each species venom gland transcriptome by calculating the number of ESTs representing each venom toxin family and expressing them as a percentage of the total number of ESTs and total number of venom toxin ESTs. Clusters identified as non-toxins were assessed individually in order to confirm their putative annotation as non-toxin ESTs; clusters containing >10 ESTs and exhibiting annotations to proteins that are not widely assumed to be involved in cellular biosynthetic processes were noted and analysed for the presence of a signal peptide in SignalP (version 3.0) (Bendtsen *et al.* 2004), implying their secretion in the venom gland. Non-significant BLAST annotated

clusters were assessed for the presence of novel toxin families unique to individual species venom gland transcriptomes; cluster-specific contigs were nucleotide BLAST searched against all other cluster contigs present in the 'all_echis' database, significant hits (>$1e^{-05}$ and a sequence overlap of >42bp) were subsequently analysed. In order to determine if any unidentified clusters represent novel toxin families unique to the genus *Echis*, the Serpentes nucleotide and protein databases used for BLAST annotation were modified to exclude previous sequence information derived from the genus *Echis* and then used for BLAST annotation of the *Echis* venom gland transcriptomes. Bioinformatic searches of the postgreSQL 'all_echis' database were undertaken to identify specific clusters that had hits in the Serpentes databases but not in the Serpentes databases excluding the *Echis*-derived sequences.

The SQL command used was:

```
SELECT where clus_id FROM blast
WHERE db = 'database name including Echis' AND id! = ''
AND clus_id NOT IN (
    SELECT clus_id FROM blast
    WHERE db = 'database name excluding Echis' AND id! = '');
```

2.13 Full length toxin sequencing

ESTs encoding toxin families that represent >4% of total toxin encoding ESTs (snake venom metalloproteinases (SVMP), C-type lectins (CTL), phospholipases A_2 (PLA_2), serine proteases (SP) and cysteine rich secretory proteins (CRISP)) were aligned using CLUSTAL W (Thompson *et al.* 1994) implemented in MEGA (Molecular Evolutionary Genetics Analysis) (version 4.0) (Tamura *et al.* 2007), followed by manual adjustments by eye. Observations of the aligned toxin family datasets and their translated amino acid sequences revealed further sequencing of SVMP and SP ESTs was required in order to achieve full length protein coding sequences. Identical EST clones were excluded in order to remove redundant sequences. Reverse sequencing of all remaining SP ESTs was obtained using

generic M13 reverse primers for sequencing as described previously. The reverse sequencing success rate for the SP datasets was 87%. Individual EST forward and reverse DNA sequences were stitched in SeqMan (LaserGene software suite, www.dnastar.com) to provide full length coding regions; non-homologous base pairs in overlapping regions were correctly assigned according to trace chromatogram quality or marked as unknown (n) if remained undetermined. Due to the quantity of SVMP ESTs present in each venom gland transcriptome (240-405 ESTs), coupled with the size of the maximal SVMP coding region (~1870bp), a modified primer walking strategy was adopted in order to produce full length EST clones. Membership of SVMP clusters were assessed using CLUSTAL W and the viewing interface Jalview (version 2.2.1) (Waterhouse *et al.* 2009) to identify non-identical ESTs that exhibited the presence of the catalytic site (H-box domain – HEXGHXXGXXHD) that characterises metalloproteinases (Fox and Serrano, 2005). ESTs exhibiting the presence of this domain were typically intact at the 5' end and therefore provided the opportunity to derive full length sequencing of the coding region. A total of 439 SVMP clones from the four *Echis* species were selected for further sequencing. In order to provide full length reads of the SVMPs two primer sites were required. Primer design was carried out using the generated DNA alignment and the primer design program PrimerSelect (LaserGene software suite, www.dnastar.com). Due to the sequence variation observed between the different sub-classes of SVMPs the primer sites were designed at conserved domains (Figures 2.6-2.7). Selected clones were prepared for sequencing by ice crystal picking the relevant 96 well of the original sequenced cDNA library plates with a pipette tip before incubation in 150µl LB broth (containing 8% glycerol and 50µl/ml of kanamycin) for 10-15 minutes. Plates containing LB broth were covered and incubated at 37°C overnight, sealed with self-adhesive plate sealers and stored at -80°C prior to sequencing. Full length SVMP nucleotide sequences were derived from the multiple reads (original sequencing and reads derived from primer 1 and primer 2) as described for the SP ESTs. The sequencing success rate was calculated as 94%.

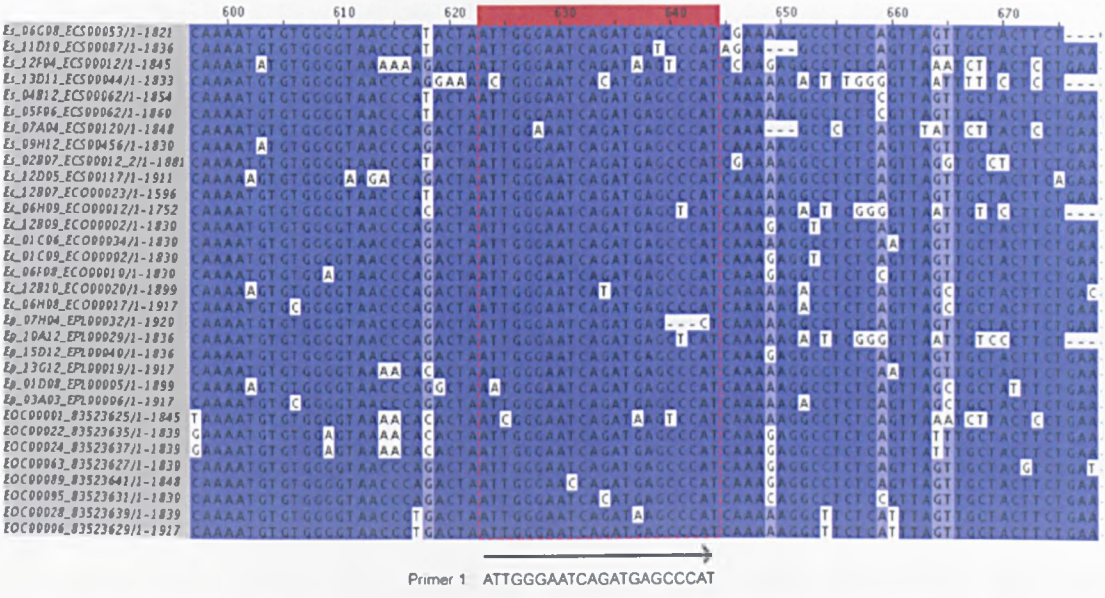

Figure 2.6. *Echis* SVMP alignment highlighting the first primer design site.

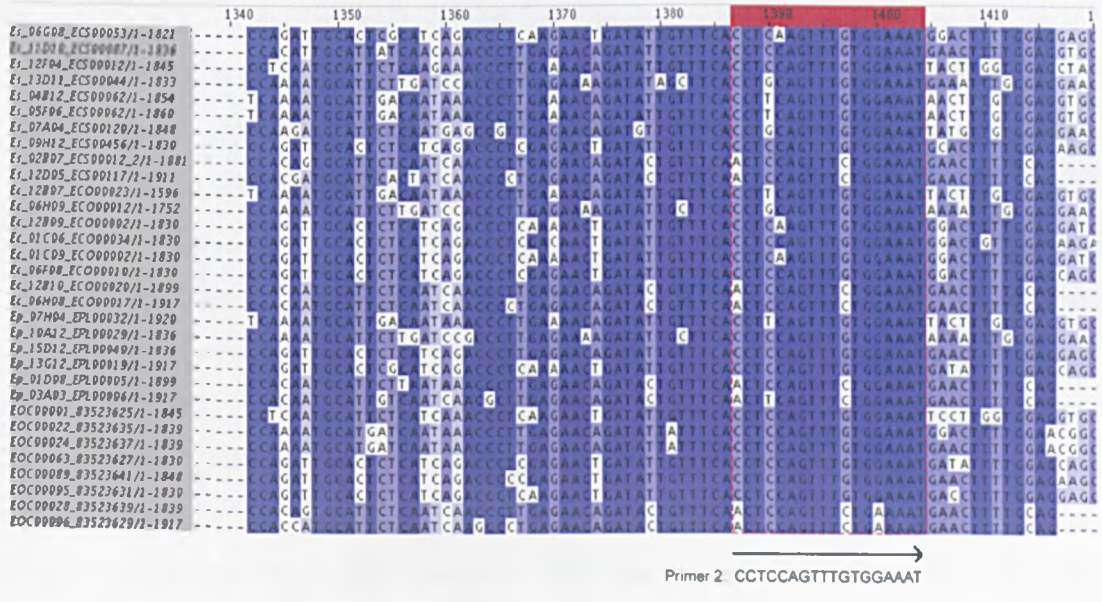

Figure 2.7. *Echis* SVMP alignment highlighting the second primer design site.

2.14 Ethical declaration

All animal experimentation conducted during the course of this work was undertaken using standard protocols approved by the University of Liverpool Animal Welfare Committee and performed with the approval of the UK Home Office under project licence #40/3216.

CHAPTER 3

Clustering expressed sequence tags: assessments of CLOBB2 (cluster on the basis of BLAST similarity) and modified CLOBB algorithms reveals substantial diversity in venom gland derived EST cluster formation

3.1 Abstract

The generation of expressed sequence tags (ESTs) from cDNA libraries provide a cost-effective discovery method that produces a wealth of molecular data for tissue-specific gene discovery. A fundamental step of bioinformatic processing of ESTs is clustering, where ESTs are grouped into putative gene objects based upon sequence similarity. Clustering using the CLOBB (cluster on the basis of BLAST similarity) algorithm has previously demonstrated advantages over alternative clustering methodologies, including the rejection of chimeric clusters and the recording of cluster merging and splitting events as incremental additions of ESTs occur. Here, the clustering integrity of two differing CLOBB algorithms (CLOBB2 and CLOBB modified to increase clustering stringencies to 95%) are assessed using a test snake venom gland derived EST dataset to determine the most efficient method to generate venom gland transcriptome profiles. Clustering was assessed using a number of bioinformatic tools including CLUSTAL W, PHRAP, PreGap4 and Gap4 alongside manual analysis. Modified CLOBB demonstrated increased clustering stringencies over CLOBB2, leading to an increase in the number of singleton and clusters containing more than one EST and a reduction in the size of the largest cluster. Analysis of cluster TES00002 demonstrated that Modified CLOBB provided the optimum agreement of EST clustering with manual analysis and conferred increased EST discrimination compared to CLOBB2. Efficiency assessments of a clustering method are fundamental for the production of transcriptomic data; efficient clustering underpins the integrity of a dataset and the conclusions that are drawn thereafter. These results strongly support the use of Modified CLOBB as the optimal algorithm for clustering snake venom gland derived ESTs.

3.2 Introduction

The construction of cDNA libraries coupled with the generation of expressed sequence tags provides a wealth of molecular data adequate for the creation of a partial genome or organ/tissue-specific gene discovery (e.g. Adams *et al.* 1991; Wagstaff and Harrison, 2006). This highly cost-effective discovery method provides a representative fraction of the genes present in the starting material, although the generation of redundant and partial sequence data provides downstream data management challenges (Parkinson *et al.* 2002). Bioinformatic processing of cDNA library generated ESTs facilitates the exclusion of poor quality sequences and contaminating vector and adaptor sequences, before clustering ESTs into putative gene objects in order to manage sequence redundancy (e.g. Parkinson *et al.* 2002, 2004). Subsequently, contiguous sequences (contigs) of clustered gene objects are generated and typically annotated via BLAST (basic local alignment search tool) similarity to existing annotated sequences present in multiple DNA and protein databases (e.g. Parkinson *et al.* 2004). A critical step in the processing of EST data is that of clustering. The process of clustering is fundamental for the generation of a replicable, manageable dataset; efficient clustering underpins the integrity of a dataset and the conclusions that are drawn thereafter. For example, if EST sequences are incorrectly grouped into multiple clusters that are homologous, duplication of the data occurs. Conversely, incorrect cluster placement of non-homologous ESTs produces a loss of data; single BLAST annotations are provided for each cluster based on homology to a cluster's generated contiguous sequence, thereby masking the sequence variation present in incorrectly placed non-homologous ESTs. Furthermore, obtaining an appropriate clustering stringency is fundamental for down-stream sequence analysis; the generation of large clusters containing gene products from similar genes or multi-locus gene families is often undesirable for subsequent data manipulation, whilst excessive increases in cluster stringency can produce multiple clusters which separate polymorphic homologous genes based on minimal base pair differences, leading to the creation of unwarranted novel clusters. Obtaining the optimal clustering stringency of an algorithm is highly desirable, whilst the ability to appropriately modify an algorithm to variable EST datasets is particularly advantageous.

There are a number of bioinformatic clustering tools available for EST datasets generated by non-'next generation' sequencing technologies, including primitive scripts which run and parse the results of sequence database searches, e.g. REX (Yee and Conklin, 1998), INCA (Graul and Sadée, 1997) and SEALS (Walker and Koonin, 1997) and programs which operate on non-alignment based algorithms, e.g. d2_cluster (Burke *et al.* 1999). Separate from these stand alone resolutions are dedicated database systems which have been implemented to process EST databases using gene indices, these include, UniGene (Boguski and Schuler, 1995) and TIGR (Adams *et al.* 1995; Sutton *et al.* 1995; White and Kervalage, 1996; Pertea *et al.* 2003). The UniGene system operates by analysing pair-wise comparisons of mRNAs and genomic DNA fragments before matching by similarity. TIGR uses a gene indices system created by WU-BLAST (Altschul *et al.* 1990); this analysis is also based upon a series of pair-wise comparisons, EST sequences are grouped together if they share 95% sequence similarity over 40 base pairs, before the sequence data is subjected to a round of clustering by the program CAP3 (Huang, 1996; Huang and Madan, 1999) to generate initial consensus sequences. In contrast, the PartiGene pipeline uses the PERL scripting language to drive a fully automated, integrated pipeline consisting of three major scripts which are fully customisable; Trace2dbEST processes raw trace sequences, PartiGene incorporates the clustering algorithm CLOBB (cluster on the basis of BLAST similarity) and BLAST for deriving sequence annotations, whilst Prot4EST derives peptide predictions from the processed EST sequences (Parkinson *et al.* 2002; 2004; Wasmuth and Blaxter, 2004).

Comparison analyses of these clustering methodologies determined that CLOBB and TIGR produce datasets with both a greater number of clusters and singletons than UniGene, thereby implying these algorithms are more discriminating (Parkinson *et al.* 2002). In addition, the CLOBB algorithm was shown to be more capable at finding potential matches for sequences than the TIGR algorithm (Parkinson *et al.* 2002). Another advantage of CLOBB is the way it deals with large clusters termed 'superclusters'. Clustering algorithms produce results that vary according to the order in which sequences are added due to the unidirectional order in which

sequences are processed (Parkinson *et al.* 2002), i.e. further growth of a cluster is reliant on which sequences it encounters next. The problems that can arise from this include the formation of superclusters from the merging of two unsuitable clusters via an intermediate chimeric sequence (Parkinson *et al.* 2002). The CLOBB algorithm prevents the merging of these clusters and automatically identifies these issues (Parkinson *et al.* 2002). Although this action leads to an increased division of related clusters, and hence putative genes, compared to other methods, it is preferable to have two or more related clusters than a chimeric cluster. Post-CLOBB assembly uses the supercluster information to allow the merging of related clusters manually prior to assembly and therefore reduce the number of putative genes to a more acceptable level (Parkinson *et al.* 2002).

The CLOBB algorithm clusters processed ESTs into groups of putative gene objects according to BLAST similarity (Parkinson *et al.* 2002). Once ESTs are assigned to clusters of putative gene objects, consensus contiguous sequences can be derived which increase both the length and overall sequence quality of a transcript (Parkinson *et al.* 2002), therefore reducing the common problems of reliability associated with EST datasets. Initially, CLOBB reads ESTs individually and compares them to the current cluster database using BLASTN. The BLAST output is subsequently parsed for high-scoring segment pairs (HSPs); those with a sequence identity of >95% and an overlap length of >30 base pairs are designated as type I matches, whilst those with a sequence identity of <95% are placed into a new cluster (Parkinson *et al.* 2002). Type I matches are subsequently checked for integrity before being further characterised into type II or type III matches. Type II matches occur when sequences do not contain high quality overlap extensions of more than 30 base pairs beyond the HSPs, whilst type III matches are assigned when high quality extensions do occur (Parkinson *et al.* 2002). Cluster assignment then checks the identified type II and type III matches to ensure that no conflicts arise; if a sequence forms both a type II and type III match with different members of a particular cluster then the query sequence is assigned to a new cluster to prevent the creation of undesirable chimeric clusters (Parkinson *et al.* 2002). Clustering conflict can arise via multiple type II matches with distinct clusters. If the HSPs of these matches occur in overlapping regions the query sequence is likely to be a spliced

variant of one gene or a closely related member of a gene family; as such the sequence is assigned to the cluster with the highest BLAST score and noted as a 'supercluster' for subsequent manual analysis (Parkinson *et al.* 2002). When the HSPs of the matching sequences do not occur in overlapping regions the query links the clusters together and advocates merging the two clusters (Parkinson *et al.* 2002). Following the resolution of EST clustering, the sequence database contains two types of generated clusters; those containing one EST (singleton clusters or singletons) and those containing more than one EST (clusters). One of the most useful features of CLOBB algorithms is the presence of multiple variables that can be easily modified; factors such as minimum length of HSP, maximum allowable non-HSP overlap and percentage identity in overlap, can be tuned to produce the most satisfactory clustering results for any particular dataset (Parkinson *et al.* 2002). Furthermore, CLOBB records any merging or splitting events as the number of ESTs in a dataset increases, allowing the cluster membership of a dataset to be tracked as incremental additions of sequences occur (Parkinson *et al.* 2002).

In order to assess the optimal CLOBB-derived clustering methodology for venom gland generated EST data, two CLOBB algorithm variants were used to cluster 1440 ESTs generated from the *Echis coloratus* (Serpentes: Viperidae) venom gland cDNA library. The CLOBB2 algorithm, an unmodified, standard clustering script pre-installed with PartiGene on Bio-Linux v4.0, was compared to a modified CLOBB algorithm. Modified CLOBB is a variant of the original CLOBB algorithm pre-installed on the Bio-Linux v4.0 predecessor Bio-Linux v3.0. This script was provided by S. C. Wagstaff pre-modified to increase sequence clustering stringencies to 95%. Modified CLOBB was demonstrated to significantly increase clustering stringency over the original CLOBB algorithm for venom gland generated EST data (Wagstaff, S. C., personal communication), and was therefore implemented for bioinformatic processing of the *E. ocellatus* venom gland transcriptome (Wagstaff and Harrison, 2006). In order to evaluate whether the modified CLOBB script retains an increase in cluster stringency over the newer CLOBB2 script, bioinformatic processing of a venom gland generated EST dataset was undertaken using both clustering scripts prior to comparative analysis in order to determine the optimal clustering strategy for constructing venom gland transcriptome databases.

3.3 Methods

The venom gland cDNA library was constructed from ten wild-caught specimens of *Echis coloratus* (Egypt), using identical protocols to those described for the construction of the *E. ocellatus* venom gland cDNA library (Wagstaff and Harrison, 2006). 1440 clones from the cDNA library were picked randomly and sequenced (NERC Molecular Genetics Facility, UK) using M13 forward primers. Bioinformatic processing was carried out using the PartiGene pipeline (www.nematodes.org). Sequences were processed (to exclude low quality, contaminating vector sequences and poly A+ tracts) using Trace2dbEST v.2.1.1 (Parkinson *et al.* 2004). The Trace2dbEST output was parsed through PartiGene v3.0 incorporating the PERL language clustering script CLOBB2 (Parkinson *et al.* 2002). The generated CLOBB2 output was given the three letter library identifier TES (test). The Trace2dbEST output was subsequently processed through PartiGene v2.2.0, using identical parameters to PartiGene v3.0 apart from the implementation of the modified CLOBB script (provided by Wagstaff, S. C.) to cluster the raw sequences. The library identifier provided for modified CLOBB-processed ESTs was ECO (*Echis coloratus*). The clustering results of both scripts were analysed using the multiple alignment program CLUSTAL W (v1.82) (Thompson *et al.* 1994) and the pre-genome assembly program PreGAP4 (v1.5) (Bonfield *et al.* 1995) by implementing PHRAP (Green, 1995) under standard settings. The CLUSTAL W output was analysed using the Jalview (v2.2.1) viewing interface (Waterhouse *et al.* 2009), whilst PreGAP4 output was analysed in GAP4 (v4.10) (Bonfield *et al.* 1995) by invoking the join editor. Manual analysis of DNA sequence variation was assessed by the number of base pair differences between, (i) contigs and (ii) individual ESTs and contigs, over the length of the contiguous sequence and expressed as a percentage.

3.4 Results and Discussion

3.4.1 CLOBB2 versus modified CLOBB

Trace2dbEST sequence processing produced 1070 high quality submissible EST sequences. The CLOBB2 script clustered ESTs into a total of 389 clusters, of which 291 clusters were singleton clusters (Table 3.1). 98 clusters containing more than

one member were derived from 779 ESTs; the average cluster size (excluding singleton clusters) was 7.95 transcripts per cluster. The largest cluster, TES00002, was noted due to its size (131 ESTs), representing 12.24% of total ESTs. The remaining clusters ranged in size between two and fifty-eight transcripts per cluster. In contrast, the modified CLOBB script clustered the dataset into 425 clusters, of which 324 clusters were singletons (Table 3.1). 101 clusters, derived from 746 sequences, contained more than one transcript yielding an average cluster size of 7.39 ESTs. Cluster sizes ranged from two ESTs to the largest cluster, ECO00011, which contained 60 transcripts. The modified CLOBB algorithm appeared to cluster ESTs more stringently than CLOBB2; fewer sequences are included in clusters containing >1 EST, leading to an increase in singleton clusters (Table 3.1). I hypothesise that singleton ESTs were excluded from non-singleton clusters because of increased clustering stringencies based on a lack of significant sequence similarity. However, it remains unclear whether: (i) the CLOBB2 algorithm over-clusters the ESTs, (ii) the modified CLOBB algorithm under-clusters or (iii) whether a combination of both factors is occurring. In order to determine the factors responsible for the variation in cluster distribution generated by the two CLOBB algorithms, the largest cluster created by CLOBB2, TES00002, was analysed using multiple DNA sequence alignment and clustering tools.

	CLOBB2	Modified CLOBB
Number of ESTs	1070	1070
Number of singleton clusters	291	324
Number of clusters (ESTs >1)	98	101
Number of ESTs that form clusters >1	779	746
Average cluster size	7.95	7.39
EST size of largest cluster	131	60

Table 3.1. A comparison of the clustering statistics produced by the two CLOBB algorithms.

3.4.2 PHRAP analysis of cluster TES00002

Cluster TES00002 was generated from the PartiGene output by the PERL script CLOBB2. Members of the cluster were aligned in CLUSTAL W and viewed in Jalview. The majority of ESTs exhibited a high sequence identity apart from 14 ESTs which aligned separately. This group of ESTs exhibited poor sequence similarity to the remaining members of TES00002 (117 ESTs) and induced multiple insertions in the CLUSTAL W alignment. In order to validate the membership of TES00002, members of the cluster were parsed into PreGAP4, clustered using PHRAP and viewed in GAP4. Surprisingly, PHRAP clustered members of TES00002 into four disparate contiguous sequences of varying length (Table 3.2). Manually viewing the membership of the contigs revealed the ESTs that failed to align in CLUSTAL W were transcripts that began coding a minimum of 250bp from the 5' end of the gene. Whilst CLUSTAL W was unable to align these sequences due to a lack of sequence overlap, GAP4 demonstrated these transcripts were homologous to the remaining ESTs in the cluster.

	Number of ESTs	Length of contiguous sequence (bp)
Contig 1	5	822
Contig 2	23	1034
Contig 3	44	921
Contig 4	59	1974

Table 3.2. A comparison of the number of ESTs and the length of coding sequence of four contiguous sequences created by PreGAP4 analysis of cluster TES00002.

3.4.3 Inter-contig comparisons

The separation of cluster TES00002 into four separate contiguous sequences strongly implies over-clustering by the CLOBB2 algorithm. To confirm this theory the join editor in GAP4 was invoked to align the four generated contigs for manual analysis of sequence similarity based upon base pair and base pair percentage differences (Table 3.3). The smallest base pair percentage difference (3.66%) was observed between contig 3 and contig 4; this equated to 33 base pair substitutions or insertion/deletions (indels) over the entire contig. In contrast, the greatest difference,

between contig 2 and contig 4 was substantially higher (9.62% - 100bp). An average difference of 6.38% (57bp) was found between the four contiguous sequences. The variation observed between the four contigs is sufficiently divergent to advocate their separation into distinct clusters; the implication is that the CLOBB2 algorithm is ineffective at splitting clusters in the absence of a >10% sequence similarity difference is notable.

	Percentage difference between contiguous sequences				
Base pair difference between contiguous sequences		Contig 1	Contig 2	Contig 3	Contig 4
	Contig 1	-	5.54	5.90	7.14
	Contig 2	46	-	6.42	9.62
	Contig 3	49	58	-	3.66
	Contig 4	59	100	33	-

Table 3.3. The base pair percentage and base pair differences between the four contiguous sequences created by PreGAP4 and GAP4 analysis of cluster TES00002.

3.4.4 Intra-contig comparisons

The membership of individual contigs generated by PreGAP4 were analysed in order to check the membership integrity of individual contigs and to determine whether sufficient intra-contig sequence variation exists to justify further separation of EST contig membership; analyses were undertaken to test the clustering stringency implemented by PHRAP in PreGAP4. Contig 1 contained five ESTs exhibiting minimal base pair variation; in total only four base pairs differed from the consensus sequence and all occurred within 30bp of the 3' end of the EST which may be more susceptible to base calling errors due to a reduction in trace sequence quality. The maximum percentage base pair difference observed was 0.27%, thereby justifying the membership of this contig.

Contig 2 contained 23 ESTs and exhibited considerable variation compared to contig 1. The greatest sequence variation observed was 15bp and 9bp equating to a 2.72% and 1.66% difference. The remaining sequences exhibited a sequence identity of at

least 99.35%. The average percentage difference between members of the contig and the contiguous sequence was 0.38%; therefore the exclusion of the two ESTs exhibiting the greatest variation (04B03 and 11B12) from the cluster could be justified. Transcript 04B03 displayed four sequence alignment insertions when compared to other members of the contig, whilst transcript 11B12 exhibited nine unique base pair substitutions. The lack of similarity between these two transcripts advocates their separation into individual singleton clusters.

The average base pair percentage difference between the 44 members of contig 3 was 0.45%. Six ESTs were determined to have >1% base pair divergence from the consensus sequence, with five of the transcripts (01A03, 01C08, 04B02, 09F07 and 09G08) displaying nine identical base pair substitutions, indicating their non-homology to other members of the contig. However, transcript 01A03 contains an additional seven base pair substitutions that were not observed in the remaining variants, representing a further 1.2% base pair divergence to the other contig 3 variants. These observations advocate the placement of transcript 01A03 into a novel singleton cluster and the remaining four ESTs (01C08, 04B02, 09F07 and 09G08) into a separate novel cluster. The final transcript (04G10) exhibited substantial variation from the contiguous sequence (11bp substitutions - 2.07%); two substitutions were homologous to the transcripts described above, whilst nine are unique within contig 3, advocating the placement of 04G10 into a novel singleton cluster.

Contig 4 contains 59 sequences which vary from the consensus sequence by an average of 0.19% (2.07bp). This is the lowest amount of intra-contig sequence variation found in any of the contigs containing more than 10 ESTs (Table 3.4). However, four sequences differ from the consensus sequence by greater than 1%; the highest base pair variation observed represented 1.46%. Upon visual analysis of these variants, only transcripts 13F04 and 02F12 exhibited distinct variation that justified removal from the contig. Transcript 13F04 exhibited four unique base pair substitutions and created two insertions (1.39%) in the consensus sequence, whilst transcript 02F12 displayed three unique base pair substitutions and created two

insertions (1.17%). As these transcripts do not exhibit homology to each other, their removal from contig 4 into singleton clusters is advocated. The remaining ESTs exhibiting >1% base pair difference displayed the majority of variation at the 3' end of the EST sequence. For this reason further separation of the contig cannot be advocated until primer walking the 3' end of the ESTs of interest determines the quality of base calling. Interestingly, the remaining variation in contig 4 focuses primarily around five base pair positions. The majority (31) of the ESTs contain the base pair motif G-T-A-G-G at these sites. However, eight ESTs exhibit four substitutions at these base pair positions exhibiting the motif C-C-T-G-T, whilst six ESTs exhibit two base pair substitutions yielding the motif C-T-A-T-G. The remaining fourteen sequences begin downstream of these base pair positions and can therefore not be classified by this motif variation. For this reason further splitting of contig 4 cannot be advocated until full length sequence data is generated.

	Contig 1	Contig 2	Contig 3	Contig 4
Number of ESTs	5	23	44	59
Average EST difference from consensus (base pairs)	0.80	2.20	2.52	2.07
Average EST difference from consensus (%)	0.11	0.38	0.45	0.19
Greatest EST difference from consensus (%)	0.27	2.72	3.85	1.46

Table 3.4. Summary statistics from manual contiguous sequence analysis in GAP4.

Manual analysis of the four TES00002 contiguous sequences generated by PHRAP in PreGAP4 infers the presence of five distinct clusters (ESTs >1), with one cluster containing three isoforms which may be distinct gene products (Table 3.5). In addition, six ESTs should be separated into individual singleton clusters on the basis of dissimilarity.

	Summary of manual contiguous sequence analysis
Contig 1	Contig resolves completely into one cluster.
Contig 2	Contig resolves into one cluster containing 21 sequences and two singleton clusters (11B12 and 04B03).
Contig 3	Contig resolves into two clusters containing 38 and four sequences and two singleton clusters (04G10 and 01A03).
Contig 4	Contig resolves into one cluster containing three distinct isoforms and two singleton clusters (13F04 and 02F12). Splitting of the cluster into three isoform clusters may be advocated following extensive sequencing.

Table 3.5. A summary of recommendations following manual intra-contig sequences analysis of cluster TES00002 in PreGAP4 and GAP4.

3.4.5 Modified CLOBB

An initial overview of the statistics generated following the clustering of *E. coloratus* derived ESTs implied the modified CLOBB algorithm exhibited increased clustering stringencies compared to CLOBB2; increases in the number of singleton clusters and clusters containing >1 EST were observed (Table 3.1). The largest cluster in the ECO EST database contained 60 transcripts, indicating that cluster TES00002, generated by CLOBB2, has been split into two or more clusters. Manual analysis of the EST clone identifiers revealed that modified CLOBB split cluster TES00002 into five clusters (ECO00011, 00017, 00020, 00044 and 00047) and six singleton clusters. Manual analysis was undertaken to correlate cluster memberships generated by modified CLOBB with PreGAP4 generated contigs and their subsequent manual analysis (Table 3.6).

Cluster ECO00011 contained 60 ESTs, including 57 of the 59 ESTs found in contig 4 generated by PreGAP4. The remaining two ESTs from contig 4 (13F04 and 02F12) were identified in individual singleton clusters (ECO00072 and ECO00421), as advocated by manual analysis. Interestingly, modified CLOBB clustered three additional transcripts (07C12, 08C05 and 08C04), which were absent in TES00002, into this cluster; two of these ESTs are identical to the consensus sequence, whilst

07C12 differs by three base pair substitutions (0.49%) towards the 3' end of the sequence. Given the high level of sequence identity between these additional ESTs and the consensus sequence of contig 4, it is not apparent why they were originally excluded from cluster TES00002 by CLOBB2. Notably, the isoforms observed in contig 4 by manual analysis are all retained within cluster ECO00011.

CLOBB2	PHRAP in PreGAP4	Manual analysis in GAP4	Modified CLOBB
TES00002 (*131*)	Contig 1 (*5*)	One cluster (*5*)	Cluster ECO00047 (*5*)
	Contig 2 (*23*)	One cluster (*21*) Two singletons (04B03 and 11B12)	Cluster ECO00017 (*27*) Singletons ECO00174 (04B03) and ECO00391 (11B12)
	Contig 3 (*44*)	Two clusters (*38* and *4*) Two singletons (01A03 and 04G10)	Clusters ECO00020 (*38*) and ECO00044 (*4*) Singletons ECO00058 (01A03) and ECO00150 (04G10)
	Contig 4 (*59*)	One cluster (*57*) Two singletons (13F04 and 02F12)	Cluster ECO00011 (*60*) Singletons ECO00072 (13F04) and ECO00421 (02F12)

Table 3.6. A summary of the membership of cluster TES00002 using multiple different clustering methods. The number of ESTs present in each cluster are italicised in parentheses.

Cluster ECO00017 contained 27 ESTs, including 21 of the 23 transcripts observed in contig 2. The remaining two members of contig 2 (04B03 and 11B12) were separated from this cluster by modified CLOBB, as suggested by manual analysis, into singleton clusters (ECO00174 and ECO00391). As in cluster ECO00011, the modified CLOBB algorithm has identified and clustered additional ESTs that were not present in the CLOBB2 generated cluster TES00002. The additional six ESTs

present in ECO00017 are 5' truncated and extend the 3' end of the consensus sequence by ~700bp. The overlap between these ESTs and the consensus sequence is 130bp and no sequence variation was observed in the overlapping region, thereby strongly supporting their cluster membership as homologous ESTs. These observations imply the overlapping region observed is insufficient for CLOBB2 clustering; this is particularly surprising considering CLOBB2 appears to be less stringent than modified CLOBB.

Contig 3 (44 ESTs) generated by PreGAP4 completely resolved into clusters ECO00020 and ECO00044 and two singleton clusters (ECO00058 and ECO00150). Cluster ECO00020 contains the majority of ESTs observed in contig 3 (38 ESTs) and contains no additional sequences as observed in the modified CLOBB generated clusters ECO00011 and ECO00017. The remaining six sequences present in contig 3 were split by modified CLOBB into a separate cluster containing four transcripts (ECO00044) and two singleton clusters ECO00058 and ECO00150. Cluster ECO00044 contains the sequences that manual analysis advocated separating from the contig on the basis of nine identical base pair substitutions (01C08, 04B02, 09F07 and 09G08). Notably, transcript 01A03 was not included in this cluster by modified CLOBB and has been assigned its own singleton cluster (ECO00058), as previously suggested by manual analysis, despite the observed sequence variation it shared with members of cluster ECO00044. Modified CLOBB placed the remaining EST (04G10) into a singleton cluster (ECO00150) as advocated by manual analysis. Cluster ECO00047 exhibited an identical cluster membership to contig 1 generated by PreGAP4 analysis of cluster TES00002. PreGAP4, manual analysis and modified CLOBB all advocated the separation of this cluster, strongly suggesting these transcripts are non-homologous to the remaining members of cluster TES00002.

3.5 Conclusions

The use of bioinformatic algorithms to process cDNA library generated EST data is particularly valuable. The CLOBB algorithms have a number of advantages over other clustering algorithms, including the rejection of chimeric clusters, recording cluster merging or splitting events as incremental additions of ESTs occur and an

easily modifiable, freely available, script (Parkinson *et al.* 2002). Previous analyses between the original CLOBB algorithm and CLOBB modified to increase clustering stringencies to 95% demonstrated modified CLOBB increased the clustering proficiency amenable for venom gland derived EST data (Wagstaff, S. C., personal communication). In order to determine whether the proficiencies generated by modified CLOBB remain following the creation of the CLOBB2 algorithm, a direct clustering comparison of *E. coloratus* venom gland ESTs was undertaken. The modified CLOBB algorithm provided the optimum agreement of predicted open reading frames out of the methods undertaken. In comparison, the CLOBB2 script proved to be less effective at manipulating venom gland derived ESTs into distinct, putative gene products. Decreases in the number of clusters and singleton clusters, combined with an increase in the average number of ESTs present in each cluster implied CLOBB2 was less discriminatory and over-clustered the *E. coloratus* dataset. Automated and manual analysis, facilitated by PHRAP, PreGAP4 and GAP4 confirmed this hypothesis and demonstrated the proficiency of modified CLOBB to create stringent clusters in general accordance with manual analysis.

The integrity of an automated clustering method is fundamental to producing a reliable dataset for down-stream sequence analysis. Whilst manual analysis of EST derived sequence data is useful for analysing the proficiency of a clustering mechanism, it is not viable for large scale sequence processing; an automated system provides a significant reduction in time and reduces experimental bias. Accurate clustering of ESTs, as demonstrated by modified CLOBB, separates sequence data into manageable putative gene objects for subsequent sequence analysis. An optimal clustering stringency partitions the dataset into an optimal number of gene objects, which in turn increases the accuracy of BLAST-derived cluster-specific annotations; non-homologous ESTs are separated into novel clusters, thereby providing diversity in cluster-derived consensus sequences. The successful clustering of venom gland derived ESTs generated from *E. coloratus*, alongside previous analyses with *E. ocellatus* venom gland data (Wagstaff and Harrison, 2006), strongly support the use of modified CLOBB as the optimal algorithm for clustering snake venom gland derived ESTs.

CHAPTER 4

Comparative venom gland transcriptome surveys of the saw-scaled vipers (Viperidae: *Echis*) reveal substantial intra-family gene diversity and novel venom transcripts

4.1 Abstract

Venom variation occurs at all taxonomical levels and can impact significantly upon the clinical manifestations and efficacy of antivenom therapy following snakebite. Variation in snake venom composition is thought to be subject to strong natural selection as a result of adaptation towards specific diets. Members of the medically important genus *Echis* exhibit considerable variation in venom composition, which has been demonstrated to co-evolve with evolutionary shifts in diet. A venom gland transcriptome approach was adopted to investigate the diversity of toxins in the genus and elucidate the mechanisms which result in prey-specific adaptations of venom composition.

Venom gland transcriptomes were created for *E. pyramidum leakeyi*, *E. coloratus* and *E. carinatus sochureki* by sequencing ~1000 expressed sequence tags from venom gland cDNA libraries. A standardised methodology allowed a comprehensive intra-genus comparison of the venom gland profiles to be undertaken, including the previously described *E. ocellatus* transcriptome. BLAST annotation revealed the presence of snake venom metalloproteinases, C-type lectins, group II phopholipases A_2, serine proteases, L-amino oxidases and growth factors in all transcriptomes throughout the genus. Transcripts encoding disintegrins, cysteine-rich secretory proteins and hyaluronidases were obtained from at least one, but not all, species. A representative group of novel venom transcripts exhibiting similarity to lysosomal acid lipase were identified from the *E. coloratus* transcriptome, whilst novel metallopeptidases exhibiting similarity to neprilysin and dipeptidyl peptidase III were identified from *E. p. leakeyi* and *E. coloratus* respectively.

The comparison of *Echis* venom gland transcriptomes revealed substantial intrageneric venom variation in representations and cluster numbers of the most

abundant venom toxin families. The expression profiles of established toxin groups exhibit little obvious association with venom-related adaptations to diet described from this genus. I therefore hypothesise that alterations in isoform diversity or transcript expression levels within the major venom protein families are likely to be responsible for prey specificity, rather than differences in the representation of entire toxin families or the recruitment of novel toxin families, although the recruitment of lysosomal acid lipase as a response to vertebrate feeding cannot be excluded. Evidence of marked intrageneric venom variation within the medically important genus *Echis* strongly advocates further investigations into the medical significance of venom variation in this genus and its impact upon antivenom therapy.

4.2 Introduction

Snake venoms contain a complex mix of components, with biologically active proteins and peptides comprising the vast majority (Aird, 2002). Variation in the composition of venom occurs at several taxonomical levels in multiple snake lineages (reviewed in Chippaux *et al.* 1991; Gutiérrez *et al.* 2009). The view that variation in venom composition evolves primarily through neutral evolutionary processes (Sasa, 1999a, 1999b; Mebs, 2001) is not supported by other reports that snake venom composition is subject to strong natural selection as a result of adaptation towards specific diets (e.g. Daltry *et al.* 1996a; Kordiš and Gubenšek, 2000; Jorge-da-Silva and Aird, 2001). Since the primary role of venom is to aid prey capture (Chippaux *et al.* 1991), it is perhaps unsurprising that variation in the protein composition of venom has been associated with significant dietary shifts in a number of genera (Jorge-da-Silva and Aird, 2001; Barlow *et al.* 2009; Creer *et al.* 2003; Sanz *et al.* 2006). Irrespective of the evolutionary forces underpinning venom protein composition, variation in venom components can significantly impact upon the clinical manifestations of snake envenoming (Warrell, 1989; Prasad *et al.* 1999; Shashidharamurthy *et al.* 2002) and, because the clinical efficacy of an antivenom may be largely restricted to the venom used in its manufacture, the success of antivenom therapy (Theakston *et al.* 1989; Galán *et al.* 2004; Visser *et al.* 2008).

Envenoming by saw-scaled viper (Viperidae: *Echis*) species is thought to be responsible for more snakebite deaths worldwide than any other snake genus (Warrell *et al.* 1977). Envenomed victims typically suffer a combination of systemic and local haemorrhagic symptomatologies and up to 20% mortality rates without antivenom treatment (Warrell *et al.* 1977, Warrell, 1995; Habib *et al.* 2001). Whilst the clinical symptoms are largely consistent throughout this widely distributed genus (Warrell, 1995), cases of incomplete intrageneric antivenom efficacy have been documented, implying substantial inter-species venom variation (Gillissen *et al.* 1994; Kochar *et al.* 2007; Visser *et al.* 2008; Warrell, 2008). The four species complexes making up this genus, the *E. carinatus*, *E. ocellatus*, *E. pyramidum* and *E. coloratus* species groups (Barlow *et al.* 2009; Pook *et al.* 2009), exhibit considerable vertebrate or invertebrate dietary preferences, *E. coloratus* being a vertebrate specialist whereas invertebrates feature prominently in the diet of the others (Barlow *et al.* 2009). Since the proportions of consumed invertebrates correlated strongly with alterations in venom toxicity to scorpions, the toxicity of the venom from these species appears to have co-evolved alongside evolutionary shifts in diet (Barlow *et al.* 2009). A preliminary venom protein analysis using reduced SDS-PAGE failed to identify an obvious link between venom composition and diet (Barlow *et al.* 2009), justifying the use of a more comprehensive venom composition analysis in order to elucidate the mechanisms driving venom adaptations within the *Echis* viper genus.

Based on earlier work with *E. ocellatus* (Wagstaff and Harrison, 2006), a comparative venom gland transcriptome approach was elected and venom gland cDNA libraries from *E. coloratus*, *E. pyramidum leakeyi* and *E. carinatus sochureki* were generated. Together with the existing *E. ocellatus* database, these provided DNA sequence data representing the venom gland transcriptomes for each of the four major species groups within the genus. The production of multiple *Echis* venom gland expressed sequence tag databases (vgDbEST) provided an unbiased overview of the transcriptional activity during venom synthesis in the venom glands of four species in this genus. This, the first comprehensive compilation of venom gland transcriptomes of congeneric snake species, was then interrogated to determine whether the mechanisms resulting in prey-specific adaptation of venom composition

involve (i) the recruitment of novel prey-specific venom toxin transcripts, (ii) major changes in the expression levels of established toxin families, (iii) the diversification of functional isoforms within established toxin families or (iv) a combination of these factors.

4.3 Methods

Venom gland cDNA libraries were constructed from ten wild-caught specimens of *Echis coloratus* (Egypt), *E. p. leakeyi* (Kenya) and *E. c. sochureki* (Sharjah, UAE), maintained in the herpetarium of the Liverpool School of Tropical Medicine, using identical protocols described for the construction of the venom gland cDNA library from *E. ocellatus* (Wagstaff and Harrison, 2006). Clones from the cDNA libraries were picked randomly and sequenced (NERC Molecular Genetics Facility, UK) using M13 forward primers.

Bioinformatic processing was carried out using the PartiGene pipeline with the same protocols used previously (Wagstaff and Harrison, 2006). Briefly, sequences were processed (to exclude low quality, contaminating vector sequences and poly A+ tracts) using Trace2dbEST (Parkinson *et al.* 2004). Subsequently, assembly was undertaken in PartiGene version 3.0, using high stringency clustering parameters (Parkinson *et al.* 2004; Wagstaff and Harrison, 2006). A total of 1070 (*E. coloratus*), 1078 (*E. p. leakeyi*) and 1156 (*E. c. sochureki*) processed ESTs were entered into respective species databases alongside the 883 ESTs generated from the *E. ocellatus* vgDbEST (Wagstaff and Harrison, 2006). Assembled ESTs were BLAST annotated against UniProt (v56.2), TrEMBL (v39.2) and separate databases containing only Serpentes nucleotide and protein sequences derived from the same Uniprot/TrEMBL release versions.

Clustering was performed incrementally (96 sequences per round) to determine the number of sequences required to construct a representative transcriptome (i.e. the point where further sequencing only adds to existing clusters). It was estimated that

a minimum of 800 EST sequences were required to provide an accurate representation of the three vgDbESTs (Appendix II Figure 1). For longer clones (i.e. SVMPs), representatives of each cluster were subject to primer walking to acquire sufficient sequence data for isoform classification. SVMPs were characterised based upon the presence or absence of additional domains extending from the metalloproteinase domain (Fox and Serrano, 2005). PIVs were distinguished from PIIIs by the presence of an additional cysteine residue in the cysteine-rich region at positions 397 or 400 (Fox and Serrano, 2005; Wagstaff *et al.* 2009 (numbering from Fox and Serrano, 2005)).

Appendix II Table 1 displays the catalogue of venom toxin transcripts present in each of the four *Echis* vgDbESTs based upon significant (>1e-05) BLAST annotation. Presentation of the fully assembled and annotated vgDbESTs can be viewed at http://venoms.liv.ac.uk. The sequences reported in this paper have also been submitted into dbEST division of the public database GenBank: *E. coloratus* [GenBank: GR947900-GR948969], *E. c. sochureki* [GenBank: GR948970-GR950126] and *E. p. leakeyi* [GenBank: GR950127-GR951204].

4.4 Results

EST data provide a powerful insight into the transcriptional activity of a tissue at a particular time point. The protocols used for the generation of venom gland EST databases provide a snapshot of transcriptional activity in the venom gland 3 days after venom expulsion, when transcription peaks (Paine *et al.* 1992) in preparation for new venom synthesis. Although each individual venom transcript cannot be correlated with the mature venom proteome without considerable extra experimental verification, previous work with *E. ocellatus* (Wagstaff *et al.* 2009) shows there is a good general accordance between the venom proteome and that predicted from the venom gland transcriptome. Thus, whilst a cautionary approach is required when interpreting a correlation between transcriptome and proteome, the sensitivity and unbiased nature of venom gland transcriptome surveys can be valuable in the

identification of rare, unusual or potentially novel toxins and their isoforms that are difficult to detect in the proteome (e.g. Harrison *et al.* 2007).

To provide a representative overview of the transcriptional variation in venom components in each species, whilst minimising compositional bias arising from intraspecific variation in venom composition, venom gland cDNA libraries were based on ten specimens of variable size and gender. Generated ESTs were clustered under high stringency conditions to assemble overlapping single sequence reads into full length gene objects where possible. Using BLAST, 80-93% of gene objects for each library were assigned a functional annotation based upon significant (>1e-05) scores against multiple databases. The majority of annotated ESTs (61-74%) were assigned to clusters representing distinct gene objects (Appendix II Table 2). The proportion of toxin encoding transcripts (enzymes and non-enzymatic toxins) assigned by BLAST homology, was typically greater than those encoding non-toxin transcripts (for example, those involved in cellular biosynthetic processes) and unidentified components (i.e. with no significant hit against the databases) (Figure 4.1). There were twice the numbers of unidentified ESTs in the *E. c. sochureki* vgDbESTs than in any of the other *Echis* vgDbESTs. As the bulk of these unidentified ESTs were singletons, not clustered gene objects, I interpret this to result from increases in unidentified 3' untranslated regions rather than unidentified novel toxin transcripts. The annotated venom toxin encoding profiles for the four *Echis* species revealed substantial variation in (i) the inferred expression levels and (ii) the cluster diversity within many toxin families (Figure 4.2 and Appendix II Table 1). The details and potential implications of this species-specific variation in the representation of each toxin family will be discussed in turn.

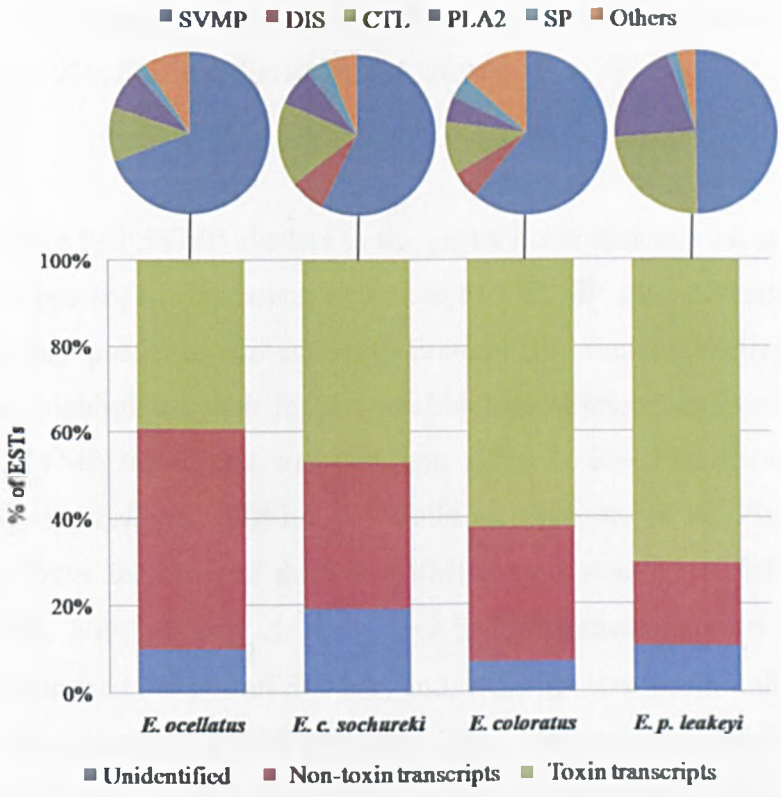

Figure 4.1. The relative expression of annotated venom gland transcriptomes from four members of the genus *Echis*. Bar charts represent the proportions of BLAST-annotated ESTs; unidentified = non-significant hits. Toxin encoding transcripts are expanded as pie charts illustrating the proportional representation of snake venom metalloproteinases (SVMP), short coding disintegrins (DIS), C-type lectins (CTL), group II phospholipases A_2 (PLA_2), serine proteases (SP) and other less represented venom toxins (Others) in the transcriptomes of each *Echis* species.

4.4.1 Snake venom metalloproteinases (SVMP)

The SVMP transcripts were the most abundant and divergent (in terms of cluster numbers) *Echis* venom toxin family (Figure 4.2) and comprised roughly half of the total toxin transcripts (Figure 4.1). The SVMPs are a diverse group of enzymes classified into those comprising only the metalloproteinase domain (PI) and those sequentially extended by a disintegrin domain (PII), a disintegrin-like and cysteine-rich domain (PIII) and the latter co-valently linked to C-type lectin-like components (PIV) (Fox and Serrano, 2005). Known and suspected modifications in domain structure are thought to account for the wide range of SVMP pathological activities,

including haemorrhage, coagulopathy, fibrinolysis and prothrombin activation (Warrell *et al.* 1976; Fox and Serrano, 2005, 2008).

There were more PIII SVMP clusters in the genus *Echis* than any other toxin family clusters. The presence of apparent, extensive PIII SVMP gene diversification hints that evolutionary pressures are acting to increase the functional diversity of this SVMP group, highlighting their fundamental biological importance to the genus. In contrast, PI SVMP transcripts were present, albeit at low levels, only in the *E. coloratus* and *E. ocellatus* vgDbESTs. While the diversity of the PII SVMPs was substantially lower than that of the PIII SVMPs, their abundance differed between species. Thus, 80% of total *E. p. leakeyi* SVMP transcripts were PIIs (cluster EPL00005 comprised 38% of all SVMPs) and, although less numerically significant, 38% of the *E. coloratus* SVMPs were also PIIs. Despite intrageneric variation in abundance and diversity, analysis of PII contiguous sequences throughout the genus revealed the ubiquitous representation of motifs (RGD, KGD and VGD) involved in binding to the $\alpha_{IIb}\beta_3$, $\alpha_v\beta_3$ and $\alpha_5\beta_1$ integrins implicated in platelet aggregation inhibition (Huang *et al.* 1987; Calvete *et al.* 2005). The RGD-only representation of *E. p. leakeyi* PII SVMPs implies evolutionary conservation of this particular disintegrin motif, in contrast to the gene diversification observed in the PIIIs. I assigned some PIII SVMP transcripts as putative PIV SVMPs according to the presence of an additional cysteine residue in the cysteine-rich region at positions 397 or 400 (Fox and Serrano, 2005; Wagstaff *et al.* 2009, (numbering from Fox and Serrano, 2005)). These transcripts also form strongly supported monophyletic groups (data not shown) with homologues of SVMP PIVs previously characterised from venom proteomes; two of the three putative *E. coloratus* PIVs (ECO00075 & ECO00144) show the greatest sequence similarity to PIV SVMPs characterised from *Macrovipera lebetina* and *Daboia russelii* respectively [UniProt:Q7T046 and Q7LZ61], whereas all other *Echis* PIVs showed greatest similarity to the previously characterised *E. ocellatus* PIV SVMP, EOC00024 (Wagstaff *et al.* 2009). The relative representation of these putative PIV SVMPs was substantially greater in *E. ocellatus* (EOC00024 - 23% and EOC00022 - 7%) than *E. coloratus* and *E. c. sochureki* (<4%); no PIV SVMPs were found in the *E. p. leakeyi* vgDbEST. Taken

together, this implies that two divergent forms of PIV SVMPs may be uniquely present in *E. coloratus*, despite their low representation in this species.

