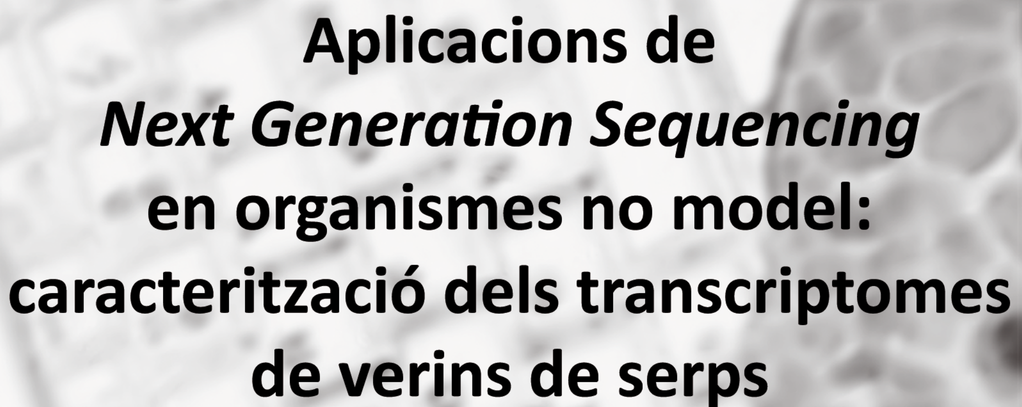
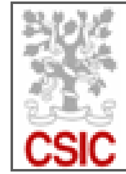


Aplicacions de *Next Generation Sequencing* en organismes no model: caracterització dels transcriptomes de verins de serps

JORDI DURBAN i SANCHEZ

TESI DOCTORAL 2014



VNIVERSITAT DE VALÈNCIA

DEPARTAMENT DE BIOQUÍMICA I BIOLOGIA MOLECULAR

Aplicacions de *Next Generation Sequencing* en organismes no model : caracterització dels transcriptomes de verins de serps

Jordi Durban i Sanchez

Tesi Doctoral 2014

VNIVERSITAT DE VALÈNCIA

DEPARTAMENT DE BIOQUÍMICA I BIOLOGIA MOLECULAR

Aplicacions de *Next Generation Sequencing* en organismes no model : caracterització dels transcriptomes de verins de serps

Tesi dirigida pel Dr. Juan José Calvete Chornet

Memòria presentada per Jordi Durban i Sanchez per a optar al grau de Doctor per la Universitat de València

Programa de doctorat: 030E Bioquímica Clinicomèdica i Immunologia
Tutor de Tesi: Dr. Juan Viña Ribes

Prof. Juan J. Calvete
Director del Laboratorio de Venómica Estructural y Funcional
Instituto de Biomedicina de Valencia, C.S.I.C.
Jaime Roig, 11
46010 Valencia
Email: jcalvete@ibv.csic.es

JUAN JOSÉ CALVETE CHORNET, Doctor en Biologia i Professor d'Investigació a l'Institut de Biomedicina del Consell Superior d'Investigacions Científiques,

INFORMA que en Jordi Durban i Sanchez, llicenciat en Bioquímica per la Universitat de València ha realitzat sota la meua direcció al Laboratori de Venòmica Estructural i Funcional (Institut de Biomedicina de València) la memòria titulada "*Aplicacions de Next Generation Sequencing en organismes no model: caracterització dels transcriptomes de verins de serps*", que presenta per a optar al títol de Doctor per la Universitat de València.

I perquè així conste i tinga els efectes oportuns, signe el present document en València el 31 d'Octubre de 2014.