A new *E. ocellatus* cDNA precursor, encoding numerous QKW tripeptides and a polyH/G peptide that have potent SVMP-inhibiting activities, was recently identified (Wagstaff *et al.* 2008). Representatives of this SVMP inhibitory transcript were identified in each *Echis* vgDbEST (data not shown), but no correlation was identified between the proportional representation of the *Echis* SVMPs and their SVMP inhibitory transcripts.

4.4.2 Disintegrins

Snake venom disintegrins are derived either from proteolytic processing of PII SVMP precursors (Shimokawa *et al.* 1996) or are encoded by discreet PII-derived disintegrin-only genes, containing only a signal peptide and a disintegrin domain – previously described as 'short coding' disintegrins (Okuda *et al.* 2002; Francischetti *et al.* 2004). Representation of short coding disintegrins in the *Echis* genus is variable; small clusters were found in *E. c. sochureki* (4% and 3% of toxin transcripts) and *E. coloratus* (5%), whilst only a singleton transcript was found in *E. p. leakeyi*. Despite not being represented in the original *E. ocellatus* vgDbEST, a sequence encoding the short coding disintegrin ocellatusin has previously been identified from this species by PCR (Juárez *et al.* 2006b), confirming the presence of short coding disintegrin transcripts throughout the *Echis* genus.

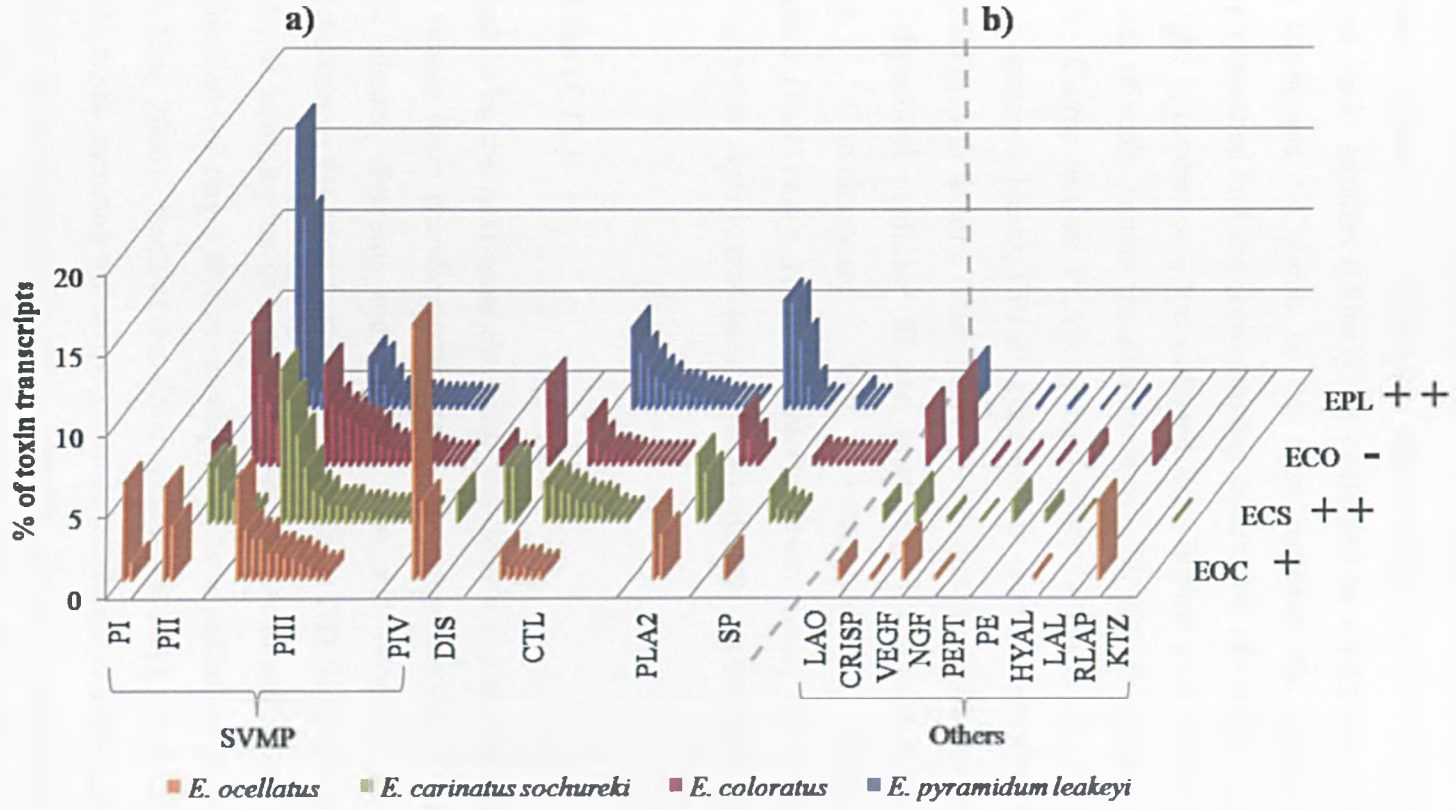
a)
b)
% of toxin transcripts
20
15
10
5
0
PI
PII
PIII
PIV
DIS
CTL
PLA2
SP
LAO
CRISP
VEGF
NGF
PEPT
PE
HYAL
LAL
RLAP
KTZ
SVMP
Others
EPL ++
ECO -
ECS ++
EOC +
E. ocellatus
E. carinatus sochurecki
E. coloratus
E. pyramidum leakeyi

Figure 4.2 (previous page). The relative abundance and diversity of each *Echis* genus venom toxin family. a) Relative expression levels of non-singleton clusters of the most representative venom toxin families and b) Relative expression levels of total non-singleton clusters and singletons representing the less numerically represented venom toxin families (Others) are expressed as a percentage of total toxin encoding transcripts. Column to the right indicates the proportion of invertebrate prey consumed and the corresponding correlation of venom toxicity to scorpions: ++, high; +, moderate; -, low (adapted from Barlow *et al.* 2009). Key – PI-PIV: sub-classes of snake venom metalloproteinases (SVMP); DIS: short coding disintegrins; CTL: C-type lectins; PLA2: group II phospholipases A_2; SP: serine proteases; LAO: L-amino oxidases; CRISP: cysteine-rich secretory proteins; VEGF: vascular endothelial growth factors; NGF: nerve growth factors; PEPT: peptidases – aminopeptidase, dipeptidyl peptidase III and neprilysin; PE: Purine liberators - phosphdiesterase, 5'-nucleotidase and ectonucleoside triphosphate diphosphohydrolase (E-NTPase); HYAL: hyaluronidases; LAL: lysosomal acid lipases; RLAP: renin-like aspartic proteases; KTZ: kunitz-type protease inhibitors.

4.4.3 C-type lectins (CTL)

The CTLs proved to be the next most abundant and diverse (by cluster numbers) group of *Echis* venom toxin encoding transcripts. As argued for the SVMPs, the substantial CTL cluster diversity and implied functional diversity would be consistent with the known variation in CTL activity. Thus, CTL isoforms typically act synergistically as homologous or heterologous multimers to promote or inhibit platelet aggregation and/or target distinct elements of the coagulation cascade (see Markland 1998; Kini, 2006). Each of the *Echis* species showed considerable CTL diversity (10-24% toxin encoding transcripts) with *E. p. leakeyi* exhibiting both the largest number of ESTs and cluster-diversity. Notably, clusters showing similarity to echicetin α and β, a platelet aggregation-inhibitor isolated from *E. c. sochureki* (Peng *et al.* 1993; Polgár *et al.* 1997), were found throughout the *Echis* genus and are the most represented CTLs in both *E. c. sochureki* and *E. p. leakeyi*. Recently, *E. ocellatus* echicetin-like CTLs were demonstrated to be associated with forming the quaternary structure of PIV *E. ocellatus* SVMPs (Wagstaff *et al.* 2009). However,

PIV SVMPs are absent from the *E. p. leakeyi* vgDbEST and present in only small numbers in *E. c. sochureki* (2%), implying that PIV-related binding may not be the sole function of echicetin. In contrast, each of the *Echis* vgDbESTs (except for *E. p. leakeyi*) contained clusters showing high sequence similarity to another PIV-related CTL, Factor X activator light chain 2 from *M. lebetina* (Siigur *et al.* 2004), producing an *Echis* representational profile of CTLs matching that of the PIV SVMPs.

4.4.4 Phospholipase A_2 (PLA_2)

Group II PLA_2s are ubiquitously expressed in *Echis* species (Bharati *et al.* 2003). *Echis* PLA_2s have been demonstrated to inhibit platelet aggregation and induce oedema, neurotoxicity and myotoxicity through multiple isoforms exhibiting high (Asp^{49}) and low (Ser^{49}) enzymatic activity (Kemparaju *et al.* 1994; 1999; Jasti *et al.* 2004a; Zhou *et al.* 2008). Despite low representation and diversity in *E. coloratus*, *E. ocellatus* and *E. c. sochureki* (5-8% of toxin transcripts), an increase in representation (21%) and cluster diversity was observed in *E. p. leakeyi*, suggesting an important role for PLA_2 activity in the venom of this species. Furthermore, both enzymatic PLA_2 variants are conserved throughout the genus, highlighting the apparent importance of these functionally-distinct isoforms - presumably for prey capture. Given that Ser^{49} PLA_2s have only been isolated from the genera *Vipera* (Petan *et al.* 2007) and *Echis* (Zhou *et al.* 2008), which are not sister taxa (Wüster *et al.* 2008), the presence of this isoform would be expected in other members of the Viperinae. However, considering the absence of Ser^{49} PLA_2s from a *Bitis gabonica* vgDbEST (Francischetti *et al.* 2004), I cannot rule out convergent evolution of this myotoxic PLA_2 type and its consequent functional importance in these genera.

4.4.5 Serine proteases (SP)

The snake venom serine proteases are a multi-gene enzyme family that act upon platelet aggregation, blood coagulation and fibrinolytic pathways (reviewed in Kini, 2006). Considering the severe coagulopathy observed in victims of *Echis* envenoming (Warrell *et al.* 1976; 1977), SPs are represented in amounts lower than

predicted (2-5% of toxin encoding transcripts), particularly given their high representation in other, albeit distantly related, Viperidae species (Cidade *et al.* 2006; Pahari *et al.* 2007). Interestingly, variations in cluster diversity are considerable, with nine clusters of low representation identified in *E. coloratus* compared to one in *E. ocellatus*. Despite low levels of representation, the unique variation in cluster diversity observed in *E. coloratus* implies multiple gene duplication events within this lineage; a process that underpins functional diversification in multi-gene venom proteins (Kordiš and Gubenšek, 2000; Župunski *et al.* 2003).

4.4.6 L-amino oxidases (LAO)

Snake venom LAOs have been demonstrated to induce apoptosis and inhibit platelet function (reviewed in Du and Clemetson, 2002). While the mechanisms for these actions remain predominately uncharacterised, it seems clear that, unlike other snake venom toxin families, isoform diversity is not a requirement. Thus, the low representation (1-4% of toxin transcripts) observed in the *Echis* vgDbESTs is consistent with other viperid venom gland transcriptomes (Junqueira-de-Azevedo and Ho, 2002; Francischetti *et al.* 2004; Kashima *et al.* 2004; Cidade *et al.* 2006; Junqueira-de-Azevedo *et al.* 2006; Wagstaff and Harrison, 2006; Zhang *et al.* 2006; Pahari *et al.* 2007). Indeed, the atypically high level of sequence conservation between all the *Echis* LAOs and those from other viperid genera (>80%) implies a conserved mechanism of action, whereby evolutionary pressures act to constrain diversification.

4.4.7 Cysteine-rich secretory proteins (CRISP)

Members of the snake venom CRISP family interact with ion channels and exhibit the potential to block arterial smooth muscle contraction and nicotinic acetylcholine receptors (e.g. Yamazaki and Morita, 2004; Gorbacheva *et al.* 2008). The relative CRISP expression profiles vary considerably in the genus *Echis*, ranging from 5% of toxin encoding transcripts in *E. coloratus*, less than 2% in *E. c. sochureki* and *E. ocellatus* and none in *E. p. leakeyi*. Given that CRISPs are typically underrepresented toxin transcripts in Viperidae vgDbESTs (Junqueira-de-Azevedo

and Ho, 2002; Francischetti *et al.* 2004; Kashima *et al.* 2004; Cidade *et al.* 2006; Junqueira-de-Azevedo *et al.* 2006; Wagstaff and Harrison, 2006; Zhang *et al.* 2006), the abundant representation observed in *E. coloratus* implies an unidentified evolutionary pressure favouring transcriptional expression in this species. Its potential biological significance is further highlighted by the apparent absence of these toxins in the transcriptome of the most closely related species, *E. p. leakeyi*, which differs strongly in diet from *E. coloratus* (Barlow *et al.* 2009).

4.4.8 Other venom components

Clusters encoding vascular endothelial growth factors and nerve growth factors were identified in small numbers (Appendix II Table 1) throughout the genus and, like the LAOs, each showed a high degree of sequence conservation. Similarly, and consistent with previous reports (Kemparaju and Girish, 2006), the sequence homology of the new hyaluronidase singleton ESTs of *E. c. sochureki* and *E. ocellatus* was also considerable, and extended to hyaluronidase sequences of other genera. It is apparent that evolutionary forces exist to conserve the sequence of this group of venom proteins, presumably because their role in disseminating venom toxins by reducing the viscosity of the extracellular matrix (Harrison *et al.* 2007) is a universal requirement for prey 'knock-down'. Another singleton EST from the *E. c. sochureki* vgDbEST exhibited 81% identity to a kunitz-type protease inhibitor isolated from the elapid snake *Austrelaps labialis* (Doley *et al.* 2008a). Given the phylogenetic distance between these species, homology between these haemostatic disruptors is surprising, particularly since the singleton exhibited only 38% identity to kunitz-type protease inhibitors identified from the *Bitis gabonica* vgDbEST (Francischetti *et al.* 2004), a species closely related to *Echis*. An additional number of peptidases and purine liberators were identified as minor components in all but the *E. ocellatus* vgDbEST (Table 4.1). Despite their low representation and inconsistent conservation throughout the genus, the distinct biological activities of these components have been reported to play a role in the pathology of viper envenoming (Table 4.1), although these claims require experimental confirmation.

Identification	No. of ESTs	Species present	Activity	Possible venom function
Aminopeptidase	8	*E. c. sochureki*	Hydrolysis of the N-terminal region of peptides (Glenner and Folk, 1961).	Potential interference with angiogenesis and blood pressure control (Marchio *et al.* 2004; Fournie-Zaluski *et al.* 2004).
	1	*E. coloratus*		
Ectonucletotide pyrophosphatase/ phosphodiesterase	2	*E. coloratus*	Hydrolysis of nucleotides and nucleic acids (Fürstenau *et al.* 2006).	Interaction with platelet function (Fürstenau *et al.* 2006). Activity previously described in *Echis carinatus* (Taborska, 1971).
	3	*E. c. sochureki*		
5'-nucleotidase	3	*E. coloratus*	Cleavage of a wide variety of ribose and deoxyribose nucleotides (Aird, 2002).	Potential inhibitor of platelet aggregation (Aird, 2002). Activity identified in a number of different lineages including *Echis carinatus* (Taborska, 1971).
	2	*E. p. leakeyi*		
	1	*E. c. sochureki*		
Ectonucleoside triphosphate diphosphohydrola-se 2 (E-NTPase 2)	2	*E. coloratus*	Hydrolysis of nucleoside-5'-triphosphates/ diphosphates (Sales and Santoro, 2008)	Potential inhibitor of platelet aggregation (Champagne, 2005; Sales and Santoro, 2008).

Table 4.1. Under-represented toxin encoding transcripts from the *Echis* vgDbESTs potentially associated with venom function.

4.4.9 Novel venom gland transcriptome components

I identified a cluster from the *E. coloratus* vgDbEST that exhibited 64% identity to mammalian lysosomal acid lipase/cholesteryl ester hydrolase (LAL) [UniProt:Q4R4S5]. The most critical function of LAL is to modulate intracellular cholesterol metabolism by degrading cholesterol esters and triglycerides derived from low density lipoproteins that are transported, via specific receptors, into most cells (Li *et al.* 2007; Qu *et al.* 2009). Although LAL is a common enzyme in many lineages, this is the first time it has been identified from a venomous animal. The vgDbESTs were interrogated for other transcripts with annotations related to lysosomal processes and singleton transcripts were identified in multiple species (data not shown). However, their quantities were considerably lower than LAL suggesting to us that an association between venom gland LAL and intracellular processes was unlikely. Furthermore, the identification of a signal peptide using SignalP v3.0 (Bendtsen *et al.* 2004) and the comparable representation of this enzyme (2%) with other venom toxin encoding transcripts (e.g. SPs, LAOs, growth factors), strongly implies these transcripts are a novel group of secreted venom components. Their biological contribution to the activity of *E. coloratus* venom and the venom gland and expression in other venomous snake genera will be the subject of future research.

In addition to the discovery of LAL, two singleton transcripts were identified (Appendix II Table 1) from the *Echis* vgDbESTs as novel Serpentes zinc-dependent metallopeptidases (Baral *et al.* 2008). A transcript exhibiting 67% identity to human dipeptidyl peptidase III (DPPIII) [UniProt:Q53GT4] was identified in *E. coloratus* and a related EST exhibiting 84% similarity to Neprilysin from *Gallus gallus* [Uniprot:Q67BJ2] was identified in the *E. p. leakeyi* vgDbEST. While signal peptides were absent from these ESTs due to EST N-terminal truncation, the constitutive physiological targets of their mammalian analogues indicate that these metallopeptidases may contribute to pathology. Mammalian DPPIII exhibits particular affinity for the degradation of hypertension-inducing peptides via the inactivation and degradation of angiotensin II to angiotensin III; the consequential reduction in vasoconstrictor activity likely induces hypotension alongside

thrombolysis, by reducing the activity of plasminogen activator inhibitors that constrain fibrinolysis (Lee and Snyder, 1982; Abramić *et al.* 1988; Skurk *et al.* 2001). It was previously reported that the *E. ocellatus* vgDbEST contained a substantial number of novel, potentially hypotensive, venom toxins termed the renin-like aspartic proteases (Wagstaff and Harrison, 2006). Neprilysin demonstrates affinity for a broader range of physiological targets, including natriuretic, vasodilatory and neuro peptides (Turner *et al.* 2001). Specific functional interactions include the termination of brain neuropeptides, such as enkephalins and substance P, at peptidergic synapses (Matsas *et al.* 1983), and the degradation of the hypotension-inducing atrial natriuretic peptide (ANP) (Turner *et al.* 2001). It is notable that Neprilysin has been implicated in the inactivation of peptide transmitters and their modulators in vertebrates and invertebrates (Turner *et al.* 2001; Isaac, 1988), suggesting the potential for conserved neurotoxic activity across a range of prey species.

4.5 Discussion

The most numerically abundant venom toxin families in the four *Echis* species were the SVMPs, CTLs, PLA_2s, and SPs. This is broadly consistent with previous viperid venom gland analyses, although considerable inter-generic variations in the EST-inferred expression levels of these toxin families have been observed (Junqueira-de-Azevedo and Ho, 2002; Francischetti *et al.* 2004; Kashima *et al.* 2004; Cidade *et al.* 2006; Junqueira-de-Azevedo *et al.* 2006; Wagstaff and Harrison, 2006; Zhang *et al.* 2006; Pahari *et al.* 2007). The correlation of toxin families identified from the genus *Echis* and other viperid species support current theories of early venom toxin recruitment prior to the radiation of the Viperidae (Fry *et al.* 2008). The absence of three finger toxins from the *Echis* vgDbESTs is particularly notable as their recent identification in other viper species (Junqueira-de-Azevedo *et al.* 2006; Pahari *et al.* 2007) implies the venom gland recruitment of these toxins occurred prior to the divergence of the Viperidae; presumably these toxins have subsequently been lost in an ancestor of *Echis*. Consistent with the early, PCR-driven, reports of accelerated evolution of venom serine proteases (Deshimaru *et al.* 1996), CTLs (Ogawa *et al.* 2005) and PLA_2s (Nakashima *et al.* 1993), it is apparent from the *Echis* genus

vgDbESTs and those of other vipers that the evolutionary forces driving venom toxin recruitment in the genus *Echis* have served to promote diversification in some toxin lineages (PII and PIII SVMPs, CTLs) while in comparison relatively low diversification exists in others (PI and PIV SVMPs, PLA_2s, LAOs, the growth factors, and remaining minor venom components). Prey capture is considered a major biological imperative driving the venom toxin selection process. This project was undertaken to identify correlations between intrageneric dietary preferences and transcript expression in order to elucidate the influence dietary selection pressures may have on the toxin composition of snake venoms.

(i) Recruitment of novel venom toxins and diet. The *Echis* vgDbESTs reveal the recruitment of novel renin-like aspartic proteases in *E. ocellatus* (Wagstaff and Harrison, 2006), LAL and DPPIII in *E. coloratus* and Neprilysin in *E. p. leakeyi*. The potential hypotensive role of venom aspartic proteases has been discussed previously (Wagstaff and Harrison, 2006). Whilst expression in the venom proteome requires experimental verification, the presence of a signal peptide suggests that LAL is more likely to be secreted in the venom gland rather than acting as an intracellular protein. LAL has been implicated in severe alveolar destruction following over-expression of these enzymes in the lungs of mice (Li *et al.* 2007). Lipases such as LAL and lipoprotein lipase may also contribute to an influx of fatty acids into the brain by hydrolysing lipoproteins in the microvascular system of the cerebral cortex (Brecher and Kuan, 1979). The suggestion that these fatty acids are then intra-cellularly internalised within lysosomes (Brecher and Kuan, 1979) correlates with intriguing observations from *E. coloratus* induced pathology, where increases in the size and numbers of lysosomes within the neuronal tissue of guinea pigs were implicated in neuron lysis and cerebral damage (Sandbank and Djaldetti, 1966). I infer from the predominately vertebrate-only diet of *E. coloratus* and the exclusive, yet substantial, representation of LAL in this species (2% – equivalent to the SPs, LAOs and growth factors) that LALs may play a contributory, albeit not yet understood, role in prey envenoming. As singletons, it is more difficult to argue that the novel recruitments of DPPIII and Neprilysin represent additional adaptations to prey preference; as they are found in such low numbers it is impossible to determine whether they are indeed novel species-specific venom gland recruitments or are rare

transcripts that remain undetected in other snake species. Barlow *et al.* (2009) previously reported that invertebrate feeding likely evolved as a basal trait in the genus *Echis*. The absence of genus-wide transcripts encoding novel putative venom toxin families implies that the adaptation to invertebrate feeding in *Echis* did not evolve as a consequence of recruiting novel invertebrate-specific venom toxins. However, I cannot exclude the possibility that the novel recruitment of LAL into the *E. coloratus* venom gland transcriptome may result from the subsequent reversion to vertebrate feeding observed in this species (Barlow *et al.* 2009), particularly given the absence of these well represented putative toxin transcripts in other members of the genus.

(ii) Changes in toxin family expression and diet. All the major *Echis* venom toxin families (SVMP, CTL, PLA_2, and SP) exhibited considerable intrageneric variation in transcriptional representation. Thus, the *E. p. leakeyi* vgDbEST was notable for its absence of PI and PIV SVMPs, short coding disintegrins and CRISPs and atypically abundant representation of PII SVMPs, CTLs and PLA_2s. The CRISPs were only represented by clusters in *E. c. sochureki* and *E. coloratus*, species whose vgDbESTs draw similarities, particularly in their high comparative expression of PIII SVMPs and short coding disintegrins. The only distinguishing feature (in terms of transcript abundance) in the *E. ocellatus* vgDbEST was the atypically high number of PIV SVMPs. However, none of these toxin encoding expression profiles showed a clear association with diet. Most notably, *E. p. leakeyi* and *E. c. sochureki* exhibit distinct toxin encoding profiles (Figure 4.2), despite both species feeding predominately on invertebrates and exhibiting highly invertebrate-lethal venom (Barlow *et al.* 2009).

(iii) Diversification of venom toxins and diet. The above observations imply adaptations to diet are occurring within venom toxin families rather than resulting from changes in expression levels of entire toxin families. Evidence supporting this hypothesis is provided by substantial increases in representation of echicetin-like CTLs (relative to other CTLs) in both *E. p. leakeyi* and *E. c. sochureki*, implying perhaps a significant role for these inhibitors of platelet aggregation in invertebrate

prey capture. The absence of PI SVMPs in these species perhaps suggests that this SVMP isoform is more associated with a vertebrate diet. Furthermore, a number of atypical observations identified from the *E. coloratus* vgDbEST may be associated with a reversion to vertebrate feeding (Barlow *et al.* 2009), including; (i) increases in the representation of CRISPs, (ii) increases in cluster diversity of the SPs and (iii) the identification of putative novel venom toxins (LAL and DPPIII). However, the general similarity between the toxin encoding expression profiles of *E. c. sochureki* and *E. coloratus* (Figure 4.2), despite *E. coloratus* exhibiting a significant reduction in venom toxicity to invertebrates (Barlow *et al.* 2009), indicates that more analytical molecular tools are required to determine whether snake prey specificity is achieved through subtle alterations in isoform expression levels within the major venom toxin families. I am subjecting the *Echis* genus vgDbEST data generated here to a phylogenetic analysis on each toxin class to determine species-specific trends in diversification, which will determine whether multiple levels of gene control in the *Echis* genus venom gland (switching of transcriptional expression, gene duplication conferring functional diversification and novel gene expression) are responsible for evolutionary responses to dietary pressures.

Correlations between variation in venom gland toxin encoding profiles and snakebite symptomatologies from the genus *Echis* are unclear, particularly given the similar, predominately incoagulable and haemorrhagic, clinical outcomes observed throughout the genus (Warrell *et al.* 1977; Warrell, 1995; Habib *et al.* 2001) and the presence of multiple isoforms of toxin families implicated in haemorrhage and coagulopathy. However, some observations of atypical symptoms can be tentatively explained; substantial increases in PLA_2 representation and the unique presence of Neprilysin may correlate with the rare manifestation of neurotoxicity observed in an *E. pyramidum* envenomation (Gillissen *et al.* 1994), whilst the putative function of DPPIII may imply a contributory role in cases of hypotension observed following *E. coloratus* snakebite (Warrell, 1995).

Venom gland transcriptome surveys provide a comprehensive description of the venom composition of each major *Echis* lineage, which, using proteomic

(antivenomic) techniques (Gutiérrez *et al.* 2009), will identify the extent to which intrageneric variation in venom composition impacts on the preclinical efficacy of commercially available antivenoms. Such analyses may (i) explain past antivenom failures described following snakebite by members of this medically important genus (Gillissen *et al.* 2004; Kochar *et al.* 2007; Visser *et al.* 2008; Warrell, 2008) and (ii) identify the venom toxin mix required to generate an antivenom with continent-wide clinical effectiveness against *Echis* envenoming.

4.6 Conclusions

The first comprehensive comparison of intrageneric venom gland transcriptomes reveals substantial venom variation in the genus *Echis*. The observed variations in venom toxin encoding profiles reveal little association with venom adaptations to diet previously described from this genus. I hypothesise that relatively subtle alterations in toxin expression levels within the major venom toxin families are likely to be predominately responsible for prey specificity, although I cannot rule out a contributory role for novel putative venom toxins, such as lysosomal acid lipase. The observation of substantial venom variation within the medically important genus *Echis* strongly advocates further investigations into the medical significance of venom variation and its potential impact upon antivenom therapy.

4.7 Authorship order and contributions

Nicholas R Casewell, Robert A Harrison, Wolfgang Wüster and Simon C Wagstaff. I undertook the experiments, the comparative analysis and drafted the publication manuscript (see Appendix VI) that forms the basis of this chapter. RAH participated in the venom gland dissections and SCW provided guidance and assistance with cDNA library construction and bioinformatic analysis. All authors were involved in the critical review of the manuscript.

CHAPTER 5

Selective snake venom: genomic basis of adaptation of venom composition in saw-scaled vipers (Serpentes: Viperidae: *Echis*) as a response to alterations in diet

5.1 Abstract

Variation in snake venom occurs at multiple taxonomic levels and can significantly impact upon the clinical manifestations of snakebite and the efficacy of antivenom therapy. Natural selection for the optimisation of venom to differing prey items has frequently been invoked as the most likely evolutionary force driving variation in venom components, although the genetic basis for these adaptations remains incompletely understood. Here, I investigate the influence of diet upon the evolutionary history of the five most representative toxin families present in the venom glands of the medically important saw-scaled vipers (Serpentes: Viperidae: *Echis*). Gene tree parsimony analyses provide the first evidence of the genomic basis of snake venom adaptations as a response to alterations in a venom-required diet, with snake venom metalloproteinase and serine protease toxin families exhibiting diet-associated gene events that correspond with a reversion to vertebrate-feeding in *E. coloratus*. Furthermore, the diversification and retention of these coagulopathic and haemorrhagic toxins in the venom gland of *E. coloratus* correlate with significant differences in venom function in the form of *in vivo* haemorrhage. These results provide genetic and functional evidence of coevolution between diet and venom components and highlight the selective influence alterations in diet can have upon venom composition. Understanding the selective processes that underpin venom variation is of fundamental importance to understand the pathologies induced by snakebites and for the rational design of future antivenom therapies that aim to treat the ~0.4-2.6 million people who suffer snake envenomations each year.

5.2 Introduction

The evolution of gene families is widely regarded as an important means by which organisms evolve adaptively (see Ohno, 1970; Ohta, 1991; Zhang, 2003). Through gene duplication the functional constraints of a gene can be released, facilitating the neofunctionalization of the redundant copy by positive selection, whilst others are inactivated or deleted from the genome (see Nei and Hughes; 1992; Ohta, 2000; Zhang, 2003; Nei and Rooney, 2005; Lynch, 2007). This 'birth-and-death' model of gene evolution (Nei and Hughes, 1992) is thought to be advantageous by promoting increases in the diversity and complexity of gene function (Ohta, 1991) and is predominately responsible for the evolution of large multi-gene families, including the major histocompatability complex and snake venom toxins; both of which suffer rapid functional and structural gene diversifications alongside enhanced rates of sequence evolution (see Kini and Chan, 1999; Kordiš and Gubenšek, 2000; Župunski *et al.* 2001; Nei and Rooney, 2005).

Snake venoms primarily comprise of a complex mix of biologically active proteins and peptides (toxins) (Aird, 2002) which are known to vary at multiple taxonomic levels, including inter-specifically and ontogenetically (see Chippaux *et al.* 1991; Gutiérrez *et al.* 2009). The rapidly-evolving nature of multi-locus toxin-encoding genes (Kini and Chan, 1999; Kordiš and Gubenšek, 2000; Župunski *et al.* 2001) provides a model system to analyse the genomic basis of selective adaptations in multi-gene families. Understanding the genetic adaptations responsible for conferring variation in venom components is particularly desirable due to the medical importance of snake venoms (Chippaux, 1998; Kasturiratne *et al.* 2008), which often exhibit distinct functional and pathological activities induced by toxins encoded by the same multi-gene family (see Fry *et al.* 2003; Fox and Serrano, 2005; Lynch, 2007). Consequently, venom variation can significantly impact upon the clinical manifestations of envenoming (Warrell *et al.* 1989; Prasad *et al.* 1999; Shashidharamurthy *et al.* 2002) and the clinical efficacy of antivenom therapy (Theakston *et al.* 1989; Galán *et al.* 2004; Visser *et al.* 2008).

As the primary role of snake venom is to aid prey capture (Chippaux *et al.* 1991), natural selection for the optimisation of venom to differing prey items has frequently been invoked as the most likely evolutionary driving force responsible for venom variation (see Daltry *et al.* 1996; Li *et al.* 2005; Barlow *et al.* 2009; Gibbs and Mackessy, 2009), although the genetic basis for these adaptations remains incompletely understood. Well documented cases of resistance to envenoming in prey species (Poran *et al.* 1987; Heatwole and Poran, 1995; Biardi *et al.* 2006) highlights the potential for coevolutionary 'arms races' to occur between venom toxicity and prey, whereby selective pressures act to overcome prey resistance. The high metabolic cost of venom production likely produces a trade-off between venom synthesis and foraging efficiency (McCue, 2006); the production of a reduced volume of highly toxic venom likely represents a metabolic advantage over an excessive injection of less toxic venom. Evidence of venom 'metering', whereby the amount of venom injected varies according to prey size (Hayes *et al.* 1995), alongside the selective loss of functional toxin-encoding genes and cases of atrophied venom delivery apparatus following dietary shifts to egg-eating (Li *et al.* 2005; Fry *et al.* 2008), provide further support for a trade-off between venom production and foraging. Correlations between venom composition, prey-specific toxicity and diet have been observed in a number of genera (Daltry *et al.* 1996; Jorge-da-Silva and Aird, 2001; Creer *et al.* 2003; Sanz *et al.* 2006; Barlow *et al.* 2009; Gibbs and Mackessy, 2009) and attributed to adaptive venom evolution. However, the accumulation of deleterious mutations in toxin-encoding genes following the loss of venom-dependent predation in *Aipysurus eydouxii* (Li *et al.* 2005) provides the only genetic evidence for dietary venom evolution, though the genetic mechanisms underpinning toxin-specific adaptations responsible for conferring increases in venom toxicity to natural prey items have yet to be elucidated.

The saw-scaled vipers (Serpentes: Viperidae: *Echis*) are a group of medically important viperid snakes that exhibit considerable variation in venom components, prey lethality and dietary composition (Barlow *et al.* 2009; Casewell *et al.* 2009 – Chapter 4), and thus represent a model system to analyse the influence dietary selection pressures have upon toxin-encoding genes. Envenomings by *Echis sp.* are

thought to be responsible for more snakebite deaths worldwide than any other snake genus (Warrell *et al.* 1977). Envenomed victims typically suffer a consistent combination of systemic and local haemorrhagic symptomatologies, but cases of incomplete intrageneric antivenom efficacy have been documented, implying substantial medically-relevant inter-species venom variation (Visser *et al.* 2008; Gillissen *et al.* 1994; Kochar *et al.* 2007; Warrell, 2008). Furthermore, representatives of the four *Echis* species complexes (Pook *et al.* 2009) exhibit considerable variation in diet, *E. coloratus* being a vertebrate specialist whilst arthropods represent a substantial proportion of the diet of *E. ocellatus*, *E. p. leakeyi* and *E. c. sochureki* (Barlow *et al.* 2009). Proportions of consumed arthropods correlated with venom toxicity to scorpions, strongly suggesting coevolution of venom toxicity and diet in the evolutionary history of this genus (Barlow *et al.* 2009). The phylogenetic mapping of dietary prey preferences to a well-supported mitochondrial and nuclear gene derived phylogeny support a dietary shift towards arthropod-feeding at the base of the genus *Echis*, followed by a subsequent reversion to vertebrate feeding in the *E. coloratus* species group (Barlow *et al.* 2009; Pook *et al.* 2009). Considering the profound physiological differences between arthropod and vertebrate prey items (see Krem and Di Cera, 2002; Muñoz-Chápuli *et al.* 2005) and the corresponding differences in venom toxicity to arthropods (Barlow *et al.* 2009), it would appear that evolutionary pressures may induce variation in venom components towards prey specificity. However, preliminary protein analyses failed to identify obvious links between venom composition and diet (Barlow *et al.* 2009), whilst a venom gland transcriptomic survey of multiple *Echis* species identified substantial intra-generic toxin family variation, but little obvious association with venom-related adaptations to diet (Casewell *et al.* 2009 – Chapter 4). It was suggested that alterations in isoform diversity and their respective representation within major toxin families were likely to be responsible for prey specificity rather than alterations in entire toxin family representation or the recruitment of novel toxins (Casewell *et al.* 2009 – Chapter 4).

Here I examine the genomic basis of dietary adaptations in the genus *Echis*; I hypothesise that shifts in the diversity of toxin gene families will coincide with the phylogenetic placement of shifts in diet, and I expect those toxin families most likely

to be relevant to different prey items to be most affected. I assessed the diversity of venom components isolated from members of the genus *Echis* by phylogenetic analysis of toxin encoding expressed sequence tags (ESTs) generated from venom gland cDNA libraries of *E. ocellatus*, *E. coloratus*, *E. pyramidum leakeyi* and *E. carinatus sochureki* (Wagstaff and Harrison, 2006; Casewell *et al.* 2009 – Chapter 4). The evolutionary history of the five most important venom toxin families, snake venom metalloproteinases (SVMP), C-type lectins (CTL), phospholipases A_2 (PLA_2), serine proteases (SP) and cysteine rich secretory proteins (CRISP), were elucidated using optimised models of sequence evolution coupled with Bayesian inference. I predict that the reconciliation of toxin gene trees with known species phylogenies will reveal significant differences in gene duplication or loss events that coincide with the evolutionary timing of dietary shifts.

5.3 Methods

5.3.1 cDNA library synthesis, bioinformatics and sequencing

Venom gland cDNA libraries were constructed using procedures previously outlined (Wagstaff and Harrison, 2006; Casewell *et al.* 2009 – Chapter 4). Briefly, multiple cDNA libraries were constructed from ten wild-caught specimens of *Echis ocellatus* (Nigeria), *E. p. leakeyi* (Kenya), *E. coloratus* (Egypt) and *E. c. sochureki* (UAE); ~1000 random clones per species were picked for sequencing using M13 forward primers. ESTs were bioinformatically processed using the PartiGene pipeline (Parkinson *et al.* 2004) with high stringency CLOBB clustering (Parkinson *et al.* 2002; Wagstaff and Harrison, 2006) and BLAST annotation against multiple databases (see Casewell *et al.* 2009 – Chapter 4). Full length sequencing of BLAST annotated PLA_2, CTL and CRISP clones were obtained during the initial round of sequencing. Reverse sequencing, using M13 reverse primers, was undertaken to generate full length DNA sequences of SP clones. Due to the frequency of SVMP annotated sequences, near full length sequence information was gained via primer walking all non-redundant, non-truncated SVMP clones which demonstrated sequence similarity to the catalytic site (H-box) of the metalloproteinase domain (Fox and Serrano, 2005).

5.3.2 Toxin family gene trees

Full length *Echis* ESTs annotated as SVMPs, CTLs, PLA_2s, SPs and CRISPs were compiled into nucleotide toxin family datasets alongside all existing non-redundant Viperidae sequences identified by sequence database searches in GenBank, EMBL, dbEST and UniProt. Alignments were generated using Clustal W (Thompson *et al.* 1994), implemented in MEGA4 (Molecular Evolutionary Genetics Analysis) (Tamura *et al.* 2007), followed by manual adjustments by eye. Non-Serpentes outgroup sequences for each family were identified by sequence similarity searches against a number of non-Serpentes databases before inclusion in the datasets (see Appendix III Figures 1-6). Datasets were translated and trimmed to the open reading frame of the proteins in MEGA4; redundant sequences and those containing frameshifts or truncations as the result of indels were excluded. Gene trees were produced using optimised models of sequence evolution combined with Bayesian inference; translated DNA datasets were subjected to analysis in ModelGenerator v0.85 (Keane *et al.* 2006) to select appropriate models of evolution for maximal extraction of phylogenetic signal (see Castoe *et al.* 2005; Castoe and Parkinson, 2006), with the model favoured under the Akaike Information Criterion (AIC) selected (Posada and Buckley, 2004). Bayesian inference analyses were undertaken in MrBayes v3.1 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) on the freely available bioinformatic platform Bioportal (www.bioportal.uio.no). Each dataset was run in duplicate using four chains for $5x10^6$ generations and sampling every 500^{th} tree.

5.3.3 Tree reconciliation

Bayesian generated consensus gene trees were edited in PhyloWidget (Jordan and Piel, 2008) to remove excessive non-*Echis* Viperidae nodes from each toxin family tree; representative outgroup taxa for each *Echis* clade were retained, and where possible kept consistent, for subsequent species tree reconciliation (see Appendix III Figures 1-6). Consensus gene trees topologies were parsed through PAUP* v4.0b10 (Swofford, 2002) to remove branch lengths and internal node labels. Gene tree topologies were subsequently edited to GeneTree v1.0 (Page, 1998) input specifications alongside species tree topologies of the remaining taxa inferred from

previous phylogenetic studies (Garrigues *et al.* 2005; Castoe and Parkinson, 2006; Wüster *et al.* 2008). The reconciliation option in GeneTree was used to map the multiple toxin family loci onto the inferred species tree, thereby elucidating the evolutionary pattern of gene duplication and loss events for each toxin family. Gene events occurring within the genus *Echis* were observed and mapped to the previously determined saw-scaled viper species tree (Barlow *et al.* 2009; Pook *et al.* 2009).

5.3.4 *In vivo* assessments of haemorrhage

Modified minimum haemorrhagic dose (MHD) experiments (see Theakston and Reid, 1983; Gutierrez *et al.* 1985) were undertaken to compare the haemorrhagicity of pooled venom milked from the four species of the genus *Echis* used to construct the venom gland transcriptomes. Following manual extraction, venom was frozen, lyophilised and stored at 4°C prior to reconstitution at 0.2mg/ml in 1X phosphate-buffered saline (PBS). 10μg doses (previously determined MHD for *E. ocellatus* – Cook *et al.* in press) of each venom were injected intradermally into the shaved dorsal skin of groups of six male CD-1 mice (18-20g - Charles River) under halothane anaesthesia. After 24 hours the dorsal skin was removed and the size of the lesion on the inner surface of the skin measured in two directions at right angles using callipers and background illumination. The mean diameter of each lesion was calculated prior to one-way analysis of variance and pair-wise comparison statistical assessments in Minitab 15.

5.4 Results

The translated DNA datasets comprised a total of 714 amino acids of SVMP (n=220) [GenBank: AM039691-AM039701, GU012123-GU012315 and GU594192-GU594224], 260 amino acids of SP (n=27) [GenBank: GU012092-GU012122], 173 amino acids of CTL (n=116), 144 amino acids of PLA_2 (n=33) (see Appendix III Table 1 for CTL and PLA_2 GenBank accession numbers) and 245 amino acids of CRISP (n=6) [GenBank: DW361159, GR948128, GR948365, GR948728, GR949534 and GR950013] *Echis*-derived EST sequence data. For Bayesian inference, ModelGenerator v0.85 identified the WAG + Γ model for all amino acid

datasets except the SVMP gene family, where a mixed model of evolution was implemented as the size of this dataset prevented model selection. Consensus gene trees generated by Bayesian inference for each toxin family are displayed in the supporting information (Appendix III Figures 1-6).

SVMPs are classified into four sub-classes (PI-PIV) based upon the presence of additional domains extending sequentially from the metalloproteinase domain (Fox and Serrano, 2005, 2008). Prior to gene/species tree reconciliation, the SVMP toxin family was separated into two separate tree reconciliation analyses (PI/PII and PIII/PIV) due to the distinction between sub-classes; *Echis*-derived representatives of PI and PII SVMP sub-classes form a strongly supported monophyletic group distinct from the PIII/PIV sub-classes (Appendix III Figure 1), whilst the non-monophyly of the PIV sub-class supported their inclusion within the PIIIs (Appendix III Figure 2).

Gene tree reconciliation with the previously determined *Echis* species tree (Barlow *et al.* 2009; Pook *et al.* 2009) revealed the evolutionary pattern of gene duplication and loss events for each toxin family (Figure 5.1). Considerable variation in gene duplication events was noted in a number of toxin families; notably in *E. p. leakeyi* and *E. coloratus* in the SVMP PI/PII sub-classes (Figure 5.1A), *E. coloratus* in the SVMP PIII/PIV sub-classes (Figure 5.1B) and *E. p. leakeyi* in both the CTLs and PLA_2s (Figures 5.1C and 5.1D). Variations in loss events were less pronounced, although substantially fewer gene losses were observed in *E. coloratus* in the SVMP PIII/PIV (Figure 5.1B), PLA_2 (Figure 5.1D) and SP (Figure 5.1E) reconciled trees.

One-way analysis of variance demonstrated significant intra-generic differences ($=p<0.05$) in *in vivo* haemorrhagic lesions induced by venom from four species of the genus *Echis*; *E. coloratus* venom produced the largest and *E. p. leakeyi* venom the smallest haemorrhagic lesions after 24 hours (Figure 5.2).

5.5 Discussion

The variation in venom inherent to multiple taxonomic levels of the Serpentes has previously been correlated with influencing factors such as phylogenetic position, geography and diet (see Chippaux *et al.* 1991). Whilst such studies have inferred the selective influence diet can have upon venom variation, through the correlation of venom toxicity and electrophoretic and proteomic venom profiles with dietary composition (Daltry *et al.* 1996; Sanz *et al.* 2006; Barlow *et al.* 2009; Gibbs and Mackessy, 2009), no such link has been determined at the level of the gene. Here, the reconciliation of toxin family gene trees with the genus *Echis* phylogeny provides the first evidence of the genomic basis of snake venom adaptations as a response to alterations in diet; toxin families exhibiting diet-associated gene events correlate with a reversion to vertebrate-feeding in *E. coloratus*. Furthermore, significant differences in the degree of haemorrhage induced by *Echis* venoms strongly correlates with toxin family gene events, providing a functional association between toxin family evolution and diet.

Reconciled gene/species trees for the PIII/PIV SVMP classes and the serine proteases exhibit strong correlations with the dietary shifts described previously (Barlow et al. 2009) (Figure 5.3). The PIII/PIV SVMP reconciled tree exhibits considerable numbers of gene duplication and loss events in each member of the genus (Figure 5.1B). However, a substantial increase in the number of gene duplications (alongside a smaller reduction in loss events) is evident in *E. coloratus* when compared to the other members of the genus. By mapping the net result of these duplications and losses to the *Echis* phylogeny, it is apparent that the substantial increase in PIII/PIV SVMP gene duplication events correlates strongly with the evolutionary timing of a reversion to vertebrate-feeding by *E. coloratus* (Figure 5.3A). In contrast, the remaining predominately arthropod-feeding species (Barlow *et al.* 2009) exhibit only a modest increase in net gene duplication/loss events; the greatest remaining increase was observed in *E. ocellatus* (Figure 5.3A) which feeds on vertebrates to a greater extent than either *E. p. leakeyi* or *E. c. sochureki* (Barlow *et al.* 2009).

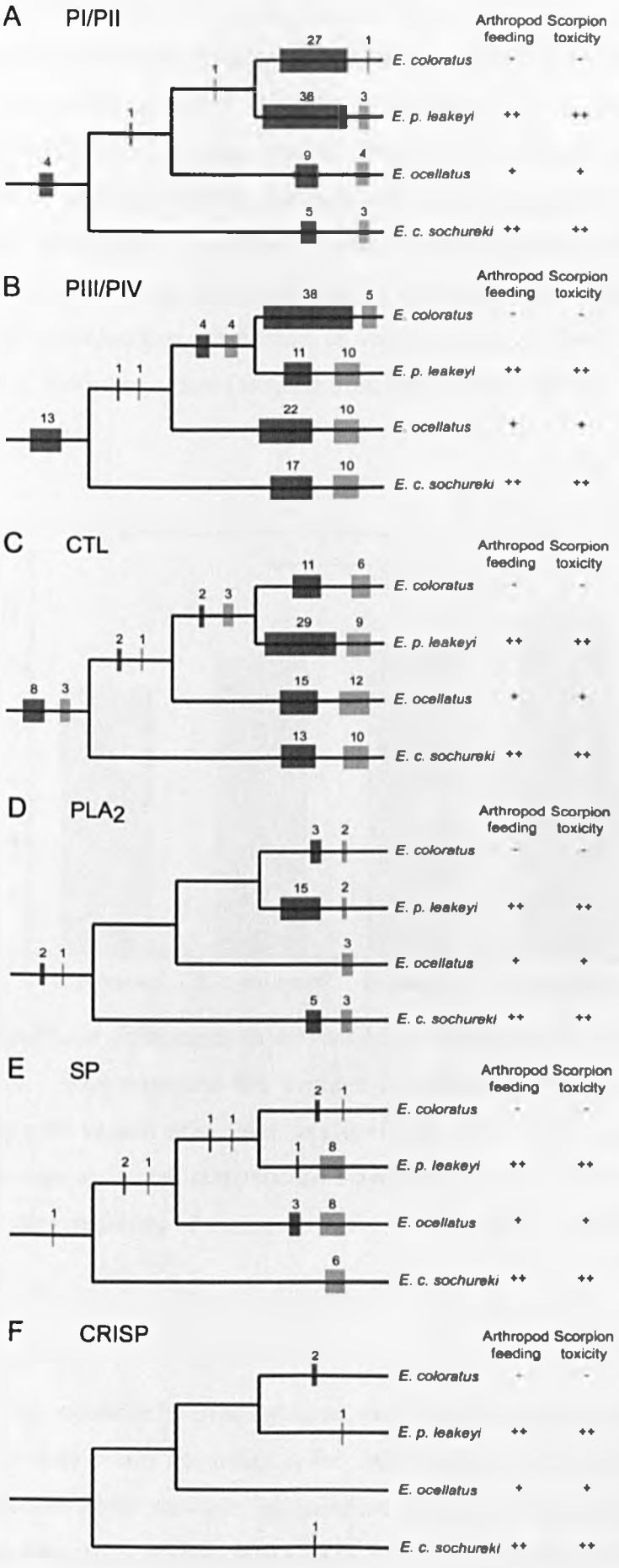
A PI/PII
Arthropod feeding
Scorpion toxicity
E. coloratus
E. p. leakeyi
E. ocellatus
E. c. sochureki
B PIII/PIV
C CTL
D PLA2
E SP
F CRISP

Figure 5.1 (previous page). Reconciled gene and species trees displaying gene duplication and loss events for representative *Echis*-derived toxin families. A) PI/PII SVMP sub-classes, B) PIII/PIV SVMP sub-classes, C) C-type lectin, D) phospholipase A_2, E) serine protease and F) cysteine-rich secretory proteins. Dark grey bars represent gene duplications and light grey represent gene losses. The width of bars visually represents the number of gene events annotated above each bar. Columns to the right indicate the proportion of arthropod prey consumed by the species and the corresponding correlation of venom toxicity to an arthropod prey item: ++, high; +, moderate; -, low (adapted from Barlow *et al.* 2009).

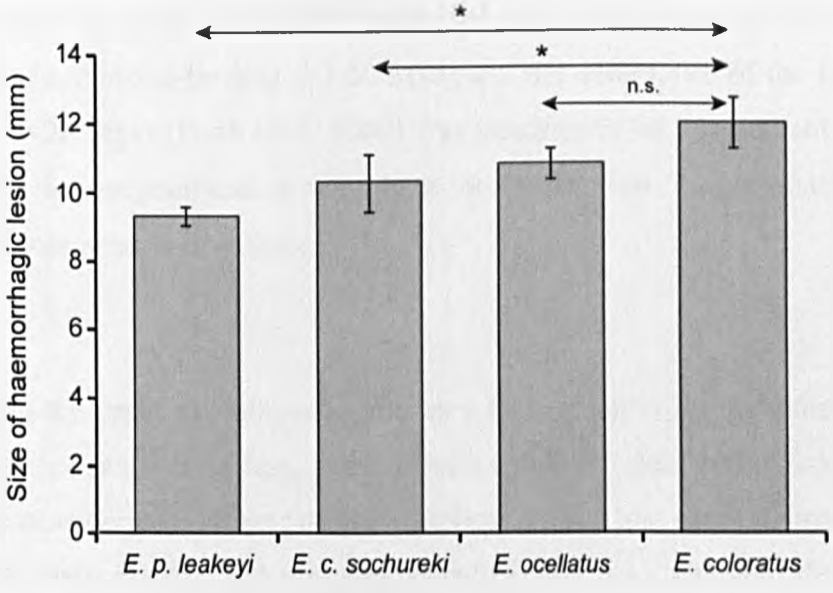

Figure 5.2. Significant differences in *in vivo* haemorrhagic activity of four *Echis* venoms in mice. Bars represent the average haemorrhagic lesion size ± s.e.m. induced by 10μg of venom after 24 hours ($p=0.043$, $n=6$ - one way analysis of variance). Pair-wise statistical comparisons between the activity of venom from *E. coloratus* and other members of the genus *Echis* are shown: * = $p<0.05$ and n.s. = not significant.

Contrastingly, the reconciled serine protease tree exhibits little evidence of gene diversification or loss events occurring in the venom gland of *E. coloratus*; instead the tree is characterised by multiple, independent gene loss events occurring in the remaining representatives of the genus (Figure 5.1E). The net result of these

duplications/losses is little change in the history of the genes in respect to a reversion to vertebrate-feeding: *E. coloratus* retains the majority of genes ancestrally present in the genus and exhibits little diversification (Figure 5.3B). In contrast, the other members of the genus *Echis* feed substantially on the ancestral prey item, arthropods (Barlow *et al.* 2009). This feeding strategy strongly correlates with the independent loss of multiple serine protease genes in each lineage (Figure 5.3B), inferring that serine protease gene products are not of functional importance for the capture of arthropod prey. Surprisingly, the loss of the majority of serine protease genes has not occurred at the base of the genus *Echis* where the shift to arthropod-feeding was inferred to have arisen (Barlow *et al.* 2009); independent losses in each arthropod-feeding lineage implies loss events have occurred following the divergence of the four *Echis* species groups. I therefore infer that either the evolutionary time between the origin of arthropod-feeding (22-30 Mya) and the divergence of the four species lineages (19-22 Mya) (Pook et al. 2009) was insufficient for considerable gene loss or that the selection pressures driving these losses were low, corresponding with the low rate of loss prior to divergence.

Considering the rapid evolutionary processes that underpin the evolution of snake venom proteins (Kini and Chan, 1999; Kordiš and Gubenšek, 2000; Župunski et al. 2003), the considerable number of gene duplication and loss events observed within most *Echis* toxin families are not unexpected (Figure 5.1), although the remaining toxin family reconciliation analyses reveal little correlation between *Echis*-derived toxin families and the pattern of dietary shifts previously described (Barlow et al. 2009). Whilst these results imply that a number of toxin families do not suffer substantial dietary selection pressures, notable increases in gene duplication events were observed in *E. p. leakeyi* for the CTLs, PLA_2s, and alongside *E. coloratus* for the PI/PII SVMPs (Figure 5.1), though the evolutionary pressures responsible for the radiation of these duplication events remains undetermined. The remaining toxin family, the CRISPs, exhibited a surprising lack of gene diversity (Figure 5.1F) and therefore may not suffer selective pressures to the same extent as the other toxin families.

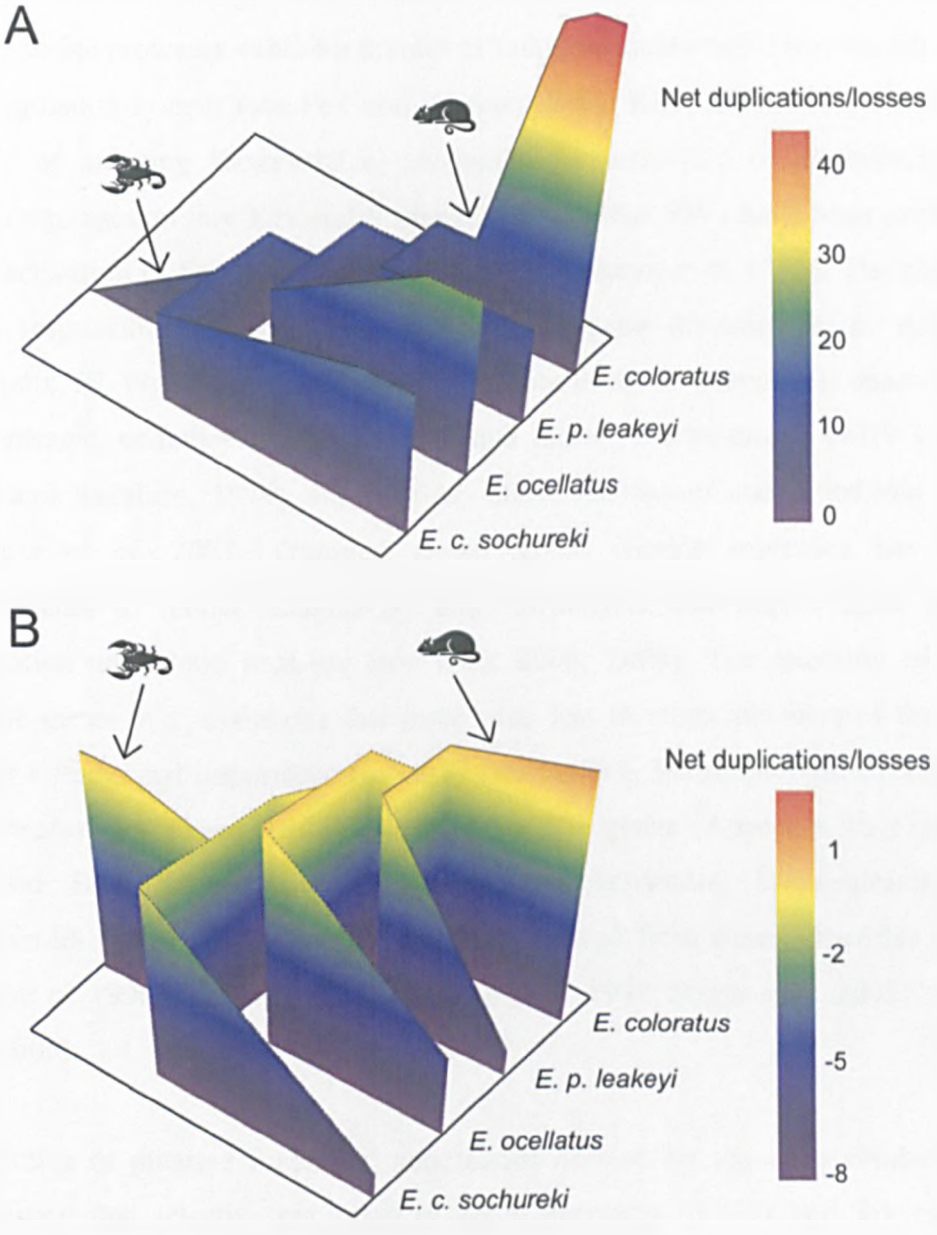

Figure 5.3. The net result of toxin family gene duplication and loss events mapped in 3D to the genus *Echis* species phylogeny. A) PIII/PIV sub-classes of SVMPs and B) serine proteases. The height of the branch reflects the net result of gene duplication and loss events in each lineage. Symbols indicate the phylogenetic position of the origin of arthropod-feeding (scorpion) and the reversion to vertebrate-feeding (mouse) in the genus *Echis* (adapted from Barlow *et al.* 2009).

Understanding the functional importance of lineage specific gene diversifications and losses within multi-gene toxin families is extremely complex. Both the SVMPs and the serine proteases exhibit a myriad of functional activities that primarily affect the coagulation system (see Fox and Serrano, 2005; Kini, 2006). PIII SVMPs are capable of inducing haemorrhage, apoptosis, the activation of prothrombin and platelet aggregation (see Fox and Serrano, 2005), whilst PIVs have been implicated in the activation of Factor X (Siigur *et al.* 2001; Takeya *et al.* 1992). The gene tree clades responsible for conferring increases in gene diversity in *E. coloratus* (Appendix III Figure 2) exhibited BLAST similarity to previously characterised haemorrhagic, endothelial cell apoptotic and factor X activating SVMPs (Omori-Satoh and Sadahiro, 1979; Siigur *et al.* 2001; Kishimoto and Takahashi, 2002; Assakura *et al.* 2003; Trummal *et al.* 2005). Serine proteases have been demonstrated to induce coagulation and fibrinolysis and impact upon platelet aggregation and blood pressure (see Kini, 2005; 2006). The retention of serine protease genes in *E. coloratus* that have been lost in other members of the genus implies a functional importance for vertebrate-feeding. Serine proteases retained by *E. coloratus*, yet absent in other members of the genus (Appendix III Figure 5), exhibited BLAST similarity to plasminogen activators, kinin-releasing and thrombin-like fibrinogenase serine proteases isolated from other Viperidae species (Hahn *et al.* 1996; Park *et al.* 1998; Serrano *et al.* 1998; Siigur *et al.* 2003; Sanchez *et al.* 2006).

Irrespective of putative functional annotations derived by sequence similarity, the observation that selective evolution of genes encoding SVMPs and SPs correlate with a reversion (or lack thereof) to vertebrate-feeding strongly implies that toxins acting upon multiple points in the coagulation cascade and capable of inducing haemorrhage are of functional importance for a predominately vertebrate-feeding strategy. Comparative haemorrhagic lesions induced by the four *Echis* venoms provides functional evidence that strongly supports this theory; significant intra-generic differences in haemorrhagicity were observed, with the vertebrate feeding species *E. coloratus* exhibiting the most haemorrhagic pathology (Figure 5.2). Interestingly, the difference in haemorrhage was greatest between the sister taxa of *E. coloratus* and *E. p. leakeyi*, highlighting considerable functional deviation following their divergence. The association between gene diversification/retention

of coagulopathic and haemorrhagic toxin families and the severity of *in vivo* haemorrhage provides strong evidence that dietary-induced venom adaptations have occurred as a response to a reversion to vertebrate feeding in *E. coloratus*. Considering the substantial differences between the circulatory systems and coagulation pathways of vertebrates and invertebrates (see Krem and Di Cera, 2002; Theopold *et al.* 2004; Muñoz-Chápuli *et al.* 2005), these observations are perhaps not unexpected. Whilst the open circulatory system in invertebrates shares a number of coagulatory components that exhibit similarity to their vertebrate counterparts, they are not true orthologues (Krem and Di Cera, 2002; Theopold *et al.* 2004). Therefore the absence of multiple SVMP and serine protease genes from the predominately arthropod-feeding saw-scaled vipers is likely a result of the differences in molecular targets present in the coagulatory pathways of these prey items. I suggest that dietary selection pressures are capable of driving the evolution of venom components by promoting the functional diversification of toxin families through the birth-and-death model of gene evolution (Ohta, 1991), thereby facilitating the neofunctionilization of genes which can assist in overcoming prey defences. Whilst it has previously been suggested some species may generate a suite of toxins to allow snake predators to adapt to a variety of prey species (Fry *et al.* 2003), here it appears that *E. coloratus* has selectively promoted the evolution of specific components which are functionally relevant for natural prey capture following an alteration in diet.

5.6 Conclusions

The first identification of differing selective genetic mechanisms that are acting independently upon multiple toxin families to confer functional alterations provides a key insight into the evolutionary adaptations responsible for variations in snake venom composition. Furthermore, the identification of adaptive processes that are acting to optimise the composition of venom to differing prey items highlights the potential influence of selective venom variation upon antivenom therapy. Considering venom components suffer evolutionary pressures independent to phylogenetic position, understanding the life history of a species becomes fundamental to comprehending the venom variation that exists between species and is therefore of utmost importance when selecting appropriate venoms for antivenom

production. Understanding the evolutionary processes that underpin the nature of venom variation will not only help us to understand the various pathologies induced by snakebites, but also aid the rational design of antivenom therapies that aim to confer increases in efficacy to the ~0.4-2.6 million people who suffer snake envenomations each year (Chippaux, 1998; Kasturiratne *et al.* 2008).

5.7 Authorship order and contributions

Nicholas R Casewell, Robert A Harrison, Darren AN Cook, Simon C Wagstaff and Wolfgang Wüster. I undertook the bioinformatic processing, preparation of clones for full length sequencing and all phylogenetic and gene tree parsimony analyses. I also undertook the experimental preparations for the *in vivo* assays and carried out the necessary observations and statistical analyses. RAH and DANC performed the animal experiments and DANC measured the haemorrhagic lesions. SCW provided bioinformatic guidance and WW provided assistance and expertise for the phylogenetic and gene tree parsimony analyses. I wrote the publication manuscript that forms the basis of this chapter.

CHAPTER 6

Bayesian gene tree parsimony of multi-gene snake venom protein families reveals species tree conflict as a result of multiple parallel gene loss

6.1 Abstract

The potential for gene tree parsimony to successfully recover species relationships from gene trees has increasing relevance considering the substantial generation of sequence data produced by recent genomic and transcriptomic studies. Previous studies have implemented bootstrap methodologies or Bayesian posterior distributions as a strategy to account for the uncertainty present in gene trees when inferring species trees. Here I implement a Bayesian methodology on multiple copy gene family datasets in the form of snake venom proteins for two separate groups of taxa. Bayesian gene tree parsimony largely failed to infer species trees congruent with each other or with robustly supported phylogenies derived from mitochondrial and single-locus nuclear sequences. Analysis of four toxin gene families from a large expressed sequence tag dataset from the viper genus *Echis* failed to produce a consistent topology, and re-analysis of a previously published gene tree parsimony dataset, from the family Elapidae, suggested that species tree topologies were predominantly unsupported. I propose that gene tree parsimony failure in the family Elapidae is likely the result of unequal and/or incomplete sampling of paralogous genes, and demonstrate that multiple parallel gene losses are likely responsible for the significant species tree conflict observed in the genus *Echis*. These results highlight the potential for gene tree parsimony analyses to be undermined by rapidly evolving multi-locus gene families experiencing non-random evolutionary pressures.

6.2 Introduction

The key assumption of molecular systematics is that the generation of gene phylogenies provides information about the evolutionary relationship of the organisms from which the genes have been isolated (Cotton and Page, 2002). It is often simply assumed that a gene phylogeny (gene tree) accurately represents the organismal phylogeny (species tree) of the species sampled (e.g. Okuda *et al.* 2001; Tsai *et al.* 2004, 2007). However, the suggestion that a species tree can be obtained simply by sampling a specific gene across a range of species is often erroneous (Page and Cotton, 2000; Cotton and Page, 2002), particularly if the gene is of multiple copy origin rather than having a single chromosomal locus. Correctly inferred gene trees do not always correspond to species trees due to evolutionary processes such as duplication and loss, deep coalescence and horizontal transfer (Goodman *et al.* 1979; Doyle, 1992; Slowinski and Page, 1999; Galtier and Daubin, 2008). The combination of gene duplication and loss can produce conflicts with a species tree when paralogous sequences are sampled and treated as orthologous, a common occurrence in under-sampled datasets (Figure 6.1A) (Slowinski *et al.* 1997; Page and Cotton, 2000). Deep coalescence (or ancestral polymorphism) is an event at a single locus where a sequence from a less related species coalesces with one of the descendents of the deep coalescence (Figure 6.1B) (Slowinski *et al.* 1997; Slowinski and Page, 1999). Deep coalescence can produce an analogous situation to duplication and loss because paralogous sequences are simply sequences that have coalesced prior to the ancestor of the species from which they were sampled (Slowinski *et al.* 1997; Slowinski and Page, 1999). Sequencing both loci of a duplicated gene should resolve the discordance between species and gene trees due to paralogous sequences (Doyle, 1992), highlighting the fundamental importance of substantial gene sampling. Horizontal transfer, including processes such as hybridisation and gene transfer between species, is widely assumed to be more common in prokaryotes and of lesser importance in eukaryotic datasets (Figure 6.1C) (Syvanen, 1994; Galtier and Daubin, 2008).

The reconciliation of species and gene trees was first implemented by Goodman *et al.* (1979) and has subsequently been progressed by a number of different

approaches over the years (e.g. Page, 1994; Eulenstein, 1997; Ronquist, 1997). An extension of tree reconciliation, gene tree parsimony, aims to identify the species tree that minimises the assumptions of evolutionary events (duplications, losses and/or deep coalescences) necessary to fit a given gene tree to the species tree (Slowinski *et al.* 1997; Slowinski and Page, 1999), a considerable challenge given the frequency with which these events occur, particularly within rapidly diversifying gene families (Page and Cotton, 2000). GeneTree (Page, 1998) was the first program to implement this logical strategy by using simple, standard tree search heuristics to infer species trees from gene trees under three independent optimality criteria: duplications and losses, duplications-only and deep coalescences. Subsequent programs and models have attempted to improve the biological realism of gene processes through time (e.g. Liu and Pearl, 2007; Liu *et al.* 2010) or improve the implementation of gene tree parsimony (Sanderson and McMahon, 2007; Oliver, 2008; Wehe *et al.* 2008). Nevertheless, despite its simple search strategy, GeneTree remains the only widely available software that implements independent analyses for gene duplication and loss, gene duplication only and deep coalescence optimality criteria. Ideally, a strategy that uses heuristics to search for multiple gene processes simultaneously would be applied, although such a method has yet to be implemented due to the fundamental problem of how to weight duplications, losses and coalescences against each other. Despite this issue, gene tree parsimony has been reported to obtain results consistent with other analyses in snakes (Slowinski *et al.* 1997) and vertebrates (Cotton and Page, 2002), and performed well against a known species tree in an Angiosperm dataset (Sanderson and McMahon, 2007). Considering the substantial increases in the generation of sequence data by recent genomic and transcriptomic studies, assessing the potential for gene tree parsimony to successfully recover species relationships from comprehensively sampled datasets has become a particularly timely exercise.

A major criticism of the gene tree parsimony methodology is that it fails to quantify confidence levels in the reconciled species tree by disregarding any uncertainty in the gene tree (Page and Cotton, 2000; Sanderson and McMahon, 2007). In order to account for gene tree uncertainty when inferring species trees, methodologies that incorporate the bootstrap have been implemented (Cotton and Page, 2002; Sanderson

and McMahon, 2007) and the use of Bayesian posterior distributions has been advocated (Buckley *et al.* 2006; Oliver, 2008). The use of Bayesian Markov Chain Monte Carlo (MCMC) analyses is particularly valuable, as they produce less biased predictions of phylogenetic accuracy, accommodate the inherent uncertainty present in gene genealogies, provide easy interpretation of results, and have computational advantages over other techniques (Larget and Simon, 1999; Huelsenbeck and Ronquist, 2001; Alfaro *et al.* 2003; Ronquist and Huelsenbeck, 2003). Furthermore, it has been demonstrated that subjecting substantial numbers of gene tree Bayesian posterior distributions to multiple species tree searches prior to generating a majority rule consensus tree can provide a rigorous assessment of node uncertainty within an inferred species tree (Buckley *et al.* 2006; Oliver, 2008).

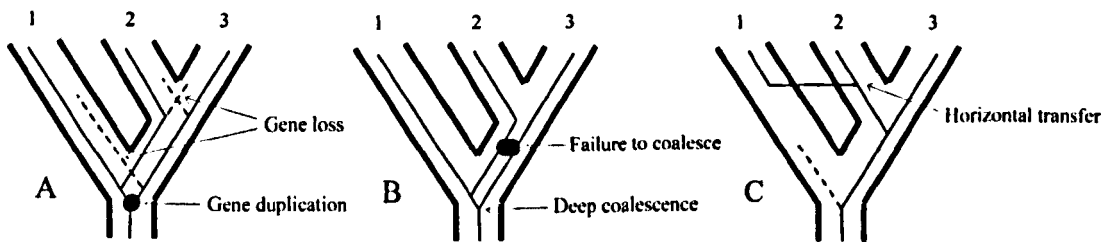

Figure 6.1. Examples of gene trees embedded in a species tree, demonstrating sources of gene tree and species tree conflict (adapted from Slowinski and Page, 1999). A: Duplication and loss. B: Deep coalescence. C: Horizontal transfer. In each case, the gene tree groups species 1 and 2 together despite them not being each other's closest relatives.

Snake venoms are a complex mixture of proteins and peptides; they exhibit a high level of biological activity and a diverse array of actions on both natural prey items and humans (Chippaux, 1991, Aird, 2002). The majority of venom proteins appear to have been recruited into the venom gland from multi-gene protein families normally expressed in a variety of bodily tissues for ordinary physiological 'housekeeping' purposes (Fry, 2005). Following their recruitment, venom proteins evolve rapidly via a 'birth and death' model of evolution, whereby frequent duplications of protein-encoding genes permit rapid functional and structural diversification alongside enhanced rates of sequence evolution (Nei *et al.* 1997; Kini

and Chan, 1999; Kordiš and Gubenšek, 2000; Župunski *et al.* 2003). Whilst some genes become deleted from the genome or degenerate into pseudogenes, others undergo neofunctionalization, resulting in the generation of a range of proteins that exhibit distinct functional diversification (Fry *et al.* 2003b; Lynch, 2007). These rapidly evolving gene families provide an ideal model to investigate whether species tree relationships can be predicted from rapidly evolving multiple copy genes using gene tree parsimony.

Here I use a novel dataset containing snake venom gland expressed sequence tags (ESTs) from four closely related species of saw-scaled vipers (Serpentes: Viperidae: *Echis*) alongside a strongly supported phylogeny, based on mitochondrial and single copy nuclear gene sequences (Figure 6.2) (Barlow *et al.* 2009; Pook *et al.* 2009), to implement entire Bayesian posterior distributions in gene tree parsimony. These identically generated, multi-species EST datasets provide an unbiased, directly comparable sampling resource and supply comprehensive multiple copy data for multiple gene families, whilst the rigorous generation of node support values using Bayesian posterior distributions provides a measure of confidence for species tree interpretation. Furthermore, the EST data, together with a quantitative measure of species tree support, permit a direct comparison between species trees inferred from nucleotide and translated nucleotides, thereby allowing the relationship between multiple copy gene trees and species trees to be investigated in greater detail. Snake venom protein families have previously been analysed using gene tree parsimony: Slowinski *et al.* (1997) recovered species relationships consistent with other analyses inferred from phospholipase A_2 (PLA_2) and short neurotoxin (NXS) venom proteins isolated from members of the Elapidae (Serpentes). However, this investigation did not take into account the substantial uncertainty observed in the gene trees; for this reason I revisit this dataset and apply Bayesian posterior distributions in order to interpret the inferred species trees alongside rigorously generated node support values.

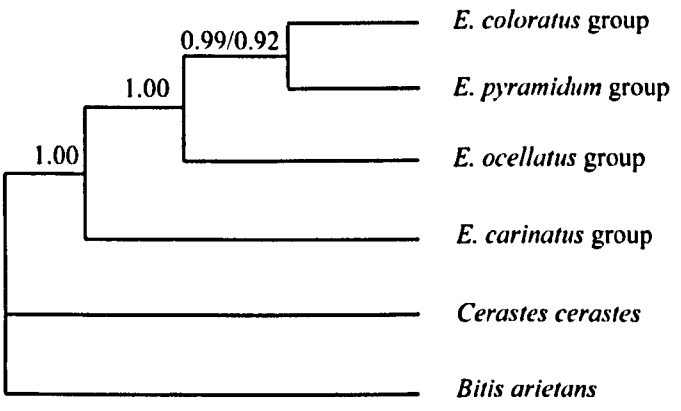

Figure 6.2. Bayesian phylogeny of the major *Echis* species groups inferred by four mitochondrial genes and one nuclear gene (adapted from Barlow *et al.* 2009; Pook *et al.* 2009). Bayesian posterior probabilities are shown for relevant nodes. Outgroup taxa are *Cerastes cerastes* and *Bitis arietans*.

6.3 Methods

6.3.1 Venom protein sequences

Venom gland cDNA libraries were constructed using procedures previously outlined (Wagstaff and Harrison, 2006; Casewell *et al.* 2009 – Chapter 4). Briefly, multiple cDNA libraries were constructed from ten wild-caught specimens of *Echis ocellatus* (Nigeria), *E. pyramidum leakeyi* (Kenya), *E. coloratus* (Egypt) and *E. carinatus sochureki* (UAE); ~1000 random clones per species were picked for sequencing using M13 forward primers. ESTs were bioinformatically processed using the PartiGene pipeline (Parkinson *et al.* 2004) with high stringency CLOBB clustering (Parkinson *et al.* 2002; Wagstaff and Harrison, 2006) and BLAST annotation against multiple databases (see Casewell *et al.* 2009 – Chapter 4). ESTs exhibiting significant (>1e-05) BLAST annotation to the most representative venom proteins present in the venom gland transcriptomes, the snake venom metalloproteinase (SVMP), C-type lectin (CTL), PLA_2 and serine protease (SP) protein families, were identified prior to alignment in Clustal W (Thompson *et al.* 1994). Full length sequencing of PLA_2 and CTL clones were obtained during the initial round of sequencing, whilst reverse sequencing, using M13 reverse primers, was carried out on all SP clones to generate full length sequences. Due to the frequency of SVMP annotated sequences, full length sequence information was gained via primer

walking a non-redundant set of SVMP clones which demonstrated sequence similarity to the catalytic site (H-box) of the metalloproteinase domain (Fox and Serrano, 2005). Outgroup sequences for each gene family were identified by sequence similarity searches against a number of non-Serpentes databases. The datasets were trimmed to the open reading frame of the translated proteins; identical sequences and those containing truncations or frameshifts as the result of insertions or deletions were excluded in MEGA v4.0.2 (Tamura *et al.* 2007). The alignment of full length variants using Clustal W preceded additional manual adjustments. The finalised DNA datasets were then translated into amino acids (AA) and realigned before the exclusion of any remaining identical sequences.

The Elapidae PLA_2 (59 sequences from 25 species) and NXS datasets (42 sequences from 27 species) analysed by Slowinski *et al.* (1997) were retrieved from the protein database SWISS-PROT using the NCBI browser. Signal sequences were removed in MEGA v4.0.2 prior to alignment in Clustal W and subsequent manual adjustments.

6.3.2 Gene tree analysis

Gene trees were produced using optimised models of sequence evolution combined with Bayesian inference. Given that complex models of sequence evolution have been demonstrated to extract additional phylogenetic signal from data (Castoe *et al.* 2005; Castoe and Parkinson, 2006), I subjected the DNA datasets to analysis in MrModeltest v2.3 (Nylander, 2004) and the AA datasets in ModelGenerator v0.85 (Keane *et al.* 2006). Prior to analysis of the DNA datasets, sequences were partitioned into first, second and third codon partitions to incorporate any differences in patterns of sequence evolution. The model favoured under the Akaike Information Criterion (AIC) was selected for all partitions. Bayesian inference analyses were undertaken in MrBayes v3.1 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) on the freely available bioinformatic platform Bioportal (www.bioportal.uio.no). Each dataset was run in duplicate using four chains simultaneously (three heated and one cold) for $5x10^6$ generations, sampling every 500th cycle from the chain and using default settings in regards to priors.

Plots of ln(L) against generation were constructed to determine the burnin period; trees generated prior to the completion of burnin were discarded.

6.3.3 Tree reconciliation

To infer species trees from gene trees, I implemented a gene tree parsimony strategy similar to that described by Buckley *et al.* (2006) and Oliver (2008) using a novel bioinformatic pipeline consisting of GeneTree v1.0 (Page, 1998) and PAUP* v4.0b10 (Swofford, 2002). Tree topologies of the total post-burnin trees (36004) generated for each Bayesian dataset were extracted in PAUP using the savetrees command and by removing branch lengths and internal node labels. The tree topologies were edited to GeneTree input specifications before each of the trees was subjected to heuristic species tree searches in GeneTree using the steepest ascent option. Each analysis was run using fifty heuristic searches in order to undertake a comprehensive search of the tree space and to account for extraneous random starting trees, whilst branch swapping was carried out using the most effective option (ALT), which alternates between nearest-neighbour interchanges and subtree pruning and regrafting (Page and Charleston, 1997). The individual species trees inferred from each of the post-burnin gene trees were summarised into a single consensus species tree using the majority rule consensus tree function in PAUP. The frequency of each node recovered from the 36004 inferred species trees thus represents a measure of the uncertainty for the relationships present in the consensus species tree.

The generation of reconciled species trees was undertaken separately for each venom protein family, as they represent independent non-homologous gene families and therefore likely exhibit different gene histories and rates of change between unlinked loci (Takahata, 1989; Maddison, 1997). Furthermore, for each venom protein family the heuristic searches in GeneTree were implemented for three separate optimality criteria for both DNA and AA datasets: (i) duplications and losses, (ii) duplications only and (iii) deep coalescences.

6.3.4 Alterations in methodology: Elapidae datasets

The number of species in the Elapidae datasets caused computational problems when implementing total Bayesian posterior distributions. To reduce GeneTree processing times, heuristic searches were reduced to one; analyses minimising duplications and losses were successfully processed, whilst alternate post-burnin trees were implemented for the deep coalescence criterion in order to maintain GeneTree computational time at a manageable level. Whilst sampling in the EST dataset appears to be comprehensive (Wagstaff and Harrison, 2006; Casewell *et al.* 2009 – Chapter 4), sampling of the Elapidae protein families was non-systematic and therefore almost certainly highly incomplete. When gene sampling is incomplete, it is difficult to distinguish gene loss from the absence of data, suggesting that implementing gene tree parsimony to minimise gene duplications only is more realistic and appropriate than seeking to minimise both duplications and losses (Wehe *et al.* 2008). Computational constraints prevented processing the duplications only criterion for the Elapidae datasets in GeneTree; I therefore employed a restricted Bayesian posterior distribution strategy using the faster heuristic searches implemented in DupTree (Wehe *et al.* 2008). As DupTree generates a single inferred species tree for multiple gene trees, in this case multiple Bayesian posterior distributions, I partitioned the total post-burnin trees into 100 partitions of 360 trees and ran the standard analysis for each. I subsequently summarised the inferred species trees into a consensus species tree as described above. Whilst this method has obvious limitations compared to inferring species trees from individual post-burnin trees, it is the most rigorous methodology available considering the computer limitations associated with implementing Bayesian posterior distributions for datasets containing large species numbers in GeneTree.

6.4 Results

6.4.1 Sequence data and Bayesian inference

The *Echis* datasets comprised of a total of 2004bp of SVMP (n=209) [GenBank: GU012123-GU012315 and AM039691-AM039701], 780bp of SP (n=32) [GenBank: GU012092-GU012122], 519bp of CTL (n=130) and 444bp of PLA_2 (n=42) aligned

sequence data (CTL and PLA_2 GenBank accession numbers can be found in Appendix III Table 1). The aligned *Echis* amino acid datasets represented 667 amino acids of SVMP (n=194), 260 amino acids of SP (n=27), 173 amino acids of CTL (n=116) and 144 amino acids of PLA_2 (n=33) sequence. The Elapidae datasets implemented by Slowinski *et al.* (1997) were aligned into 126 AA of PLA_2 (n=59) and 65 AA of NXS (n=42) sequence data. For Bayesian inference, MrModelTest v2.3 identified the following models of sequence evolution for the DNA data partitions: GTR + I + Γ for SVMP and CTL codon position 1 and CTL codon position 2, HKY + I + Γ for SVMP codon position 2, GTR + Γ for PLA_2 and SP codon position 1 and SVMP and PLA_2 codon position 3, HKY + Γ for PLA_2 codon position 2 and CTL and SP codon position 3 and SYM + I + Γ for SP codon position 2. ModelGenerator v0.85 selected the WAG + Γ model for all AA datasets except the *Echis* SVMP gene family, where a mixed model of evolution was implemented as the size of this dataset prevented model selection. Following Bayesian inference, visual inspection of the plots of tree ln(*L*) vs. generation indicated that burnin was complete in all datasets after approximately 100,000 generations, although I discarded the first 500,000 generations as an additional safety margin.

6.4.2 Gene tree parsimony in the genus *Echis*

The majority rule consensus trees generated by gene tree parsimony analyses for duplication and loss, duplications-only and deep coalescences are shown in Figure 6.3 (DNA) and Figure 6.4 (amino acid). Notably, considerable variation was observed in the species trees recovered from the different venom protein families; no less than nine differing species tree topologies (out of 26 possible) were inferred from the twenty-four analyses. Furthermore, only two fully resolved species trees, generated using the SVMP protein family amino acid gene trees under the duplication and loss and duplications-only optimality criteria, matched the strongly supported mitochondrial and single-locus nuclear gene phylogeny for this genus (Barlow *et al.* 2009; Pook *et al.*, 2009); only one node, supporting the monophyly of *E. coloratus* and *E. p. leakeyi*, was strongly supported (>95%) in both trees. While this node represented the most frequent node observed in the *Echis* species trees, it was not ubiquitous and only strongly supported in 25% (DNA) and 58% (AA) of the

inferred trees. However, many other species trees contained nodes incongruent both with the phylogeny of Barlow *et al.* (2009) and with species trees recovered from other venom protein gene trees, although only one of these was strongly supported (the monophyly of *E. ocellatus* and *E. p. leakeyi* in the serine proteases under the duplication and loss and deep coalescence criteria). A number of other nodes were unresolved or weakly supported, highlighting the lack of topological consistency observed throughout the inferred species trees.

6.4.3 Gene tree parsimony in the family Elapidae

The species trees inferred from gene tree parsimony analysis of the PLA_2 family are displayed in Figure 6.5; despite the differences in the optimality criterion employed by gene tree parsimony, the inferred species trees display similar topologies. The NXS data (Figure 6.6) produced largely unresolved species topologies, except when implementing the duplications only criterion in DupTree.

6.5 Discussion

6.5.1 Gene tree parsimony in the genus *Echis*

A total of nine distinct species trees were generated from the gene tree parsimony analyses of the four *Echis* venom protein families. The incongruence among species trees derived from these datasets is highlighted by the fact that the most frequent species tree topology represents only 25% of the total number of inferred species trees. Furthermore, the most commonly observed species tree is only partially resolved. This incongruence is particularly surprising, with only the duplications and loss analyses of DNA sequences for the SVMPs and CTLs producing fully resolved identical topologies, although neither tree exhibits strong support (>95%) for every node. The lack of consistency among species trees is important, as it highlights the absence of any strong signal opposing that of the mitochondrial and

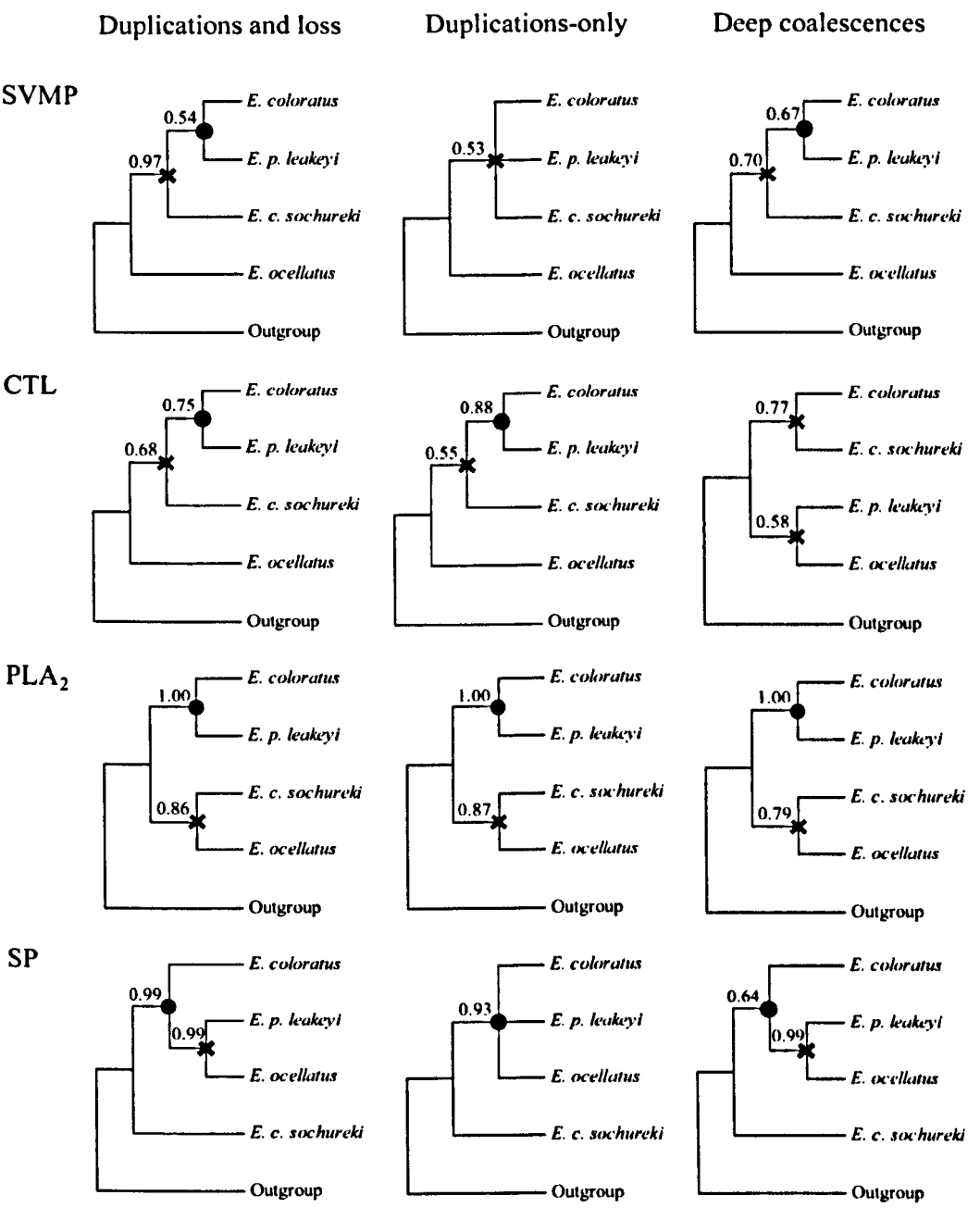

Figure 6.3. Majority rule consensus trees for four DNA datasets of venom protein families using gene tree parsimony. Separate analyses were implemented to minimise duplications and loss, duplications-only and deep coalescences. Circles indicate nodes congruent and crosses indicate nodes incongruent with the mitochondrial and nuclear phylogeny of the genus *Echis* (Barlow *et al.* 2009; Pook *et al.* 2009). Black circles and crosses indicate that a node is robustly supported (>95%), grey symbols indicate insignificant node support.

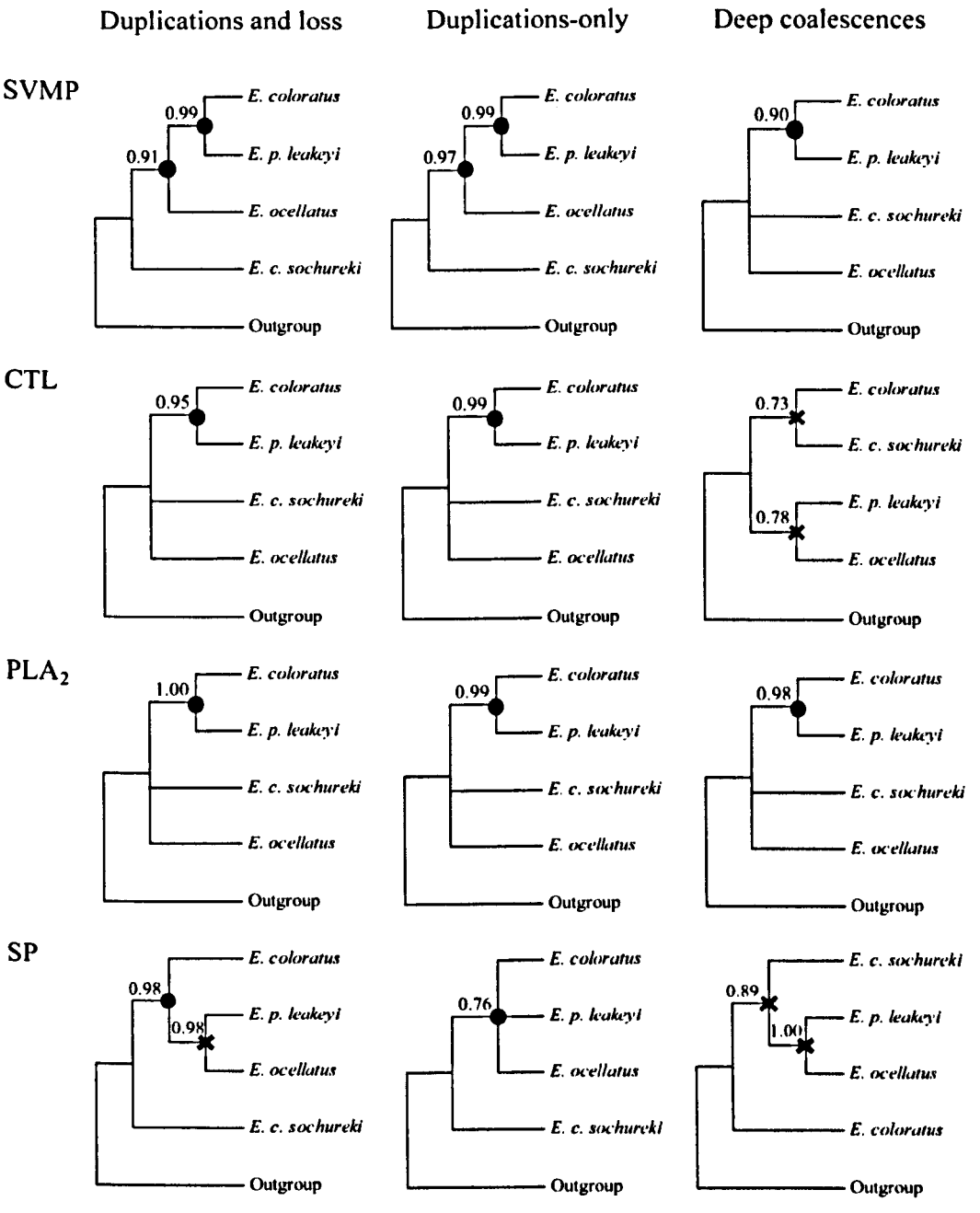

Figure 6.4. Majority rule consensus trees for four amino acid datasets of venom protein families using gene tree parsimony. Separate analyses were implemented to minimise duplications and loss, duplications-only and deep coalescences. Circles indicate nodes congruent and crosses indicate nodes incongruent with the mitochondrial and nuclear phylogeny of the genus *Echis* (Barlow *et al.* 2009; Pook *et al.* 2009). Black circles and crosses indicate that a node is robustly supported (>95%), grey symbols indicate insignificant node support.

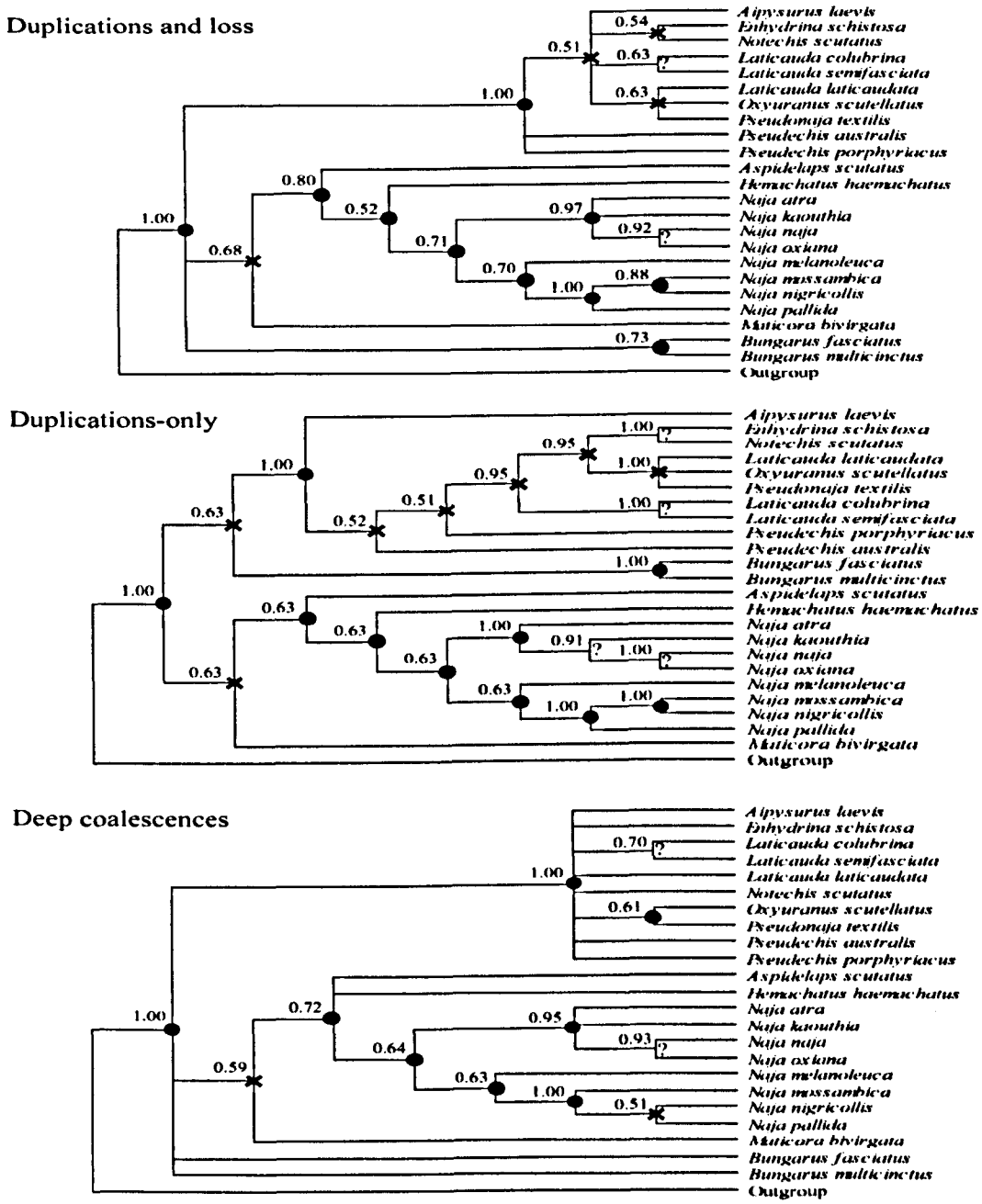

Figure 6.5. Majority rule consensus trees for the Elapidae PLA_2 venom protein family using gene tree parsimony. Separate analyses were implemented to minimise duplications and loss, duplications-only and deep coalescences. Circles indicate nodes congruent and crosses indicate nodes incongruent with mitochondrial analyses (Slowinski and Keogh, 2000; Lukoschek and Keogh, 2006; Wüster and Broadley, 2007; Wüster *et al.* 2007; Sanders *et al.* 2008). Black circles and crosses indicate that a node is robustly supported (>95%), grey symbols indicate insignificant node support. Question marks represent nodes for which the species relationships remain undetermined.

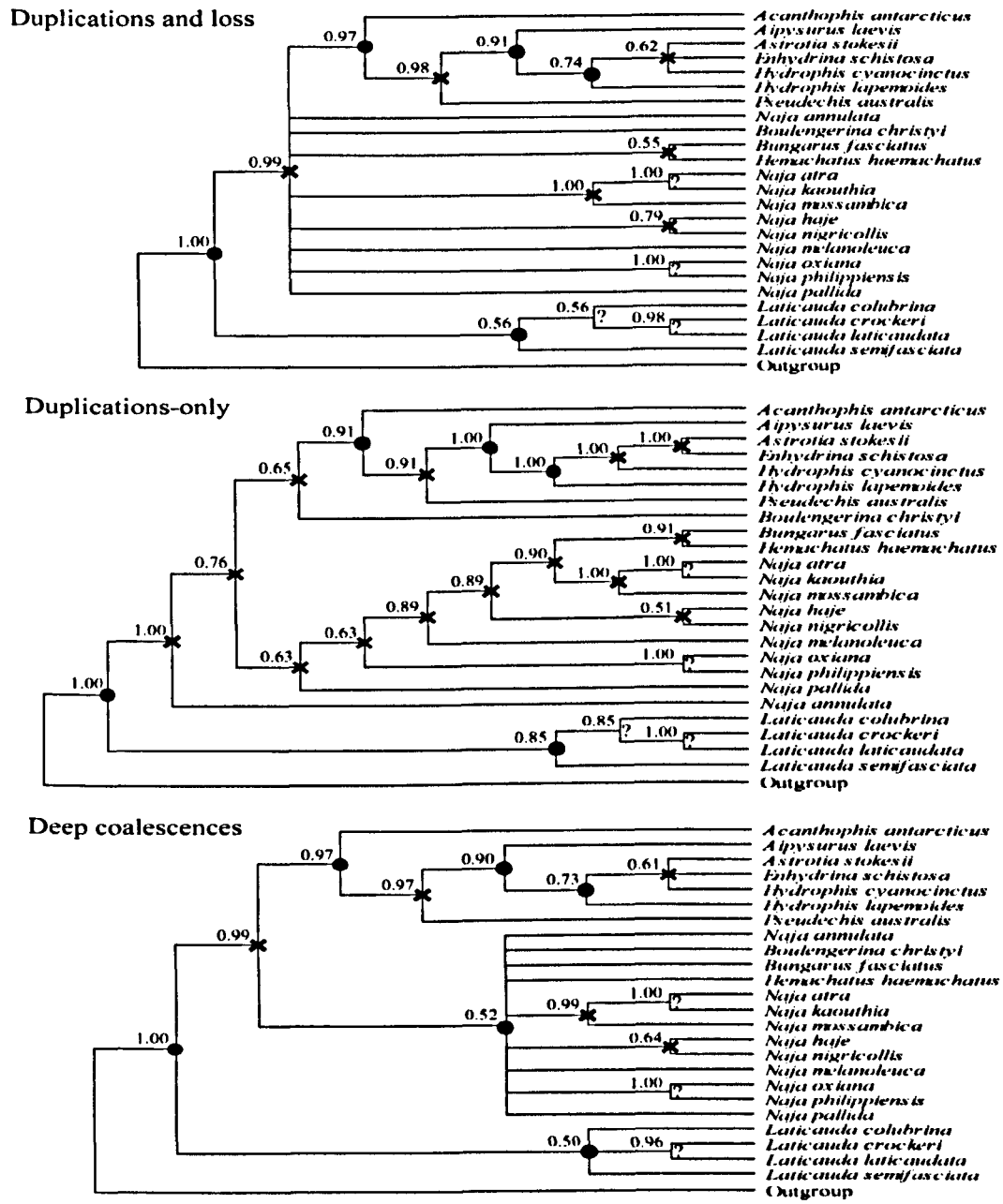

Figure 6.6. Majority rule consensus trees for the Elapidae NXS venom protein family using gene tree parsimony. Separate analyses were implemented to minimise duplications and loss, duplications-only and deep coalescences. Circles indicate nodes congruent and crosses indicate nodes incongruent with mitochondrial analyses (Slowinski and Keogh, 2000; Lukoschek and Keogh, 2006; Wüster and Broadley, 2007; Wüster *et al.* 2007; Sanders *et al.* 2008). Black circles and crosses indicate that a node is robustly supported (>95%), grey symbols indicate insignificant node support. Question marks represent nodes for which the species relationships remain undetermined.

nuclear phylogeny derived by Barlow *et al.* (2009) and Pook *et al.* (2009); consistent conflict between reconciled trees and the species phylogeny might suggest that the latter tree is in error; however, I did not uncover any consistent pattern of conflict. Moreover, the two nodes present in the inferred species trees that are congruent with the mitochondrial and nuclear phylogeny are only strongly supported in 42% (monophyly of *E. coloratus* and *E. p. leakeyi*) and 13% (monophyly of *E. coloratus*, *E. p. leakeyi* and *E. ocellatus*) of the total derived trees. Interestingly, analyses of the serine proteases showed that it was the only venom protein family that strongly contradicts the monophyly of *E. p. leakeyi* and *E. coloratus*; instead the monophyly of *E. p. leakeyi* and *E. ocellatus* is observed, except in the duplications-only analysis which fails to resolve the relationships among *E. coloratus*, *E. p. leakeyi* and *E. ocellatus*.

The majority of GeneTree analyses of Bayesian posterior distributions generated from the *Echis* DNA datasets produced fully resolved inferred species trees that are typically supported by higher node values than their amino acid counterparts. This is not unexpected given the increase in the number of phylogenetically-informative characters used to resolve the DNA gene trees and clearly emphasises the preferred use of nucleotide datasets for gene tree parsimony where available. Although many amino acid species trees exhibit unresolved nodes, the majority of the resolved clades display topologies identical to those inferred by the corresponding DNA datasets. The main exception is that the AA duplication and loss and duplications-only SVMP species trees; both exhibit different tree topologies from their DNA counterparts with greater node support. Coincidentally, these species trees are unique in inferring the topology of the genus *Echis* as predicted by the mitochondrial and nuclear phylogeny (Barlow *et al.* 2009; Pook *et al.* 2009).

Altering the optimality criteria implemented in GeneTree also resulted in alterations to the inferred species tree topologies: trees minimising deep coalescences are often incongruent with those inferred by minimising duplications and losses and duplications-only. Moreover, they display lower node support values in the majority of trees, indicating that duplications and losses may be more important in the

evolutionary history of snake venom proteins. In general, major changes in the species tree topologies are not observed between the duplication and loss and duplications-only analyses, consistent with the assumption that losses are informative in the EST datasets as a consequence of comprehensive gene sampling. However, in contrast to the duplications and loss analyses, none of the species trees derived from the duplications-only criteria exhibit strongly supported nodes that conflict with the mitochondrial/nuclear DNA phylogeny (Barlow *et al.*, 2009; Pook *et al.*, 2009), implying that the inclusion of loss events may be partially responsible for gene tree parsimony incongruence in this dataset. Notably, the duplications-only analyses for the SP venom protein family reveals a consistent change in tree topology from fully resolved to partially resolved, the collapsed node being the grouping of *E. ocellatus* and *E. p. leakeyi* that is inconsistent with the "known" species tree. Considering the comprehensive sampling methodology for the four venom protein families, this observation implies that gene loss in the serine proteases has a greater influence on the outcome of species tree reconstruction than in the other venom protein families.

Considering the lack of congruence between reconciled venom protein gene trees and the genus *Echis* phylogeny, gene tree parsimony was subsequently undertaken by simultaneously considering multiple gene loci derived from the SVMP, CTL, PLA_2 and SP DNA consensus gene trees using the deep coalescences multiple loci analysis in Mesquite (Maddison and Knowles, 2006; Maddison and Maddison, 2008). This approach also failed to infer a species tree congruent with the species phylogeny determined by Barlow *et al.* (2009) and Pook *et al.* (2009) (Figure 6.7A). Despite simultaneously incorporating data from the four gene loci, the reconciled tree was incongruent with the *Echis* phylogeny and included the monophyly of *E. p. leakeyi* and *E. ocellatus*. Considering different venom proteins represent independent non-homologous gene families these results are not unexpected; differing gene families likely exhibit different gene histories and rates of change (Takahata, 1989; Maddison, 1997). However, it is notable that when excluding the SP protein family from the deep coalescence multiple loci analysis, the reconciled tree supports the monophyly of *E. p. leakeyi* and *E. coloratus* (Figure 6.7B), a node congruent with Barlow *et al.*'s (2009) and Pook *et al.*'s (2009) analyses. Whilst the subsequently

reconciled tree remains partially incongruent with the previously determined species phylogeny, these results further highlight the potential conflicting influence the serine protease gene family has on gene tree parsimony in the genus *Echis*.

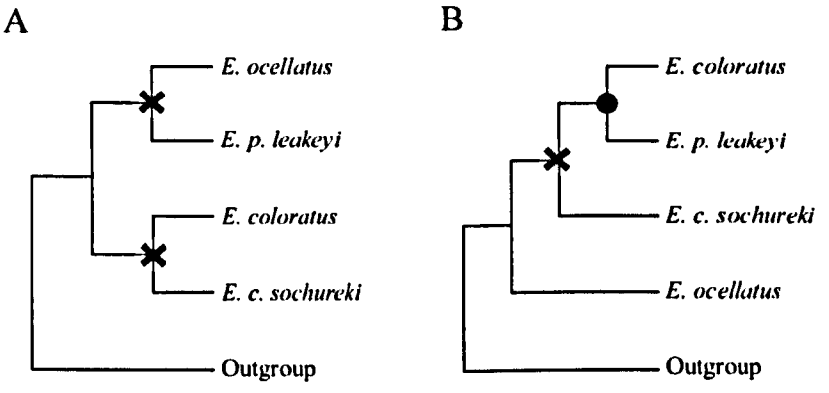

Figure 6.7. Reconciled trees derived from multiple loci deep coalescence analyses of *Echis* venom protein families. A: SVMP, CTL, PLA_2 and SP loci. B: SVMP, CTL and PLA_2 loci. Circles indicate nodes congruent and crosses indicate nodes incongruent with the mitochondrial and nuclear phylogeny of the genus *Echis* (Barlow *et al.* 2009; Pook *et al.* 2009).

6.5.2 Gene tree parsimony in the family Elapidae

The inferred species trees generated from the Elapidae PLA_2 gene family provided strong support for the monophyly of the Australian and marine elapid radiation throughout the varying gene optimality criterions (gene duplication, loss and deep coalescence). The analysis minimising duplications only was alone in resolving the relationships within this clade with any significant support, inferring both the monophyly of the Australian and marine elapids to the exclusion of *Aipysurus* and *Pseudechis* and the non-monophyly of *Laticauda*; however, both observations are strongly contradicted by recent molecular phylogenetic studies using mitochondrial and single-locus nuclear gene sequences (Slowinski and Keogh, 2000; Sanders *et al.* 2008). The monophyly of marine and terrestrial Australian species was also recovered by Slowinski *et al.* (1997), with their analysis suggesting a fully resolved

topology incongruent with these analyses, likely reflecting an unsupported species relationship. In the context of the African and Asian Elapids, the monophyly of *Aspidelaps*, *Hemachatus* and *Naja* established by the consensus trees matched those of Slowinski *et al.* (1997); despite this consistency, support values are insufficient to exclude the possibility of an alternate topology. Furthermore, Slowinski *et al.*'s (1997) placement of *Bungarus* as outgroup to *Aspidelaps*, *Hemachatus* and *Naja*, is unsupported by this PLA_2 analyses, with different gene optimality criteria producing contrasting topologies. However, within the genus *Naja*, the consensus trees are consistent with both Slowinski *et al.* (1997) and recent mitochondrial phylogenies (Wüster and Broadley, 2007; Wüster *et al.* 2007), exhibiting strong support for the monophyly of the African spitting cobras (*N. mossambica*, *N. pallida* and *N. nigricollis*) and the Asian cobras (*N. kaouthia*, *N. atra*, *N. naja* and *N. oxiana*).

The NXS consensus trees produced largely unresolved species topologies, except when implementing the duplications only criterion in DupTree (Figure 6.6). The observed unresolved topologies and corresponding low node support values are perhaps unsurprising given that the NXS dataset contains less sequence data (65 amino acids) and thus fewer characters than the other venom proteins, due to the short length of the NXS genes. All three gene analyses provided strong support for *Laticauda* as the sister taxon of all other Elapidae, conflicting with Slowinski *et al.*'s analyses (1997) and a recent multi-gene phylogeny (Sanders *et al.* 2008); both placed *Laticauda* at the base of the marine and terrestrial Australian elapids. The significant support values associated with the placement of *Laticauda* suggest that the topology obtained by Slowinksi *et al.* (1997) may not have been strongly supported, despite its consistency with Sanders *et al.* (2008). Nevertheless, the monophyly of the marine and terrestrial Australian snakes, excluding *Laticauda*, is supported in each consensus tree (all nodes >90%), displaying a topology similar to that described previously (Slowinski *et al.* 1997), although inconsistencies with recent molecular phylogenies, including the non-monophyly of (i) *Hydrophis lapemoides* and *H. cyanocinctus* and (ii) *Acanthophis* and *Pseudechis*, exist within this clade (Lukoschek and Keogh, 2006; Sanders *et al.* 2008). The NXS consensus trees also fail to significantly resolve the species relationship of the African and Asian elapids, except when minimising duplications-only. This analysis provided

support for i) the paraphyly of *Naja* due to the inclusion of *Bungarus* and ii) the exclusion of *Naja* (formerly *Boulengerina*) *annulata* from this clade; neither observation is supported by a recent mitochondrial analysis (Wüster *et al.* 2007). In contrast to results here, Slowinski *et al.* (1997) described a predominately resolved clade for the African and Asian Elapids; this incongruence suggests that the topology provided by Slowinski *et al.* (1997) is largely unsupported. Only two *Naja* clades previously described (Slowinski *et al.* 1997) exhibit significant node support values in the NXS consensus trees; i) the monophyly of *N. oxiana* and *N. philippinensis* and ii) the monophyly of *N. mossambica*, *N. kaouthia* and *N. atra*. Despite strong support for these two clades in this analysis, the latter is refuted by a recent mitochondrial phylogeny (Wüster *et al.* 2007).

6.5.3 The basis of unsuccessful tree reconciliation

Here, gene tree parsimony analyses were largely unsuccessful at reconstructing species trees from multiple copy venom protein families. Despite the previous apparent success of venom protein gene tree parsimony (Slowinski *et al.* 1997), these results show that significant changes in inferred Elapidae tree topologies occur when incorporating gene tree uncertainty. Furthermore, a number of relationships recovered by Slowinski *et al.* (1997) are not significantly supported in the species trees, which suggests that their results were only weakly supported, and emphasises the importance of assessing node support in species trees obtained through gene tree parsimony. Despite partial species tree congruence between the analysis of Slowinski *et al.* (1997) and more recent molecular studies, little consensus can be derived from the inferred species trees, with different venom protein families predicting different evolutionary histories within the family Elapidae. It is highly plausible that the failure of gene tree parsimony in the elapid datasets is a result of unequal and/or highly incomplete sampling of paralogous genes (mean number of sequences per species=1.6 [NXS] and 2.0 [PLA_2]) preventing the correct species tree being extracted. Given that the duplications only optimality criterion attempts to account for incomplete sampling, the higher node support values associated with these analyses support this hypothesis. The observed conflict between species trees obtained from different protein families, and between them and those from single

locus genes, is the inevitable result of using sequence data collected non-systematically during the course of diverse toxinological studies.

However, gene tree parsimony was also unsuccessful at inferring the species relationship in the genus *Echis*; despite using substantially greater numbers of sequences, base pairs, venom protein families and fewer species, only two reconciled species trees correctly inferred the topology determined from a strongly supported mitochondrial and single locus nuclear gene phylogeny (Barlow *et al.* 2009; Pook *et al.* 2009). In addition, the unbiased EST sampling method incorporated for the *Echis* dataset more likely reflects the true multiple copy nature of these venom protein families and has been demonstrated to be representative of proteomic venom expression (Wagstaff *et al.* 2009). Notably, nodes throughout the *Echis* species trees are typically weakly supported, whether they are consistent or inconsistent with the species phylogeny determined by Barlow *et al.* (2009) and Pook *et al.* (2009). Conflict arising from species tree reconciliation is likely to be a result of this weak signal; the majority of nodes responsible for causing species tree incongruence with the species phylogeny are unsupported (<95%). The consistent exception to this observation occurs in the serine protease venom protein family, where the duplications and loss inferred species trees produced a strongly supported (>95%) topology incongruent with the species phylogeny (Barlow *et al.* 2009; Pook *et al.* 2009), both for amino acid and DNA-based gene trees.

Recombination was excluded as a factor confounding species tree reconciliation following the analysis of the four *Echis* DNA datasets in the Recombination Detection Program v.3.34 (RDP3) (Heath *et al.* 2006). The results of a standard RDP3 analysis revealed only false positive results in the CTL, PLA_2 and SP datasets (data not shown) which exhibited significance scores similar to those obtained from a vertebrate mitochondrial cytochrome b dataset [GenBank: AB185152, AB253437, AP003423-AP003425, AP003428, AY487676, AY137598, EU035750, EU165259, EU380953, EU798758, EU856453, EU934483, FJ457612, FJ997847, GQ142135] devoid of recombination, and much lower than a snake venom protein dataset that has previously been demonstrated to contain recombinants [GenBank: AY861138,

AY861382, AY861383] (Zha *et al.* 2006). Although the SVMP dataset exhibited four sequences (out of 209) containing evidence of apparent recombination [GenBank: GU012190, GU012203, GU012213, GU012261], all but one of these recombinants [GenBank: GU012203] are nested within monophyletic species-specific clades, and would therefore not have influenced the reconstruction of the species tree. Furthermore, as all of the recombinant sequences are from *E. coloratus* and *E. p. leakeyi*, yet the relationship between these two species is correctly inferred in five of the six SVMP gene tree parsimony analyses, I exclude recombination as a factor responsible for confounding gene tree parsimony. Venom protein families may also be subjected to additional evolutionary phenomenon such as accelerated segment switches in exons to alter targeting (ASSET), where exons are radically changed to unrelated sequences leading to rapid functional evolution (Doley *et al.* 2008b, 2009). Recent analyses demonstrated that ASSET may play a significant role in the evolution of certain venom protein families, including SVMPs, PLA_2s and SPs (Doley *et al.* 2009). In order to exclude the potential role of ASSET confounding gene tree parsimony, I repeated the analyses for the venom protein families described above but excluding the regions of DNA and corresponding AA sequence demonstrated to be under the influence of ASSET (Doley *et al.* 2009). All of these analyses produced inferred species tree topologies consistent with the original analyses (data not shown).

The composition of snake venom proteins is under strong natural selection for adaptation towards specific diets (e.g. Daltry *et al.* 1996a; Kordiš and Gubenšek, 2000; Jorge da Silva and Aird, 2001; Barlow *et al.* 2009). Consequently, the effect of selection on patterns of gene duplication and loss cannot be excluded as a factor confounding gene tree parsimony by influencing gene events within lineages with divergent diets. Since members of the genus *Echis* exhibit considerable variation in prey preference (Barlow *et al.* 2009), adaptive selection pressures may be responsible for generating the strongly supported serine protease species trees that are incongruent with the *Echis* mitochondrial and nuclear phylogeny (Barlow *et al.* 2009). The presence of repeated selective loss in one lineage (Figure 6.8A), or multiple parallel loss in multiple lineages (Figure 6.8B) can confound gene tree parsimony; in both cases the most parsimonious explanation for the species

relationship can require fewer gene events than that of the true species tree (Figure 6.8). In the case of the serine proteases, the gene trees (Appendix IV Figure 1) exhibit minimal representation of clades containing *E. ocellatus* and *E. p. leakeyi* SPs, suggesting that multiple parallel gene losses may have occurred in these two species. Consequently, any gene tree parsimony analyses seeking to minimise the required number of assumptions of gene loss would result in a species tree grouping these taxa together (e.g. Figure 6.8B). This hypothesis was tested by analysing the serine protease gene data and the *Echis* phylogeny (Barlow *et al.* 2009; Pook *et al.* 2009) in GeneTree by implementing the reconciliation option. Reconciling the gene tree with the correct species tree elucidated the evolutionary history of gene duplication and loss events in the serine protease gene family and revealed multiple parallel gene loss events occurring in each lineage with the exclusion of *E. coloratus* (Figure 6.9). It therefore appears that gene tree parsimony is failing to produce a species tree topology congruent with Barlow *et al.* (2009) and Pook *et al.* (2009) as a result of multiple parallel losses; the incongruent monophyly of *E. ocellatus* and *E. p. leakeyi* occurs as parsimony minimises the number of gene events required to reconcile the gene tree to a species tree (see Figure 6.8B). These results explain the gene processes that are responsible for the presence of strongly supported incongruent nodes in the *Echis* serine protease reconciled trees and highlight the method by which gene tree parsimony can be undermined by non-random gene events in rapidly evolving multi-gene families.

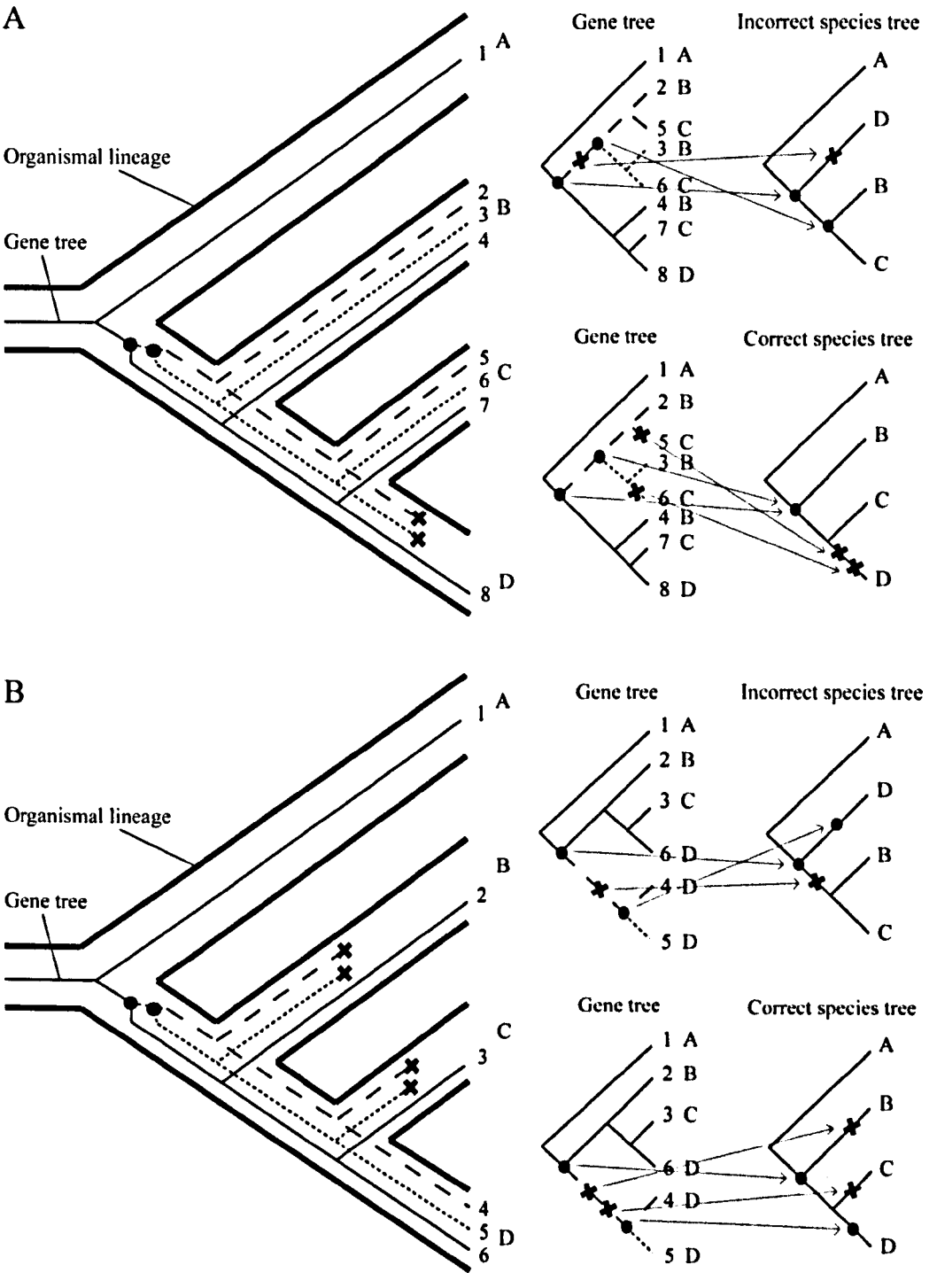

Figure 6.8. Selective and parallel loss events preventing correct species tree reconciliation. Numbers refer to alleles and letters A-D refer to species. Circles indicate duplication events and crosses indicate loss events. A: Repeated selective loss in species D leads to gene tree parsimony inferring the incorrect species tree if duplications and losses are taken into account. B: Multiple parallel gene loss in species B and C leads to gene tree parsimony inferring the incorrect species tree if

duplications and losses are taken into account. In both cases the number of events required to derive the correct species tree is four (two duplications and two losses), whilst the most parsimonious explanation infers an incorrect species tree with only three gene events (two duplications and one loss). Note also that, in both cases, gene tree parsimony will underestimate the number of gene losses.

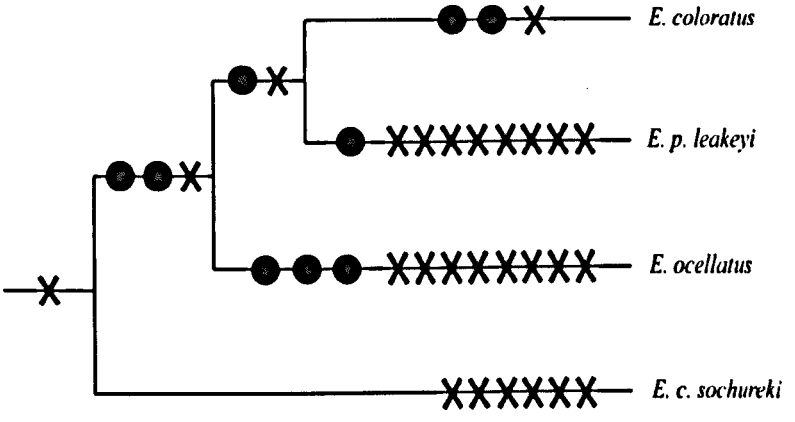

Figure 6.9. Serine protease gene tree reconciled with the species phylogeny of Barlow *et al.* (2009) and Pook *et al.* (2009) displaying lineage specific gene duplication (circles) and loss (crosses) events.

6.6 Conclusions

These results demonstrate the importance of rigorously assessing node support values for inferred species trees generated by gene tree parsimony. The implementation of Bayesian posterior distributions for multiple venom protein families allowed inferred species trees to be interpreted with confidence and highlighted a lack of support for a number of previously reconstructed evolutionary relationships in two different datasets. In this case gene tree parsimony largely failed to correctly infer strongly supported species trees from a comprehensive dataset of four multi-gene venom protein families isolated from four closely related members of the genus *Echis*, and from a smaller dataset of two venom protein families from members of the family Elapidae. It is notable, yet not unexpected, that when incorporating gene tree uncertainty for estimates of species tree inference, the

estimates of species relationships often reflect more uncertainty. I suggest that gene tree parsimony is unable to consistently resolve the elapid species relationship as a result of unequal and/or highly incomplete sampling of paralogous genes, whereas weak signal, evident by low node support values, undermines species tree reconciliation in the *Echis* datasets. I also hypothesise that the strongly supported conflict in the serine protease gene family is a result of non-random patterns of parallel gene loss and I have described how such gene process may confound gene tree parsimony. Given that the relationship between venom protein gene trees and inferred species trees has been demonstrated to be complex, I suggest that utmost caution should be employed when interpreting gene tree data generated from rapidly evolving multi-gene families likely to be suffering non-random selection pressures.

6.7 Authorship order and contributions

Nicholas R Casewell, Simon C Wagstaff, Robert A Harrison and Wolfgang Wüster. I undertook the bioinformatic processing and gene tree and tree reconciliation analyses. WW provided assistance and expertise for gene tree parsimony analyses. SCW and RAH contributed to the original sequence data production (see Chapter 4). I wrote the publication manuscript that forms the basis of this chapter.

CHAPTER 7

Intra-generic immunological and antivenomic comparisons of the saw-scaled vipers reveal paraspecific venom neutralisation of African *Echis* species by EchiTabG® antivenom

7.1 Abstract

The saw-scaled vipers (Viperidae: *Echis*) are thought to be responsible for a greater proportion of snakebite deaths worldwide than any other group of snakes. Considerable variations in venom components and toxicity have previously been identified in the genus *Echis*, alongside reports of incomplete intra-generic antivenom neutralisation. In order to investigate the confounding influence intra-generic venom variation may bestow upon antivenom cross-reactivity, immunological assessments of four monospecific antivenoms with homologous and non-homologous *Echis* venoms were compared, alongside their *in vivo* neutralisation with the *E. ocellatus* antivenom EchiTabG®. End-Point titration ELISAs, immunoblotting and small scale affinity purification revealed little difference in the cross-species immunoreactivity between homologous and non-homologous venom-antivenom mixes, although the anti-*E. ocellatus* antivenom exhibited the highest relative avidity. There was no significant difference in the lethality of the four *Echis* venoms as determined by venom LD_{50} assays. EchiTabG® neutralised the lethal effects of venom from the African *E. coloratus* and *E. pyramidum leakeyi* species with comparable efficacy as shown against the homologous *E. ocellatus* venom. However, EchiTabG® was ineffective at neutralising the lethal effects of venom from the Asian species, *E. carinatus sochureki*. Antivenomic and proteomic analysis of the complexes formed between EchiTabG® and the four venoms revealed snake venom metalloproteinases and cysteine-rich secretory proteins as venom components that failed to bind to EchiTabG®. Preclinical assessments of EchiTabG strongly suggest this antivenom will be an effective therapy in cases of envenoming by African members of the genus *Echis* and advocates the commencement of clinical trials aimed at expanding the geographic coverage of this antivenom to treat *Echis*-induced snakebite throughout the African continent.

7.2 Introduction

Envenoming by venomous snakes is estimated to cause as many as 94,000-125,000 deaths per year worldwide (Chippaux *et al.* 1998; Kasturiratne *et al.* 2008), with the saw-scaled vipers (Viperidae: *Echis*) thought to be responsible for a greater proportion of these deaths than any other single genus of snakes (Warrell *et al.* 1977). Members of the genus *Echis* have a wide distribution throughout much of Africa north of the equator, the Arabian Peninsula and India and Sri Lanka (Cherlin, 1990; Pook *et al.* 2009). Saw-scaled vipers represent the most medically significant group of snakes present throughout much of this range due to the possession of potently haemorrhagic venom (Warrell and Arnett, 1976; Warrell *et al.* 1977) combined with a high incidence of *Echis*-induced snakebite, particularly in West Africa (*E. ocellatus*) (Pugh and Theakston, 1980; Habib *et al.* 2001) and North-West India (*E. carinatus* ssp.) (Bhat, 1974; Bawaskar *et al.* 2008). Untreated mortality rates can be as high as 20% (Warrell *et al.* 1977). Envenoming by members of the genus *Echis* typically induces severe systemic symptoms such as spontaneous bleeding, disseminated intravascular coagulation and haemolysis, alongside local effects such as necrosis, swelling, blistering and oedema (Warrell *et al.* 1977; Porath *et al.* 1992; Benbassat and Shalev, 1993; Gillissen *et al.* 1994; Ali *et al.* 2004; Kochar *et al.* 2007).

The complex mix of proteins and peptides present in snake venoms is responsible for the pathology observed in cases of snakebite; they exhibit a high level of biological activity and a diverse array of actions on both natural prey items and humans (Chippaux, 1991; Aird, 2002). The venom composition of members of the genus *Echis* has been the subject of much recent research; including venom gland transcriptome surveys (Wagstaff and Harrison, 2006; Casewell *et al.* 2009 – Chapter 4) from members of the four *Echis* species groups (Pook *et al.* 2009), *E. ocellatus*, *E. coloratus*, *E. pyramidum leakeyi* and *E. carinatus sochureki*, whilst proteomic profiles of *E. ocellatus* venom components were correlated with the transcriptomic database (Wagstaff *et al.* 2009). Considerable inter- and intra-toxin family variation was observed within the major toxin families (enzymatic and non-enzymatic toxins) present in the *Echis* venom gland expressed sequence tag databases (vgDbEST)

(snake venom metalloproteinases (SVMP), C-type lectins (CTL), phospholipases A_2 (PLA_2), serine proteases (SP) and L-amino oxidases) (Casewell *et al.* 2009 – Chapter 4). Moreover, a number of less represented venom proteins were not ubiquitous throughout the genus, including short-coding disintegrins, cysteine-rich secretory proteins (CRISPs) and potentially novel venom proteins such as renin-like aspartic proteases and lysosomal acid lipase (Wagstaff and Harrison, 2006; Casewell *et al.* 2009 – Chapter 4). Venom variation observed in the genus *Echis* was hypothesised to be the result of shifts in diet following the correlation of dietary data with increases in venom toxicity to natural prey items (Barlow *et al.* 2009). Despite the lack of obvious association between diet and venom gland transcriptomic surveys (Casewell *et al.* 2009 – Chapter 4), further investigations analysing intra-toxin family variation suggest selective dietary pressures may be responsible for the diversification of specific toxin families (see Chapter 5).

Understanding the nature of venom variation is essential for therapy. The production of effective antivenom is fundamentally dependent upon the knowledge of the variability of venoms within and between specific localities and species (e.g. Theakston *et al.* 1989; Galán *et al.* 2004). A number of monospecific and polyspecific antivenoms produced against the venom of different *Echis* species have been effective at reducing mortality rates to 2-8% (e.g. Bhat, 1974; Warrell *et al.* 1977). Nevertheless there are increasing reports that antivenom availability and cross-reactivity are a problem within this genus (Warrell and Arnett, 1976; Visser et al 2008; Warrell, 2008), as demonstrated by the ineffectiveness of *E. carinatus* antivenom to treat patients envenomed by *E. carinatus sochureki* and *E. ocellatus* (Kochar *et al.* 2007; Visser *et al.* 2008) and antivenom raised against West and East African species to treat bites from a north African member of the *E. pyramidum* complex (Gillissen *et al.* 1994). Recent assessments of the polyspecific antivenom EchiTab-Plus-ICP®, generated against the venom of *E. ocellatus*, *Bitis arietans* and *Naja nigricollis*, demonstrated effective cross-neutralisation of the lethal activity of homologous and non-homologous venoms, including *E. leucogaster*, *E. p. leakeyi* and members of the genus *Bitis* (Segura *et al.* 2010). In order to assess the immunoreactivity of antivenoms against specific venom components 'antivenomic' techniques, focusing on the proteomic analysis of non-immunoprecipitated venom

components, have recently been applied (Lomonte *et al.* 2008; Gutiérrez *et al.* 2008, 2009; Calvete *et al.* 2009). This technique revealed EchiTab-Plus-ICP® failed to completely immunodeplete a number of venom components, particularly disintegrins and PLA_2s, despite effectively neutralising the lethal activity of the venoms (Calvete *et al.* in press). The implication of specific venom components exhibiting poor immunogenicity, however important in pathogenesis, highlights the potential for antivenom supplementation in order to enhance the immune response against specific venom toxins (Calvete *et al.* in press).

In order to further investigate antivenom cross-reactivity within the genus *Echis* and to assess whether intra-generic transcriptomic venom variation impacts upon therapeutic outcomes, I compared: (i) the lethal activity of venoms from four geographically distinct species of *Echis*, (ii) their immunological cross-reactivity with four monospecific antivenoms raised against each of the venoms and (iii) their *in vivo* neutralisation by the monospecific *E. ocellatus* antivenom EchiTabG®. In order to elucidate a case of incomplete, non-homologous venom neutralisation and the potential for antivenom supplementation, modified 'antivenomic' techniques (e.g. Lomonte *et al.* 2008; Calvete *et al.* 2009; Gutiérrez *et al.* 2009) were utilised to identify venom components that were non-immunodepleted by EchiTabG®.

7.3 Methods

7.3.1 Venom extraction

Pooled venom was extracted from wild-caught specimens of *E. ocellatus* (Nigeria), *E. coloratus* (Egypt), *E. pyramidum leakeyi* (Kenya) and *E. carinatus sochureki* (United Arab Emirates) used to create the previously described venom gland transcriptomes (Wagstaff and Harrison, 2006; Casewell *et al.* 2009 – Chapter 4). Following manual extraction, venom was frozen, lyophilised and stored at 4°C prior to reconstitution at 10mg/ml in 1X phosphate-buffered saline (PBS). Snakes were maintained in the Herpetarium at the Liverpool School of Tropical Medicine.

7.3.2 Immunisation and antiserum production

Antisera were generated against venom from *E. p. leakeyi*, *E. coloratus* and *E. c. sochureki* using protocols identical to the production of the *E. ocellatus* antivenom EchiTabG®. Six sheep (two per venom) were initially immunised with 0.5mg of venom emulsified with Freund's complete adjuvant followed by subsequent immunisations of 1.0mg of venom emulsified with Freund's incomplete adjuvant every 28 days. Venom doses were injected sub-cutaneously at six sites in the neck and groin. Sheep were bled every 14 days after immunisation and final sera was taken once the optimal time of the immune response was reached at 16 weeks (Landon, J., personal communication). Blood was centrifuged for 40 minutes at 4543 x g prior to the removal of sera and frozen at -20°C. Ovine IgG was extracted by the addition of caprylic acid (Sigma, UK) to a final concentration of 5%, stirred vigorously for two hours to precipitate non-IgG proteins, spun at 4543 x g for 60 min and dialysed overnight with sodium phosphate buffer pH 7.4. Purified IgG was diluted to 30mg/ml in 1X PBS and stored at -20°C. IgG generated against *E. ocellatus* venom and the *E. ocellatus* antivenom EchiTabG® were obtained from MicroPharm Ltd (UK).

7.3.3 End point and relative avidity ELISAs

Assays were prepared using 100ng of venom from the four *Echis* species per well. Ninety-six (96) well plates were blocked with 5% nonfat milk (diluted with TBST – 0.01M Tris-HCl, pH 8.5; 0.15M NaCl; 1% Tween 20) for 3 h at room temperature (RT), washed six times in TBST and incubated in each of the four species-specific IgG antivenoms (1:100 followed by 1:5 serial dilutions for end point and 1:10000 for relative avidity) overnight at 4°C. Plates were washed again in TBST and incubated in horseradish peroxidise-conjugated goat anti-sheep IgG (1:1000; Sigma, UK) for 3 h at RT. Relative avidity plates were incubated with 0.1ml of varying concentrations (1M-8M) of ammonium thiocyanate for 15 min, followed by washing in TBST prior to the addition of the secondary antibody. Results were visualized by addition of substrate (0.2% 2,2/-azino-bis (2-ethylbenzthiazoline-6-sulphonic acid) in citrate buffer, pH 4.0 containing 0.015% hydrogen peroxide; Sigma, UK) and measurement of optical density (OD) at 405nm. End point titres were determined by the IgG

antivenom titre that exhibited OD readings greater than two standard deviations of the control, whilst relative avidity was expressed as the percentage reduction in OD from the control to the highest concentration (8M) of ammonium thiocyanate.

7.3.4 Small scale affinity purification

In order to assess the cross-reactivity of the four IgG antivenoms raised against the four *Echis* venoms, small scale affinity columns were prepared for each of the venoms. 1g of CNBr-activated 4 Fast Flow Sepharose (GE Healthcare, UK) was swollen and washed with 1mM HCl, transferred to a 3.5ml column (Bio-Rad, UK) and washed twice with 0.1M sodium hydrogen carbonate pH 8.3. 5mg of venom (1mg/ml 0.1M sodium hydrogen carbonate pH 8.3 solution) was coupled with the Sepharose by end-over-end mixing at 4°C overnight. Columns were drained and active groups blocked by end-over-end mixing for 2 hours with 1M Ethanolamine-Cl pH 9.0, washed (0.1M sodium phosphate pH 7.5 containing 0.5M NaCl) and eluted (0.1M glycine pH 2.5 containing 0.1M HCl) before storage at 4°C. Columns were equilibrated at RT, washed with washing buffer, before 3mg of monospecific IgG (1mg/ml in washing buffer) was added to the column and mixed overnight. Columns were subsequently washed and eluted. The eluate was concentrated using 5kDa cut-off Vivaspin columns (Sartorius Stedim Biotech, UK) and quantified using a LD1000 series NanoDrop spectrophotometer (Thermo Scientific, USA).

7.3.5 Venom lethality and neutralisation by EchiTabG®

Determinations of the intravenous (i.v.) median lethal dose (LD_{50}) for each of the four *Echis* venoms were carried out as described by Laing *et al.* (1992) except for a reduction in observation time to 7 h. Briefly, groups of five male CD-1 mice (18-20g - Charles River) received an i.v. tail injection of varying doses of venom in 100µl 1X PBS; LD_{50}s were estimated at 7 h after injection by recording the number of deaths in each group of mice. The LD_{50} and 95% confidence limits were calculated using probit analysis (Finney, 1971). Tests for estimating the neutralising effects of the *E. ocellatus* antivenom EchiTabG® against the lethal effects (5×i.v. LD_{50}) of the four venoms were carried out using protocols previously described (e.g.

Laing *et al.* 1992; Laing *et al.* 1995; Theakston *et al.* 1995), again with a reduction in observation time to 7 h; groups of mice received i.v. injections of various doses of EchiTabG® antivenom mixed with 5xLD_{50}s of venom in 200µl 1X PBS preincubated at 37°C for 30 minutes. Deaths at 7 h were counted and the median effective dose (ED_{50}) and 95% confidence limits were estimated using probit analysis (Finney, 1971). The reduced observation time prevented unnecessary mouse-venom exposure; previous assays revealed that >98% *Echis* envenoming deaths occurred within 7 hours of the injection of the venom/antivenom mixture (Cook DAN and Harrison RA, personal communication).

7.3.6 EchiTabG® affinity purification 'antivenomics'

In order to assess whether the *E. ocellatus* antivenom EchiTabG® fails to bind venom proteins from members of the genus *Echis*, 10mg of EchiTabG® was coupled to a 1ml HiTrap NHS-activated HP affinity column using the manufacturer's protocol (GE Healthcare, UK). Varying concentrations of reconstituted venom in 1ml PBS solution were bound to the column. Unbound material was washed from the column, using an ÄKTAprime plus (GE Healthcare, UK), with 0.1M sodium phosphate pH 7.5 containing 0.5M NaCl at a flow rate of 0.1ml/min, prior to elution with 0.1M glycine pH 2.5 containing 0.1M HCl at 1ml/min. 0.5ml fractions containing the unbound and bound material were collected.

7.3.7 Electrophoretic analysis and immunoblotting

Reconstituted venoms were diluted to 1mg/ml in reducing SDS-PAGE sample buffer and boiled for ten minutes. Samples were separated on 1mm 15% SDS-PAGE gels according to the manufacturer's recommendations (BioRad, UK) and stained overnight using Coomassie Blue R-250. Venom, bound fractions and unbound fractions collected from the EchiTabG® column run with 0.5mg venom were separated by SDS-PAGE as described above under reduced and unreduced conditions, alongside native PAGE separation in 1:1 native sample buffer (5mM Tris-Cl pH 6.8, 33% glycerol). Gels were electro-blotted to 0.45µm nitrocellulose membranes using the manufacturer's protocols (Bio-Rad, UK). Following transfer

and visualisation by Ponceau S, membranes were incubated overnight in blocking buffer (5% nonfat milk in PBS), followed by six washes of TBST over 90 minutes and incubation overnight with primary antibodies (EchiTabG® and the species-specific IgG raised against individual venoms from *E. p. leakeyi*, *E. coloratus* and *E. c. sochureki*) at 1:5000 dilution in blocking buffer. Blots were washed as above with TBST and incubated for 2 hours with donkey anti-sheep secondary antibody (1:2000 dilution) coupled to horseradish peroxidise, prior to a final wash with TBST and visualisation after the addition of DAB peroxidase substrate (Sigma, UK).

7.3.8 LC-MS and protein identification by MS/MS

LC-MS and MS/MS protein identification was undertaken using previously described protocols (Currier *et al.* 2010). Briefly, proteins observed in the SDS-PAGE profiles that failed to bind to the EchiTabG® column were excised, de-stained and in-gel trypsin-digested (Hayter *et al.* 2003) before rehydration and sonication. Samples were fractionated in the first dimension over a gradient (600-900mM NaCl in 0.1% formic acid, pH 2.3) at a flow rate of 60µl/min before second dimension fractionation over a gradient (2-90% acetonitrile in 0.1% formic acid over 50 minutes) at a flow rate of 300nl/min. Eluted peptides were analysed on a LCQ Deca XP Plus Mass Spectrometer (ThermoFisher, UK) operating on a 'triple play' mode (zoom scan followed by MS/MS) before identification against Uniprot databases and the translated *Echis* vgDbESTs (Wagstaff and Harrison, 2006; Casewell *et al.* 2009 – Chapter 4) using Proteome Discoverer 1.0.0 software (ThermoScientific) incorporating both Sequest and Mascot search algorithms. Tolerances and search stringencies were as previously described (Currier *et al.* 2010).

7.4 Results

7.4.1 Immuno-comparisons of species-specific IgG antivenoms

The reduced SDS-PAGE profiles of venom extracted from the four *Echis* species demonstrated considerable protein variation (Figure 7.1A). However, substantial cross-reactivity between homologous and non-homologous venom-antivenom mixes

was observed in reduced immunoblots (Figure 7.1B-E); slight increases in reactivity were observed between homologous venoms and antivenoms. Comparisons of the four antivenom end point titres revealed little variation; each antivenom exhibited titres against the four venoms that varied by a maximum of one dilution factor, whilst comparisons between the four antivenoms demonstrate they are comparable (Table 7.1 and Appendix V Figure 1). Small scale affinity purification revealed the percentage of IgG that binds to venom coupled affinity columns; in all cases the highest binding occurred between an antivenom and its homologous venom (Table 7.2). Interestingly, *E. p. leakeyi* venom-derived IgG bound *E. coloratus* venom at comparable levels to its homologous venom, whilst the *E. ocellatus* and *E. coloratus* antivenoms displayed little variation in the percentage of IgG that bound to the non-homologous venoms. Relative avidity assays demonstrated homologous venom-antivenom mixes exhibited the highest avidity (Figure 7.2), consistent with results obtained from immunoblotting and affinity purification. However, *E. c. sochureki* antivenom displayed a similar avidity against *E. ocellatus* venom to its homologous venom, whilst avidities of *E. p. leakeyi* and *E. coloratus* venom with the *E. p. leakeyi* antivenom were not comparable. Comparisons between the antivenoms revealed the *E. ocellatus* antivenom EchiTabG® exhibited the highest relative avidities except against venom from *E. c. sochureki* (Figure 7.2).

7.4.2 Lethality of *Echis* venoms and neutralisation with EchiTabG®

Venom lethalities, expressed as LD_{50}s, ranged from 9.81 (μg venom per mouse) for *E. coloratus* to 15.10 for *E. c. sochureki*; 95% confidence limits indicate there is no significant difference between the venom lethality of the four *Echis* species (Table 7.3). The *E. ocellatus* antivenom EchiTabG® was effective at neutralising the venom lethality ($5xLD_{50}$) of the three African *Echis* species (*E. ocellatus*, *E. p. leakeyi* and *E. coloratus*), but was ineffective against the Asian species *E. c. sochureki* (Table 7.3). ED_{50}s ranged from 44.25 (μl antivenom per mouse) for *E. coloratus* venom to 64.87 for *E. p. leakeyi* venom, although 95% confidence limits indicate there is no significant difference between the effective ED_{50}s (Table 7.3). Interestingly, the EchiTabG® ED_{50} against the homologous venom, *E. ocellatus*, is higher than previously reported (Abubakar *et al.* 2010; Segura *et al.* 2010); similar values to

those reported here were generated from repeated experiments with different batches of EchiTabG® antivenom in order to confirm this apparent anomaly (Cook DAN, personal communication). Effective neutralisation of *E. c. sochureki* venom was achieved with the homologous *E. c. sochureki* antivenom with an ED_{50} (54.42µl/mouse) comparable to those obtained with EchiTabG® (Table 7.3).

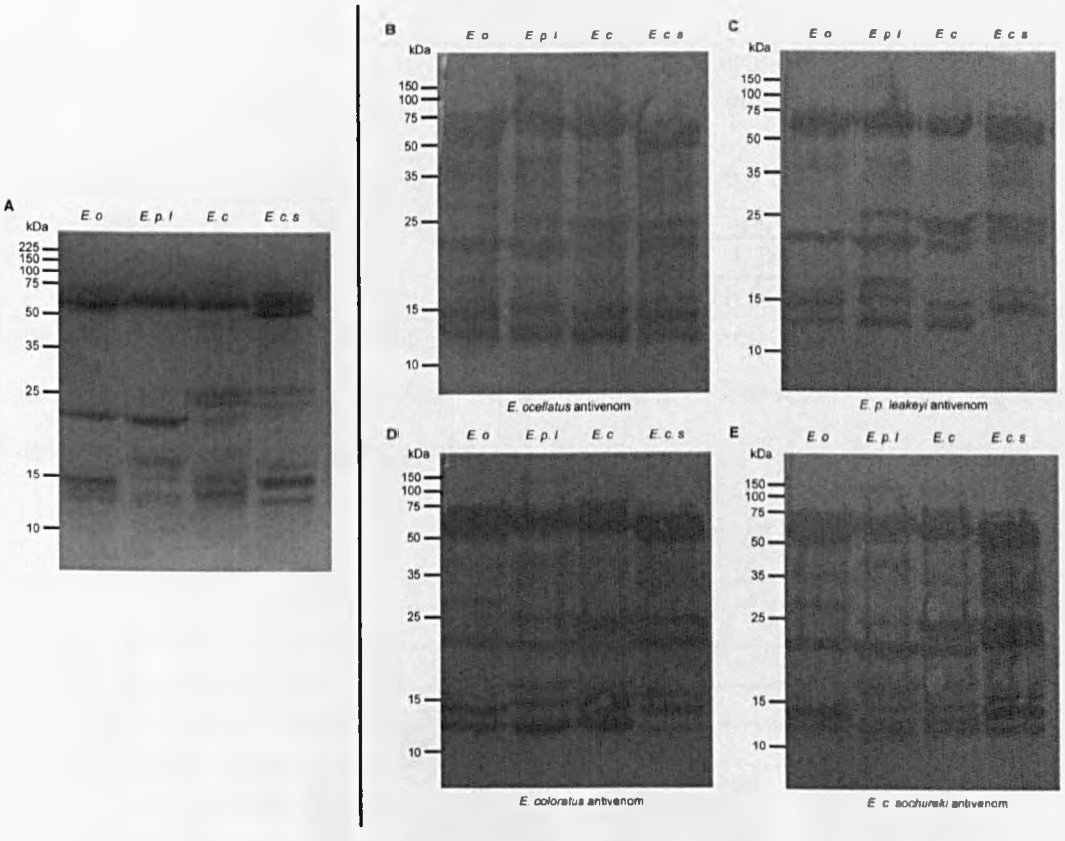

Figure 7.1. A) Reduced SDS-PAGE profiles of four venoms from the genus *Echis*. *E.o – E. ocellatus*, *E. p. l – E. p. leakeyi*, *E. c – E. coloratus*, *E. c. s – E. c. sochureki*. B-E) Reduced SDS-PAGE immunoblotting of the four *Echis* venoms with four species-specific IgG antivenoms. B) *E. ocellatus* antivenom, C) *E. p. leakeyi* antivenom, D) *E. coloratus* antivenom, E) *E. c. sochureki* antivenom.

Venom	Species-specific IgG antivenom			
	E. ocellatus	*E. p. leakeyi*	*E. coloratus*	*E. c. sochureki*
E. ocellatus	1.56×10^{-06}	3.12×10^{-05}	1.56×10^{-06}	1.56×10^{-06}
E. p. leakeyi	1.56×10^{-06}	3.12×10^{-05}	1.56×10^{-06}	1.56×10^{-06}
E. coloratus	7.81×10^{-06}	1.56×10^{-06}	1.56×10^{-06}	1.56×10^{-06}
E. c. sochureki	1.56×10^{-06}	1.56×10^{-06}	1.56×10^{-06}	1.56×10^{-06}

Table 7.1. The end point titres of four species-specific IgG antivenoms against four *Echis* venoms. Bordered values highlight homologous venom-antivenom results.

Venom	Species-specific IgG antivenom			
	E. ocellatus	*E. p. leakeyi*	*E. coloratus*	*E. c. sochureki*
E. ocellatus	10.23	5.12	6.77	7.51
E. p. leakeyi	8.32	8.02	6.95	7.53
E. coloratus	8.44	7.71	9.28	9.38
E. c. sochureki	8.11	4.95	6.90	11.12

Table 7.2. The percentage of four species-specific IgG antivenoms bound by small scale affinity purification to four *Echis* venoms. Bordered values highlight homologous venom-antivenom results.

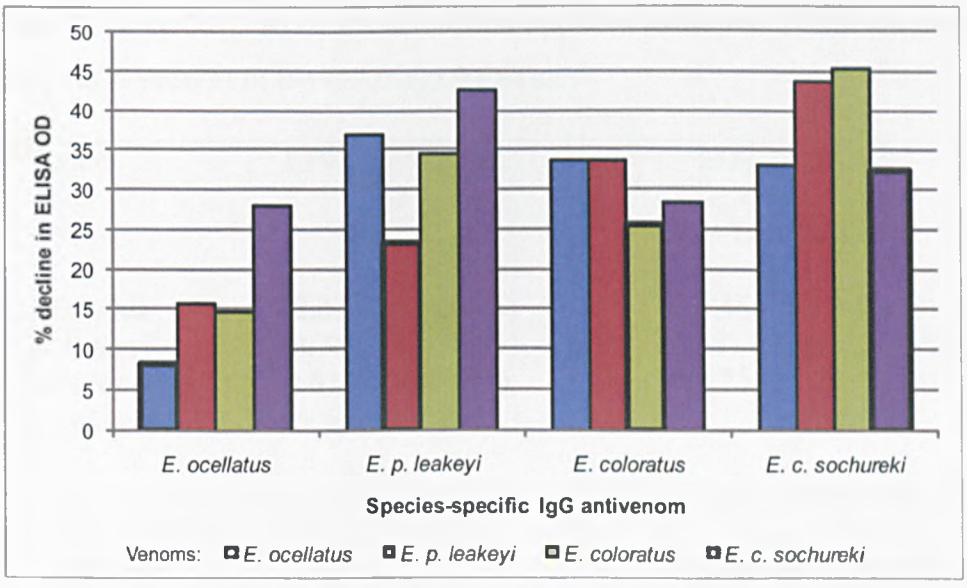

Figure 7.2. The relative avidity of four species-specific IgG antivenoms against four *Echis* venoms expressed as the percentage decline in ELISA optical density (405nm) from the control to incubation with 8M ammonium thiocyanate. Bordered values highlight homologous venom-antivenom results.

7.4.3 EchiTabG® 'antivenomics'

Affinity purified fractions of the four *Echis* venoms with EchiTabG® were visualised by SDS-PAGE (Appendix V Figure 2). The concentration of venom added to the column was decreased until proteins observed in the bound fractions were depleted from the unbound fractions to exclude the influence of antibody saturation. Protein bands remaining in the unbound fractions (Appendix V Figure 2) were not observed in the bound fractions at any venom concentration. Unbound fractions and crude venoms were subsequently subjected to reduced SDS-PAGE and native PAGE immunoblotting with EchiTabG® in order to confirm the absence of immunoreactivity. In reduced form the unbound proteins displayed high immunoreactivity with EchiTabG®; all ten protein bands were recognised by the antivenom antibodies (Figure 7.3A). Contrastingly, native PAGE immunoblotting demonstrated complete absence of immunoreactivity in the unbound fractions, yet high reactivity with the crude venom samples (Figure 7.3B). Peptide sequencing facilitated the identification of eight of the ten unbound protein bands (annotated in Figure 7.3A) via BLAST similarity to the translated *Echis* vgDbESTs (Wagstaff and Harrison, 2006; Casewell *et al.* 2009 – Chapter 4). The identifications revealed members of two venom protein families, SVMPs and CRISPs, failed to bind to EchiTabG® (Table 7.4); all of the identified peptides exhibited 100% identity with translated ESTs present in the *Echis* vgDbESTs.

Venom	LD_{50} (µg/mouse)	ED_{50} (µl/mouse)
		EchiTabG®
E. ocellatus	12.43 (9.00-20.45)	58.46 (35.32-90.92)
E. p. leakeyi	13.55 (8.98-38.33)	64.87 (23.86-129.65)
E. coloratus	9.81 (6.06-19.25)	44.25 (21.90-58.29)
E. c. sochureki	15.10 (6.49-19.70)	NE
		α *E. c. sochureki*
E. c. sochureki		54.42 (43.93-58.33)

Table 7.3. Median lethal doses of four *Echis* venoms, their corresponding median effective doses with the *E. ocellatus* antivenom EchiTabG® and the median effective dose of *E. c. sochureki* antivenom against *E. c. sochureki* venom. 95% confidence limits are displayed in parentheses. NE = Not effective.

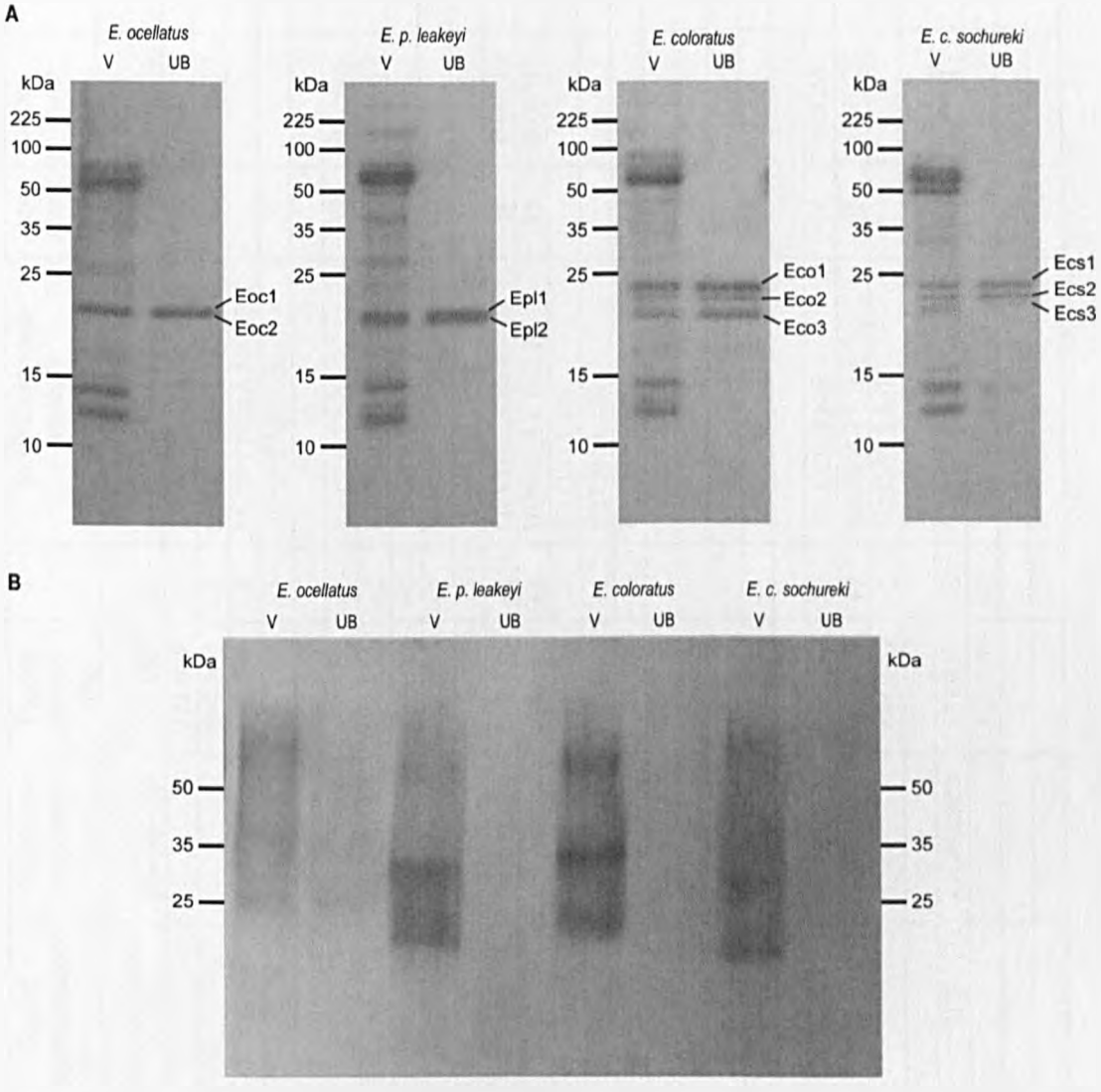

Figure 7.3. Immunoblotting of four *Echis* venoms (V) and their respective affinity purified unbound fractions (UB) with the *E. ocellatus* antivenom EchiTabG®. A) Reduced SDS-PAGE and B) native PAGE. Species number identifiers correspond to the unbound bands for each species that were excised from SDS-PAGE gels for protein identification.

Species	Band	Protein family	Cluster identified	Cluster representation	Accession number	Peptide ion m/z (Da)	z	MS/MS derived sequence	Mascot		Sequest	
									Ion score	Exp value	Proba-bility	XCorr
E. ocellatus	Eoc1	CRISP	EOC00029	0.29%	DW361159	595.055	+2	SVNPTASNMLR	37	0.00176		
						777.605	+2	MEWYPEAAANAER	34	0.00199		
						603.100	+2	SVNPTASNMLR	32	0.00419		
						777.605	+2	MEWYPEAAANAER			37.62	2.92
						603.100	+2	SVNPTASNMLR			37.62	2.26
	Eoc2	PII-SVMP	ECO00011	5.74%	GU012238	487.510	+2	NNGDLTAIR	52	0.00005		
			EPL00005	17.65%	GU012274	487.050	+2	NNGDLTAIR	46	0.00017		
						487.050	+2	NNGDLTAIR			43.39	2.68
						487.510	+2	NNGDLTAIR			31.58	2.41
E. p. leakeyi	Epl1	PII-SVMP	EPL00005	17.65%	GU012274	606.235	+2	QSVGIIENHSK	38	0.00110		
						620.650	+2	HDNTQLLTGLK	35	0.00247		
						515.990	+2	EYQSYLTK			19.20	2.01
						1032.325	+1	EYQSYLTK			6.49	1.47
						1031.295	+1	EYQSYLTK			21.84	1.42
	Epl2	No sig. hit	-	-	-	-	-	-	-	-	-	-
E. coloratus	Eco1	No sig. hit	-	-	-	-	-	-	-	-	-	-
	Eco2	PII-SVMP	ECO00020	5.74%	GU012246	494.190	+2	NKGDLTAIR	36	0.00244		
						494.625	+2	NKGDLTAIR	35	0.00275		
						494.190	+2	NKGDLTAIR			5.72	2.59
	Eco3	PI-SVMP	ECO00047	1.51%	GU012229	527.010	+2	YNSDLTAIR	47	0.00017		
						527.010	+2	YNSDLTAIR			32.94	2.28

									Mascot		Sequest	
Species	Band	Protein family	Cluster identified	Cluster representation	Accession number	Peptide ion m/z (Da)	z	MS/MS derived sequence	Ion score	Exp value	Proba-bility	XCorr
E. c. sochureki	Ecs1	CRISP	ECS00168	1.83%	GR950013	603.065	+2	SVNPTASNMLR	29	0.00915		
						603.065	+2	SVNPTASNMLR			34.70	2.39
						595.570	+2	SVNPTASNMLR			21.21	2.22
	Ecs2	PII-SVMP	ECS00253	1.47%	GU012265	754.025	+2	DLINVVSSSSDTLR	33	0.00350		
						754.025	+2	DLINVVSSSSDTLR			26.75	2.90
	Ecs3	PIII-SVMP	EOC00001	2.60%	AM039691	1535.880	+3	XNHDNTQLLTGMN FDGPTAGLGYVGT MCHPQFSAAVVQD HNK	21	0.00910		

Table 7.4. Identification of venom proteins from the venom of four *Echis* species which failed to bind to the *E. ocellatus* antivenom EchiTabG®. Cluster identifications arise by BLAST sequence similarity to translated expressed sequence tags derived from the four *Echis* venom gland transcriptomes (Wagstaff and Harrison, 2006; Casewell *et al.* 2009 – Chapter 4). In all cases 100% sequence similarity was observed. Cluster representation is expressed as the percentage of toxin encoding ESTs each cluster represents in the respective species venom gland transcriptome (Casewell *et al.* 2009 – Chapter 4).

7.5 Discussion

EchiTabG® antivenom is generated by immunising sheep with the venom of the West African saw-scaled viper *Echis ocellatus*. Pre-clinical and randomised controlled clinical studies have demonstrated this antivenom effectively neutralises the toxic activities of *E. ocellatus* venom with a low minimum effective dose (Abubakar *et al.* 2010), providing a cost effective therapy for *Echis*-induced snakebite in West Africa. The effective neutralisation of *E. p. leakeyi* and *E. coloratus* venom by EchiTabG®, at similar levels to the homologous venom of *E. ocellatus*, implies this antivenom is capable of neutralising the lethal components present in these species, despite the variation in toxin components observed from transcriptomic, proteomic and invertebrate lethality studies (Wagstaff and Harrison, 2006; Barlow *et al.* 2009; Casewell *et al.* 2009 – Chapter 4; Wagstaff *et al.* 2009). Whilst pre-clinical assays do not necessarily imply therapeutic neutralisation in cases of human envenoming, these results strongly advocate the geographic expansion of this venom for clinical testing in other regions of Africa. The provision of an antivenom capable of neutralising venom from multiple *Echis* species would provide a valuable therapeutic tool, particularly in areas where congeneric species overlap given the homogenous morphology of this genus (Cherlin, 1990). However, despite successful neutralisation of lethality in the African *Echis* species, EchiTabG® failed to completely neutralise the lethal effect of the Asian species *E. c. sochureki*. The neutralisation of *E. c. sochureki* venom with homologous antivenom implies that the failure of EchiTabG® is a result of variation in the toxic components present in the venom of these two species.

It is notable that immunological investigations comparing the four antivenoms were unable to predict the failure of EchiTabG® to neutralise *E. c. sochureki* venom. The immunoreactivity of EchiTabG® with non-homologous venoms was comparable, whilst reactivity against major protein bands present in *E. c. sochureki* venom SDS-PAGE profiles was observed (Figure 7.1). EchiTabG® exhibited comparable end point titres against all *Echis* venoms including identical titres with *E. ocellatus* and *E. c. sochureki* (Table 7.1), whilst the percentage of IgG bound by the non-homologous venoms exhibited little variation (Table 7.2). However, assessments of

relative avidity provided correlations with the EchiTabG® ED_{50}s results; the antivenom exhibited a ~55% drop in avidity when binding venom from *E. c. sochureki* compared to other members of the genus. The results of these various immunological investigations highlight the complex nature of quantifying venom-antibody interactions; assessments of immunoreactivity and antivenom binding may not be representative predictors of pre-clinical assays.

In order to further investigate the nature of venom-antibody binding and the failure of EchiTabG® antivenom to neutralise the venom of *E. c. sochureki*, a modified 'antivenomics' approach was implemented. Previous 'antivenomic' approaches involve the incubation of antivenom and venom prior to the immunoprecipitation of resulting complexes and subsequent identification by proteomic analysis (e.g. Lomonte *et al.* 2008; Gutiérrez *et al.* 2009; Calvete *et al.* in press). Here I adopted an alternative approach using column chromatography; antivenom is coupled to affinity columns, venom proteins are allowed to bind, unbound proteins are washed and bound proteins eluted. Using EchiTabG®, at least one venom component was identified from each *Echis* species that failed to bind to the antivenom; these components were confirmed as non-binding through the absence of immunoreactivity in native immunoblotting (Figure 7.3B). Surprisingly, the unbound components identified were all recognised by reduced immunoblotting with EchiTabG® (Figure 7.3A), suggesting denaturation of these proteins exposes epitopes recognised by antibodies.

The antivenomic results differ considerably from those of Calvete *et al.* (in press), who identified disintegrins and PLA_2s as incompletely immunoprecipitated by the polyspecific antivenom EchiTab-Plus-ICP® (generated against *E. ocellatus*, *B. arietans* and *N. nigricollis*). Notably, EchiTabG® appears to effectively bind the majority of venom proteins including, PLA_2s, CTLs, SPs, disintegrins, L-amino oxidases and a number of other minor venom components. However, I identified specific SVMPs and CRISPs that were not found to bind to EchiTabG®. SVMPs are a diverse group of enzymes classified into those comprising only the metalloproteinase domain (PI) and those sequentially extended by a disintegrin

domain (PII), a disintegrin-like and cysteine-rich domain (PIII) and the latter co-valently linked to C-type lectin-like components (PIV) (Fox and Serrano, 2005, 2008). Unbound PII-SVMP proteins were identified from each species, whilst additional SVMPs were identified in *E. coloratus* (PI-SVMP) and *E. c. sochureki* (PIII-SVMP). In all cases the peptide sequences exhibited 100% sequence similarity to the metalloproteinase domain of translated SVMP ESTs; this observation combined with the molecular weight of the identified bands (20-25kDa), implies these proteins are the processed metalloproteinase domains of SVMPs (effectively PI-SVMPs), devoid of disintegrin, disintegrin-like and cysteine-rich domain extensions thought to be largely responsible for their biological activity (Wagstaff *et al.* 2009). The SVMPs are the most abundant toxin family present in the *Echis* vgDbESTs and the *E. ocellatus* proteome (Wagstaff and Harrison, 2006; Casewell *et al.* 2009 – Chapter 4; Wagstaff *et al.* 2009) and are widely assumed to be predominately responsible for serious pathological manifestations occurring in human envenoming, including local and systemic haemorrhage (Gutiérrez *et al.* 2005; Fox and Serrano, 2005, 2008). Snake venom CRISPs have been demonstrated to interact with ion channels and exhibit the potential to block arterial smooth muscle contraction and nicotinic acetylcholine receptors (Yamazaki and Morita, 2004; Gorbacheva *et al.* 2008), however the functional significance of CRISPs in saw-scaled viper venom remains unclear. Unbound CRISP peptides identified in *E. ocellatus* and *E. c. sochureki* fractions exhibited 100% sequence similarity to those present in the *Echis* vgDbESTs. Cysteine-rich secretory proteins appear to be minor venom components in both *E. ocellatus* and *E. c. sochureki*; representing 0.29% and 1.83% of toxin ESTs in their respective vgDbESTs and 1.7% proteomically in *E. ocellatus* (Casewell *et al.* 2009 – Chapter 4; Wagstaff *et al.* 2009). Surprisingly, CRISPs were not identified from the unbound fractions of *E. coloratus*, despite considerable representation in the vgDbEST (5.28%) and high sequence similarity between these *Echis* proteins (~87%); it is conceivable that the unbound protein band Eco1, which failed to yield quality peptide sequences, represents this venom protein family.

The presence of similar toxin family isoforms identified in the unbound fractions of venoms that exhibit disparate neutralisation efficacies impedes determining the

proteins responsible for incomplete *E. c. sochureki* venom neutralisation by EchiTabG®. Whilst it is tempting to speculate that the unique presence of an unbound processed PIII-SVMP in *E. c. sochureki* venom may be responsible for conferring incomplete neutralisation, the presence of an SVMP gene analogue in the *E. ocellatus* vgDbEST, coupled with the absence of this peptide in unbound fractions of *E. ocellatus* venom, implies that these peptides are present in the immunising material. Nevertheless, the disparate representation of these PIII-SVMPs in the *E. ocellatus* (2.60%) and *E. c. sochureki* (9.54%) toxin encoding vgDbESTs (Casewell *et al.* 2009 – Chapter 4) implies that the expression of this SVMP isoform may be of greater functional importance in *E. c. sochureki.* Further investigations are required to determine if this venom component remains only partially neutralised by EchiTabG® due to insufficient antibody generation. Thorough investigations into specific antibody-toxin interactions are required alongside assessments of the sensitivity of antivenomic approaches in order to elucidate the significance of the results obtained here.

7.6 Conclusions

Antivenomic techniques have proven to be useful tools to assess antibody-toxin isoform interactions occurring between homologous and non-homologous venoms (Lomonte *et al.* 2008; Gutiérrez *et al.* 2008, 2009; Calvete *et al.* 2009). Whilst these results fail to explain the observed incomplete *E. c. sochureki* venom neutralisation by the *E. ocellatus* antivenom EchiTabG®, they provide identifications of specific venom components that are not recognised by the antivenom for future investigation. Moreover, the identification of two specific toxin types, processed SVMPs and CRISPs, that failed to bind to EchiTabG® highlights the potential for increasing the efficacy and cross-reactivity of antivenoms by supplementation with antibodies against specific antigens known to elicit poor immune responses or that are absent from the immunising venom. Nevertheless, preclinical assessments of EchiTabG® strongly suggest that this antivenom is effective at neutralising the venoms of multiple African *Echis* species and robustly advocates the commencement of clinical trials aimed at expanding the geographic coverage of EchiTabG® to treat *Echis*-induced snakebite throughout the African continent.

7.7 Author contributions

Nicholas R Casewell, Darren AN Cook, Rachel B Currier, Gavin D Laing, Wolfgang Wüster, Simon C Wagstaff and Robert A Harrison. I undertook the majority of experiments: including all ELISAs, affinity purification, electrophoresis and immunoblotting. I also undertook the experimental preparations for the *in vivo* assays and carried out the necessary observations and statistical analyses - RAH and DANC performed the animal experiments. I undertook excision and trypsin digestion of protein bands for protein identification - RBC and GDL performed LC-MS and MS/MS. SCW and RAH provided guidance and assistance for the immunological assessments. I wrote the publication manuscript that forms the basis of this chapter.

CHAPTER 8

DISCUSSION

8.1 Discussion

The construction of cDNA libraries coupled with the generation of expressed sequence tags have proven to be particularly powerful tools for generating an overview of the diversity and inferred expression levels of toxin family secretion in the venom gland, whilst also facilitating the discovery of novel toxin families (e.g., Junqueira-de-Azevedo and Ho, 2002; Fry *et al.* 2006, 2008; Wagstaff and Harrison, 2006). Furthermore, transcriptomic data was been demonstrated to be representative of the proteomic expression of venom components (Wagstaff *et al.* 2009). In the case of the genus *Echis*, the production of multiple transcriptomes generated from four closely related species provided a unique opportunity to compare and analyse the nature of inter-specific venom variation at the genomic level. The identification of SVMPs, CTLs, PLA_2s and SPs as the most heavily represented venom components in *Echis sp.* is unsurprising considering previous work on *E. ocellatus* (Wagstaff and Harrison, 2006) and other members of the Viperidae (Junqueira-de-Azevedo and Ho, 2002; Francischetti *et al.* 2004; Kashima *et al.* 2004; Cidade *et al.* 2006; Junqueira-de-Azevedo *et al.* 2006; Zhang *et al.* 2006; Pahari *et al.* 2007). However, following the optimisation of clustering algorithms, considerable intra-generic variation (in the form of cluster representation and diversity) was observed in a number of these toxin families, particularly in PII and PIII SVMPs, CTLs and SPs. Detailed analyses of the *Echis* transcriptomes also revealed a number of novel putative venom toxins: renin-like aspartic proteases (Wagstaff and Harrison, 2006), lysosomal acid lipase/cholesteryl ester hydrolase and the metallopeptidases dipeptidyl peptidase III and neprilysin. Despite a number of potential physiological roles for these putative toxins in envenoming, experimental evidence of their functional activity and presence in venom remains essential for toxin confirmation.

Whilst comparative transcriptomic data provides a useful tool to assess venom variation at the intra-generic level, the over-reaching aim of this study was to analyse the potential selective role of diet upon the evolution of venom components. Previous work by Barlow *et al.* (2009) revealed the apparent co-evolution of venom toxicity and diet in the genus *Echis*. The generation of molecular gene data for multiple venom components provided a unique model system to assess whether dietary selection pressures, i) generate the recruitment of novel toxin components or ii) confer variation in the diversity or representation of existing venom components, to generate increases in venom toxicity. Principal comparative analyses revealed little correlation between the representation of entire toxin families and dietary data, particularly when considering the contrasting toxin encoding profiles between the predominately invertebrate feeding species *E. p. leakeyi* and *E. c. sochureki*. Considering dietary shifts in the genus *Echis* were inferred to have occurred prior to the divergence of the genus (switch to invertebrate feeding) and in the *E. coloratus* lineage (reversion to vertebrate feeding) (Barlow *et al.* 2009), the absence of novel toxins present throughout the genus implies adaptations to invertebrate feeding are unlikely to be the consequence of novel toxin recruitment. However, I cannot exclude the possibility that the exclusive presence of lysosomal acid lipase in *E. coloratus* may represent a direct adaptation to the reversion to vertebrate feeding. These initial observations inferred adaptations to diet are likely occurring within venom toxin families; to test this hypothesis phylogenetic analyses of the most represented toxin families was undertaken prior to tree reconciliation analyses using gene tree parsimony. The reconciliation of complex multi-locus toxin family gene trees with known species trees previously generated from members of the genus *Echis* (Barlow *et al.* 2009; Pook *et al.* 2009) facilitated tracing the evolutionary history of toxin family gene events. Notably, reconciled gene and species trees revealed strong correlations between PIII/PIV SVMP and serine protease gene events and the reversion to vertebrate feeding in *E. coloratus*. These results provide the first evidence of the genomic basis of venom adaptations as a response to alterations in diet. Interestingly, these adaptations appear to be the result of multiple genetic mechanisms, with substantial increases in SVMP gene diversifications occurring in *E. coloratus*, whilst the loss of multiple serine protease genes has occurred independently in the predominately invertebrate feeding species when compared to the retention of SP genes in *E. coloratus*. These results correlated with

significant differences in *in vivo* haemorrhage and therefore strongly imply a functional importance for haemorrhagic and coagulopathic SVMPs and SPs in vertebrate prey capture. The loss of coagulopathic serine protease genes in the invertebrate feeding members of the genus *Echis* correlates with the substantial difference that exists in the coagulation systems present in invertebrates and vertebrates. Whilst venom from members of the genus *Echis* exhibit significant differences in haemorrhagicity, comparable venom LD_{50} values in mice were exhibited; no significant differences were observed between species despite *E. coloratus* exhibiting the highest toxicity. However, this may imply that: i) the venom components suffering dietary selection pressures in *E. coloratus* have not yet evolved sufficiently to confer a significant increase in venom toxicity, ii) the ancestral components that remain in the invertebrate feeding species are sufficient to confer a high toxicity to vertebrates or iii) that the limitations of the LD_{50} test (particularly the number of mice used and that white mice are not natural prey items for *Echis* species) are sufficient to prevent significance being detected.

Correlations between toxin family gene events and the evolution of invertebrate feeding remain undetected. The inclusion of a closely-related vertebrate-feeding outgroup species would greatly enhance any subsequent analysis. For example, the presence of equally representative data from a closely related species (e.g. *Cerastes cerastes* or *Bitis arietans*) would determine toxin clades within the major toxin families that are unique to the genus *Echis*; any such gene diversifications would therefore correlate with a dietary shift to invertebrate feeding. The identified toxin clades would subsequently provide ideal targets to functionally investigate the venom components responsible for increases in toxicity to invertebrates. Alternatively, the toxins responsible for these differences may, i) not be well represented in the venom gland transcriptome and therefore excluded from the previous analyses, ii) be a combination of specific toxin isoforms from different toxin families or iii) be members of different toxin families as a result of the independent evolution of invertebrate-feeding in each of the three lineages. Despite previous successful correlations between the *E. ocellatus* transcriptome and proteome (Wagstaff *et al.* 2009), it is conceivable that transcriptomic representation of components in the venom gland does not accurately represent true venom protein

expression. Ideally, the combination of both techniques is desirable, with use of the transcriptomic databases to identify the toxin isoforms partially determined in the proteome. Such studies would ensure toxins well represented proteomically were not excluded from the phylogenetic analyses; however these additional analyses were outside both the scope and technical expertise of this study. The less represented toxin families remain targets for conferring increases in toxicity, although their low transcriptomic representation (and proteomic in *E. ocellatus* (Wagstaff *et al.* 2009)) and predominately unknown functionalities imply they likely play a minor role in envenoming; subsequent proteomic analyses alongside functional characterisation of any identified components may be revealing. To test the hypothesis that increases in toxicity to invertebrates has evolved as the result of independent mechanisms in each invertebrate feeding lineage, the inclusion of transcriptomic data generated from multiple representatives of each genus *Echis* species group (see Pook *et al.* 2009) would be required. Subsequently, gene tree parsimony would more accurately trace toxin family gene histories following the divergence of species and their alterations in diet. An alternative functional approach to determine the mechanism by which invertebrate-specific adaptations are conferred would be the use of size exclusion techniques, such as gel filtration and/or anion exchange chromatography, to fractionate whole venom into its constituents. Subsequently, generated fractions could be used in invertebrate LD_{50} experiments (as per Barlow *et al.* 2009) to determine fractions conveying lethal activity, prior to their protein identification by LC-MS, MS/MS and BLAST similarity to the transcriptomic databases. Unfortunately such a method would be particularly costly as a result of the large quantity of venom and live animals required, particularly if multiple venom components are working synergistically to confer increases in toxicity.

The identification of selective pressures responsible for driving the molecular evolution of venom components partially explains the intra-generic variation in venom components observed in the genus *Echis* (Táborská, 1971; Casewell *et al.* 2009 – Chapter 4). Whilst dietary selection pressures are likely responsible for conferring substantial variation in venom components in a number of additional snake genera, other factors, such as geographical variation and phylogenetic position

are also likely contributing factors (reviewed in Chippaux *et al.* 1991). Irrespective of the mechanism driving venom variation, a number of medically important snake genera have been observed to exhibit considerable variation in venom components, the symptomatologies these components confer and the subsequent efficacy of antivenom therapy (e.g. Tan *et al.* 1989; Theakston *et al,* 1989; Chippaux *et al.* 1991; Prasad *et al.* 1999; Shashidharamurthy *et al.* 2002; Galán *et al,* 2004; Gowda *et al.* 2006a). The generation of four monospecific *Echis* antivenoms provided a model system to test the immunological cross-reactivity of homologous and non-homologous intra-generic antivenoms. Surprisingly, little variation in immunological cross-reactivity, end-point titre and the percentage of bound IgG was observed between homologous and non-homologous venom-antivenom mixes, predicting high levels of intra-generic cross-reactivity. The neutralisation of four *Echis* venoms with the monospecific *E. ocellatus* antivenom EchiTabG® revealed cross-neutralisation of three African *Echis* species but failure to completely neutralise *E. c. sochureki* venom. It is therefore notable that the prior immunological assessments of the monospecific antivenoms predominately failed to predict the neutralisation failure of EchiTabG® against *E. c. sochureki*; assessments of immunoreactivity and antivenom binding may not be representative predictors of pre-clinical antivenom neutralisation assays.

EchiTabG® has previously been demonstrated to effectively neutralise the toxic activities of *E. ocellatus* venom in pre-clinical and randomised controlled clinical studies (Abubakar *et al.* 2010). The effective neutralisation of venom from other African *Echis* species, to similar levels as the homologous venom (*E. ocellatus*), strongly advocate the geographical expansion of this antivenom to treat *Echis*-induced snakebite throughout the African continent. Whilst *E. ocellatus* is responsible for significant snakebite mortality in West Africa (Pugh and Theakston, 1980; Habib *et al,* 2001), other African *Echis* species are responsible for a substantial proportion of snakebite incidences and mortalities throughout the African continent north of the equator (see Warrell, 1995). Furthermore, the expansion of an existing antivenom, currently in use in West Africa, to cover the entire continent for cases of *Echis*-induced snakebite is an attractive proposition, particularly in areas where congeneric cryptic species overlap. Furthermore, EchiTabG® has been

demonstrated to be effective at a low minimum dose, thereby reducing the cost of therapy (Abubakar *et al.* 2010), whilst production and distribution issues are likely reduced due to the current existence of the product on the African continent. These factors strongly advocate the commencement of randomised controlled clinical studies in other African countries where *Echis* snakebite is a serious health issue.

The results of venom neutralisation studies demonstrate that even when substantial variation in venom components is observed at the transcriptomic level, immunological cross-reactivity of epitopes can be sufficient to generate complete venom neutralisation, with efficacies comparable to that of the immunising material. Nevertheless, transcriptomic variation existing in the genus *Echis* derived sufficient proteomic variation to prevent the neutralisation of *E. c. sochureki* venom by EchiTabG®. Attempts to identify the venom components responsible for conveying this incomplete venom neutralisation, using 'antivenomic' techniques (see Lomonte *et al.* 2008; Gutiérrez *et al.* 2008; 2009; Calvete *et al.* 2009), identified members of the SVMPs and CRISPs as venom proteins that failed to bind to EchiTabG®. Nevertheless, the identification of these unbound venom components does not completely explain incomplete venom neutralisation, particularly considering members of these protein families were identified as unbound in the venom of other members of the genus. Furthermore, previous antivenomic approaches, using immunoprecipitation and *E. ocellatus* and *E. p. leakeyi* venoms, identified PLA_2s and disintegrins as the toxin families incompletely neutralised by the polyspecific (*E. ocellatus*, *B. arietans* and *N. nigricollis*) antivenom EchiTab-Plus-ICP® (Calvete *et al.* in press). Whilst the difference between antivenoms may be responsible for the distinct difference in antivenomic results generated by these two studies, it would be imprudent to ignore the difference between the antivenomic techniques themselves. Future comparative assessments of both techniques would be greatly beneficial to elucidate the complex nature of toxin-antibody binding and its role in complete or partial non-homologous venom neutralisation. In particular, repetition of the techniques described here using the venoms and antivenom tested by Calvete *et al.* (in press) and the converse, using immunoprecipitation techniques for the *Echis* venoms and antivenoms, would likely provide valuable insights into the strength and

reliability of these techniques. As the field of antivenomics is still in infancy, such methodological assessments are integral for the future interpretations of results.

The combination of transcriptomic data, full-length toxin sequences and assessments of venom-antivenom interactions have provided a substantial increase in our knowledge of the evolution, composition and antivenom cross-reactivity of venom in the genus *Echis*. However, the generation of substantial numbers of full-length toxin encoding DNA sequences also provided a model system to test whether the selective processes that influence the evolution of rapidly-evolving multi-gene toxin families can also prevent the correct derivation of species trees from gene trees. The incorporation of rigorous assessments of gene tree uncertainty, through species tree searches of entire Bayesian posterior distributions, provided node support values in reconciled trees that could be interpreted with confidence (Buckley *et al.* 2006; Oliver, 2008). Subsequent assessments of *Echis* species trees derived from full length transcriptomic data from four toxin families failed to produce a consistent topology; only two of the twelve species trees produced a topology congruent with the *Echis* phylogeny derived from mitochondrial and nuclear loci (Barlow *et al.* 2009; Pook *et al.* 2009). Furthermore, reassessments of a previously tested Elapidae dataset (Slowinski *et al.* 1997) with the incorporation of node support values revealed that the species tree topologies previously determined were largely unsupported. The limitations of gene tree parsimony to resolve the Elapidae dataset are unsurprising, particularly when considering the likely use of paralogous genes as a result of unequal and/or incomplete sampling. However, the *Echis* sequences represent a large unbiased representative dataset, yet the derived species trees lacked a consistent topology and were predominately unsupported. It is notable that by incorporating gene tree uncertainty the estimates of species relationships reflect more uncertainty; I hypothesise that generation of this weak signal is predominately responsible for undermining gene tree parsimony in the majority of *Echis* datasets. However, the serine protease analyses uniquely produced species trees with strongly supported nodes incongruent to the species phylogeny (Barlow *et al.* 2009; Pook *et al.* 2009). Whilst the role of recombination and accelerated segment switches in exons (Doley *et al.* 2008b, 2009) were excluded as confounding influences, the selective role of diet appears to be responsible for producing this incongruence;

multiple parallel gene losses occurring in *E. ocellatus* and *E. p. leakeyi* cause parsimony to group these species together to the exclusion of *E. coloratus*. The previous demonstration that dietary selection pressures are driving the loss of serine protease genes in these invertebrate feeding species highlights the confounding influence non-random gene events can have upon gene tree parsimony. For these reasons utmost caution should be employed when interpreting complex gene tree data generated from gene families that suffer non-random genetic pressures.

The identification of selective pressures that can influence the evolution of venom components and subsequently confound the derivation of species relationships from toxin data, raises questions about the use of venom profiles as species identifiers (e.g. Calvete *et al.* 2007; Angulo *et al.* 2008). Whilst immunological or proteomic profiles may be valid between species separated by large evolutionary distances (e.g. Detrait and Saint Girons, 1979; Saint Girons and Detrait, 1980), their use at the intra-generic level may be more problematic. For example, whilst distinct venom profiles may exist between closely-related morphologically indistinguishable species (Angulo *et al.* 2008), the use of these profiles as species identifiers assumes that the venom profiles observed are solely driven by phylogenetic distance and ignores the potential selective influence of evolutionary pressures such as diet. Furthermore, these previous observations also ignore the potential role of factors such as geography influencing inter- and intra-specific venom variation (Jimenez-Porras, 1964; Chippaux *et al.* 1991); different populations of the same species may exhibit considerable venom variation, causing species identification to be based solely upon information from single populations (or even individuals) which are not representative for the species. Because selective pressures can influence the evolution of venom composition independently to phylogenetic position, I advocate the use of venom profiles solely as a secondary species identifier after the primary use of traditional phylogenetic markers and morphological characters.

8.2 Future work

The production of multiple venom gland transcriptomes from representative species of the genus *Echis* has not only greatly improved our knowledge of the venom gland composition of these medically important species, but provided a model system to investigate: i) the selective influence of diet upon venom evolution, ii) the use of multi-gene families as predictors of organismal relationships and iii) the impact transcriptomic variation may have upon antivenom neutralisation. Whilst this project has delivered key insights into these areas of research, there are a number of future experiments that would further the data generated and any subsequent conclusions. Proteomic assessments of the venoms isolated from the species used to construct the venom gland transcriptomes would provide the tools for a comprehensive comparison between the composition of venom glands and expelled venoms. Furthermore, such studies may provide confirmation of the presence of novel toxins identified in the venom gland (e.g. lysosomal acid lipase, neprilysin and dipeptidyl peptidase III) as secreted venom components. Isolation and functional characterisation of any identified putative toxins is particularly desirable, considering their potential role in envenoming inferred from the biological activity of gene homologues. The inclusion of toxin gene data from closely related outgroup species may elucidate the genomic basis of increases in venom toxicity to invertebrates. Alternative functional approaches, based on size exclusion separation of venom components, will likely identify the toxins responsible for invertebrate lethality prior to subsequent correlations with toxin gene data. Identifying the proteins and the genes that encode them that are responsible for adaptations to invertebrate feeding would provide a valuable comparison with the two identified genetic mechanisms that facilitate adaptations to vertebrate feeding in *E. coloratus* (gene diversification and retention), thereby furthering our understanding of the genetic controls responsible for conferring alterations in venom composition. Finally, experimental evidence of venom neutralisation from other African *Echis* species by EchiTabG® (e.g. *E. jogeri*, *E. leucogaster* and *E. p. pyramidum*), would provide further justification for the geographical expansion of this antivenom to the entire African continent.

8.3 Summary

The first evidence for the genomic basis of venom composition adaptations as a response to selection pressures represents a considerable step to understanding the mechanisms that underpin the evolution of snake venoms. Clear evidence that selective pressures can influence the composition of venom components is of potential significance when assessing the nature of venom variation between both closely and distantly related species, the symptomatologies induced by snake envenomations and the appropriate selection of venoms for antivenom production. However, in the genera *Echis*, dietary induced venom variation does not appear to prevent the successful neutralisation of venom by a non-homologous antivenom. Nevertheless, these results may not represent the rule for such investigations; in this case antivenom cross-reactivity likely occurs due to venom variation being primarily limited to the diversification of existing, intra-generically conserved toxin families. In cases where venom variation occurs as the result of the recruitment of novel functionally active toxin families, I would expect antivenom cross-reactivity to be substantial reduced. Notably, the identification of EchiTabG® antivenom cross-reactivity against venom from African members of the medically important genus *Echis* represents a significant step for the production and distribution of an effective therapy to combat a substantial proportion of the ~400,000 snake envenomations occurring throughout this continent annually (Kastiruratne *et al.* 2008).

REFERENCES

Abramić M, Zubanović M and Vitale L (1988) Dipeptidyl peptidase III from human erythrocytes. *Biol. Chem. Hoppe- Seyler.* **369:** 29-38.

Abubakar SB, Abubakar IS, Habib AG, Nasidi A, Durfa N, Yusuf PO, Larnyang S, Garnvwa J, Sokomba E, Salako L, Laing GD, Theakston RDG, Juszczak E, Alder N, Warrell DA and for the Nigeria-UK EchiTab study group. 2010. Pre-clinical and preliminary dose-finding and safety studies to identify candidate antivenoms for treatment of envenoming by saw-scaled or carpet vipers (*Echis ocellatus*) in northern Nigeria. *Toxicon* **55:** 719-723.

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropolous MH, Xiao H, Merril CR, Wu A, Olde B and Moreno RF (1991) Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science.* **252:** 1651-1656.

Adams MD, Kerlavage AR, Fleischmann RD, Fuldner RA, Bult CJ, Lee NH, Kirkness RF, Weinstock KG, Gocayne JD, White O, Sutton G, Blake JA, Brandon RC, Man-Wai C, Clayton RA, Cline RT, Cotton MD, Earle-Hughes J, Fine LD, Fitzgerald LM, Fitzhugh WM, Fritchman JL, Geoghagen NSM, Glodek A, Gnehm CL, Hanna MC, Hedblom E, Hinkle PS, Kelley JM, Klimek KM, Kelley JC, Li-Ing L, Marmaros SM, Merrick JM, Moreno-Palanques RF, McDonald LA, Nguyen DT, Pellegrino SM, Phillips CA, Ryder SE, Scott JL, Saudek DM, Shirley R, Small KV, Spriggs TA, Utterback TR, Weidman JF, Yi L, Barthlow R, Bednarik DP, Liang C, Cepeda MA, Coleman TA, Collins EJ, Dimke D, Ping F, Ferrie A, Fischer C, Hastings GA and Wei-Wu H (1995) Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature.* **377:** 3-174.

Aird SD (2002) Ophidian envenomation and the role of purines. *Toxicon.* **40:** 335-393.

Akaike H (1973) Information theory as an extension of the maximum likelihood principle. Petrov BN and Csaki F (Editors) Second International Symposium on information theory. Akademia Kiado, Budapest. pp 267-281.

Alfaro ME, Zoller S and Lutzoni F (2003) Bayes or Bootstrap? A simulation study comparing the performance of Bayesian Markov Chain Monte Carlo sampling

and Bootstrapping in assessing phylogenetic confidence. *Mol. Biol. Evol.* **20(2):** 255-266.

Ali G, Kak M, Kumar M, Bali SK, Tak SI, Hassan G and Wadhwa MB (2004) Acute renal failure following *Echis carinatus* (saw-scaled viper) envenomation. *Indian J. Nephrol.* **14:** 177-181.

Altschul SF, Gish W, Miller W, Myers MW and Lipman DJ (1990) Basic local alignment search tool. *J. Mol. Biol.* **215:** 403-410.

Angulo Y, Escolano J, Lomonte B, Gutiérrez JM, Sanz L and Calvete JJ (2008) Snake venomics of Central American pitvipers: clues for rationalizing the distinct envenomation profile of *Atropoides nummifer* and *Atropoides picadoi*. *J. Proteome Res.* **7(2):** 708-719.

Annobil SH (1993) Complications of *Echis colorata* snake bites in the Asir region of SaudiArabia. *Annals of Trop. Paediatrics.* **13(1):** 39-44.

Aragon-Ortiz F and Gubenšek F (1981) *Bothrops asper* venom from the Atlantic and Pacific zones of Costa Rica. *Toxicon.* **19:** 797-805.

Arnold EN, Robinson MD and Carranza S (2009) A preliminary analysis of phylogenetic relationships and biogeography of the dangerously venomous Carpet Vipers, *Echis* (Squamata, Serpentes, Viperidae) based on mitochondrial DNA sequences. *Amphibia-Reptilia* **30:** 273-282.

Assakura MT, Silva CA, Mentele R, Camargo AC and Serrano SM (2003) Molecular cloning and expression of structural domains of bothropasin, a P-III metalloproteinase from the venom of *Bothrops jararaca*. *Toxicon.* **41:** 217-227.

Auffenberg W and Rehman H (1991) Studies on Pakistan reptiles: Pt.1. The genus *Echis* (Viperidae). *Bull. Florida. Mus. Nat. Hist.* **35:** 263-314.

Baral PK, Jajčanin-Jozić N, Deller S, Macheroux P, Abramić M and Gruber K (2008) The first structure of dipeptidyl-peptidase III provides insight into the catalytic mechanism and mode of substrate binding. *J. Biol. Chem.* **283(32):** 22316-22324.

Barlow A, Pook CE, Harrison RA and Wüster W (2009) Co-evolution of diet and prey-specific venom activity supports the role of selection in snake venom evolution. *Proc. R. Soc. B.* **276:** 2443-2449.

Barrio A and Brazil OV (1951) Neuromuscular action of the *Crotalus terrificus terrificus* poisons. *Acta Physiol. Lat.-Am.* **1:** 291-308.

Bawaskar HS, Bawaskar PH, Punde DP, Inamdar MK, Dongare RB and Bhoite RR (2008) Profile of snakebite envenoming in rural Maharashtra, *India. J. Assoc. Physicians. India.* **56:** 88-95.

Bazaa A, Marrakchi N, El Ayeh M, Sanz L and Calvete JJ (2005) Snake venomics: Comparative analysis of the venom proteomes of the Tunisian snakes *Cerastes cerastes*, *Cerastes vipera* and *Macrovipera lebetina. Proteomics* **5:** 4223-4235.

Benbassat J and Shalev O (1993) Envenomation by *Echis coloratus* (mid-east saw-scaled viper): a review of the literature and indication for treatment. *Isr. J. Med. Soc.* **29:** 239-250.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340:** 783-795.

Bernadsky G, Bdolah A and Kochva E (1986) Gel permeation patterns of venoms from eleven species of the genera *Vipera. Toxicon.* **24:** 721-725.

Bertke EM, Watt DD and Tu T (1966) Electrophoretic pattens of venom from species of *Crotalidae* and *Elapidae* snakes. *Toxicon.* **4:** 73-76.

Bharati K, Hasson SS, Oliver J, Laing GD, Theakston RDG and Harrison RA (2003) Molecular cloning of phospholipases A_2 from venom glands of *Echis* carpet vipers. *Toxicon.* **41:** 941-947.

Bhat RN (1974) Viperine snake poisoning in Jammu. *J. Ind. Med. Asso.* **63(12):** 383-392.

Biardi JE, Coss RG and Smith DG (2000) California ground squirrel (*Spermophilus beecheyi*) blood sera inhibits crotalid venom proteolytic activity. *Toxicon.* **38:** 713-721.

Biardi JE, Chien DC and Coss RG (2006) California ground squirrel (*Spermophilus beecheyi*) defenses against rattlesnake venom digestive and hemostatic toxins. *J. Chemical Ecology* **32:** 137-154.

Bjarnason JB and Fox JW (1994) Hemorrhagic metalloproteinases from snake venoms. *Pharmaology & Therpeutics.* **62:** 325-372.

Bjarnason JB and Fox JW (1995) Snake venom metalloendopeptidases: reprolysins. *Methods Enzymol.* **248:** 345-368.

Boche J, Chippaux JP and Courtois B (1981) Contribution à l'étude des variations biochemiques des venins de serpents d'Afrique de l'Ouest. *Bull. Soc. Path. Exot.* **74:** 356-366.

Boguski MS and Schuler GD (1995) ESTablishing a human transcript map. *Nature genetics.* **10:** 369-371.

Bonfield JK, Smith KF and Staden R (1995) A new DNA sequence assembly program. *Nucleic Acid Res.* **23:** 4992-4999.

Braga MDM, Martins AMC, Amora DN, de Menezes DB, Toyama MH, Toyama DO, Marangoni S, Barbosa PSF, de Sousa-Alves R, Fonteles MC and Monteiro HSA (2006) Purification and biological effects of C-type lectin isolated from *Bothrops insularis* venom. *Toxicon.* **47:** 859-867.

Brecher P and Kuan HT (1979) Lipoprotein lipase and acid lipase activity in rabbit brain microvessels. *J. Lipid Res.* **20:** 464-471.

Buckley TR, Cordeiro M, Marshall DC and Simon C (2006) Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (*Maoricicada* Dugdale). *Syst. Biol.* **55(3):** 411-425.

Burke J, Davidson D and Hide W (1999) d2_cluster: A validated method for clustering EST and full-length cDNA sequences. *Genome research.* **9:** 1135-1142.

Calvete JJ, Marcinkiewicz C, Monleón D, Esteve V, Celda B, Juárez P and Sanz L (2005) Snake venom disintegrins: evolution of structure and function. *Toxicon.* **45:** 1063-1074.

Calvete JJ, Escolano J and Sanz L (2007) Snake venomics of *Bitis* species reveals large intragenus venom toxin composition variation: Application to taxonomy of congeneric taxa. *J. Proteome Res.* **6:** 2732-2745.

Calvete JJ, Sanz L, Angulo Y, Lomonte B and Gutiérrez JM (2009) Venoms, venomics, antivenomics. *FEBS Lett.* **583:** 1736-1743.

Calvete JJ, Cid P, Sanz L, Segura A, Villalta M, Herrera M, León G, Harrison RA, Durfa N, Nasidi A, Theakston RDG, Warrell DA and Gutiérrez JM (Forthcoming) Antivenomic assessment of the immunological reactivity of EchiTAb-Plus-ICP®, an antivenom for the treatment of snakebite envenoming in sub-Saharan Africa. *J. Proteomics.* In press.

Casewell NR, Harrison RA, Wüster W and Wagstaff SC (2009) Comparative venom gland transcriptome surveys of the saw-scaled vipers (Viperidae: *Echis*) reveal substantial intra-family gene diversity and novel venom transcripts. *BMC Genomics* **10:** 564.

Castoe TC, Sasa M and Parkinson CL (2005) Modelling nucleotide evolution at the mesoscale: the phylogeny of the Neotropical pit vipers of the *Porthidium* group (Viperidae: Crotalinae). *Mol. Phylogenet. Evol.* **37:** 881-898.

Castoe TC and Parkinson CL (2006) Bayesian mixed models and the phylogeny of pitvipers (Viperidae: Serpentes). *Mol. Phylogenet. Evol.* **39:** 91-110.

Champagne DE (2005) Antihemostatic molecules from saliva of blood-feeding arthropods. *Pathophysiol Haemos Thromb* **34:** 221-227.

Charpentier I, Pillet L, Karlsson E, Couderc J and Menez A (1990) Recognition of the acetycholine receptor binding site of a long chain neurotoxin by toxin specific monoclonal antibodies. *J. Mol. Recog.* **3:** 74-81.

Chen YL and Tsai IH (1996) Functional and sequence characterization of coagulation factor IX/factor X binding protein from the venom of *Echis carinatus leucogaste*. *Biochemistry* **35:** 5264–5271

Cherlin VA (1983) New facts on the taxonomy of snakes of the genus *Echis* (in Russian). *Vestnik Zoologii.* **1983(2):** 42-26. Translation by Owusu FSH, edited by Hughes B and Zug GR (1984) *Smithsonian Herp. Info. Serv.* **61.**

Cherlin VA (1990) Taxonomic revision of the snake genus *Echis* (Viperidae) II. An analysis of taxonomy and description of new forms. *Tr. Zool. Inst. Akad. Nauk. SSSR.* **207:** 193-223.

Chippaux JP, Williams V and White J (1991) Snake venom variability: Methods of study, results and interpretation. *Toxicon.* **29:** 1279-1303.

Chippaux JP (1998) Snake-bites: appraisal of the global situation. *Bull. World Health Organ.* **76:** 515-524.

Chippaux JP (2006) The toxicology of the venoms. In: *Snake venoms and envenomations.* Krieger Publishing Company, Malabar, Florida, USA. pp 75-124.

Chugh KS (1989) Snakebite induced acute renal failure in India. *Kidney International.* **35:** 891-907.

Cidade DAP, Simão TA, Dávila AMR, Wagner G, Junqueira-de-Azevedo ILM, Ho PL, Bon C, Zingali RB and Albano RM (2006) *Bothrops jararaca* venom gland transcriptome: Analysis of the gene expression pattern. *Toxicon* **48:** 437-461.

Cook DAN, Owen T, Wagstaff SC, Kinne J, Wernery U and Harrison RA (Forthcoming) Analysis of camelid antibodies for antivenom development: neutralisation of venom-induced pathology. *Toxicon.* In press.

Cotton JA and Page RDM (2002) Going nuclear: gene family evolution and vertebrate phylogeny reconciled. *Proc. R. Soc. B.* **269:** 1555-1561.

Creer S, Malhotra A, Thorpe RS, Stőcklin R, Favreau P and Chou WH (2003) Genetic and ecological correlates of intraspecific variation in pitviper venom composition detected using matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF-MS) and isoelectric focusing. *J. Mol. Evol.* **56:** 317-329.

Currier RB, Harrison RA, Rowley PD, Laing GD and Wagstaff SC (2010) Intra-specific variation in venom of the African puff adder (*Bitis arietans*): differential expression and activity of snake venom metalloproteinases (SVMPs). *Toxicon.* **55:** 864-873.

Daltry JC, Wüster W and Thorpe RS (1996a) Diet and snake venom evolution. *Nature.* **379:** 537-540.

Daltry JC, Ponnuburai G, Shin CK, Tan NH, Thorpe RS and Wüster W (1996b). Electrophoretic profiles and biological activities: Intraspecific variation in the venom of the Malayan pit viper (*Calloselasma rhodostoma*). *Toxicon.* **34:** 67-79.

David P and Ineich I (1999) Les serpents venimeux du monde: systématique et répartition. *Dumerilia* **3:** 3–499.

de Oliveira L, Cunha AOS, Mortari MR, Coimbra NC and Dos Santos WF (2006) Cataleptic activity of the denatured venom of the social wasp *Agelaia vicina* (Hymenoptera, Vespidae) in *Rattus norvegicus* (Rodentia, Muridae). *Prog. Neuro-Psychopharm. Biol. Psych.* **30:** 198-203.

Deshimaru M, Ogawa T, Nakashima KI, Nobuhisa I, Chijiwa T, Shimohigashi Y, Fukumaki Y, Niwa M, Yamashina I, Hattori and Ohno M (1996) Accelerated evolution of crotalinae snake venom gland serine proteases. *FEBS Letters.* **397:** 83-88.

Detrait J and Saint Girons H (1979) Communautés antigéniques des venins et systématique des Viperidae. *Bijdr. Dierk.* **49:** 71-80.

Diaz C, Gutierrez JM, Lomonte B and Gene JA (1991) The effect of myotoxins isolated from *Bothrops* snake venoms on multilamelar liposomes: Relationship

to phospholipase A_2, anticoagulant and myotoxic activities. *Biochim Biophys Acta.* **1070:** 455-460.

Doley R, Tram NNB, Reza MA and Kini RM (2008a) Unusual accelerated rate of deletions and insertions in toxin genes in the venom glands of the pygmy copperhead (*Austrelaps labialis*) from Kangaroo Island. *BMC Evol. Biol.* **8:** 70.

Doley R, Pahari S, Mackessy SP and Kini RM (2008b) Accelerated exchange of exon segments in Viperid three-finger toxin genes (*Sistrurus catenatus edwardsi*; Desert Massasauga). *BMC Evol. Biol.* **8:** 196.

Doley R, Mackessy SP and Kini RM (2009) Role of accelerated segment switch in exons to alter targeting (ASSET) in the molecular evolution of snake venom proteins. *BMC Evol. Biol.* **9:** 146.

Du XY and Clemetson KJ (2002) Snake venom L-amino acid oxidases. *Toxicon.* **40:** 659-665.

Escoubas P, Sollod B and King GF (2006) Venom landscapes: Mining the complexity of spider venoms via a combined cDNA and mass spectrometric approach. *Toxicon* **47:** 650-663.

Eulenstein O, Mirkin B and Vingron M (1997) Comparison of annotating duplications, tree mapping, and copying as methods to compare gene trees with species trees. In: Mirkin B, McMorris FR, Roberts FS and Rzhetsky A (editors). Mathematical hierarchies in biology. American Mathematical Society, Providence, Rhode Island, USA. p. 71-93.

Ewing B and Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8:** 186-194.

Ewing B, Hillier L, Wendl MC and Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8:** 175-185.

Finney DJ (1971) Probit analysis. London: Cambridge University Press: Third Edition.

Fournie-Zaluski MC, Fassot C, Valentin B, Djordjijevic D, Reaux-Le Goazigo A, Corvol P, Roques BP and Llorens-Cortes C (2004) Brain renin-angiotensin system blockade by systemically active aminopeptidase A inhibitors: a potential treatment of salt-dependent hypertension. *Proc. Natl. Acad. Sci. USA.* **101:** 7775-7780.

Fox JW and Serrano SMT (2005) Structural considerations of the snake venom metalloproteinases, key members of the M12 reprolysin family of metalloproteinases. *Toxicon.* **45:** 969-985.

Fox JW and Serrano SMT (2008) Insights into and speculations about snake venom metalloproteinase (SVMP) synthesis, folding and disulfide bond formation and their contribution to venom complexity. *FEBS J.* **275:** 3016-3030.

Francischetti IMB, My-Pharm V, Harrison J, Garfield MK and Ribeiro JMC (2004) *Bitis gabonica* (Gaboon viper) snake venom gland: toward a catalog for the full-length transcripts (cDNA) and proteins. *Gene.* **337:** 55-69.

Fry BG, Wickramaratna JC, Hodgson WC, Alewood PF, Kini RM, Ho H and Wüster W (2002) Electrospray liquid chromatography/mass spectrometry fingerprinting of *Acanthophis* (death adder) venoms: taxonomic and toxinological implications. *Rapid Commun. Mass Spectrom.* **16:** 600-608.

Fry BG, Lumsden NG, Wüster W, Wickramaratna JC, Hodgson WC, and Kini RM (2003a) Isolation of a Neurotoxin (α-colubritoxin) from a nonvenomous Colubrid: Evidence for early origin of venom in snakes. *J. Mol. Evol.* **57:** 446-452.

Fry BG, Wüster W, Kini RM, Brusic V, Khan A, Venkataraman D and Rooney AP (2003b) Molecular evolution and phylogeny of the Elapid snake venom three-finger toxins. *J. Mol. Evol.* **57:** 110-129.

Fry BG (2005) From genome to "venome": Molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. *Genome Res.* **15:** 403-420.

Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SFR, Kuruppu S, Fung K, Hedges SB, Richardson MK, Hodgson WC, Ignjatovic V, Summerhayes R and Kochva E (2006) Early evolution of the venom system in lizards and snakes. *Nature.* **439:** 584-588.

Fry BG, Scheib H, van der Weerd L, Young B, McNaughtan J, Ramjan SFR, Vidal N, Poelmann RE and Norman JA (2008) Evolution of an arsenal. Structural and Functional Diversification of the Venom System in the Advanced Snakes (Caenophidia). *Mol. Cell Prot.* **7:** 215-246.

Fry BG, Wroe S, Teeuwisse W, van Osch MJP, Moreno K, Ingle J, McHenry C, Ferrara T, Clausen P, Scheib H, Winter KL, Greisman L, Roelants K, van der Weerd L, Clemente CJ, Giannakis E, Hodgson WC, Luz S, Martelli P,

Krishnasamy K, Kochva E, Kwok HF, Scanlon D, Karas K, Citron DM, Goldstein EJC, Mcnaughton JE and Norman JA (2009) A central role for venom in predation by *Varanus komodoensis* (Komodo Dragon) and the extinct giant *Varanus* (*Megalania*) *priscus*. *Proc. Natl. Acad. Sci. USA* **106:** 8969-8974.

Fürstenau CR, Trentin DDS, Barreto-Chaves MLM and Sarkis JJF (2006) Ecto-nucleotide pyrophosphate/phosphodiesterase as part of a multiple system for nucleotide hydrolysis by platelets from rats: Kinetic characterization and biochemical properties. *Platelets.* **17(2):** 84-91.

Galán JA, Sánchez EE, Rodríguez-Acosta A and Pérez JC (2004) Neutralization of venoms from two Southern Pacific Rattlesnakes (*Crotalus helleri*) with commercial antivenoms and endothermic animal sera. *Toxicon.* **43:** 791-799.

Galtier N and Daubin V (2008) Dealing with incongruence in phylogenomic analyses. *Phil. Trans. R. Soc. B.* **363:** 4023-4029.

Garrigues T, Dauga C, Ferquel E, Choumet V and Failloux A-B (2005) Molecular phylogeny of *Vipera* Laurenti, 1768 and the related genera *Macrovipera* (Reuss, 1927) and *Daboia* (Gray, 1842), with comments about neurotoxic *Vipera aspis aspis* populations. *Mol. Phylogenet. Evol.* **35:** 35-47.

Gawade SP (2004) Snake venom neurotoxins: Pharmacological classification. *J. Toxicol. Toxin Rev.* **23:** 37-96.

Gibbs HL and Mackessy SP (2009) Functional basis of a molecular adaptation: Prey-specific toxic effects of venom from *Sistrurus* rattlesnakes. *Toxicon* **53:** 672-679.

Gillissen A, Theakston RDG, Barth J, May B, Krieg M and Warrell DA (1994) Neurotoxicity, haemostatic disturbances and haemolytic anaemia after a bite by a Tunisian saw-scaled or carpet viper (*Echis* '*pyramidum*'-complex): Failure of antivenom treatment. *Toxicon* **32:** 937-944.

Githens T (1935) Studies on the venom of North American pit vipers. *J. Immunol.* **29:** 165-173.

Glenn JL and Straight RC (1977) The midget faded rattlesnake (*Crotalus viridis concolor*), venom: lethal toxicity and individual variability. *Toxicon* **15:** 129-133.

Glenn JL and Straight RC (1978) Mojave rattlesnake *Crotalus scutulatus scutulatus* venom: variation in toxicity with geographical origin. *Toxicon* **16:** 81-84.

Glenn JL and Straight RC (1989) Intergradation of two different venom populations of the Mojave rattlesnake (*Crotalus scutulatus scutulatus*) in Arizona. *Toxicon.* **27:** 411-481.

Glenn JL, Straight RC, Wolfe MC and Hardy DL (1983) Geographical variation in *Crotalus scutulatus scutulatus* (Mojave rattlesnake) venom properties. *Toxicon.* **27:** 411-481.

Glenner GG and Folk JE (1961) Glutamyl peptidases in rat and guinea pig kidney slices. *Nature.* **192:** 338-340.

Goodman M, Czelusniak J, Moore GW, Romero-Herrera AE and Matsuda G (1979) Fitting the gene lineage into its species lineage: a parsimony strategy illustrated by cladograms constructed from globin sequences. *Syst. Zool.* **28:** 132-168.

Goonetilleke A and Harris JB (2002) Envenomation and consumption of poisonous seafood. *J. Neurol. Neurosurg. Psychiat.* **73:** 103-109.

Gorbacheva EV, Starkov VG, Tsetlin VI, Utkin YN and Vulfius CA (2008) Viperidae snake venoms block nicotinic acetylcholine receptors and voltage-gated Ca2+ channels in identified neurons of fresh-water snail *Lymnaea stagnalis. Biochem. (Moscow) A. Membrane Cell Biol.* **2:** 14-18.

Gowda CDR, Nataraju A, Rajesh R, Dhananjaya BL, Sharath BK and Vishwanath BS (2006a) Differential action of proteases from *Trimeresurus malabaricus*, *Naja naja* and *Daboia russellii* venoms on hemostasis. *Comp. Biochem. & Physiol. Part C: Toxicon. & Pharm.* **143:** 295-302.

Gowda CD, Rajesh R, Nataraju A, Dhananjaya BL, Raghupathi AR, Gowda TV, Sharath BK and Vishwanath BS (2006b) Strong myotoxic activity of *Trimeresurus malabaricus* venom: role of metalloproteinases. *Mol. Cell. Biochem.* **282:** 147-155.

Graul RC and Sadée W (1997) Evolutionary relationships among proteins probed by an iterative neighbourhood cluster analysis (INCA). Alignment of bacteriorhodopsins with the yeast sequence YRO2. *Pharma. Res.* **14:** 1533-1541.

Green P (1995) Documentation for PHRAP. Genome Center, University of Washington. http://www.phrap.org/phrap.docs/phrap.html.

Gutiérrez JM, Gené A, Rodas G and Cerdas L (1985) Neutralization of proteolytic and hemorrhagic activities of Costa Rican snake venoms by a polyvalent antivenom. *Toxicon.* **23:** 887-893.

Gutiérrez JM, Romero M, Díaz C, Borkow G and Ovadia M (1995) Isolation and characterization of a metalloproteinase with weak hemorrhagic activity from the venom of the snake *Bothrops asper* (terciopelo). *Toxicon* **33:** 19-29.

Gutiérrez JM, Rucavado A, Escalante T and Díaz C (2005) Hemorrhage induced by snake venom metalloproteinases: biochemical and biophysical mechanisms involved in microvessel damage. *Toxicon* **45:** 997-1011.

Gutiérrez JM, Sanz L, Escolano J, Fernández J, Lomonte B, Angulo Y, Rucavado A, Warrell DA and Calvete JJ (2008) Snake venomics of the Lesser Antillean pit vipers *Bothrops caribbaeus* and *Bothrops lanceolatus*: correlation with toxicological activities and immunoreactivity of a heterologous antivenom. *J. Proteome Res.* **7:** 4396-4408.

Gutiérrez JM, Lomonte B, León G, Alape-Girón A, Flores-Díaz M, Sanz L, Angulo Y and Calvete JJ (2009) Snake venomics and antivenomics: proteomic tools in the design and control of antivenoms for the treatment of snakebite envenoming. *J Proteomics* **72:** 165-182.

Habib AG, Gebi UI and Onyemelukwe GC (2001) Snake bite in Nigeria. *Afr. J. Med. Med. Sci.* **30:** 171-178.

Hahn B-S, Yang K-Y, Park E-M, Chang M and Kim Y-S (1996) Purification and molecular cloning of Calobin, a thrombin-like enzyme from *Agkistrodon caliginosus* (Korean viper). *J. Biochem.* **119:** 835-843.

Harris JB and Goonetilleke A (2004) Animal poisons and the nervous system: What the neurologist needs to know. *J. Neurol. Neurosurg. Psychiat.* **75:** 40-46.

Harrison RA, Wüster W and Theakston RDG (2003) The conserved structure of snake venom toxins confers extensive immunological cross-reactivity to toxin-specific antibody. *Toxicon.* **41:** 441-449.

Harrison RA, Ibison F, Wilbraham D and Wagstaff SC (2007) Identification of cDNAs encoding viper venom hyaluronidases: cross-generic sequence conservation of full-length and unusually short variant transcripts. *Gene.* **392:** 22-33.

Harvey AL and Karlsson E (1980) Dedrotoxin from the venom of the green mamba *Dendroaspis angusticeps*: a neurotoxin that enhances acetylcholine release at neuromuscular junctions. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **312:** 1-6.

Harvey AL, Barfaraz A, Thomson E, Faiz A, Preston S and Harris JB (1994) Screening of snake venoms for neurotoxic and myotoxic effects using simple *in vitro* preperations from rodents and chicks. *Toxicon.* **32:** 257-265.

Harvey AL (2001) Twenty years of dendrotoxins. *Toxicon.* **39:** 15-26.

Hayes WK (1991) Ontogeny of striking, prey-handling and envenomation behaviour of prairie rattlesnakes (*Crotalus v. viridis*). *Toxicon.* **29:** 867-875.

Hayes WK, Lavín-Murcio P and Kardong KV (1995) Northern Pacific rattlesnakes (*Crotalus viridis oreganus*) meter venom when feeding on prey of different sizes. *Copeia* **2:** 337-343.

Hayter JR, Robertson DHL, Gaskell SJ and Beynon RJ (2003) Proteome analysis of intact proteins in complex mixtures. *Mol. Cell. Proteomics.* **2:** 85-95.

Heath L, van der Walt E, Varsnai A and Martin DP (2006) Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *J. Virol.* **80:** 11827-11832.

Heatwole H and Poran NS (1995) Resistances of sympatric and allopatric eels to sea snake venoms. *Copeia.* **1:** 136-147.

Heatwole H and Powell J (1998) Resistance of eels (*Gymnothorax*) to the venom of sea kraits (*Laticauda colubrina*): a test of coevolution. *Toxicon.* **36:** 619-625.

Huang TF, Holt JC, Lukasiewicz H and Niewiarowski S (1987) Trigamin. A low molecular weight peptide inhibiting fibrinogen with platelet receptors expressed on glycoprotein IIb-IIIa complex. *J. Biol. Chem.* **262:** 16157-16163.

Huang X (1996) An improved sequence assembly program. *Genomics.* **33:** 21-31.

Huang X and Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Research.* **9:** 868-877.

Huelsenbeck JP and Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics.* **17:** 754-755.

Isaac RE (1988) Neuropeptide-degrading endopeptidase activity of locust (*Schistocerca gregaria*) synaptic membranes. *Biochem. J.* **255:** 843-847.

Jackson K (2003) The evolution of venom-delivery systems in snakes. *Zool. J. Linn. Soc.* **137:** 337-354.

Jan V, Maroun RC, Robbe-Vincent A, De Haro L and Choumet V (2002) Toxicity evolution of *Vipera aspis aspis* venom: identification and molecular modeling of a novel phospholipase A_2 heterodimer neurotoxin. *FEBS Letters* **527:** 263-268.

Jasti J, Paramasivam M, Srinivasan A and Singh TP (2004a) Structure of an acidic phospholipase A_2 from Indian saw-scaled viper (*Echis carinatus*) at 2.6 Å resolution reveals a novel intermolecular interaction. *Acta Cryst.* **D60:** 66-72.

Jasti J, Paramasivam M, Srinivasan A and Singh TP (2004b) Crystal structure of echicetin from *Echis carinatus* (Indian saw-scaled viper) at 2.4 Å resolution. *J. Mol. Biol.* **335:** 167-176.

Jimenéz-Porras JM (1964) Intraspecific variations in composition of venom of the jumping viper, *Bothrops nummifer*. *Toxicon* **2:** 187-190.

Jimenéz-Porras JM (1967) Differentiation between *Bothrops nummifer* and *Bothrops picadoi* by means of the biochemical properties of their venoms. In: Russell FE and Saunders PR, (Editors). *Animal Toxins,* Oxford, Pergamon Press, pp 307-321.

Jordan GE and Piel, WH (2008) PhyloWidget: web-based visualizations for the tree of life. *Bioinformatics* **24:** 1641-1642.

Jorge-da-Silva N and Aird SD (2001) Prey specificity, comparative lethality and compositional differences of coral snake venoms. *Comp. Biochem. Physiol. Part C: Toxicol. Pharmacol.* **128:** 425-456.

Juárez P, Sanz L and Calvete JJ (2004) Snake venomics: characterization of protein families in *Sistrurus barbouri* venom by cysteine mapping, N-terminal sequencing and tandem mass spectrometry analysis. *Proteomics* **4:** 327-338.

Juárez P, Wagstaff SC, Oliver J, Sanz L, Harrison RA and Calvete JJ (2006a) Molecular cloning of disintegrin-like transcript BA-5A from a *Bitis arietans* venom gland cDNA library: A putative intermediate in the evolution of the long-chain disintegrin bitistatin. *J. Mol. Evol.* **63:** 142-152.

Juárez P, Wagstaff SC, Sanz L, Harrison RA and Calvete JJ (2006b) Molecular cloning of *Echis ocellatus* disintegrins reveals non-venom secreted proteins and a pathway for the evolution of ocellatusin. *J. Mol. Evol.* **63:** 183-193.

Junqueira-de-Azevedo ILM, Ho PL (2002) A survey of gene expression and diversity in the venom glands of the pit viper snake *Bothrops insularis* through the generation of expressed sequence tags (ESTs). *Gene.* **299:** 279-291.

Junqueira-de-Azevedo ILM, Ching ATC, Carvalho E, Faria F, Nishiyama ML, Ho PL and Diniz MRV (2006) *Lachesis muta* (Viperidae) cDNAs reveal diverging pit viper molecules and scaffolds typical of Cobra (Elapidae) venoms: Implications for snake toxin repertoire evolution. *Genetics.* **173:** 877-889.

Karlsson E (1979) Chemistry of protein toxins in snake venoms. In: Lee CY (Editor), (1979) *Snake Venoms, Handbook of Exp. Pharm.* **52,** Springer-Verlag, Berlin, pp. 159-212.

Kashima S, Roberto PG, Soares AM, Astolfi-Filho S, Pereira JO, Giuliati S, Faria M, Xavier MAS, Fontes MRM, Giglio JR and Franca SC (2004) Analysis of *Bothrops jararacussu* venomous gland transcriptome focusing on structural and functional aspects: I – gene expression profile of highly expressed phospholipases A_2. *Biochimie.* **86:** 211-219.

Kasturiratne A, Wickremasinghe AR, de Silva N, Gunawardena NK, Pathmeswaran A, Premaratna R, Savioli L, Lalloo DG and de Silva HJ (2008) The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths. *PLOS Med.* **5(11):** e218.

Keane TM, Creevey, CJ, Pentony MM, Naughton TJ and McInerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol. Biol.* **6:** 29.

Kemparaju K, Prasad BN and Gowda VT (1994) Purification of a basic phospholipase A_2 from Indian saw-scaled viper (*Echis carinatus*) venom: characterization of antigenic, catalytic and pharmacological properties. *Toxicon.* **32:** 1187-1196.

Kemparaju K, Krishnakanth TP and Gowda VT (1999) Purification and characterization of a platelet aggregation inhibitor acidic phospholipase A_2 from Indian saw-scaled viper (*Echis carinatus*) venom. *Toxicon.* **37:** 1659-1671.

Kemparaju K and Girish KS (2006) Snake venom hyaluronidase: a therapeutic target. *Cell Biochem. Funct.* **24:** 7-12.

Kini RM and Chan YM (1999) Accelerated evolution and molecular surface of venom phospholipase A(2) enzymes. *J. Mol. Evol.* **48:** 125-132.

Kini RM (2005) Serine proteases affecting blood coagulation and fibrinolysis from snake venoms. *Pathophysiol Haemost Thromb.* **34:** 200-204.

Kini RM (2006) Anticoagulant proteins from snake venoms: structure, function and mechanism. *Biochem. J.* **397:** 377-387.

Kishimoto M and Takahashi T (2002) Molecular cloning of HR1a and HR1b, high molecular hemorrhagic factors, from *Trimeresurus flavoviridis* venom. *Toxicon.* **40:** 1369-1375.

Kochar DK, Tanwar PD, Norris RL, Sabir M, Nayak KC, Agrawal TD, Purohit VP, Kochar A and Simpson ID (2007) Rediscovery of severe saw-scaled viper (*Echis sochureki*) envenoming in the Thar Desert region of Rajasthan, India. *Wilderness Enviro. Med.* **18:** 75-85.

Kordiš D and Gubenšek F (2000) Adaptive evolution of animal toxin multigene families. *Gene.* **261:** 43-52.

Kornalik F and Master RWP (1964) A comparative examination of yellow and white venoms of *Vipera ammodytes*. *Toxicon.* **2:** 109-111.

Kornalik F and Táborská E (1988) Intraspecies variability in the composition of the coagulant active snake venoms. In: Pirkle, H. and Markland, F. S. (Eds). *Haemostasis and Animal Venoms.* New York, Dekker. Vol 7, pp. 503-513.

Kornalik F and Táborská E (1989) Cross reactivity of mono- and polyvalent antivenoms with *Viperidae* and *Crotalidae* snake venoms. *Toxicon* **27:** 1135-1142.

Krem MM and Di Cera E (2002) Evolution of enzyme cascades from embryonic development to blood coagulation. *Trends Biochem. Sci.* **27:** 67-74.

Kularatne SAM and Ratnatunga N (1999) Severe systemic effects of Merrem's hump-nosed viper bite. *Ceylon Medical Journal.* **44(4):** 169-170.

Laing GD, Theakston RDG, Leite RP, Dias da Silva WD, Warrell DA (1992) Comparison of the potency of three Brazilian *Bothrops* antivenoms using *in vivo* rodent and *in vitro* assays. *Toxicon.* **30:** 1219-1225.

Laing GD, Lee L, Smith DC, Landon J and Theakston RDG (1995) Experimental assessment of a new, low-cost antivenom for treatment of carpet viper (*Echis ocellatus*) envenoming. *Toxicon.* **33:** 307-313.

Lamb G (1902) On the precipitin of cobra venom: a means of distinguishing between the proteins of different snake poisons. *Lancet.* **II:** 431-435.

Lamb G (1904) On the precipitin of cobra venom. *Lancet.* **I:** 916-921.

Larget B and Simon DL (1999) Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* **16(6):** 750-759.

Lee CM and Snyder SH (1982) Dipeptidyl-aminopeptidase III of rat brain: selective affinity for enkephalin and angiotensin. *J. Biol. Chem.* **257(20):** 12043-12050.

Lee CY (Editor) (1979) *Snake venoms, Handbook of Exp. Pharm.* **52,** Springer-Verlag, Berlin.

Li M, Fry BG and Kini RM (2005) Eggs-only diet: Its implications for the toxin profile changes and ecology of the marbled sea snake (*Aipysurus eydouxii*). *J. Mol. Evol.* **60:** 81-89.

Li Y, Qin Y, Li H, Wu R, Yan C and Du H (2007) Lysosomal acid lipase over-expression disrupts lamellar body genesis and alveolar structure in the lung. *Int. J. Exp. Path.* **88:** 427-436.

Liu L and Pearl DK (2007) Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Syst. Biol.* **56:** 504-514.

Liu L, Yu L and Pearl DK (2010) Maximum tree: a consistent estimator of the species tree. *J. Math. Biol.* **60:** 95–106.

Lomonte B, Escolano J, Fernández J, Sanz L, Angulo Y, Gutiérrez JM and Calvete JJ (2008) Snake venomics and antivenomics of the arboreal neotropical pitvipers *Bothriechis lateralis* and *Bothriechis schlegelii*. *J. Proteome Res.* **7:** 2445-2457.

Lukoschek V and Keogh JS (2006) Molecular phylogeny of sea snakes reveals a rapidly diverged adaptive radiation. *Biol. J. Linn. Soc.* **89:** 523-539.

Lynch VJ (2007) Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. *BMC Evolutionary Biology.* **7:** 2.

Machado O, Oliveira Carvalho AL, Zingali RB and Carlini CR (1993) Purification, physiochemical characterization and N-terminal amino acid sequence of a phopholipase A_2 from *Bothrops jararaca* venom. *Brazilian J. Med. Biolog. Research.* **26:** 163-166.

Maddison WP (1997) Gene trees in species trees. *Syst. Biol.* **46:** 523-536.

Maddison WP and Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting. *Syst Biol.* **55:** 21-30.

Maddison WP and Maddison DR (2008) Mesquite: a modular system for evolutionary analysis. Version 2.5, build j55. Available at http://mesquiteproject.org.

Marchio S, Lahdenranta J, Schlingemann RO, Valdembri D, Wesseling P, Arap MA, Hajitou A, Ozawa MG, Trepel M, Giordano RJ, Nanus DM, Dijkman HB, Ooserwijk E, Sidman RL, Cooper MD, Bussolino F, Pasqualini R and Arap W (2004) Aminopeptidase A is a functional target in angiogenic blood vessels. *Cancer Cell.* **5:** 151-162.

Markland FS (1998) Snake venoms and the haemostatic system. *Toxicon.* **36:** 1749-1800.

Master RWP and Kornalik F (1965) Biochemical differences in yellow and white venoms of *Vipera ammodytes* and Russell's viper. *J. Biol. Chem.* **240:** 139-142.

Matsas R, Fulcher IS, Kenny AJ and Turner AJ (1983) Substance P and (Leu)enkephalin are hydrolysed by an enzyme in pig caudate synaptic membranes that is identical with the endopeptidase of kidney microvilli. *Proc. Natl. Acad. Sci. USA* **80:** 3111-3115.

McCue MD (2006) Cost of producing venom in three North American pitviper species. *Copeia* **4:** 818-825.

Mebs D (1999) Snake venom composition and evolution of Viperidae. *Kaupia.* **8:** 145-148.

Mebs D (2001) Toxicity in animals. Trends in evolution? *Toxicon* **39:** 87-96.

Minton SA (1956) Some properties of North American pit vipers and their correlation with phylogeny. In: Buckley EE and Porges N (Editors). *Venoms.* Washington AAAS. pp.145-151.

Montecucco C and Rossetto O (2000) How do presynaptic PLA2 neurotoxins block nerve terminals? *Trends in Biochemical Sciences.* **25:** 266-270.

Morita T (2005) Structures and functions of snake venom CLPs (C-type lecin-like proteins) with anticoagulant-, procoagulant-, and platelet modulating activities. *Toxicon.* **45:** 1099-1114.

Moura-da-Silva AM, Paine MJ, Diniz MR, Theakston RD and Crampton JM (1995) The molecular cloning of a phospholipase A_2 from *Bothrops jararacussu* snake venom: Evolution of venom group II phospholipase A_2's may imply gene duplications. *J. Mol. Evol.* **41(2):** 174-179.

Moura-da-Silva AM, Theakston RDG and Crampton JM (1996) Evolution of disintegrin cysteine-rich and mammalian matrix-degrading

metalloproteinases:gene duplication and divergence of a common ancestor rather than convergent evolution. *J. Mol. Evol.* **43:** 263-269.

Muñoz-Chápuli R, Carmona R, Guadix JA, Macías D and Pérez-Pomares JM (2005) The origin of the endothelial cells: an evo-devo approach for the invertebrate/vertebrate transition of the circulatory system. *Evol. Dev.* **7:** 351-358.

Nakashima KI, Ogawa T, Oda N, Hattori M, Sakaki Y, Kihara H and Ohno M (1993) Accelerated evolution of *Trimeresurus flavovirids* venom gland phosphlipase A_2 isoenzymes. *Proc. Natl. Acad. Sci. USA.* **90:** 5964-5968.

Nakashima KI, Nobuhisa I, Deshimaru M, Nakai M, Ogawa T, Shimohigashi Y, Fukumaki Y, Hattori M, Sakaki Y, Hattori S and Ohno M (1995) Accelerated evolution in the protein-coding regions is universal in croalinae snake venom gland phospholipase A2 isozyme genes. *Proc. Natl. Acad. Sci. USA* **92:** 5605-5609.

Nei M and Hughes AL (1992) Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. In: Tsuji K, Aizawa M and Sasazuki T (editors). 11th Histocompatibility Workshop and Conference. Oxford University Press, Oxford, UK. p. 27-38.

Nei M, Gu X and Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. USA* **94:** 7799-7806.

Nei M and Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genetics* **39:** 121-152.

Neiva M, Arraes FBM, de Souza JV, Rádis-Baptista G, da Silva ARBP, Walter MEMT, de Macedo Brigido M, Yamane T, López-Lozano J. and Astolfi-Filho S (2009) Transcriptome analysis of the Amazonian viper *Bothrops atrox* venom gland using expressed sequence tags (ESTs). *Toxicon* **53:** 427-436.

Nishida S, Fujita T, Kohno N, Atoda H, Morita T, Takeya H, Kido I, Paine MJI, Kawabata S and Iwanaga S (1995) cDNA cloning and deduced amino acid sequence of prothrombin activation (ecarin) from Kenyan *Echis carinatus* venom. *Biochemistry* **34:** 1771-1778.

Nylander JAA (2004) MrModeltest v2. Program distributed by the author, Evolutionary Biology Centre, Uppsala University.

Ogawa T, Chijiwa T, Oda-Ueda N and Ohno M (2005) Molecular diversity and accelerated evolution of C-type lectin-like proteins from snake venom. *Toxicon.* **45:** 1-14.

Ogilve ML and Gartner TK (1984) Identification of lectins in snake venoms. *J. Herpetology.* **18:** 285-290.

Ohno M, Cheijiwa T, Oda-Ueda N, Ogawa T and Hattori S (2003) Molecular evolution of myotoxin phospholipaes A_2 from snake venom. *Toxicon.* **42:** 841-854.

Ohno S (1970) Evolution by gene duplication. Springer, New York, USA.

Ohta T (1991) Multigene families and the evolution of complexity. *J. Mol. Evol.* **33:** 34-41.

Ohta T (2000) Evolution of gene families. *Gene* **259:** 45-52.

Okuda D, Nozaki C, Sekiya F and Morita T (2001) Comparitive biochemistry of disintegrins isolated from snake venom: consideration of the taxonomy and geographical distribution of snakes in the genus *Echis*. *J. Biochem.* **129(4):** 615-620.

Okuda D, Koike H and Morita T (2002) A new gene structure of the disintegrin family: a subunit of dimeric disintegrin has a short coding region. *Biochemistry.* **41:** 14248-14254.

Oliver JC (2008) AUGIST: inferring species trees while accommodating gene tree uncertainty. *Bioinformatics.* **24:** 2932-2933.

Olivera BM, Rivier J, Clark C, Ramilo CA, Corpuz GP, Abogadie FC, Mena EE, Woodward SR, Hillyard DR and Cruz LJ (1990) Diversity of Conus neuropeptides. *Science* **4966:** 257-263.

Omori-Satoh T and Sadahiro S (1979) Resolution of the major hemorrhagic component of *Trimeresurus flavoviridis* venom into two parts. *Biochim. Biophys. Acta.* **580:** 392-404.

Page RDM (1994) Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. *Syst. Biol.* **43:** 58-77.

Page RDM and Charleston MA (1997) Reconciled trees and incongruent gene and species trees. In: Mirkin B, McMorris FR, Roberts FS and Rzhetsky A (editors). Mathematical hierarchies in biology. American Mathematical Society, Providence, Rhode Island, USA. p. 57-71.

Page RDM (1998) GeneTree: comparing gene and species phylogenies using reconciled trees. *Bioinformatics.* **14:** 819-820.

Page RDM and Cotton JA (2000) GeneTree: a tool for exploring gene family evolution. In: Sankoff D and Nadeau JH (editors). Comparative genomics: empirical and analytical approaches to gene order dynamics, map alignment and the evolution of gene families. Kluwer, Dordrecht, The Netherlands. p. 525-536.

Pahari S, Mackessy SP and Kini RM (2007) The venom gland transcriptome of the Desert Massasauga rattlesnake (*Sistrurus catenatus edwardsii*): towards an understanding of venom composition among advanced snakes (Superfamily Colubroidea). *BMC Mol. Biol.* **8:** 115.

Paine MJ, Desmond HP, Theakston RDG and Crampton JM (1992) Gene expression in *Echis carinatus* (carpet viper) venom glands following milking. *Toxicon.* **30:** 379-386.

Park D, Kim H, Chung K, Kim D-S and Yun Y (1998) Expression and characterization of a novel plasminogen activatior from *Agkistrodon halys* venom. *Toxicon.* **36:** 1807-1819.

Parkinson J, Guiliano DB and Blaxter M (2002) Making sense of EST sequences by CLOBBing them. *BMC Bioninformatics.* **3:** 31.

Parkinson J, Anthony A, Wasmuth J, Schmid R, Hedley A and Blaxter M (2004) PartiGene – Constructing partial genomes. *Bioinformatics.* **20:** 1398-1404.

Peng M, Lu W, Beviglia V, Niewiarowski S, Kirby EP (1993) Echicetin: a snake venom protein that inhibits binding of von Willebrand factor and alboaggregins to platelet glycoprotein Ib. *Blood* **81:** 2321-2328.

Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J and Quackenbush J (2003) TIGR gene indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics.* **19:** 651-652.

Petan T, Križaj I and Pungerčar J (2007) Restoration of enzymatic activity in a Ser-49 phospholipase A2 homologue decreases its Ca2+-independant membrane-damaging activity and increases its toxicity. *Biochemistry.* **46:** 12795-12809.

Polgár J, Magnenat EM, Peitsch MC, Wells TN, Saqi MS and Clemetson KJ (1997) Amino acid sequence of the alpha subunit and computer modelling of the alpha

and beta subunits of echicetin from the venom of Echis carinatus (saw-scaled viper). *Biochem. J.* **323:** 533-537.

Pook CE, Joger U, Stümpel N and Wüster W (2009) When continents collide: phylogeny, historical biogeography and systematics of the medically important viper genus *Echis* (Squamata: Serpentes: Viperidae). *Mol. Phylogenet. Evol.***53:** 792-807.

Poran NS, Coss RG and Benjamini E (1987) Resistance of California groud squirrels (*Spermophilus beecheyi*) to the venom of the northern pacific rattlesnake (*Crotalus viridis oreganus*): A study of adaptive variation. *Toxicon.* **25:** 767-777.

Porath A, Gilon D, Schulchynska-Castel H, Shalev O, Keynan A and Benbassat J (1992) Risk indicators after envenomation in humans by *Echis coloratus* (mid-east saw scaled viper). *Toxicon* **30:** 25-32.

Posada D and Buckley TR (2004) Model selection and model averaging in phylogenetics: Advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* **53:** 793-808.

Prasad NB, Uma B, Bhat SKG and Gowda TV (1999) Comparative characterization of Russell's viper (*Daboia/vipera russelli*) venoms from different regions of Indian peninsula. *Biochim. Biophys. Acta.* **1428:** 121-136.

Pugh RN and Theakston RD (1980) Incidence and mortality on snake bite in savanna Nigeria. *Lancet.* **2:** 1181-1183.

Qu P, Du H, Wilkes DS and Yan C (2009) Critical roles of lysosomal acid lipase in T cell development and function. *Am. J. Pathol.* **174:** 944-956.

Rael ED, Knight RA and Zepeda H (1984) Electrophoretic variants of Mojave rattlesnake (*Crotalus scutulatus scutulatus*) venoms and migration differences of Mojave toxin. *Toxicon.* **22:** 980-985.

Ramasamy S, Fry BG and Hodgson WC (2005) Neurotoxic effects of venoms from seven species of Australasian black snakes (*Pseudechis*): Efficacy of black and tiger snake antivenoms. *Clinical and Experimental Pharmacology and Physiology.* **32:** 7-12.

Remigio EA and Duda TF (2008) Evolution of ecological specialization and venom of a predatory marine gastropod. *Mol. Ecol.* **17:** 1156-1162.

Ronquist F (1997) Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Syst. Biol.* **46:** 195-203.

Ronquist F and Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* **19:** 1572-1574.

Rosenfeld G, Hampe OG and Kelen EMA (1959) Coagulant and fibrinolytic activity of animal venoms determination of coagulant and fibrinolytic index of different species. *Mem. Inst. Butantan.* **29:** 143-163.

Rufini S, Cesaroni P, Desideri A, Farias R, Gubenšek F, Gutierrez JM, Luly P, Massoud R, Morero R and Pedersen JZ (1992) Calcium ion independent membrane leakage induced by phospholipase-like myotoxins. *Biochemistry.* **31:** 12424-12430.

Sadahiro S and Omori-Satoh T (1980) Lack of a hemorrhagic principle in habu snake venom, *Trimeresurus flavoviridis*, from the Okinawa Islands. *Toxicon.* **18:** 366-368.

Saint Girons H and Detrait J (1980) Communautés antigéniques des venins et systématique des Elapidae. *Bijdr. Dierk.* **50:** 96-104.

Sales PBV and Santoro ML (2008) Nucleotide and DNase activities in Brazilian snake venoms. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* **147(1):** 85-95.

Sanchez EF, Felicori LF, Chavez-Olortegui C, Magalhaes HB, Hermogenes AL, Diniz MV, Junqueira-de-Azevedo IL, Magalhaes A and Richardson M (2006) Biochemical characterization and molecular cloning of a plasminogen activator proteinase (LV-PA) from bushmaster snake venom. *Biochim. Biophys. Acta.* **1760:** 1762-1771.

Sandbank U and Djaldetti M (1966) Effect of Echis colorata venom inoculation on the nervous system of the dog and guinea pig. *Acta Neuropath.* **6:** 61-69.

Sanders KL, Lee MSY, Leys R, Foster R and Keogh JS (2008) Molecular phylogeny and divergence dates for Australasian elapids and sea snakes (hydrophiinae): evidence from seven genes for rapid evolutionary radiations. *J. Evol. Biol.* **21:** 682-695.

Sanderson MJ and McMahon MM (2007) Inferring angiosperm phylogeny from EST data with widespread gene duplication. *BMC Evol. Biol.* **7(1):** S3.

Sanz L, Gibbs HL, Mackessy SP and Calvete JJ (2006) Venom proteomes of closely related *Sistrurus* rattlesnakes with divergent diets. *J. Prot. Res.* **5:** 2098-2112.

Sasa M (1999a) Diet and snake venom evolution: can local selection alone explain intraspecific venom variation? *Toxicon* **37:** 249-252.

Sasa M (1999b) Reply. *Toxicon* **37:** 259-260.

Schaeffer RC (1987) Heterogeneity of *Echis* venoms from different sources. *Toxicon.* **25:** 1343-1346.

Segura A, Villalta M, Herrera M, León G, Harrison RA, Durfa N, Nasidi A, Calvete JJ, Theakston RDG, Warrell DA and Gutiérrez JM (2010) Preclinical assessment of the efficacy of a new antivenom (EchiTAb-Plus-ICP®) for the treatment of viper envenoming in sub-Saharan Africa. *Toxicon* **55:** 369-374.

Serrano SM, Matos MF, Mandelbaum FR and Sampaio CA (1993) Basic proteinases from *Bothrops moojeni* (caissaca) venom-I. Isolation and activity of two serine proteinases, MSP 1 and MSP 2, on synthetic substrates and on platelet aggregation. *Toxicon* **31:** 471-481.

Serrano SM, Hagiwara Y, Murayama N, Higuchi S, Mentele R, Sampaio CA, Camargo AC and Fink E (1998) Purification and characterization of a kinin-releasing and fibrinogen-clotting serine proteinase (KN-BJ) from the venom of *Bothrops jararaca*, and molecular cloning and sequence analysis of its cDNA. *Eur. J. Biochem.* **251:** 845-853.

Shashidharamurthy R, Jagadeesha DK, Girish KS and Kemparaju K (2002) Variation in biochemical and pharmacological properties of Indian cobra (*Naja naja*) venom due to geographical distribution. *Mol. Cell. Biochem.* **229:** 93-101.

Shelke RR, Sathish S and Gowda TV (2002) Isolation and characterization of a novel postsynaptic/cytotoxic neurotoxin from *Daboia russelli russelli* venom. *J. Pept. Res.* **59:** 257-263.

Shimokawa K, Jai LG, Wang XM and Fox JW (1996) Expression, activation and processing of the recombinant snake venom metalloproteinase, pro-atrolysin E. *Arch. Biochem. Biophys.* **335:** 283-294.

Siigur J and Siigur E (1992) The direct acting fibrin(ogen)olytic enzymes from snake venoms. *J. Toxicol., Toxin rev.* **9:** 91-113.

Siigur E, Tõnismägi K, Trummal K, Samel M, Vija H, Subbi J, Siigur J (2001) Factor X activator from *Vipera lebetina* snake venom, molecular characterization and substrate specificity. *Biochim. Biophys. Acta.* **1568:** 90-98.

Siigur E, Aaspõllu A, Siigur J (2003) Anticoagulant serine fibrinogenases from *Vipera lebetina* venom: structure-function relationships. *Thromb. Haemost.* **89:** 826-831.

Siigur E, Aaspõllu A, Trummal K, Tõnismägi K, Tammiste I, Kalkkinen N and Siigur J (2004) Factor X activator from *Vipera lebetina* venom is synthesized from different genes. *Biochim. Biophys. Acta.* **1702:** 41-51.

Skurk T, Lee YM and Hauner H (2001) Angiotensin II and its metabolites stimulate PAI-1 protein release from human adipocytes in primary culture. *Hypertension.* **37:** 1336-1340.

Slowinski JB, Knight A and Rooney AP (1997) Inferring species trees from gene trees: a phylogenetic analysis of the Elapidae (Serpentes) based on the amino acid sequences of venom proteins. *Mol. Phylogenet.. Evol.* **8(3):** 349-362.

Slowinski JB and Page RDM (1999) How should phylogenies be inferred from sequence data? *Syst. Biol.* **48:** 814-825.

Slowinski JB and Keogh JS (2000) Phylogenetic relationships of Elapid snakes based on Cytochrome *b* mtDNA sequences. *Mol. Phylogenet. Evol.* **15:** 157-164.

Soe S, Win MM., Htwe TT, Lwin M, Thet SS and Kyaw WW (1993) Renal histopathology following Russell's viper (*Vipera russelli*) bite. *SE Asian J. Trop. Med. Pub. Health.* **24:** 193-197.

Spawls S, Howell K, Drewes R and Ashe J (2004) A field guide to the reptiles of east Africa. A & C Black Publishers Ltd, London. pp 482-485.

Starkov VG, Osipov AV and Utkin YN (2007) Toxicity of venoms from vipers of *Pelias* group to crickets *Gryllus assimilis* and its relation to snake entomophagy. *Toxicon* **49:** 995-1001.

Sutton GG, White O, Adams MD and Kervalage AR (1995) TIGR assembler: a new tool for assembling large shotgun sequencing projects. *Gen. Sci. Technol.* **1:** 9-19.

Swofford DL (2002) PAUP*-Phylogenetic Analysis Using Parsimony (* and other methods). Beta version 4.0b10. Sinauer, Sunderland, Massachusetts, USA.

Syvanen M (1994) Horizontal gene transfer: evidence and possible consequences. *A. Rev. Genet.* **28:** 237-261.

Táborská E (1971) Intraspecies variability of the venom of *Echis carinatus*. *Physiol. Bohemoslov.* **20:** 307-318.

Táborská E and Kornalik F (1985) Individual variability of *Bothrops asper venom*. *Toxicon* **23:** 612.

Takahata N (1989) Gene genealogy in three related populations: consistency probability between gene and population trees. *Genetics.* **122:** 957-966.

Takeya H, Nishida S, Miyata T, Kawada S, Saisaka Y, Morita T and Iwanaga S (1992) Coagulation factor X activating enzyme from Russell's viper venom (RVV-X). A novel metalloproteinase with disintegrin (platelet aggregation inhibitor)-like and C-type lectin-like domains. *J. Biol. Chem.* **267:** 14109-14117.

Tamura K, Dudley J, Nei M and Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24:** 1596-1599.

Tan NH, Armugam A and Tan CS (1989) A comparative study of the enzymatic and toxic properties of venoms of the Asian lance-headed pit viper (Genus *Trimeresurus*). *Comp. Biochem. Physiol. B.* **93:** 757-762.

Tan NH and Ponnudurai G (1996) The toxinology of *Calloselasma rhodostoma* (Malayan pit viper) venom. *J. Toxicol., Toxin rev.* **15:** 1-17.

Than-Than, Hutton RA, Myint-Lwin, Khin-Ei-Han, Soe-Soe, Tin-Nu-Swe, Phillips RE and Warrell DA (1988) Hemostatic disturbances in patients bitten by Russell's viper (*Vipera russelli siamensis*) in Burma. *British J. Haematology.* **69:** 513-520.

Theakston RDG and Reid HA (1983) Development of simple standard assay procedures for the characterization of snake venoms. *Bull. World Health Organ.* **61:** 949-956.

Theakston RDG, Phillips RE, Warrell DA, Galigedera Y, Abeysekera DT, Dissanayake P, Hutton RA and Aloysius D J (1989) Failure of Indian (Haffkine) antivenom in treatment of *Vipera russelli pulchella* (Russell's viper) envenoming in Sri Lanka. *Toxicon.* **27:** 82.

Theakston RDG, Laing GD, Fielding CM, Freite Lascano A, Touzet J-M, Vallejo F, Guderian RH, Nelson SJ, Wüster W, Richards AM, Rumbea Guzman J, Warrell DA (1995) Treatment of snake bites by *Bothrops* species and *Lachesis muta* in Ecuador: laboratory screening of candidate antivenoms. *Trans. R. Soc. Trop. Med. Hyg.* **89:** 550-554.

Theopold U, Schmidt O, Söderhäll K and Dushay MS (2004) Coagulation in arthropods: defence, wound closure and healing. *Trends Immunol.* **25:** 289-294.

Thompson JD, Higgins DG and Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680.

Trape JF and Mané Y (2006) Guide des serpents d'Afrique occidentale. Savane et desert. Institut de recherché pour le développement, Paris.

Trummal K, Tõnismägi K, Siigur E, Aaspõllu A, Lopp A, Sillat T, Saat R, Kasak L, Tammiste I, Kogerman P, Kalkkinen N and Siigur J (2005) A novel metalloproteinase from *Vipera lebetina* venom induce human endothelial cell apoptosis. *Toxicon.* **46:** 46-61.

Tsai IH, Chen YH and Wang YM (2004) Comparative proteomics and subtyping of venom phospholipase A_2 and disintegrins of *Protobothrops* pit vipers. *Biochim. Biophys. Acta.* **1702:** 111-119.

Tsai IH, Tsai HY, Saha A and Gomes A (2007) Sequences, geographical variations and molecular phylogeny of venom phospholipases and threefinger toxins of eastern India *Bungarus fasciatus* and kinetic analyses of its Pro31 phospholipases A_2. *FEBS Journal.* **274:** 512-525.

Tu AT (1974) Sea snake investigation in the Gulf of Thailand. *J. Herpetol.* **8**(3): 201-210.

Turner AJ, Isaac RE and Coates D (2001) The neprilysin (NEP) family of zinc metalloendopeptidases: Genomics and function. *Bioessays* 001, **23(3):** 261-269.

van den Burgh CJ, Slotboom AJ, Verheij HM and de Haas GH (1998) The role of aspartic acid-49 in the active site of phospholipase A_2. A site-specific mutagenesis study of porcine pancreatic phospholipase A2 and the rationale of the enzymatic activity of (lysine49) phospholipase A_2 from *Agkistrodon piscivorous piscivorous'* venom. *European Journal of Biochemistry.* **176:** 353-357.

Vidal N, Delmas AS, David P, Cruaud C, Couloux A and Hedges SB (2007) The phylogeny and classification of caenophidian snakes inferred from seven nuclear protein-coding genes. *C. R. Biologies* **330:** 182-187.

Visser LE, Kyei-Faried S, Belcher DW, Geelhoed DW, Schagen van Leeuwen J and van Roosmalen J (2008) Failure of a new antivenom to treat *Echis ocellatus*

snake bite in rural Ghana: the importance of quality surveillance. *Trans. R. Soc. Trop. Med. Hyg.* **102:** 445-450.

Vonk FJ, Admiraal JF, Jackson K, Reshef R, de Bakker MAG, Vanderschoot K, van den Berge I, van Atten M, Burgerhout E, Beck A, Mirtschin PJ, Kochva E, Witte F, Fry BG, Woods AE and Richardson MK (2008) Evolutionary origin and development of snake fangs. *Nature* **454:** 630-633.

Wagstaff SC and Harrison RA (2006) Venom gland EST analysis of the saw-scaled viper, *Echis ocellatus*, reveals novel $\alpha_9\beta_1$ integrin-binding motifs in venom metalloproteinases and a new group of putative toxins, renin-like aspartic proteases. *Gene.* **377:** 21-32.

Wagstaff SC, Favreau P, Cheneval O, Laing GD, Wilkinson MC, Miller RL, Stöcklin R and Harrison RA (2008) Molecular characterisation of endogenous snake venom metalloproteinase inhibitors. *Biochem. Biophys. Res. Commun.* **365(4):** 650-656.

Wagstaff SC, Sanz L, Juárez P, Harrison RA, Calvete JJ (2009) Combined snake venomics and venom gland transcriptomic analysis of the ocellated carpet viper, *Echis ocellatus. J. Proteomics* **71(6):** 609-623.

Walker DR and Koonin EV (1997) SEALS: A system for easy analysis of lots of sequences. *Int. Sys. Mol. Biol.* **5:** 333-339.

Warrell DA and Arnett C (1976) The importance of bites by the saw scaled or carpet viper (*Echis carinatus*). Epidemiological studies in Nigeria and a review of the world literature. *Acta Tropica.* **33:** 307-341.

Warrell DA, Greenwood BM, Davidson NM, Ormerod LD and Prentice CRM (1976) Necrosis, haemorrhage and complement depletion following bites by the spitting cobra (*Naja nigricollis). Q. J. Med.* **45:** 1-22.

Warrell DA, Pope HM and Prentice CRM (1976) Disseminated intravascular coagulation caused by the carpet viper (*Echis carinatus*): Trial of Heparin. *Brit. J. Haemat.* **33:** 335-342.

Warrell DA, Davidson N McD, Greenwood BM, Ormerod LD, Pope HM, Watkins BJ and Prentice CRM (1977) Poisoning by bites of the saw-scaled viper or carpet viper (*Echis carinatus*) in Nigeria. *QJM.* **46(181):** 33-62.

Warrell DA (1985) Tropical snake bite: clinical studies in south east Asia. *Toxicon.* **23:** 543.

Warrell DA (1989) Snake venoms in science and clinical medicine. 1. Russell's viper: biology, venom and treatment of bites. *Trans. R. Soc. Trop. Med. Hyg.* **83:** 732-740.

Warrell DA, Phillips RE, Theakston DG, Galigedera Y, Abeysekera DT, Dissanayake P, Hutton RA and Aloysius DJ (1989) Neurotoxic envenoming by Indian Krait (*Bungarus caeruleus*), Cobra (*Naja naja naja*) and Russell's viper (*Vipera russelli pulchella*) in Anuradhapura. *Toxicon.* **27:** 85.

Warrell DA (1995) Clinical toxicology of snakebite in Africa, the Middle East/Arabian Peninsula and Asia. In Meier J and White J (Editors). Handbook of clinical toxicology of animal venoms and poisons. CRC Press, Boca Raton, Florida pp 433-595.

Warrell DA (1996) Animal Toxin. In Cook G (Editor). Manson's Textbook of Tropical Disease. 20th Edition. London, ELBS, WB Saunders. pp 483.

Warrell DA (2008) Unscrupulous marketing of snake bite antivenoms in Africa and Papua New Guinea: choosing the right product-'What's in a name?' *Trans. Royal Soc. Trop. Med. Hygiene* **102(5):** 397-399.

Wasmuth JD and Blaxter ML (2004) Prot4EST: translating expressed sequence tags from neglected genomes. *BMC Bioinformatics.* **5:** 187.

Waterhouse AM, Procter JB, Martin DMA, Clamp M and Barton GJ (2009) Jalview version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* **25:** 1189-1191.

Wehe A, Bansal MS, Burleigh JG and Eulenstein O (2008) DupTree: a program for large-scale phylogenetic analyses using gene tree parsimony. *Bioinformatics.* **24:** 1540-1541.

Whitaker R and Captain A (2004) Snakes of India – The field guide. Draco Books, Tamil Nadu, India.

White O and Kervalage AR (1996) TDB: new databases for biological discovery. *Methods Enzymol.* **266:** 27-40.

White J (2005) Snake venoms and coagulopathy. *Toxicon.* **45:** 951-967.

Whittington CM, Koh JMS, Warren WC, Papenfuss AT, Torres AM, Kuchel PW and Belov K (2009) Understanding and utilising mammalian venom via a platypus venom transcriptome. *J. Proteomics* **72:** 155-164.

Williams V and White J (1987) Variation in venom constituents within a single isolated population of Peninsula tiger snake (*Notechis ater niger*). *Toxicon.* **25:** 1240-1243.

World Health Organisation (www.searo.who.int) 1999. The clinical management of snakebites in the South-East Asia region. *SE. Asian J. Trop. Med. Pub. Health.* **30:** Supplement 1.

Wüster W, Golay P and Warrell DA (1997) Synopsis of recent developments in venomous snake systematics. *Toxicon* **35:** 319-340.

Wüster W, Daltry JC and Thorpe RS (1999) Can diet explain intraspecific venom variation? Reply to Sasa. *Toxicon* **37:** 253-258.

Wüster W and Broadley DG (2007) Get an eyeful of this: a new species of giant spitting cobra from eastern and north-eastern Africa (Squamata: Serpentes: Elapidae: *Naja*). *Zootaxa.* **1532:** 51-68.

Wüster W, Crookes S, Ineich I, Mané Y, Pook CE, Trape JF and Broadley DG (2007) The phylogeny of cobras inferred from mitochondrial DNA sequences: Evolution of venom spitting and the phylogeography of the African spitting cobras (Serpentes: Elapidae: *Naja nigricollis* complex). *Mol. Phylogenet. Evol.* **45:** 437-453.

Wüster W, Peppin L, Pook CE and Walker DE (2008) A nesting of vipers: Phylogeny and historical biogeography of the Viperidae (Squamata: Serpentes). *Mol. Phylogenet. Evol.* **49(2):** 445-459.

Yamada D, Sekiya F and Morita T (1996) Isolation and characterization of carinactivase, a novel prothrombin activator in *Echis carinatus* venom with a unique catalytic mechanism. *J. Biol. Chem.* **271:** 5200-5207.

Yamazaki Y and Morita T (2004) Structure and function of snake venom cysteine-rich secretory proteins. *Toxicon.* **44:** 227-231.

Yamazaki Y and Morita M (2007) Snake venom components affecting blood coagulation and the vascular system: structural similarities and marked diversity. *Current Pharm. Design* **13:** 2872-2886.

Yee DP and Conklin D (1998) Automated clustering and assembly of large EST collections. *Pro.c Int. Conf. Intell. Syst. Mol. Biol.* **6:** 203-211.

Zha XD, Huang HS, Zhou LZ, Liu J and Xu KS (2006) Thrombin-like enzymes from venom gland of *Deinagkistrodon acutus*: cDNA cloning, mechanism of

diversity and phylogenetic tree construction. *Acta Pharma. Sinica.* **27:** 184-192.

Zhang B, Liu Q, Yin W, Zhang X, Huang Y, Luo Y, Qiu P, Su X, Yu J, Hu S and Yan G (2006) Transcriptomic analysis of *Deinagkistrodon acutus* venomous gland focusing on cellular structure and functional aspects using expressed sequence tags. *BMC Genomics* **7:** 152.

Zhang J (2003) Evolution by gene duplication: an update. *Trends Ecol. Evol.* **18:** 292-298.

Zhou X, Tan TC, Valiyaveettil S, Go ML, Kini RM, Velazquez-Campoy A and Sivaraman J (2008) Structural characterization of myotoxic Ecarpholin S from *Echis carinatus* venom. *Biophysical J.* **95:** 3366-3380.

Zimmerman KD, Heatwole H and Davies HI (1992) Survival times and resistance to sea snake (*Aipysurus laevis*) venom by five species of prey fish. *Toxicon.* **30:** 259-264.

Župunski V, Kordiš D and Gubenšek F (2003) Adaptive evolution in the snake venom Kunitz/BPTI protein family. *FEBS Letters.* **547:** 131-136.

APPENDICES

Appendix I: General stock solutions and buffers

cDNA construction and qualification

5X First Strand Buffer

250mM Tris-HCl, pH8.3

375 mM KCl

15mM $MgCl_2$

5X Second Strand Buffer

100mM Tris-HCl, pH6.9

450mM KCl

23mM $MgCl_2$

0.75mM β-NAD

50mM $(NH_4)_2SO_4$

5X Adapter Buffer

330mM Tris-HCl, pH7.6

50mM $MgCl_2$

5mM ATP

TEN buffer

10mM Tris-HCl, pH7.5

0.1mM EDTA

25mM NaCl

TE buffer

10mM Tris-HCl, pH8.0

1mM EDTA

6x Slow Optical buffer

25mg bromophenol blue

4g sucrose

10ml H_2O

TAE buffer

40mM Tris-acetate, pH8.2

1mM EDTA

ELISA buffers

TBST buffer

10mM Tris-HCl, pH 8.5

150mM NaCl

1% Tween 20

Citrate buffer

525mg Citric acid

50ml H_2O

Coating buffer

1.59g Na_2CO_3

2.93 $NaHCO_3$

0.2g NaN_3

1L H_2O

Affinity purification buffers

10X PBS

80g NaCl

2g KCl

14.4g Na_2HPO_4

2.4g KH_2PO_4

1L ddH_20

pH 7.4

Column washing buffer

100mM NaH_2PO_4, pH 7.5

500mM NaCl

Column elution buffer

100mM glycine, pH 2.5

100mM HCl

SDS-PAGE and Western Blotting buffers

Reducing SDS-PAGE sample buffer

62.5 mM Tris-HCL, pH 6.8

10% glycerol

2% SDS

0.01 mg/ml bromophenol blue

15% β-mecaptoethanol

Native-PAGE sample buffer

5mM Tris-Cl, pH 6.8

33% glycerol

5X TGS SDS-PAGE running buffer

151g Tris

720g glycine

50g sodium dodecyl sulphate (SDS)

10L H_2O

pH8.3

Transfer buffer

2.03g Tris

14.26g glycine

800ml H_2O

200ml methanol

SDS-PAGE gels

15% Resolving gel

3.75ml H_2O

2.5ml 1.5M Tris-SDS, pH8.8

3.75ml 40%bis-acrylamide

100µl 10% sodium dodecyl sulphate (SDS)

60µl 10% ammonium persulfate (APS)

7µl tetramethylethylenediamine (TEMED)

Stacking gel

2.5ml H_2O

1ml 500mM Tris-SDS, pH6.8

350µl 40%bis-acrylamide

30µl 10%APS

5µl TEMED

Appendix II: *Echis* transcriptomics

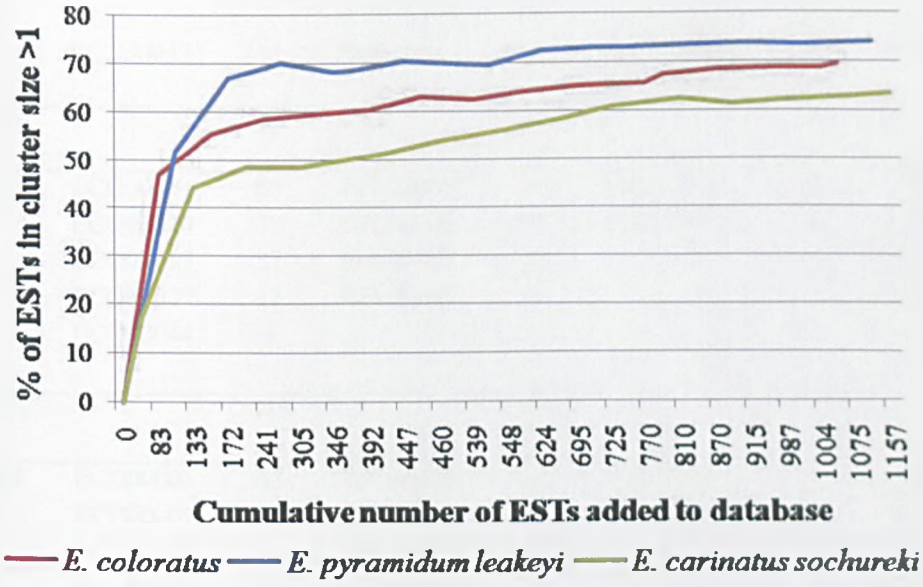

Figure 1. An overview of clustering processes for three species of the genus *Echis*. The graph demonstrates the percentage of ESTs that are added to clusters (ESTs >1) as the cumulative number of ESTs entering the database increase. In all species the number of ESTs affecting the proportion of EST clusters and singletons reaches a plateau after 800 sequences.

Venom toxin family	*E. coloratus*		*E. p. leakeyi*		*E. ocellatus*		*E. c. sochureki*	
	Cluster ID	ESTs/ cluster	Cluster ID	ESTs/ cluster	Cluster ID	ESTs/ cluster	Cluster ID	ESTs/ cluster
SVMP								
Class PI	ECO00047	10	None	-	EOC00028	21	None	-
					EOC00004	4		
Total ESTs		10		0		25		0
Class PII	ECO00011	60	EPL00005	134	EOC00006	20	ECS00117	20
	ECO00020	38	EPL00006	91	EOC00071	12	ECS00012_2	19
	ECO00017	27	EPL00056	10			ECS00114	11
	ECO00027	17	EPL00097	9			ECS00253	8
	ECO00044	4					ECS00059	3
							ECS00086	3
Total ESTs		146		244		32		64
Class PIII	ECO00002	42	EPL00008	25	EOC00063	22	ECS00012_1	52
	ECO00007	26	EPL00004	22	EOC00013	11	ECS00053	42
	ECO00023	22	EPL00002	12	EOC00001	9	ECS00031	30
	ECO00012	20	EPL00090	6	EOC00089	9	ECS00062	19
	ECO00010	18	EPL00040	5	EOC00008	6	ECS00257	11
	ECO00009	16	EPL00029	4	EOC00081	6	ECS00071	9
	ECO00067	14	EPL00061	4	EOC00086	5	ECS00003	6
	ECO00050	9	EPL00125	4	EOC00095	5	ECS00030	6
	ECO00001	7	EPL00019	3	EOC00186	4	ECS00044	6
	ECO00034	7	EPL00032	3	EOC00073	3	ECS00056	4
	ECO00004	6	EPL00044	3	EOC00016	3	ECS00163	4
	ECO00106	5	EPL00055	3	EOC00404	3	ECS00177	4
	ECO00275	5	EPL00103	3	EOC00016	3	ECS00043	3
	ECO00076	4	EPL00159	2	EOC00404	3	ECS00120	3
	ECO00406	3	EPL00396	2			ECS00251	3
	ECO00146	2					ECS00213	2
	ECO00192	2					ECS00456	2
	ECO00222	2					ECS00497	2
							ECS00678	2
Total ESTs		210		101		84		210
Class PIV	ECO00144	7	None	-	EOC00024	55	ECS00087	10
	ECO00061	2			EOC00022	17		
	ECO00075	2						
Total ESTs		11		0		72		10
ND and singletons		28		33		27		29
DIS	ECO00024	36	Singletons	1	None	-	ECS00035	20
	Singletons	1					ECS00036	19
							Singletons	1
Total ESTs		37		1		0		40

Venom toxin family	*E. coloratus*		*E. p. leakeyi*		*E. ocellatus*		*E. c. sochureki*	
	Cluster ID	ESTs/ cluster	Cluster ID	ESTs/ cluster	Cluster ID	ESTs/ cluster	Cluster ID	ESTs/ cluster
CTL	ECO00038	19	EPL00010	39	EOC00124	6	ECS00050	14
	ECO00127	10	EPL00066	27	EOC00125	3	ECS00102	13
	ECO00069	5	EPL00016	21	EOC00133	3	ECS00154	11
	ECO00108	5	EPL00038	16	EOC00334	3	ECS00230	10
	ECO00070	4	EPL00031	13	EOC00083	2	ECS00098	8
	ECO00052	3	EPL00030	9	EOC00092	2	ECS00006	7
	ECO00041	2	EPL00053	9	Singletons	18	ECS00045	7
	ECO00115	2	EPL00109	8			ECS00038	6
	ECO00153	2	EPL00034	6			ECS00140	3
	ECO00158	2	EPL00112	6			ECS00051	2
	ECO00197	2	EPL00081	5			ECS00346	2
	ECO00270	2	EPL00018	3			Singletons	8
	Singletons	10	EPL00127	3				
			EPL00060	2				
			EPL00078	2				
			EPL00282	2				
			Singletons	11				
Total ESTs		68		182		37		91
PLA_2								
Asp^{49}	ECO00086	11	EPL00071	51	EOC00079	10	ECS00002	17
	ECO00186	3	EPL00001	33				
			EPL00204	2				
Ser^{49}	ECO00035	21	EPL00012	52	EOC00015	15	ECS00014	23
			EPL00195	11				
ND	None	-	EPL00274	3	Singletons	4	Singletons	3
			Singletons	4				
Total ESTs		35		156		29		43
SP	ECO00285	4	EPL00089	6	EOC00049	5	ECS00244	11
	ECO00013	3	EPL00098	2	Singletons	3	ECS00134	5
	ECO00112	3	EPL00435	2			ECS00186	4
	ECO00117	2	Singletons	5			ECS00105	3
	ECO00119	2					Singletons	2
	ECO00135	2						
	ECO00164	2						
	ECO00182	2						
	ECO00419	2						
	Singletons	7						
Total ESTs		29		15		8		25
LAO	ECO00026	24	EPL00025	19	EOC00167	2	ECS00178	4
	Singletons	2	Singletons	1	EOC00233	2	ECS00061	2
Total ESTs		26		20		4		6
CRISP	ECO00025	33	None	-	Singletons	1	ECS00093	4
	Singletons	2					ECS00169	6
Total ESTs		35		0		1		10

Venom toxin family	*E. coloratus*		*E. p. leakeyi*		*E. ocellatus*		*E. c. sochureki*	
	Cluster ID	ESTs/ cluster	Cluster ID	ESTs/ cluster	Cluster ID	ESTs/ cluster	Cluster ID	ESTs/ cluster
VEGF	ECO00199	2	EPL00139	2	EOC00176 EOC00478	6 2	ECS00431	2
NGF	ECO00049	2	EPL00043	2	Singletons	1	Singletons	1
PEPT								
AP	Singletons	1	None	-	None	-	ECS00179 Singletons	7 1
DPP	Singletons	**1**	None	-	None	-	None	-
NEP	None	-	Singletons	**1**	None	-	None	-
PE								
PHOS	ECO00241	2	None	-	None	-	ECS00101 Singletons	2 1
5'-NUC	ECO00276 Singletons	2 1	Singletons	2	None	-	Singletons	1
E-NTPase	ECO00014	2	None	-	None	-	None	-
LAL	ECO00073 Singletons	**13** **1**	None	-	None	-	None	-
RLAP	None	-	None	-	EOC00051 EOC00123 Singletons	**10** **4** **3**	None	-
HYAL	None	-	None	-	Singletons	1	Singletons	1
KTZ	None	-	None	-	None	-	Singletons	1

Table 1. Catalogue of venom toxin encoding ESTs determined from the *Echis* vgDbESTs. Putative novel venom toxins are in bold and underlined. Key – SVMP: snake venom metalloproteinases; PI, PII, PIII, PIV: respective sub-group of SVMPs; ND: sub-class not determined; DIS: short coding disintegrins; CTL: C-type lectins; PLA_2: group II phospholipases A_2; SP: serine proteases; LAO: L-amino oxidases; CRISP: cysteine-rich secretory proteins; VEGF: vascular endothelial growth factors; NGF: nerve growth factors; PEPT: peptidases; AP: aminopeptidase; DPP: dipeptidyl peptidase III; NEP: neprilysin; PE: Purine liberators; PHOS: phosphodiesterase; 5'-NUC: 5'-nucleotidase; E-NTPase: ectonucleoside triphosphate diphosphohydrolase; LAL: lysosomal acid lipases; RLAP: renin-like aspartic proteases; HYAL: hyaluronidases; KTZ: kunitz-type protease inhibitors.

	E. coloratus			*E. p. leakeyi*			*E. c. sochureki*		
	No. of clusters	No. of ESTs	% of ESTs	No. of clusters	No. of ESTs	% of ESTs	No. of clusters	No. of ESTs	% of ESTs
Clusters >1									
- Toxin	62	612	57.20	50	717	66.51	54	502	43.39
- Non-toxin	36	135	12.62	18	84	7.79	39	209	18.06
- Unidentified	2	5	0.47	4	8	0.74	7	26	2.25
Singletons									
- Toxin	-	50	4.67	-	42	3.90	-	42	3.63
- Non-toxin	-	196	18.31	-	121	11.23	-	182	15.73
- Unidentified	-	72	6.73	-	106	9.83	-	196	16.94
Totals	100	1070	100	72	1078	100	100	1157	100

Table 2. Summary statistics following clustering and assembly of ESTs for *E. coloratus*, *E. p. leakeyi* and *E. c. sochureki*.

Appendix III: Dietary venom adaptations

	C-type lectin (CTL)		Phospholipase A_2 (PLA_2)	
	cDNA clone	GenBank accession	cDNA clone	GenBank accession
E. ocellatus	08G09	DW361405	05G03	DW361138
	02D05	DW360904	09G11	DW361491
	06B06	DW361283		
	07E01	DW361344		
	08F12	DW361413		
	03F03	DW360973		
	04D06	DW361082		
	05D01	DW361174		
	01H11	DW360768		
	02A03	DW360938		
	06H03	DW361219		
	09C01	DW361542		
	04E02	DW361075		
	04H12	DW361032		
	06D12	DW361255		
	07H03	DW361307		
	10C11	DW361620		
	08D01	DW361446		
	09A05	DW361562		
	01A10	DW360846		
	03G05	DW360959		
	02C06	DW360914		
	10F06	DW361593		
	10C09	DW361622		
	03G12	DW360952		
E. coloratus	04H11	GR947907	07F07	GR948302
	07A07	GR948156	09D06	GR947989
	06D10	GR948676	05H07	GR948205
	01A12	GR948183	02B02	GR948826
	04G05	GR948404	01C02	GR948641
	10B04	GR948638	04H07	GR948576
	13C08	GR948707		
	12G08	GR948311		
	03D01	GR948286		
	10H02	GR948540		
	11C05	GR948706		
	09D05	GR948587		
	07A09	GR948870		
	05F04	GR948817		
	14H10	GR948791		
	11G01	GR948147		
	03B08	GR948708		
	07G07	GR948762		
	05G06	GR948255		

	12F04	GR948483		
	06E11	GR948242		
	03F11	GR948161		
	11A11	GR948610		
	09G01	GR948562		
	11A07	GR948238		
	15D05	GR948376		
E. p. leakeyi	10E03	GR950261	09G08	GR950543
	04A06	GR950229	10H07	GR950962
	09D11	GR951100	10A05	GR950978
	01G07	GR951065	14B06	GR951085
	12F09	GR950961	07H08	GR950945
	05H10	GR950707	09E03	GR950437
	09G12	GR950525	08B04	GR950487
	06D08	GR950452	14E11	GR950442
	03E03	GR951078	14D03	GR950539
	14E03	GR950415	14D07	GR950571
	08E12	GR950356	01F05	GR950541
	04F11	GR950467	10H09	GR951187
	11H03	GR950383	14D04	GR951122
	09D12	GR950241	14G07	GR950535
	10D02	GR950187	09D09	GR950604
	10A04	GR950545	13F09	GR951054
	02H06	GR950176	10D09	GR950239
	02G06	GR950875	06B10	GR950630
	08C11	GR950408		
	12C06	GR950482		
	07E07	GR950472		
	01C07	GR950654		
	13C01	GR950984		
	06A03	GR950868		
	14F02	GR950787		
	10H03	GR951115		
	05A07	GR950998		
	05E10	GR950195		
	04F01	GR950497		
	01A08	GR950210		
	13B09	GR950274		
	11H08	GR950485		
	05C03	GR951174		
	01E03	GR950748		
	14F10	GR950562		
	07B12	GR950953		
	04E11	GR950389		
	08F01	GR950351		
	08H03	GR951062		
	02G01	GR950370		

	01D03	GR950367		
E. c. sochureki	05H05	GR949149	07B02	GR949587
	07G07	GR949807	02C03	GR949814
	03H12	GR949000	13C04	GR949536
	02A07	GR949655	03F06	GR949475
	08G02	GR949802	01H04	GR949348
	03C11	GR949941	05A08	GR949216
	01G03	GR949133	01G06	GR949164
	03A09	GR949711		
	06F06	GR949977		
	01H06	GR949992		
	11A04	GR949492		
	01A08	GR949041		
	11B08	GR949809		
	05F07	GR949094		
	04D10	GR949810		
	14H07	GR949902		
	12F05	GR949908		
	07F10	GR949688		
	10F02	GR949760		
	04C06	GR949132		
	05F09	GR950086		
	05H06	GR949929		
	08C12	GR949269		
	03G12	GR949137		

Table 1. GenBank accession numbers for CTL and PLA_2 sequences from four members of the genus *Echis*.

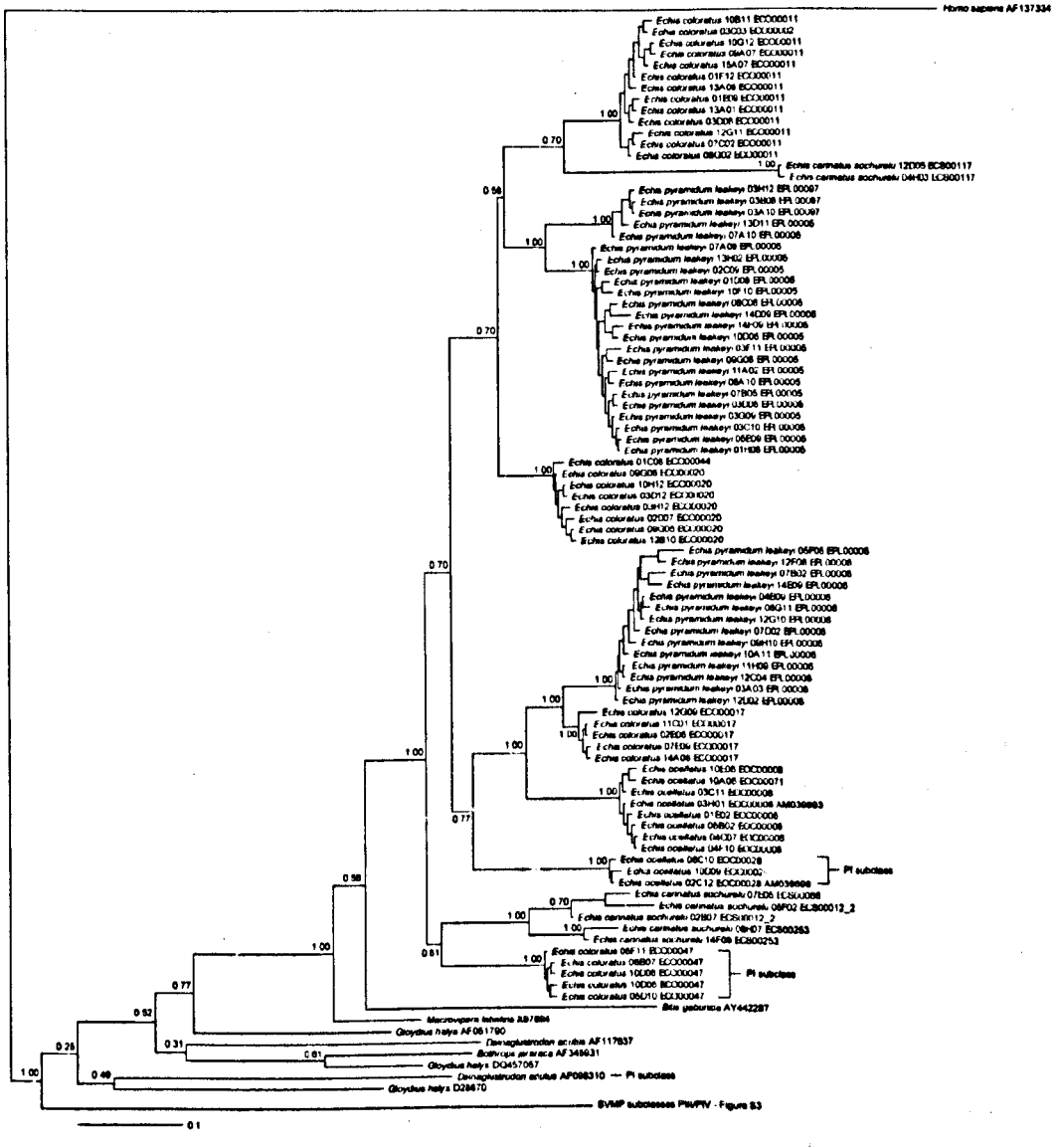

Figure 1. Bayesian snake venom metalloproteinase PI/PII amino acid gene tree. PI sub-class was identified by the absence of disintegrin or disintegrin-like domains extending the metalloproteinase domain (Fox and Serrano, 2005). Non-*Echis* sequences are labelled with corresponding UniProt or GenBank accession numbers. Outgroup sequence is *Homo sapiens* [AF137334].

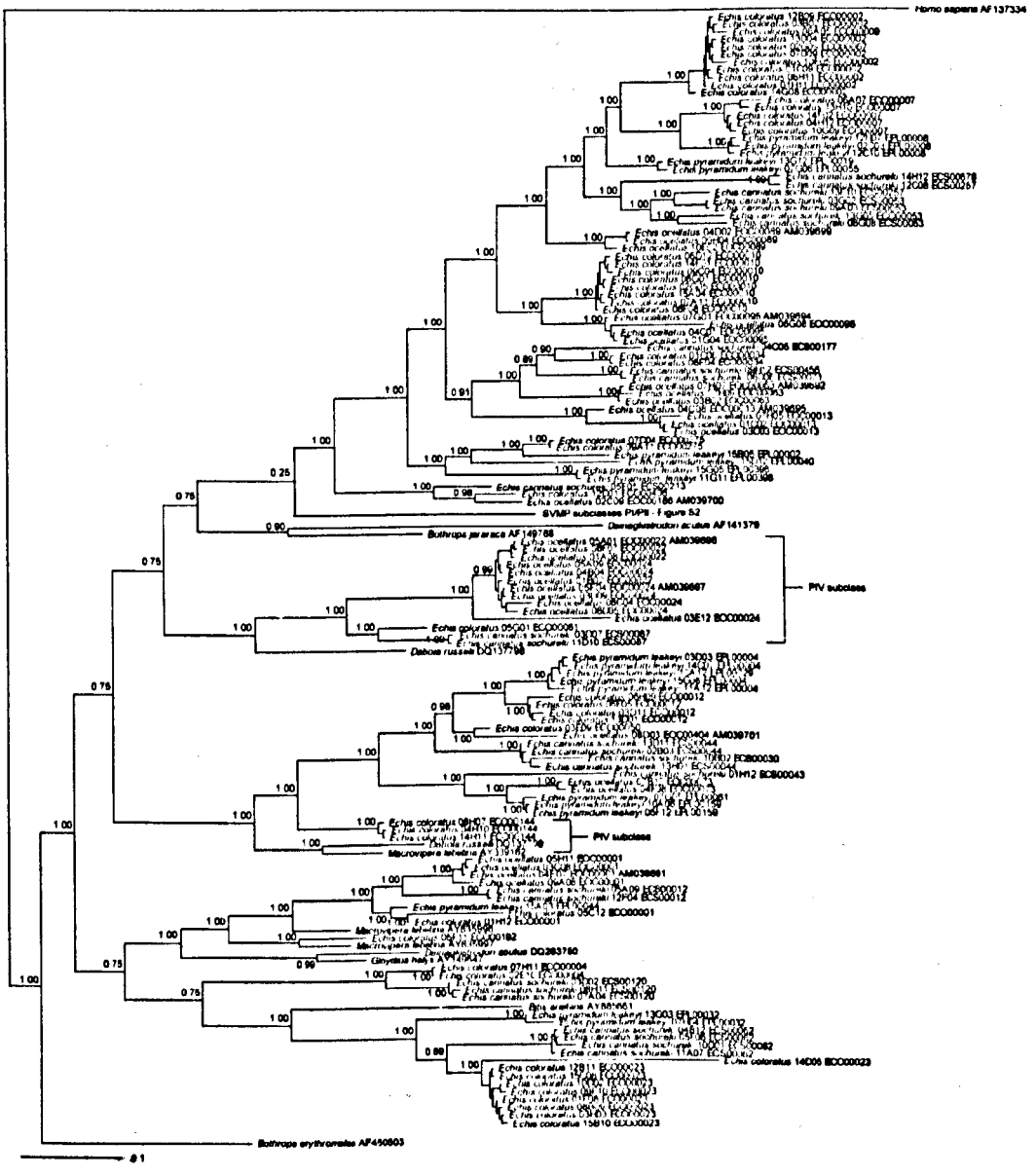

Figure 2. Bayesian snake venom metalloproteinase PIII/PIV amino acid gene tree. PIV sub-class was identified by the presence of an additional cysteine residue in the cysteine-rich domain at positions 397 or 400 (Fox and Serrano, 2005; Wagstaff *et al.* 2009 – numbering from Fox and Serrano, 2005). Non-*Echis* sequences are labelled with corresponding UniProt or GenBank accession numbers. Outgroup sequence is *Homo sapiens* [AF137334].

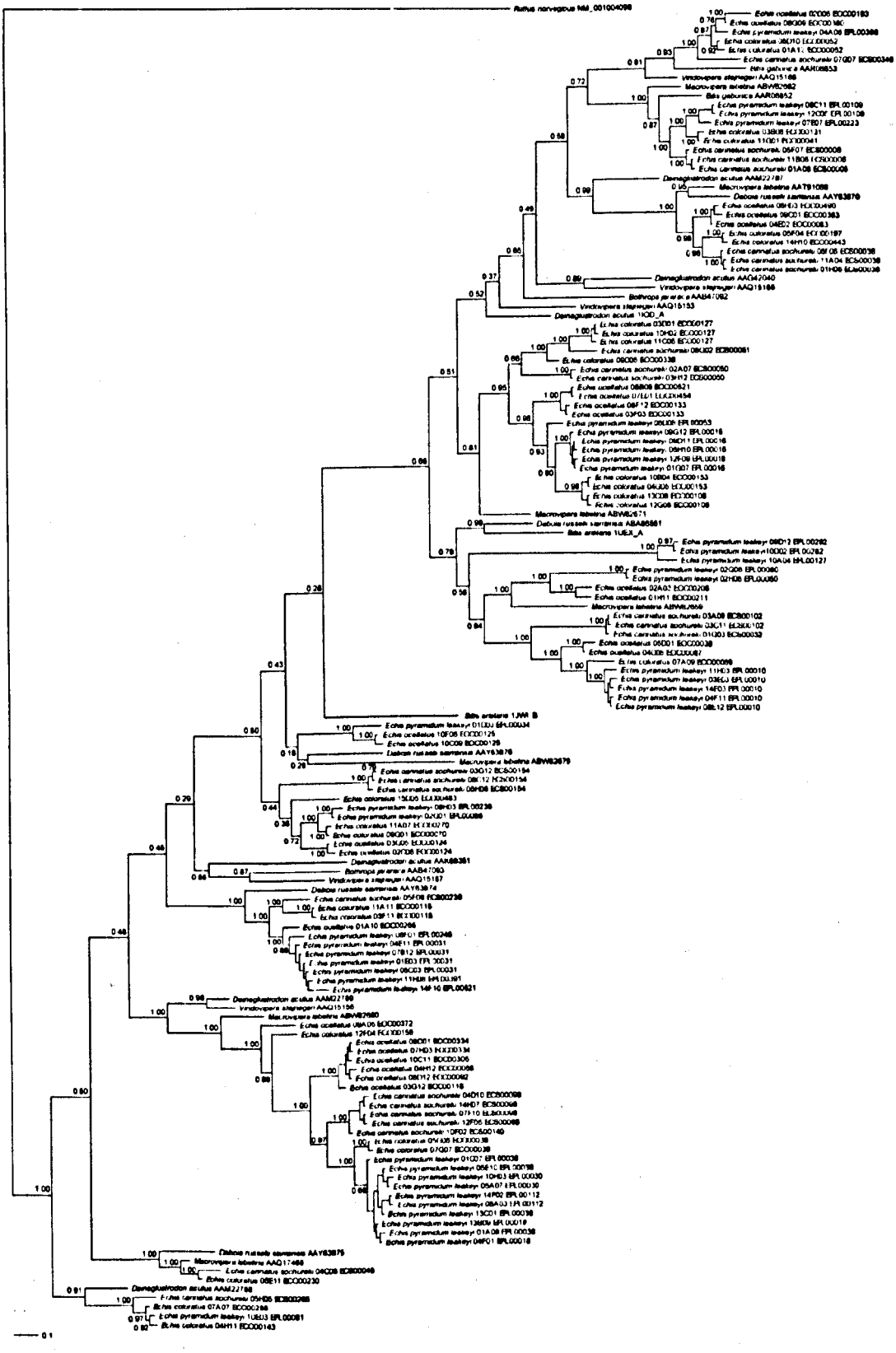

Figure 3. Bayesian C-type lectin amino acid gene tree. Non-*Echis* sequences are labelled with corresponding UniProt or GenBank accession numbers. Outgroup sequence is *Rattus norvegicus* [NM_001004096].

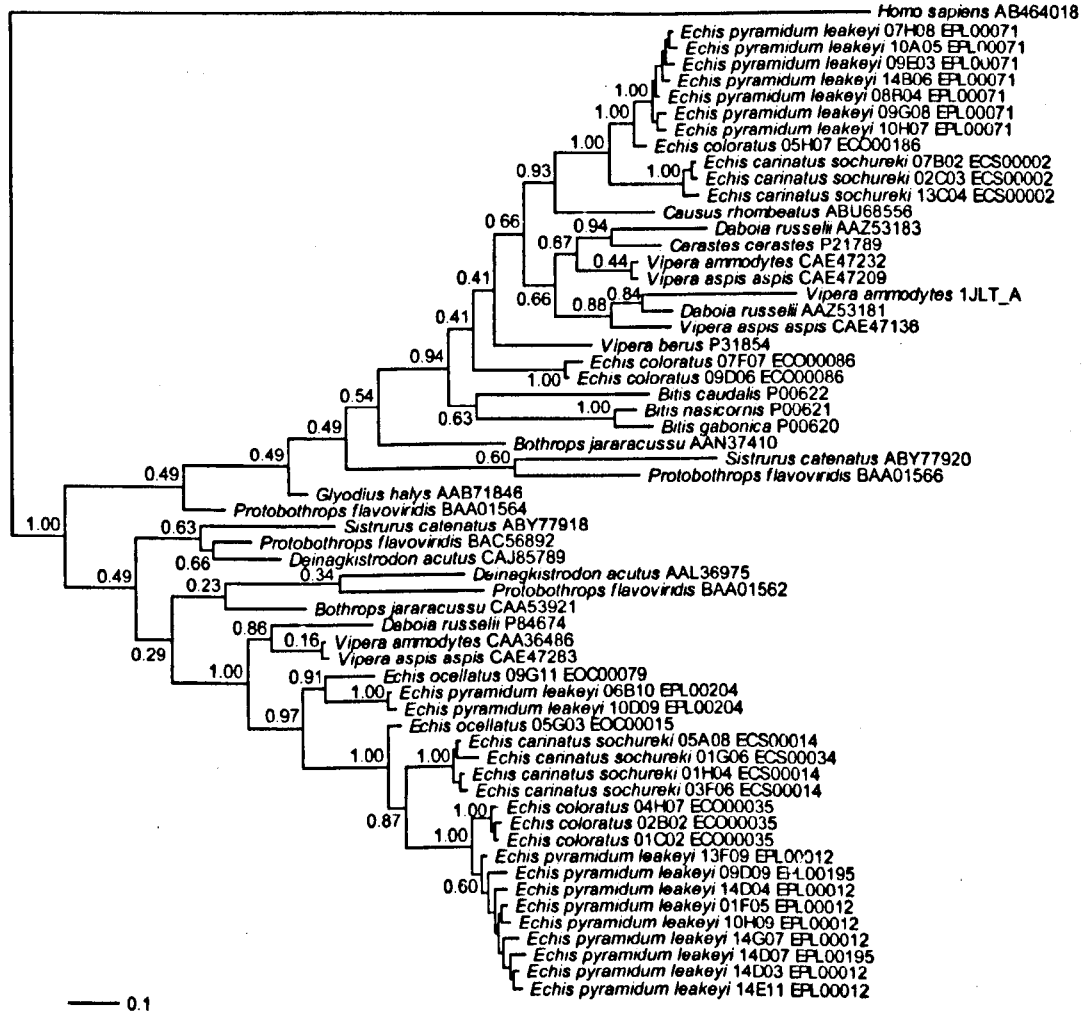

Figure 4. Bayesian phospholipase A_2 amino acid gene tree. Non-*Echis* sequences are labelled with corresponding UniProt or GenBank accession numbers. Outgroup sequence is *Homo sapiens* [AB464018].

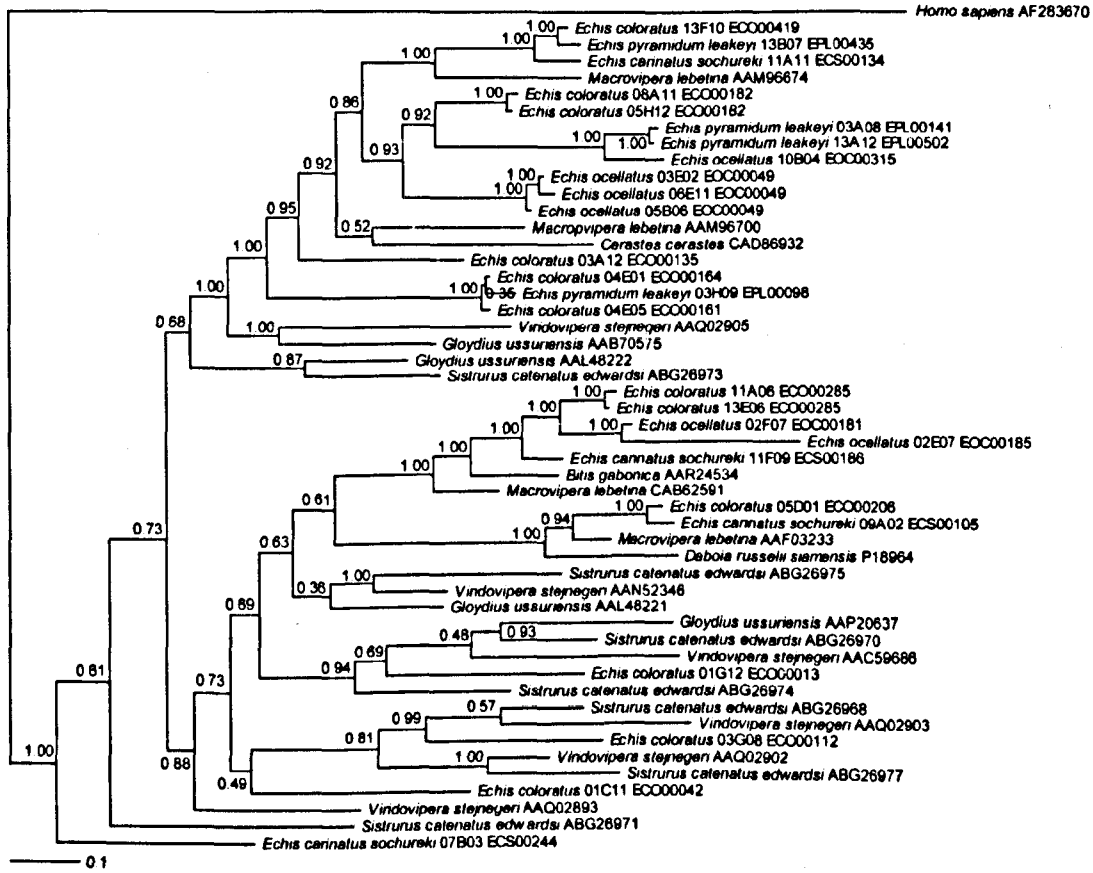

Figure 5. Bayesian serine protease amino acid gene tree. Non-*Echis* sequences are labelled with corresponding UniProt or GenBank accession numbers. Outgroup sequence is *Homo sapiens* [AF283670].

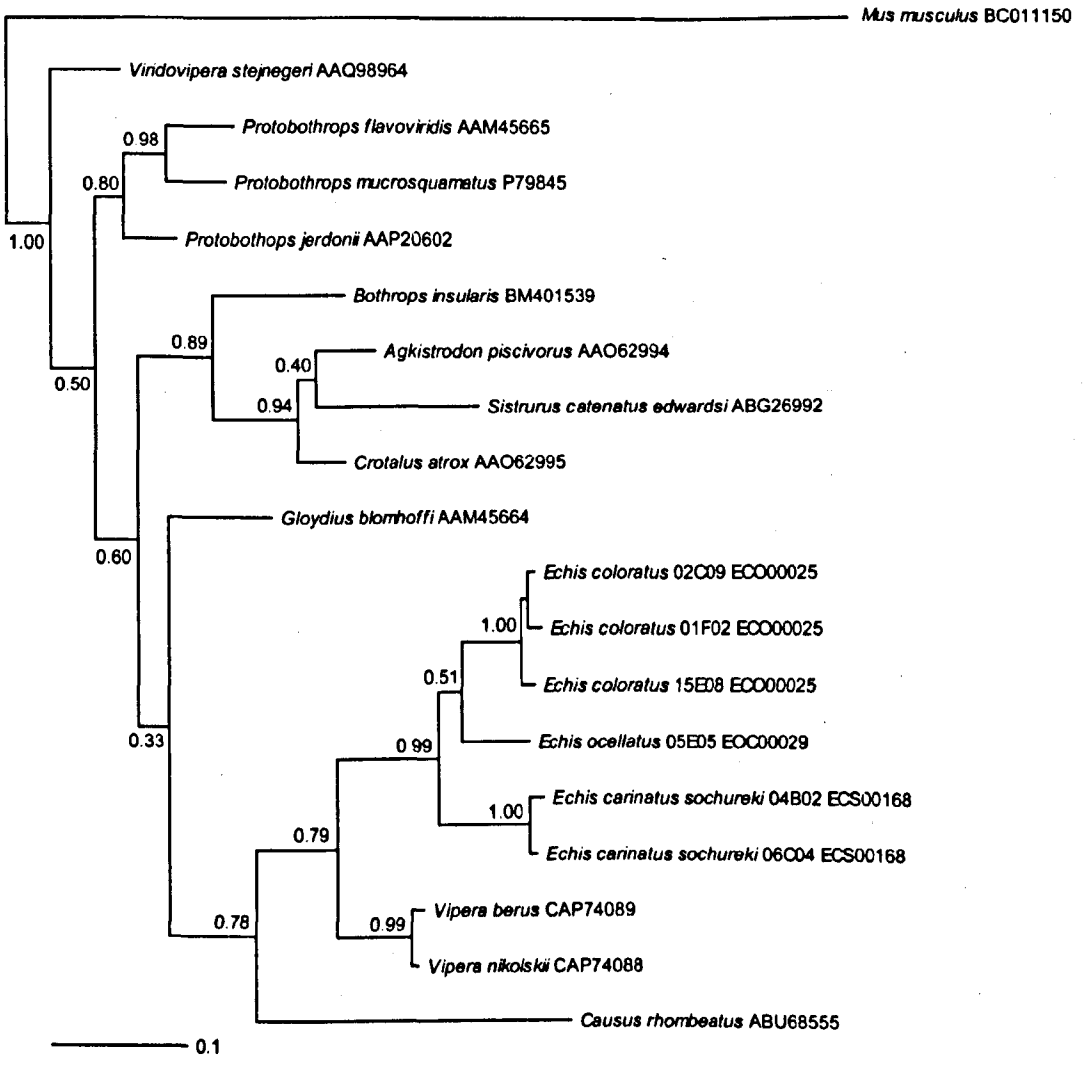

Figure 6. Bayesian cysteine-rich secretory protein amino acid gene tree. Non-*Echis* sequences are labelled with corresponding UniProt or GenBank accession numbers. Outgroup sequence is *Mus musculus* [BC011150].

Appendix IV: Venom gene tree parsimony

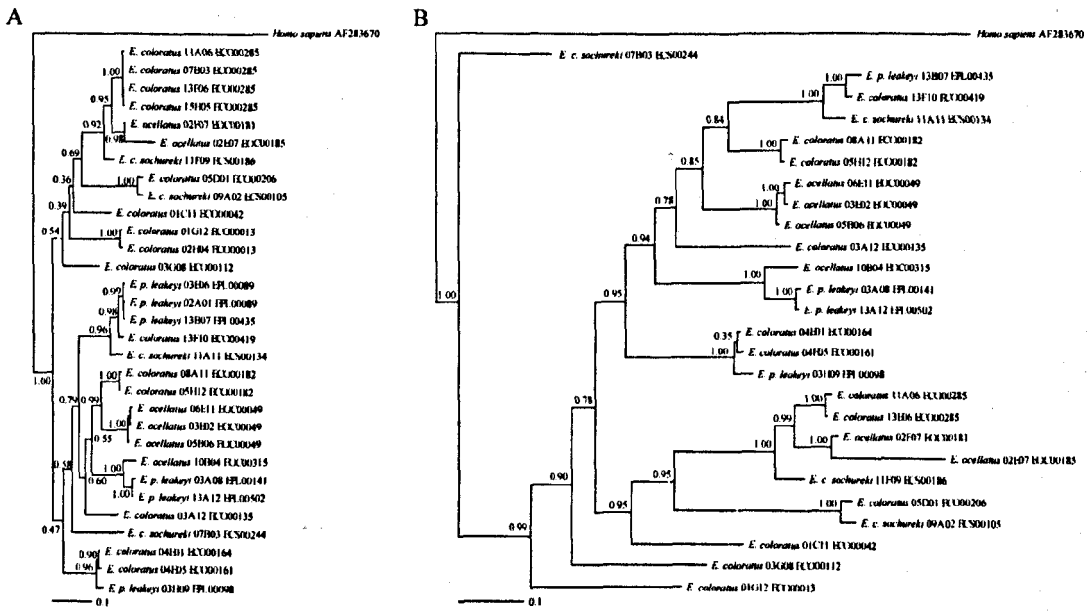

Figure 1. Bayesian serine protease gene trees for four members of the genus *Echis*. A: DNA. B: amino acid.

Appendix V: Venom neutralisation by EchiTabG®

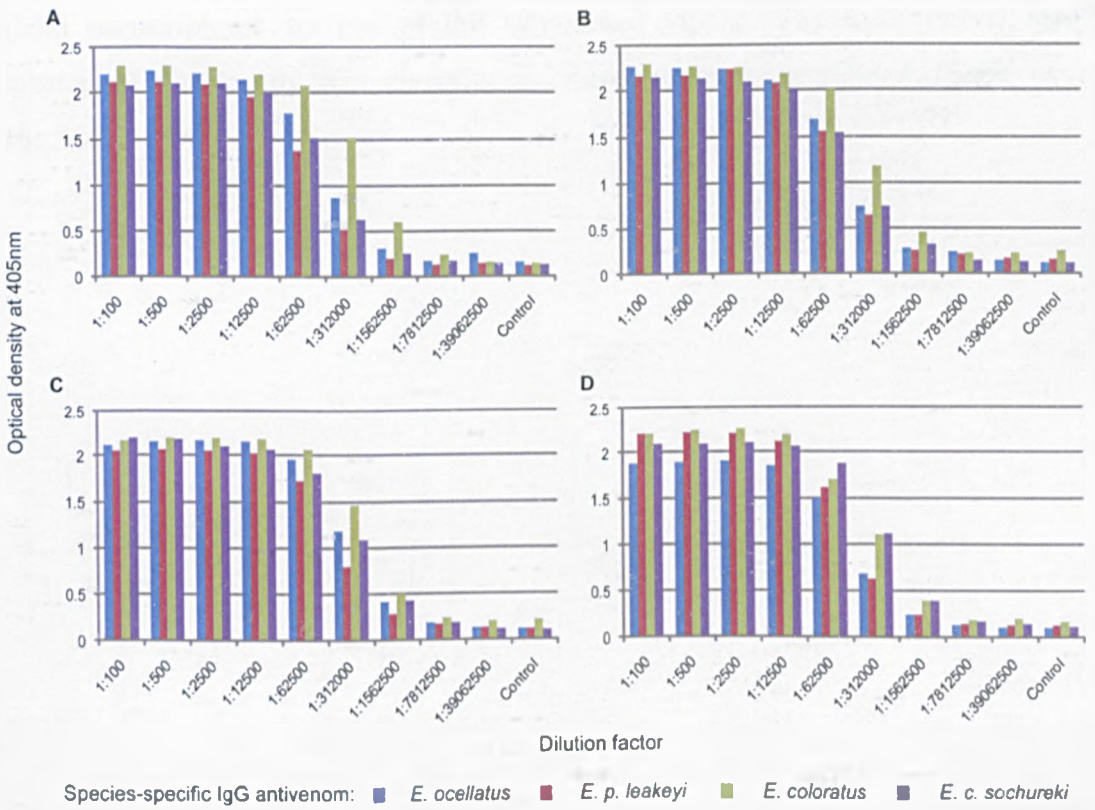

Figure 1. Bar charts demonstrating the ELISA titres of four *Echis* species-specific antivenoms against four *Echis* venoms - A) *E. ocellatus*, B) *E. p. leakeyi*, C) *E. coloratus* and D) *E. c. sochureki*.

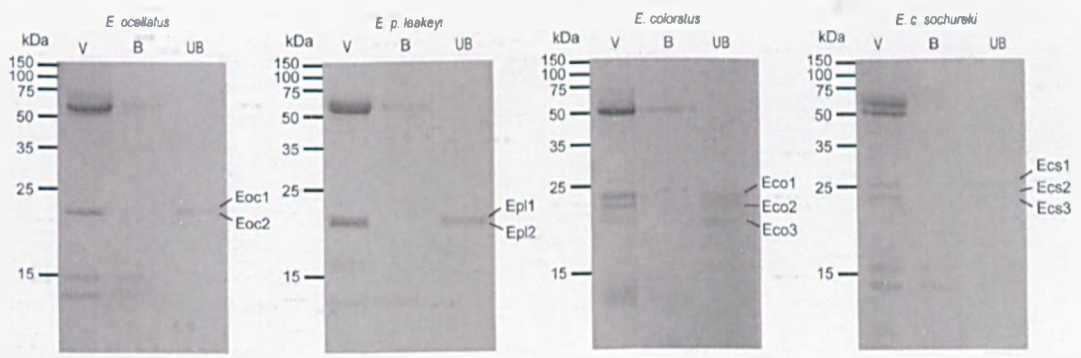

Figure 2. Reduced SDS-PAGE profiles of four *Echis* venoms and their respective bound (B) and unbound (UB) fractions following affinity purification with the *E. ocellatus* antivenom EchiTabG®. Species number identifiers indicate the unbound bands for each species that were excised for protein identification.

Appendix VI: *Echis* transcriptomics published manuscript

Casewell NR, Harrison RA, Wüster W and Wagstaff SC (2009) Comparative venom gland transcriptome surveys of the saw-scaled vipers (Viperidae: *Echis*) reveal substantial intra-family gene diversity and novel venom transcripts. *BMC Genomics* **10:** 564.

BMC Genomics

BioMed Central

Research article

Open Access

Comparative venom gland transcriptome surveys of the saw-scaled vipers (Viperidae: *Echis*) reveal substantial intra-family gene diversity and novel venom transcripts

Nicholas R Casewell*[1], Robert A Harrison[2], Wolfgang Wüster[1] and Simon C Wagstaff*[2]

Address: [1]School of Biological Sciences, Bangor University, Environment Centre Wales, Bangor, UK and [2]Alistair Reid Venom Research Unit, Liverpool School of Tropical Medicine, Liverpool, UK

Email: Nicholas R Casewell* - n.r.casewell@bangor.ac.uk; Robert A Harrison - r.harrison@liverpool.ac.uk; Wolfgang Wüster - w.wuster@bangor.ac.uk; Simon C Wagstaff* - simonw@liv.ac.uk

* Corresponding authors

Published: 30 November 2009

BMC Genomics 2009, 10:564 doi:10.1186/1471-2164-10-564

Received: 14 August 2009
Accepted: 30 November 2009

This article is available from: http://www.biomedcentral.com/1471-2164/10/564

Abstract

Background: Venom variation occurs at all taxonomical levels and can impact significantly upon the clinical manifestations and efficacy of antivenom therapy following snakebite. Variation in snake venom composition is thought to be subject to strong natural selection as a result of adaptation towards specific diets. Members of the medically important genus *Echis* exhibit considerable variation in venom composition, which has been demonstrated to co-evolve with evolutionary shifts in diet. We adopt a venom gland transcriptome approach in order to investigate the diversity of toxins in the genus and elucidate the mechanisms which result in prey-specific adaptations of venom composition.

Results: Venom gland transcriptomes were created for *E. pyramidum leakeyi*, *E. coloratus* and *E. carinatus sochureki* by sequencing ~1000 expressed sequence tags from venom gland cDNA libraries. A standardised methodology allowed a comprehensive intra-genus comparison of the venom gland profiles to be undertaken, including the previously described *E. ocellatus* transcriptome. Blast annotation revealed the presence of snake venom metalloproteinases, C-type lectins, group II phopholipases A_2, serine proteases, L-amino oxidases and growth factors in all transcriptomes throughout the genus. Transcripts encoding disintegrins, cysteine-rich secretory proteins and hyaluronidases were obtained from at least one, but not all, species. A representative group of novel venom transcripts exhibiting similarity to lysosomal acid lipase were identified from the *E. coloratus* transcriptome, whilst novel metallopeptidases exhibiting similarity to neprilysin and dipeptidyl peptidase III were identified from *E. p. leakeyi* and *E. coloratus* respectively.

Conclusion: The comparison of *Echis* venom gland transcriptomes revealed substantial intrageneric venom variation in representations and cluster numbers of the most abundant venom toxin families. The expression profiles of established toxin groups exhibit little obvious association with venom-related adaptations to diet described from this genus. We suggest therefore that alterations in isoform diversity or transcript expression levels within the major venom protein families are likely to be responsible for prey specificity, rather than differences in the representation of entire toxin families or the recruitment of novel toxin families, although the recruitment of lysosomal acid lipase as a response to vertebrate feeding cannot be excluded. Evidence of marked intrageneric venom variation within the medically important genus *Echis* strongly advocates further investigations into the medical significance of venom variation in this genus and its impact upon antivenom therapy.

Background

Snake venoms contain a complex mix of components, with biologically active proteins and peptides comprising the vast majority [1]. Variation in the composition of venom occurs at several taxonomical levels in multiple snake lineages [reviewed in [2,3]]. The view that variation in venom composition evolves primarily through neutral evolutionary processes [4-6] is not supported by other reports that snake venom composition is subject to strong natural selection as a result of adaptation towards specific diets [e.g. [7-10]]. Since the primary role of venom is to aid prey capture [2], it is perhaps unsurprising that variation in the protein composition of venom has been associated with significant dietary shifts in a number of genera [9-12]. Irrespective of the evolutionary forces underpinning venom protein composition, variation in venom components can significantly impact upon the clinical manifestations of snake envenoming [13-15] and, because the clinical efficacy of an antivenom may be largely restricted to the venom used in its manufacture, the success of antivenom therapy [16-18].

Envenoming by saw-scaled viper (Viperidae: *Echis*) species is thought to be responsible for more snakebite deaths worldwide than any other snake genus [19]. Envenomed victims typically suffer a combination of systemic and local haemorrhagic symptomatologies and up to 20% mortality rates without antivenom treatment [19-21]. Whilst the clinical symptoms are largely consistent throughout this widely distributed genus [20], cases of incomplete intrageneric antivenom efficacy have been documented, implying substantial inter-species venom variation [18,22-24]. We demonstrated that the four species complexes making up this genus, the *E. carinatus*, *E. ocellatus*, *E. pyramidum* and *E., coloratus* species groups [10,25], exhibit considerable vertebrate or invertebrate dietary preferences, *E. coloratus* being a vertebrate specialist whereas invertebrates feature prominently in the diet of the others. Since the proportions of consumed invertebrates correlated strongly with alterations in venom toxicity to scorpions, we believe the toxicity of the venom from these species to have co-evolved alongside evolutionary shifts in diet [10]. A preliminary venom protein analysis using reduced SDS-PAGE failed to identify an obvious link between venom composition and diet [10], justifying the use of a more comprehensive venom composition analysis in order to elucidate the mechanisms driving venom adaptations within the *Echis* viper genus.

Based on our earlier work with *E. ocellatus* [26], a comparative venom gland transcriptome approach was elected and we generated venom gland cDNA libraries from *E. coloratus*, *E. pyramidum leakeyi* and *E. carinatus sochureki*. Together with the existing *E. ocellatus* database, these provided DNA sequence data representing the venom gland transcriptomes for each of the four major species groups within the genus. The production of multiple *Echis* venom gland expressed sequence tag databases (vgDbEST) provides an unbiased overview of the transcriptional activity during venom synthesis in the venom glands of four species in this genus. This, the first comprehensive compilation of venom gland transcriptomes of congeneric snake species, was then interrogated to determine whether the mechanisms resulting in prey-specific adaptation of venom composition involve (i) the recruitment of novel prey-specific venom toxin transcripts, (ii) major changes in the expression levels of established toxin families, (iii) the diversification of functional isoforms within established toxin families or (iv) a combination of these factors.

Results

EST data provides a powerful insight into the transcriptional activity of a tissue at a particular time point. Our protocols for the generation of venom gland EST databases provide a snapshot of transcriptional activity in the venom gland 3 days after venom expulsion, when transcription peaks [27] in preparation for new venom synthesis. Although each individual venom transcript cannot be correlated with the mature venom proteome without considerable extra experimental verification, our own work with *E. ocellatus* [28] shows there is a good general accordance between the venom proteome and that predicted from the venom gland transcriptome. Thus, whilst a cautionary approach is required when interpreting a correlation between transcriptome and proteome, the sensitivity and unbiased nature of venom gland transcriptome surveys can be valuable in the identification of rare, unusual or potentially novel toxins and their isoforms that are difficult to detect in the proteome [29].

To provide a representative overview of the transcriptional variation in venom components in each species, whilst minimising compositional bias arising from intraspecific variation in venom composition, venom gland cDNA libraries were based on ten specimens of variable size and gender. Generated ESTs were clustered under high stringency conditions to assemble overlapping single sequence reads into full length gene objects where possible. Using BLAST, 80-93% of gene objects for each library were assigned a functional annotation based upon significant (>1e-05) scores against multiple databases. The majority of annotated ESTs (61-74%) were assigned to clusters representing distinct gene objects (additional file 1). The proportion of toxin encoding transcripts (enzymes and non-enzymatic toxins) assigned by BLAST homology, was typically greater than those encoding non-toxin transcripts (for example, those involved in cellular biosynthetic processes) and unidentified components (i.e. with no significant hit against the databases) (Figure 1). There were twice

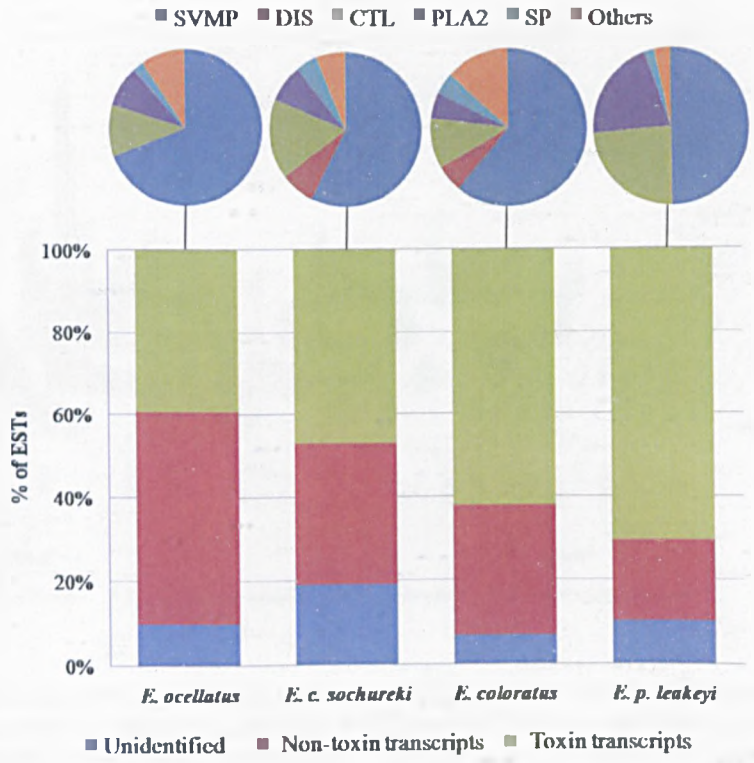

Figure 1
The relative expression of annotated venom gland transcriptomes from four members of the genus *Echis*. Bar charts represent the proportions of BLAST-annotated ESTs; unidentified = non-significant hits. Toxin encoding transcripts are expanded as pie charts illustrating the proportional representation of snake venom metalloproteinases (SVMP), short coding disintegrins (DIS), C-type lectins (CTL), group II phospholipases A_2 (PLA_2), serine proteases (SP) and other less represented venom toxins (Others) in the transcriptomes of each *Echis* species

the numbers of unidentified ESTs in the *E. c. sochureki* vgDbESTs than in any of the other *Echis* vgDbESTs. As the bulk of these unidentified ESTs were singletons, not clustered gene objects, we interpret this to result from increases in unidentified 3' untranslated regions rather than unidentified novel toxin transcripts. The annotated venom toxin encoding profiles for the four *Echis* species revealed substantial variation in (i) the inferred expression levels and (ii) the cluster diversity within many toxin families (Figure 2, additional file 2). The details and potential implications of this species-specific variation in the representation of each toxin family will be discussed in turn.

Snake venom metalloproteinases (SVMP)

The SVMP transcripts were the most abundant and divergent (in terms of cluster numbers) *Echis* venom toxin family (Figure 2) and comprised roughly half of the total toxin transcripts (Figure 1). The SVMPs are a diverse group of enzymes classified into those comprising only the metalloproteinase domain (PI) and those sequentially extended by a disintegrin domain (PII), a disintegrin-like and cysteine-rich domain (PIII) and the latter co-valently linked to C-type lectin-like components (PIV) [30]. Known and suspected modifications in domain structure are thought to account for the wide range of SVMP pathological activities, including haemorrhage, coagulopathy, fibrinolysis and prothrombin activation [30-32].

There were more PIII SVMP clusters in the genus *Echis* than any other toxin family clusters. The presence of apparent, extensive PIII SVMP gene diversification hints that evolutionary pressures are acting to increase the functional diversity of this SVMP group, highlighting their fundamental biological importance to the genus. In contrast, PI SVMP transcripts were present, albeit at low levels, only in the *E. coloratus* and *E. ocellatus* vgDbESTs. While the diversity of the PII SVMPs was substantially lower than that of the PIII SVMPs, their abundance differed between species. Thus, 80% of total *E. p. leakeyi* SVMP transcripts were PIIs (cluster EPL00005 comprised 38% of all SVMPs) and, although less numerically significant, 38% of the *E. coloratus* SVMPs were also PIIs. Despite intrageneric variation in abundance and diversity, analysis of PII contigu-

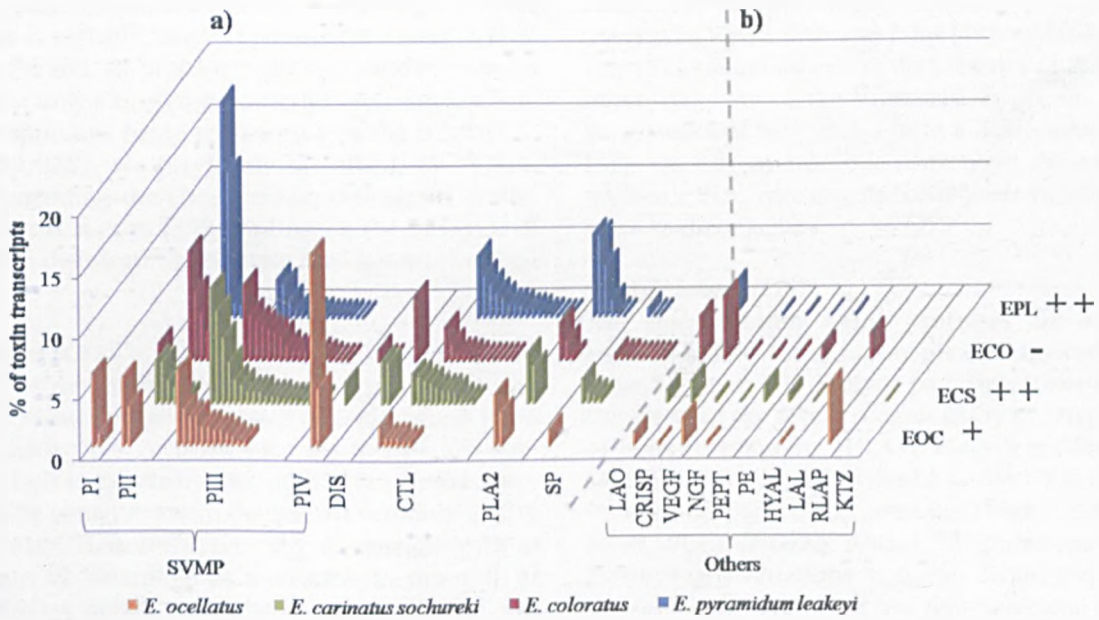

Figure 2

The relative abundance and diversity of each *Echis* genus venom toxin family. **a)** Relative expression levels of non-singleton clusters of the most representative venom toxin families and **b)** Relative expression levels of total non-singleton clusters and singletons representing the less numerically represented venom toxin families (Others) are expressed as a percentage of total toxin encoding transcripts. Column to the right indicates the proportion of invertebrate prey consumed and the corresponding correlation of venom toxicity to scorpions: ++, high; +, moderate; -, low [adapted from [10]]. Key - PI-PIV: subclasses of snake venom metalloproteinases (SVMP); DIS: short coding disintegrins; CTL: C-type lectins; PLA2: group II phospholipases A_2; SP: serine proteases; LAO: L-amino oxidases; CRISP: cysteine-rich secretory proteins; VEGF: vascular endothelial growth factors; NGF: nerve growth factors; PEPT: peptidases - aminopeptidase, dipeptidyl peptidase III and neprilysin; PE: Purine liberators - phosphdiesterase, 5'-nucleotidase and ectonucleoside triphosphate diphosphohydrolase (E-NTPase); HYAL: hyaluronidases; LAL: lysosomal acid lipases; RLAP: renin-like aspartic proteases; KTZ: kunitz-type protease inhibitors.

ous sequences throughout the genus revealed the ubiquitous representation of motifs (RGD, KGD and VGD) involved in binding to the $\alpha_{IIb}\beta_3$, $\alpha_v\beta_3$ and $\alpha_5\beta_1$integrins implicated in platelet aggregation inhibition [33,34]. The RGD-only representation of *E. p. leakeyi* PII SVMPs implies evolutionary conservation of this particular disintegrin motif, in contrast to the gene diversification observed in the PIIIs. We assigned some PIII SVMP transcripts as putative PIV SVMPs according to the presence of an additional cysteine residue in the cysteine-rich region at positions 397 or 400 [[28,30] (numbering from 30)]. These transcripts also form strongly supported monophyletic groups (data not shown) with homologues of SVMP PIVs previously characterised from venom proteomes; two of the three putative *E. coloratus* PIVs (ECO00075 & ECO00144) show the greatest sequence similarity to PIV SVMPs characterised from *Macrovipera lebetina* and *Daboia russelii* respectively [UniProt:Q7T046 and Q7LZ61], whereas all other *Echis* PIVs showed greatest similarity to the previously characterised *E. ocellatus* PIV SVMP, EOC00024 [28]. The relative representation of these putative PIV SVMPs was substantially greater in *E. ocellatus* (EOC00024 - 23% and EOC00022 -7%) than *E. coloratus* and *E. c. sochureki* (<4%); no PIV SVMPs were found in the *E. p. leakeyi* vgDbEST. Taken together, this implies that two divergent forms of PIV SVMPs may be uniquely present in *E. coloratus*, despite their low representation in this species.

We (SCW, RAH) recently identified a new *E. ocellatus* cDNA precursor encoding numerous QKW tripeptides and a polyH/G peptide that have potent SVMP-inhibiting activities [35]. Representatives of this SVMP inhibitory transcript were identified in each *Echis* vgDbEST (data not shown), but no correlation was identified between the proportional representation of the *Echis* SVMPs and their SVMP inhibitory transcripts.

Disintegrins

Snake venom disintegrins are derived either from proteolytic processing of PII SVMP precursors [36] or are encoded by discreet PII-derived disintegrin-only genes, containing only a signal peptide and a disintegrin domain - previously described as 'short coding' disintegrins

[37,38]. Representation of short coding disintegrins in the *Echis* genus is variable; small clusters were found in *E. c. sochureki* (4% and 3% of toxin transcripts) and *E. coloratus* (5%), whilst only a singleton transcript was found in *E. p. leakeyi*. Despite not being represented in the original *E. ocellatus* vgDbEST, we previously identified, by PCR, a sequence encoding the short coding disintegrin ocellatusin from this species [39], confirming the presence of short coding disintegrin transcripts throughout the *Echis* genus.

C-type lectins (CTL)

The CTLs proved to be the next most abundant and diverse (by cluster numbers) group of *Echis* venom toxin encoding transcripts. As argued for the SVMPs, the substantial CTL cluster diversity and implied functional diversity would be consistent with the known variation in CTL activity. Thus, CTL isoforms typically act synergistically as homologous or heterologous multimers to promote or inhibit platelet aggregation and/or target distinct elements of the coagulation cascade [see [40,41]]. Each of the *Echis* species showed considerable CTL diversity (10-24% toxin encoding transcripts), with *E. p. leakeyi* exhibiting both the largest number of ESTs and cluster-diversity. Notably, clusters showing similarity to echicetin α and β, a platelet aggregation-inhibitor isolated from *E. c. sochureki* [42,43], were found throughout the *Echis* genus and are the most represented CTLs in both *E. c. sochureki* and *E. p. leakeyi*. Recently, *E. ocellatus* echicetin-like CTLs were demonstrated to be associated with forming the quaternary structure of PIV *E. ocellatus* SVMPs [28]. However, PIV SVMPs are absent from the *E. p. leakeyi* vgDbEST and present in only small numbers in *E. c. sochureki* (2%), implying that PIV-related binding may not be the sole function of echicetin. In contrast, each of the *Echis* vgDbESTs (except for *E. p. leakeyi*) contained clusters showing high sequence similarity to another PIV-related CTL, Factor X activator light chain 2 from *M. lebetina* [44], producing an *Echis* representational profile of CTLs matching that of the PIV SVMPs.

Phospholipase A_2 (PLA_2)

Group II PLA_2s are ubiquitously expressed in *Echis* species [45]. *Echis* PLA_2s have been demonstrated to inhibit platelet aggregation and induce oedema, neurotoxicity and myotoxicity through multiple isoforms exhibiting high (Asp^{49}) and low (Ser^{49}) enzymatic activity [46-49]. Despite low representation and diversity in *E. coloratus*, *E. ocellatus* and *E. c. sochureki* (5-8% of toxin transcripts), an increase in representation (21%) and cluster diversity was observed in *E. p. leakeyi*, suggesting an important role for PLA_2 activity in the venom of this species. Furthermore, both enzymatic PLA_2 variants are conserved throughout the genus, highlighting the apparent importance of these functionally-distinct isoforms - presumably for prey capture. Given that Ser^{49} PLA_2s have only been isolated from the genera *Vipera* [50] and *Echis* [49], which are not sister taxa [51], we would expect the presence of this isoform in other members of the Viperinae. However, considering the absence of Ser^{49} PLA_2s from a *Bitis gabonica* vgDbEST [38], we cannot rule out convergent evolution of this myotoxic PLA_2 type and its consequent functional importance in these genera.

Serine proteases (SP)

The snake venom serine proteases are a multi-gene enzyme family acting upon platelet aggregation, blood coagulation and fibrinolytic pathways [reviewed in [41]]. Considering the severe coagulopathy observed in victims of *Echis* envenoming [19,31], the SPs are represented in amounts lower than predicted (2-5% of toxin encoding transcripts), particularly given their high representation in other, albeit distantly related, Viperidae species [52,53]. Interestingly, variations in cluster diversity are considerable, with nine clusters of low representation identified in *E. coloratus* compared to one in *E. ocellatus*. Despite low levels of representation, the unique variation in cluster diversity observed in *E. coloratus* implies multiple gene duplication events within this lineage; a process that underpins functional diversification in multi-gene venom proteins [8,54].

L-amino oxidases (LAO)

Snake venom LAOs have been demonstrated to induce apoptosis and inhibit platelet function [reviewed in [55]]. While the mechanisms for these actions remain predominately uncharacterised, it seems clear that, unlike other snake venom toxin families, isoform diversity is not a requirement. Thus, the low representation (1-4% of toxin transcripts) observed in the *Echis* vgDbESTs is consistent with other viperid venom gland transcriptomes [26,38,52,53,56-59]. Indeed, the atypically high level of sequence conservation between all the *Echis* LAOs and those from other viperid genera (>80%) implies a conserved mechanism of action, whereby evolutionary pressures act to constrain diversification.

Cysteine-rich secretory proteins (CRISP)

Members of the snake venom CRISP family interact with ion channels and exhibit the potential to block arterial smooth muscle contraction and nicotinic acetylcholine receptors [e.g. [60,61]]. The relative CRISP expression profiles vary considerably in the genus *Echis*, ranging from 5% of toxin encoding transcripts in *E. coloratus*, less than 2% in *E. c. sochureki* and *E. ocellatus* and none in *E. p. leakeyi*. Given that CRISPs are typically underrepresented toxin transcripts in Viperidae vgDbESTs [26,38,52,56-59], the abundant representation observed in *E. coloratus* implies an unidentified evolutionary pressure favouring transcriptional expression in this species. Its potential biological

significance is further highlighted by the apparent absence of these toxins in the transcriptome of the most closely related species, *E. p. leakeyi*, which differs strongly in diet from *E. coloratus* [10].

Other toxin components

Clusters encoding vascular endothelial growth factors and nerve growth factors were identified in small numbers (additional file 2) throughout the genus and, like the LAOs, each showed a high degree of sequence conservation. Similarly, and consistent with previous reports [62], the sequence homology of the new hyaluronidase singleton ESTs of *E. c. sochureki* and *E. ocellatus* was also considerable, and extended to hylauronidase sequences of other genera. It is apparent that evolutionary forces exist to conserve the sequence of this group of venom proteins, presumably because their role in disseminating venom toxins by reducing the viscosity of the extracellular matrix [29] is a universal requirement for prey 'knock-down'. Another singleton EST from the *E. c. sochureki* vgDbEST exhibited 81% identity to a kunitz-type protease inhibitor isolated from the elapid snake *Austrelaps labialis* [63]. Given the phylogenetic distance between these species, homology between these haemostatic disruptors is surprising, particularly since the singleton exhibited only 38% identity to kunitz-type protease inhibitors identified from the *Bitis gabonica* vgDbEST [38], a species closely related to *Echis*. An additional number of peptidases and purine liberators were identified as minor components in all but the *E. ocellatus* vgDbEST (Table 1). Despite their low representation and inconsistent conservation throughout the genus, the distinct biological activities of these components have been reported to play a role in the pathology of viper envenoming (Table 1), although these claims require experimental confirmation.

Novel venom gland transcriptome components

We identified a cluster from the *E. coloratus* vgDbEST that exhibited 64% identity to mammalian lysosomal acid lipase/cholesteryl ester hydrolase (LAL) [UniProt:Q4R4S5]. The most critical function of LAL is to modulate intracellular cholesterol metabolism by degrading cholesterol esters and triglycerides derived from low density lipoproteins that are transported, via specific receptors, into most cells [64,65]. Although LAL is a common enzyme in many lineages, this is the first time it has been identified from a venomous animal. We interrogated the vgDbESTs for other transcripts with annotations related to lysosomal processes and singleton transcripts were identified in multiple species (data not shown). However, their quantities were considerably lower than LAL suggesting to us that an association between venom gland LAL and intracellular processes was unlikely. Furthermore, the identification of a signal peptide using SignalP v3.0 [66] and the comparable representation of this enzyme (2%) with other venom toxin encoding transcripts (e.g. SPs, LAOs, growth factors), strongly implies these transcripts are a novel group of secreted venom components. Their biological contribution to the activity of *E. coloratus* venom and the venom gland and expression in other venomous snake genera is the subject of current research in our laboratories.

In addition to the discovery of LAL, two singleton transcripts were identified (additional file 2) from the *Echis* vgDbESTs as novel Serpentes zinc-dependent metallopeptidases [67]. A transcript exhibiting 67% identity to human dipeptidyl peptidase III (DPPIII) [UniProt:Q53GT4] was identified in *E. coloratus* and a related EST exhibiting 84% similarity to Neprilysin from *Gallus gallus* [Uniprot:Q67BJ2] was identified in the *E. p. leakeyi* vgDbEST. While signal peptides were absent from these ESTs due to EST N-terminal truncation, the constitutive physiological targets of their mammalian analogues indicate that these metallopeptidases may contribute to pathology. Mammalian DPPIII exhibits particular affinity for the degradation of hypertension-inducing peptides via the inactivation and degradation of angiotensin II to angiotensin III; the consequential reduction in vasoconstrictor activity likely induces hypotension alongside thrombolysis, by reducing the activity of plasminogen activator inhibitors that constrain fibrinolysis [68-70]. We previously reported that the *E. ocellatus* vgDbEST contained a substantial number of novel, potentially hypotensive, venom toxins termed the renin-like aspartic proteases [26]. Neprilysin demonstrates affinity for a broader range of physiological targets, including natriuretic, vasodilatory and neuro peptides [71]. Specific functional interactions include the termination of brain neuropeptides, such as enkephalins and substance P, at peptidergic synapses [72], and the degradation of the hypotension-inducing atrial natriuretic peptide (ANP) [71]. It is notable that Neprilysin has been implicated in the inactivation of peptide transmitters and their modulators in vertebrates and invertebrates [71,73], suggesting the potential for conserved neurotoxic activity across a range of prey species.

Discussion

The most numerically abundant venom toxin families in the four *Echis* species were the SVMPs, CTLs, PLA_2s, and SPs. This is broadly consistent with previous viperid venom gland analyses, although considerable intergeneric variations in the EST-inferred expression levels of these toxin families have been observed [26,38,52,53,56-59]. The correlation of toxin families identified from the genus *Echis* and other viperid species support current theories of early venom toxin recruitment prior to the radiation of the Viperidae [74]. The absence of three finger toxins from the *Echis* vgDbESTs is particularly notable as their recent identification in other viper species [53,58]

Table 1: Under-represented toxin encoding transcripts from the *Echis* vgDbESTs potentially associated with venom function.

Identification	No. of ESTs	Species present	Activty	Possible venom function
Aminopeptidase	8	*E. c. sochureki*	Hydrolysis of the N-terminal region of peptides [82].	Potential interference with angiogenesis and blood pressure control [83,84].
	1	*E. coloratus*		
Ectonucletotide pyrophosphatase/ phosphodiesterase	2	*E. coloratus*	Hydrolysis of nucleotides and nucleic acids [85].	Interaction with platelet function [85]. Activity previously described in *Echis carinatus* [86].
	3	*E. c. sochureki*		
5'-nucleotidase	3	*E. coloratus*	Cleavage of a wide variety of ribose and deoxyribose nucleotides [1].	Potential inhibitor of platelet aggregation [1]. Activity identified in a number of different lineages including *Echis carinatus* [86].
	2	*E. p. leakeyi*		
	1	*E. c. sochureki*		
Ectonucleoside triphosphate diphosphohydrolase 2 (E-NTPase 2)	2	*E. coloratus*	Hydrolysis of nucleoside-5'-triphosphates and diphosphates [87].	Potential inhibitor of platelet aggregation [87,88].

implies the venom gland recruitment of these toxins occurred prior to the divergence of the Viperidae; presumably these toxins have subsequently been lost in an ancestor of *Echis*. Consistent with the early, PCR-driven, reports of accelerated evolution of venom serine proteases [75], CTLs [76] and PLA_2s [77], it is apparent from the *Echis* genus vgDbESTs and those of other vipers that the evolutionary forces driving venom toxin recruitment in the genus *Echis* have served to promote diversification in some toxin lineages (PII and PIII SVMPs, CTLs) while in comparison relatively low diversification exists in others (PI and PIV SVMPs, PLA_2s, LAOs, the growth factors, and remaining minor venom components). Prey capture is considered a major biological imperative driving the venom toxin selection process. This project was undertaken to identify correlations between intrageneric dietary preferences and transcript expression in order to elucidate the influence dietary selection pressures may have on the toxin composition of snake venoms.

(i) Recruitment of novel venom toxins and diet. The *Echis* vgDbESTs reveal the recruitment of novel renin-like aspartic proteases in *E. ocellatus* [26], LAL and DPPIII in *E. coloratus* and Neprilysin in *E. p. leakeyi*. The potential hypotensive role of venom aspartic proteases has been discussed previously [26]. Whilst expression in the venom proteome requires experimental verification, the presence of a signal peptide suggests that LAL is more likely to be secreted in the venom gland rather than acting as an intracellular protein. LAL has been implicated in severe alveolar destruction following over-expression of these enzymes in the lungs of mice [64]. Lipases such as LAL and lipoprotein lipase may also contribute to an influx of fatty acids into the brain by hydrolysing lipoproteins in the microvascular system of the cerebral cortex [78]. The suggestion that these fatty acids are then intra-cellularly internalised within lysosomes [78] correlates with intriguing observations from *E. coloratus* induced pathology, where increases in the size and numbers of lysosomes within the neuronal tissue of guinea pigs were implicated in neuron lysis and cerebral damage [79]. We infer from the predominately vertebrate-only diet of *E. coloratus* and the exclusive, yet substantial, representation of LAL in this species (2% - equivalent to the SPs, LAOs and growth factors) that LALs may play a contributory, albeit not yet understood, role in prey envenoming. As singletons, it is more difficult to argue that the novel recruitments of DPPIII and Neprilysin represent additional adaptations to prey preference; as they are found in such low numbers it is impossible to determine whether they are indeed novel species-specific venom gland recruitments or are rare transcripts that remain undetected in other snake species. We previously reported that invertebrate feeding likely evolved as a basal trait in the genus *Echis* [10]. The absence of genus-wide transcripts encoding novel putative venom toxin families implies that the adaptation to invertebrate feeding in *Echis* did not evolve as a consequence of recruiting novel invertebrate-specific venom toxins. However,

we cannot exclude the possibility that the novel recruitment of LAL into the *E. coloratus* venom gland transcriptome may result from the subsequent reversion to vertebrate feeding observed in this species [10], particularly given the absence of these well represented putative toxin transcripts in other members of the genus.

(ii) Changes in toxin family expression and diet. All the major *Echis* venom toxin families (SVMP, CTL, PLA_2, SP) exhibited considerable intrageneric variation in transcriptional representation. Thus, the *E. p. leakeyi* vgDbEST was notable for its absence of PI and PIV SVMPs, short coding disintegrins and CRISPs and atypically abundant representation of PII SVMPs, CTLs and PLA_2s. The CRISPs were only represented by clusters in *E. c. sochureki* and *E. coloratus*, species whose vgDbESTs draw similarities, particularly in their high comparative expression of PIII SVMPs and short coding disintegrins. The only distinguishing feature (in terms of transcript abundance) in the *E. ocellatus* vgDbEST was the atypically high number of PIV SVMPs. However, none of these toxin encoding expression profiles showed a clear association with diet. Most notably, *E. p. leakeyi* and *E. c. sochureki* exhibit distinct toxin encoding profiles (Figure 2), despite both species feeding predominately on invertebrates and exhibiting highly invertebrate-lethal venom [10].

(iii) Diversification of venom toxins and diet. The above observations imply adaptations to diet are occurring within venom toxin families rather than resulting from changes in expression levels of entire toxin families. Evidence supporting this hypothesis is provided by substantial increases in representation of echicetin-like CTLs (relative to other CTLs) in both *E. p. leakeyi* and *E. c. sochureki*, implying perhaps a significant role for these platelet aggregation inhibitors in invertebrate prey capture. The absence of PI SVMPs in these species perhaps suggests that this SVMP isoform is more associated with a vertebrate diet. Furthermore, a number of atypical observations identified from the *E. coloratus* vgDbEST may be associated with a reversion to vertebrate feeding [10], including; (i) increases in the representation of CRISPs, (ii) increases in cluster diversity of the SPs and (iii) the identification of putative novel venom toxins (LAL and DPPIII). However, the general similarity between the toxin encoding expression profiles of *E. c. sochureki* and *E. coloratus* (Figure 2), despite *E. coloratus* exhibiting a significant reduction in venom toxicity to invertebrates [10], indicates that more analytical molecular tools are required to determine whether snake prey specificity is achieved through subtle alterations in isoform expression levels within the major venom toxin families. We are subjecting the *Echis* genus vgDbEST data generated here to a phylogenetic analysis on each toxin class to determine species-specific trends in diversification, which will inform us whether multiple levels of gene control in the *Echis* genus venom gland (switching of transcriptional expression, gene duplication conferring functional diversification and novel gene expression) maybe responsible for evolutionary responses to dietary pressures.

Correlations between variation in venom gland toxin encoding profiles and snakebite symptomatologies from the genus *Echis* are unclear, particularly given the similar, predominately incoagulable and haemorrhagic, clinical outcomes observed throughout the genus [19-21] and the presence of multiple isoforms of toxin families implicated in haemorrhage and coagulopathy. However, some observations of atypical symptoms can be tentatively explained; substantial increases in PLA_2 representation and the unique presence of Neprilysin may correlate with the rare manifestation of neurotoxicity observed in an *E. pyramidum* envenomation [22], whilst the putative function of DPPIII may imply a contributory role in cases of hypotension observed following *E. coloratus* snakebite [20].

Venom gland transcriptome surveys provide valuable new data that we are correlating with a proteomic analysis of the venom from each *Echis* species. With this comprehensive description of the venom composition of each major *Echis* lineage, we will identify, using proteomic (antivenomic) techniques [3], the extent to which the intrageneric variation in venom composition impacts on the preclinical efficacy of commercially available antivenoms. We hope that such analyses will (i) explain past antivenom failures described following snakebite by members of this medically important genus [18,22-24] and (ii) identify the venom toxin mix required to generate an antivenom with continent-wide clinical effectiveness against *Echis* envenoming.

Conclusion

The first comprehensive comparison of intrageneric venom gland transcriptomes reveals substantial venom variation in the genus *Echis*. The observed variations in venom toxin encoding profiles reveal little association with venom adaptations to diet previously described from this genus. We hypothesise that relatively subtle alterations in toxin expression levels within the major venom toxin families are likely to be predominately responsible for prey specificity, although we cannot rule out a contributory role for novel putative venom toxins, such as lysosomal acid lipase. The observation of substantial venom variation within the medically important genus *Echis* strongly advocates further investigations into the medical significance of venom variation and its potential impact upon antivenom therapy.

Methods

Venom gland cDNA libraries were constructed from ten wild-caught specimens of *Echis coloratus* (Egypt), *E. p. leakeyi* (Kenya)and *E. c. sochureki* (Sharjah, UAE), maintained in the herpetarium of the Liverpool School of Tropical Medicine, using identical protocols described for the construction of the venom gland cDNA library from *E. ocellatus* [26]. Clones from the cDNA libraries were picked randomly and sequenced (NERC Molecular Genetics Facility, UK) using M13 forward primers.

Bioinformatic processing was carried out using the PartiGene pipeline [80] with the same protocols used previously [26]. Briefly, sequences were processed (to exclude low quality, contaminating vector sequences and poly A+ tracts) using Trace2dbEST [81]. Subsequently, assembly was undertaken in PartiGene version 3.0, using high stringency clustering parameters [26,81]. A total of 1070 (*E. coloratus*), 1078 (*E. p. leakeyi*) and 1156 (*E. c. sochureki*) processed ESTs were entered into respective species databases alongside the 883 ESTs generated from the *E. ocellatus* vgDbEST [26]. Assembled ESTs were BLAST annotated against UniProt (v56.2), TrEMBL (v39.2) and separate databases containing only Serpentes nucleotide and protein sequences derived from the same Uniprot/TrEMBL release versions.

Clustering was performed incrementally (96 sequences per round) to determine the number of sequences required to construct a representative transcriptome (i.e. the point where further sequencing only adds to existing clusters). We estimate that a minimum of 800 EST sequences were required to provide an accurate representation of the three vgDbESTs (additional file 3). For longer clones (i.e. SVMPs), representatives of each cluster were subject to primer walking to acquire sufficient sequence data for isoform classification. SVMPs were characterised based upon the presence or absence of additional domains extending from the metalloproteinase domain [30]. PIVs were distinguished from PIIIs by the presence of an additional cysteine residue in the cysteine-rich region at positions 397 or 400 [[28,30] (numbering from 30)].

Additional file 2 displays the catalogue of venom toxin transcripts present in each of the four *Echis* vgDbESTs based upon significant (>1e-05) BLAST annotation. Presentation of the fully assembled and annotated vgDbESTs can be viewed at http://venoms.liv.ac.uk. The sequences reported in this paper have also been submitted into dbEST division of the public database GenBank: *E. coloratus* [GenBank: GR947900-GR948969], *E. c. sochureki* [GenBank: GR948970-GR950126] and *E. p. leakeyi* [GenBank: GR950127-GR951204].

All animal experimentation was conducted using standard protocols approved by the University of Liverpool Animal Welfare Committee and performed with the approval of the UK Home Office (40/3216) under project licence # 40/3216.

Authors' contributions

NRC participated in the experiments, the comparative analysis and drafted the manuscript. RAH participated in the experiments, the design of the study and reviewed the manuscript. WW participated in the design of the study and reviewed the manuscript. SCW participated in the experiments, the design of the study, the comparative analysis and reviewed the manuscript. All authors have read and approved the paper.

Additional material

Additional file 1

Summary statistics following clustering and assembling of ESTs for E. coloratus, E. p. leakeyi *and* E. c. sochureki.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-564-S1.doc]

Additional file 2

Catalogue of venom toxin encoding ESTs determined from the* Echis *vgDbESTs. *Putative novel venom toxins are in bold and underlined. Key - SVMP: snake venom metalloproteinases; PI, PII, PIII, PIV: respective sub-group of SVMPs; ND: sub-class not determined; DIS: short coding disintegrins; CTL: C-type lectins; PLA_2: group II phospholipases A_2; SP: serine proteases; LAO: L-amino oxidases; CRISP: cysteine-rich secretory proteins; VEGF: vascular endothelial growth factors; NGF: nerve growth factors; PEPT: peptidases; AP: aminopeptidase; DPP: dipeptidyl peptidase III; NEP: neprilysin; PE: Purine liberators; PHOS: phosphdiesterase; 5'-NUC: 5'-nucleotidase; E-NTPase: ectonucleoside triphosphate diphosphohydrolase; LAL: lysosomal acid lipases; RLAP: renin-like aspartic proteases; HYAL: hyaluronidases; KTZ: kunitz-type protease inhibitors.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-564-S2.doc]

Additional file 3

***An overview of clustering processes for three species of the genus* Echis.** *The graph demonstrates the percentage of ESTs that are added to clusters (ESTs >1) as the cumulative number of ESTs entering the database increase. In all species the number of ESTs affecting the proportion of EST clusters and singletons reaches a plateau after 800 sequences.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-564-S3.jpeg]

Acknowledgements

The authors wish to thank Paul Rowley for expert herpetological assistance, Damien Egan and Paul Vercammen (Breeding Centre for Endangered Arabian Wildlife, United Arab Emirates) for providing specimens of *E. c. sochureki*, Ann Hedley and Mark Blaxter (NERC Molecular Genetics Facility,

University of Edinburgh) for providing sequencing and bioinformatic advice regarding the PartiGene pipeline and Tim Booth, Bela Tiwari and Jorge Soares (NERC Environmental Bioinformatics Centre) for bioinformatic advice. This work was funded by Research Studentship NER/S/A/2006/14086 from the Natural Environmental Research Council (NERC) to NRC, access to the NERC Molecular Genetics Facility at the University of Edinburgh (ref MGF 150) to WW, the Leverhulme Trust (Grant F/00 174/I) to WW and RH and the Biotechnology and Biological Sciences Research Council (BBSRC) to RH and SCW (BB/F012675/1).

References

1. Aird SD: **Ophidian envenomation strategies and the role of purines.** *Toxicon* 2002, **40**:335-393.
2. Chippaux JP, Williams V, White J: **Snake venom variability: methods of study, results and interpretation.** *Toxicon* 1991, **29**:1279-1303.
3. Gutiérrez JM, Lomonte B, Leün G, Alape-Girün A, Flores-Díaz M, Sanz L, Angulo Y, Calvete JJ: **Snake venomics and antivenomics: proteomic tools in the design and control of antivenoms for the treatment of snakebite envenoming.** *J Proteomics* 2009, **72**:165-182.
4. Sasa M: **Diet and snake venom evolution: can local selection alone explain intraspecific venom variation?** *Toxicon* 1999, **37**:249-252.
5. Sasa M: **Reply.** *Toxicon* 1999, **37**:259-260.
6. Mebs D: **Toxicity in animals. Trends in evolution?** *Toxicon* 2001, **39**:87-96.
7. Daltry JC, Wüster W, Thorpe RS: **Diet and snake venom evolution.** *Nature* 1996, **379**:537-540.
8. Kordiš D, Gubenšek F: **Adaptive evolution of animal toxin multigene families.** *Gene* 2000, **261**:43-52.
9. Jorge da Silva N Jr, Aird SD: **Prey specificity, comparative lethality and compositional differences of coral snake venoms.** *Comp Biochem Physiol* 2001, **128C**:425-456.
10. Barlow A, Pook CE, Harrison RA, Wüster W: **Co-evolution of diet and prey-specific venom activity supports the role of selection in snake venom evolution.** *Proc R Soc B* 2009, **276**:2443-2449.
11. Creer S, Malhotra A, Thorpe RS, Stöcklin R, Favreau P, Chou WH: **Genetic and ecological correlates of intraspecific variation in pitviper venom composition detected using matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF-MS) and isoelectric focusing.** *J Mol Evol* 2003, **56**:317-329.
12. Sanz L, Gibbs HL, Mackessy SP, Calvete JJ: **Venom proteomes of closely related *Sistrurus* rattlesnakes with divergent diets.** *J Proteome Res* 2006, **5**:2098-2112.
13. Warrell DA: **Snake venoms in science and clinical medicine. 1. Russell's viper: biology, venom and treatment of bites.** *Trans R Soc Trop Med Hyg* 1989, **83**:732-740.
14. Prasad NB, Uma B, Bhat SKG, Gowda TV: **Comparative characterization of Russell's viper (*Daboia/Vipera russelli*) venoms from different regions of Indian peninsula.** *Biochim Biophys Acta* 1999, **1428**:121-136.
15. Shashidharamurthy R, Jagadeesha DK, Girish KS, Kemparaju K: **Variation in biochemical and pharmacological properties of Indian cobra (*Naja naja*) venom due to geographical distribution.** *Mol Cell Biochem* **229**:93-101.
16. Theakston RDG, Phillips RE, Warrell DA, Galigedera Y, Abeysekera DT, Dissanayake P, Hutton RA, Aloysius DJ: **Failure of India (Haffkine) antivenom in treatment of *Vipera russelli pulchella* (Russell's viper) envenoming in Sri Lanka.** *Toxicon* 1989, **27**:82.
17. Galán JA, Sánchez EE, Rodríguez-Acosta A, Pérez JC: **Neutralization of venoms from two Southern Pacific rattlesnakes (*Crotalus helleri*) with commercial antivenoms and endothermic animal sera.** *Toxicon* 2004, **43**:791-799.
18. Visser LE, Kyei-Faried S, Belcher DW, Geelhoed DW, Schagen van Leeuwen J, van Roosmalen J: **Failure of a new antivenom to treat *Echis ocellatus* snake bite in rural Ghana: the importance of quality surveillance.** *Trans R Soc Trop Med Hyg* 2008, **102**:445-450.
19. Warrell DA, Davidson NM, Greenwood BM, Ormerod LD, Pope HM, Watkins BJ, Prentice CRM: **Poisoning by bites of the saw-scaled or carpet viper (*Echis carinatus*) in Nigeria.** *QJM* 1977, **46**:33-62.
20. Warrell DA: **Clinical toxicology of snakebite in Africa, the Middle East/Arabian Peninsula and Asia.** In *Handbook of clinical toxicology of animal venoms and poisons* Edited by: Meier J, White J. Boca Raton, Florida: CRC Press; 1995:433-595.
21. Habib AG, Gebi UI, Onyemelukwe GC: **Snake bite in Nigeria.** *Afr J Med Med Sci* 2001, **30**:171-178.
22. Gillissen A, Theakston RDG, Barth J, May B, Krieg M, Warrell DA: **Neurotoxicity, haemostatic disturbances and haemolytic anaemia after a bite by a Tunisian saw-scaled or carpet viper (*Echis 'pyramidum'*-complex): Failure of antivenom treatment.** *Toxicon* 1994, **32**:937-944.
23. Kochar DK, Tanwar PD, Norris RL, Sabir M, Nayak KC, Agrawal TD, Purohit VP, Kochar A, Simpson ID: **Rediscovery of severe saw-scaled viper (*Echis sochureki*) envenoming in the Thar Desert region of Rajasthan, India.** *Wilderness Enviro Med* 2007, **18**:75-85.
24. Warrell DA: **Unscrupulous marketing of snake bite antivenoms in Africa and Papua New Guinea: choosing the right product-'What's in a name?'.** *Trans Royal Soc Trop Med Hygiene* 2008, **102(5)**:397-399.
25. Pook CE, Joger U, Stümpel N, Wüster W: **When continents collide: phylogeny, historical biogeography and systematics of the medically important viper genus *Echis* (Squamata: Serpentes: Viperidae).** *Mol Phylogenet Evol* 2009 in press. doi:10.1016/j.ympev.2009.08.002
26. Wagstaff SC, Harrison RA: **Venom gland EST analysis of the saw-scaled viper, *Echis ocellatus*, reveals novel $\alpha_9\beta_1$ integrin-binding motifs in venom metalloproteinases and a new group of putative toxins, renin-like aspartic proteases.** *Gene* 2006, **377**:21-32.
27. Paine MJ, Desmond HP, Theakston RDG, Crampton JM: **Gene expression in *Echis carinatus* (carpet viper) venom glands following milking.** *Toxicon* 1992, **30**:379-386.
28. Wagstaff SC, Sanz L, Juárez P, Harrison RA, Calvete JJ: **Combined snake venomics and venom gland transcriptomic analysis of the ocellated carpet viper, *Echis ocellatus*.** *J Proteomics* 2009, **71(6)**:609-623.
29. Harrison RA, Ibison F, Wilbraham D, Wagstaff SC: **Identification of cDNAs encoding viper venom hyaluronidases: cross-generic sequence conservation of full-length and unusually short variant transcripts.** *Gene* 2007, **392**:22-33.
30. Fox JW, Serrano SMT: **Structural considerations of the snake venom metalloproteinases, key members of the M12 reprolysin family of metalloproteinases.** *Toxicon* 2005, **45**:969-985.
31. Warrell DA, Pope HM, Prentice CRM: **Disseminated intravascular coagulation caused by the carpet viper (*Echis carinatus*): Trial of Heparin.** *Brit J Haemat* 1976, **33**:335-342.
32. Fox JW, Serrano SMT: **Insights into and speculations about snake venom metalloproteinase (SVMP) synthesis, folding and disulfide bond formation and their contribution to venom complexity.** *FEBS J* 2008, **275**:3016-3030.
33. Huang TF, Holt JC, Lukasiewicz H, Niewiarowski S: **Trigamin. A low molecular weight peptide inhibiting fibrinogen with platelet receptors expressed on glycoprotein IIb-IIIa complex.** *J Biol Chem* 1987, **262**:16157-16163.
34. Calvete JJ, Marcinkiewicz C, Monleün D, Esteve V, Celda B, Juárez P, Sanz L: **Snake venom disintegrins: evolution of structure and function.** *Toxicon* 2005, **45**:1063-1074.
35. Wagstaff SC, Favreau P, Cheneval O, Laing GD, Wilkinson MC, Miller RL, Stöcklin R, Harrison RA: **Molecular characterisation of endogenous snake venom metalloproteinase inhibitors.** *Biochem Biophys Res Commun* 2008, **365(4)**:650-656.
36. Shimokawa K, Jai LG, Wang XM, Fox JW: **Expression, activation and processing of the recombinant snake venom metalloproteinase, pro-atrolysin E.** *Arch Biochem Biophys* 1996, **335**:283-294.
37. Okuda D, Koike H, Morita T: **A new gene structure of the disintegrin family: a subunit of dimeric disintegrin has a short coding region.** *Biochemistry* 2002, **41**:14248-14254.
38. Francischetti IMB, My-Pharm V, Harrison J, Garfield MK, Ribeiro JMC: ***Bitis gabonica* (Gaboon viper) snake venom gland: toward a catalog for the full-length transcripts (cDNA) and proteins.** *Gene* 2004, **357**:55-69.
39. Juárez P, Wagstaff SC, Sanz L, Harrison RA, Calvete JJ: **Molecular cloning of *Echis ocellatus* disintegrins reveals non-venom secreted proteins and a pathway for the evolution of ocellatusin.** *J Mol Evol* 2006, **63**:183-193.

40. Markland FS: **Snake venoms and the haemostatic system.** *Toxicon* 1998, **36**:1749-1800.
41. Kini RM: **Anticoagulant proteins from snake venoms: structure, function and mechanism.** *Biochem J* 2006, **397**:377-387.
42. Peng M, Lu W, Beviglia V, Niewiarowski S, Kirby EP: **Echicetin: a snake venom protein that inhibits binding of von Willebrand factor and alboaggregins to platelet glycoprotein Ib.** *Blood* 1993, **81**:2321-2328.
43. Polgár J, Magnenat EM, Peitsch MC, Wells TN, Saqi MS, Clemetson KJ: **Amino acid sequence of the alpha subunit and computer modelling of the alpha and beta subunits of echicetin from the venom of Echis carinatus (saw-scaled viper).** *Biochem J* 1997, **323**:533-537.
44. Siigur E, Aaspõllu A, Trummal K, Tõnismägi K, Tammiste I, Kalkkinen N, Siigur J: **Factor X activator from *Vipera lebetina* venom is synthesized from different genes.** *Biochim Biophys Acta Prot Proteomics* 2004, **1702**:41-51.
45. Bharati K, Hasson SS, Oliver J, Laing GD, Theakston RDG, Harrison RA: **Molecular cloning of phospholipases A_2 from venom glands of *Echis* carpet vipers.** *Toxicon* 2003, **41**:941-947.
46. Kemparaju K, Prasad BN, Gowda VT: **Purification of a basic phospholipase A_2 from Indian saw-scaled viper (*Echis carinatus*) venom: characterization of antigenic, catalytic and pharmacological properties.** *Toxicon* 1994, **32**:1187-1196.
47. Kemparaju K, Krishnakanth TP, Gowda VT: **Purification and characterization of a platelet aggregation inhibitor acidic phospholipase A_2 from Indian saw-scaled viper (*Echis carinatus*) venom.** *Toxicon* 1999, **37**:1659-1671.
48. Jasti J, Paramasivam M, Srinivasan A, Singh TP: **Structure of an acidic phospholipase A_2 from Indian saw-scaled viper (*Echis carinatus*) at 2.6 Å resolution reveals a novel intermolecular interaction.** *Acta Cryst* 2004, **D60**:66-72.
49. Zhou X, Tan TC, Valiyaveettil S, Go ML, Kini RM, Velazquez-Campoy A, Sivaraman J: **Structural characterization of myotoxic Ecarpholin S from *Echis carinatus* venom.** *Biophysical J* 2008, **95**:3366-3380.
50. Petan T, Križaj I, Punger.ar J: **Restoration of enzymatic activity in a Ser-49 phospholipase A2 homologue decreases its Ca2+-independant membrane-damaging activity and increases its toxicity.** *Biochemistry* 2007, **46**:12795-12809.
51. Wüster W, Peppin L, Pook CE, Walker DE: **A nesting of vipers: Phylogeny and historical biogeography of the Viperidae (Squamata: Serpentes).** *Mol Phylogenet Evol* 2008, **49(2)**:445-459.
52. Cidade DAP, Simão TA, Dávila AMR, Wagner G, Junqueira-de-Azevedo ILM, Ho PL, Bon C, Zingali RB, Albano RM: ***Bothrops jararaca* venom gland transcriptome: Analysis of the gene expression pattern.** *Toxicon* 2006, **48**:437-461.
53. Pahari S, Mackessy SP, Kini RM: **The venom gland transcriptome of the Desert Massasauga rattlesnake (*Sistrurus catenatus edwardsii*): towards an understanding of venom composition among advanced snakes (Superfamily Colubroidea).** *BMC Mol Biol* 2007, **8**:115.
54. Župunski V, Kordiš D, Gubenšek F: **Adaptive evolution in the snake venom Kunitz/BPTI protein family.** *FEBS Letters* 2003, **547**:131-136.
55. Du XY, Clemetson KJ: **Snake venom L-amino acid oxidases.** *Toxicon* 2002, **40**:659-665.
56. Junqueira-de-Azevedo ILM, Ho PL: **A survey of gene expression and diversity in the venom glands of the pit viper snake *Bothrops insularis* through the generation of expressed sequence tags (ESTs).** *Gene* 2002, **299**:279-291.
57. Kashima S, Roberto PG, Soares AM, Astolfi-Filho S, Pereira JO, Giuliati S, Faria M, Xavier MAS, Fontes MRM, Giglio JR, Franca SC: **Analysis of *Bothrops jararacussu* venomous gland transcriptome focusing on structural and functional aspects: I - gene expression profile of highly expressed phospholipases A_2.** *Biochimie* 2004, **86**:211-219.
58. Junqueira-de-Azevedo ILM, Ching ATC, Carvalho E, Faria F, Nishiyama ML, Ho PL, Diniz MRV: ***Lachesis muta* (Viperidae) cDNAs reveal diverging pit viper molecules and scaffolds typical of Cobra (Elapidae) venoms: Implications for snake toxin repertoire evolution.** *Genetics* 2006, **173**:877-889.
59. Zhang B, Liu Q, Yin W, Zhang X, Huang Y, Luo Y, Qiu P, Su X, Yu J, Hu S, Yan G: **Transcriptomic analysis of *Deinagkistrodon acutus* venomous gland focusing on cellular structure and functional aspects using expressed sequence tags.** *BMC Genomics* 2006, **7**:152.
60. Yamazaki Y, Morita T: **Structure and function of snake venom cysteine-rich secretory proteins.** *Toxicon* 2004, **44**:227-231.
61. Gorbacheva EV, Starkov VG, Tsetlin VI, Utkin YN, Vulfius CA: **Viperidae snake venoms block nicotinic acetylcholine receptors and voltage-gated Ca2+ channels in identified neurons of fresh-water snail *Lymnaea stagnalis*.** *Biochem (Moscow) A Membrane Cell Biol* 2008, **2**:14-18.
62. Kemparaju K, Girish KS: **Snake venom hyaluronidase: a therapeutic target.** *Cell Biochem Funct* 2006, **24**:7-12.
63. Doley R, Tram NNB, Reza MA, Kini RM: **Unusual accelerated rate of deletions and insertions in toxin genes in the venom glands of the pygmy copperhead (*Austrelaps labialis*) from Kangaroo Island.** *BMC Evol Biol* 2008, **8**:70.
64. Li Y, Qin Y, Li H, Wu R, Yan C, Du H: **Lysosomal acid lipase overexpression disrupts lamellar body genesis and alveolar structure in the lung.** *Int J Exp Path* 2007, **88**:427-436.
65. Qu P, Du H, Wilkes DS, Yan C: **Critical roles of lysosomal acid lipase in T cell development and function.** *Am J Pathol* 2009, **174**:944-956.
66. Bendtsen JD, Nielsen H, von Heijne G, Brunak S: **Improved prediction of signal peptides: SignalP 3.0.** *J Mol Biol* 2004, **340**:783-795.
67. Baral PK, Jaj.anin-Jozi(N, Deller S, Macheroux P, Abrami(M, Gruber K: **The first structure of dipeptidyl-peptidase III provides insight into the catalytic mechanism and mode of substrate binding.** *J Biol Chem* 2008, **283(32)**:22316-22324.
68. Lee CM, Snyder SH: **Dipeptidyl-aminopeptidase III of rat brain: selective affinity for enkephalin and angiotensin.** *J Biol Chem* 1982, **257(20)**:12043-12050.
69. Abrami(M, Zubanovi(M, Vitale L: **Dipeptidyl peptidase III from human erythrocytes.** *Biol Chem Hoppe Seyler* 1988, **369**:29-38.
70. Skurk T, Lee YM, Hauner H: **Angiotensin II and its metabolites stimulate PAI-1 protein release from human adipocytes in primary culture.** *Hypertension* 2001, **37**:1336-1340.
71. Turner AJ, Isaac RE, Coates D: **The neprilysin (NEP) family of zinc metalloendopeptidases: Genomics and function.** *Bioessays* 2001, **23(3)**:261-269.
72. Matsas R, Fulcher IS, Kenny AJ, Turner AJ: **Substance P and (Leu)enkephalin are hydrolysed by an enzyme in pig caudate synaptic membranes that is identical with the endopeptidase of kidney microvilli.** *Proc Natl Acad Sci USA* 1983, **80**:3111-3115.
73. Isaac RE: **Neuropeptide-degrading endopeptidase activity of locust (*Schistocerca gregaria*) synaptic membranes.** *Biochem J* 1988, **255**:843-847.
74. Fry BG, Scheib H, Weerd L van der, Young B, McNaughtan J, Ramjan SFR, Vidal N, Poelmann RE, Norman JA: **Evolution of an arsenal.** *Mol Cell Prot* 2008, **7**:215-246.
75. Deshimaru M, Ogawa T, Nakashima KI, Nobuhisa I, Chijiwa T, Shimohigashi Y, Fukumaki Y, Niwa M, Yamashina I, Hattori S, Ohno M: **Accelerated evolution of crotalinae snake venom gland serine proteases.** *FEBS Letters* 1996, **397**:83-88.
76. Ogawa T, Chijiwa T, Oda-Ueda N, Ohno M: **Molecular diversity and accelerated evolution of C-type lectin-like proteins from snake venom.** *Toxicon* 2005, **45**:1-14.
77. Nakashima K, Ogawa T, Oda N, Hattori M, Sakaki Y, Kihara H, Ohno M: **Accelerated evolution of *Trimeresurus flavoviridis* venom gland phosphlipase A_2 isoenzymes.** *Proc Natl Acad Sci USA* 1993, **90**:5964-5968.
78. Brecher P, Kuan HT: **Lipoprotein lipase and acid lipase activity in rabbit brain microvessels.** *J Lipid Res* 1979, **20**:464-471.
79. Sandbank U, Djaldetti M: **Effect of Echis colorata venom inoculation on the nervous system of the dog and guinea pig.** *Acta Neuropath* 1966, **6**:61-69.
80. **The PartiGene EST-software pipeline at the nematode and neglected genomics database** [http://www.nematodes.org/bioinformatics/PartiGene/index.shtml]
81. Parkinson J, Anthony A, Wasmuth J, Schmid R, Hedley A, Blaxter M: **PartiGene - constructing partial genomes.** *Bioinformatics* 2004, **20**:1398-1404.
82. Glenner GG, Folk JE: **Glutamyl peptidases in rat and guinea pig kidney slices.** *Nature* 1961, **192**:338-340.
83. Marchio S, Lahdenranta J, Schlingemann RO, Valdembri D, Wesseling P, Arap MA, Hajitou A, Ozawa MG, Trepel M, Giordano RJ, Nanus DM, Dijkman HB, Ooserwijk E, Sidman RL, Cooper MD, Bussolino F,

Pasqualini R, Arap W: **Aminopeptidase A is a functional target in angiogenic blood vessels.** *Cancer Cell* 2004, **5**:151-162.
84. Fournie-Zaluski MC, Fassot C, Valentin B, Djordjijevic D, Reaux-Le Goazigo A, Corvol P, Roques BP, Llorens-Cortes C: **Brain renin-angiotensin system blockade by systemically active aminopeptidase A inhibitors: a potential treatment of salt-dependent hypertension.** *Proc Natl Acad Sci USA* 2004, **101**:7775-7780.
85. Fürstenau CR, Trentin DDS, Barreto-Chaves MLM, Sarkis JJF: **Ecto-nucleotide pyrophosphate/phosphodiesterase as part of a multiple system for nucleotide hydrolysis by platelets from rats: Kinetic characterization and biochemical properties.** *Platelets* 2006, **17(2)**:84-91.
86. Taborska E: **Intraspecies variability of the venom of *Echis carinatus*.** *Physiol Bohemoslov* 1971, **20**:307-318.
87. Sales PBV, Santoro ML: **Nucleotide and DNase activities in Brazilian snake venoms.** *Comp Biochem Physiol C Toxicol Pharmacol* 2008, **147(1)**:85-95.
88. Champagne DE: **Antihemostatic molecules from saliva of blood-feeding arthropods.** *Pathophysiol Haemos Thromb* 2005, **34**:221-227.