Dr. Juan José Calvete

C/ Serrano, 117
28006 Madrid ESPAÑA
Telf. 91 585 50 00
Fax.: 91 411 30 77

Aquest treball ha estat realitzat amb el suport econòmic dels projectes d'investigació que s'enumeren a continuació:

BFU2007-61563: Enfoque "top-down" aplicado a la venómica. Evolución de la familia de las disintegrinas y evaluación del uso de la disintegrina jerdostatina para la visualización in vivo de angiogénesis dependiente de la integrina alpha1beta1.
Ministerio de Educación y Ciencia (I.P. J.J. Calvete)

BFU2010-17373: Venómica Traslacional
Ministerio de Ciencia e Innovación (IP. J.J. Calvete)

Jordi Durban i Sanchez ha gaudit d'una beca predoctoral FPI atorgada pel Ministerio de Ciencia e Innovación associada al projecte BFU2007-61563

A Laia i Joana

TAULA DE CONTINGUTS

TAULA DE CONTINGUTS **11**

INTRODUCCIÓ **17**

Les serps **19**

Els verins **21**
Composició 22
Proteïnes amb activitat enzimàtica 24
Metaloproteases del verí de les serps 24
Fosfolipases A_2 25
L-aminoàcid oxidases 27
Serines Proteases 27
Nucleases, nucleotidases i fosfomonoesterases 28
Glutaminil ciclases 29
Proteïnes sense activitat enzimàtica 30
Pèptids natriurètics i potenciadors de la bradiquinina 30
Disintegrines 30
Proteïnes Semblants a Lectines tipus C 31
Factors de creixement 31
CRISP (Cysteine-Rich Secretory Proteins) 32
Aplicacions 32
Enverinament i símptomes 36

Transcriptòmica i proteòmica **39**
Transcriptòmica 39
Proteòmica 43

Seqüenciació de Nova Generació (NGS) **45**
Plataformes actuals 50
454 (Roche) 51
SOLiD (Life Technologies) 52
Illumina 53
Ion Torrent (Life Tecnologies) 55
Pacific Biosciences 55
NGS i transcriptòmica 56

OBJECTIUS **59**

PUBLICACIONS **63**

Capítol 1 .... **65**

Capítol 2 .... **83**

Capítol 3 .... **103**

RESULTATS 117

Capítol 1 .... **119**
Seqüenciació i ensamblatge 119
Anotació funcional 121
Estimació del nombre de còpies gèniques 124
Comparativa transcriptoma - proteoma 124

Capítol 2 .... **125**
Seqüenciació i ensamblatge 125
Anotació funcional 126
Els miRNA 128
Comparativa transcriptoma - proteoma 129

Capítol 3 .... **130**

DISCUSSIÓ 133

Capítols 1 i 2 .... **135**
Seqüenciació i ensamblatge 135
Anotació funcional 137
Composició del transcriptoma 138
Regulació post-transcripcional 140

Capítol 3 .... **143**

CONCLUSIONS 145

ANNEX I 149

Estudi de la divergència molecular entre diferents poblacions de Bothrops asper **151**
Introducció 151
Mètodes 152
Resultats i discussió 152

LAAO 153
SVSPs 154
SVMPs 155
PLA2 158

ANNEX II 159

REFERÈNCIES 185

ABREVIATURES I ACRÒNIMS 211

INTRODUCCIÓ

Les serps

Les serps pertanyen al subordre *Serpentes* dins l'ordre dels *Squamata* i, juntament amb els llangardaixos, conformen la classe *Reptilia*. Les al voltant de 3.150 espècies de serps identificades han aconseguit colonitzar un ampli ventall d'hàbitats de tot el món, excepte a Alaska i el nord de Canadà al continent americà, Groenlàndia, Islàndia, Irlanda i l'extrem nord d'Escandinàvia i Rússia a Euràsia, i l'Antàrtida i Nova Zelanda al sud (Figura 1). El coneixement de les relacions filogenètiques d'aquestes espècies esdevé una part vital en diferents àrees de la biologia com genètica de poblacions, ecologia o taxonomia. En aquest sentit, la classificació taxonòmica de les serps ha estat revisada diverses vegades en els últims anys [1–6] i esdevé un continu motiu de debat i conflicte (Figura 2). La dificultat d'obtenció de mostres o la consideració de dades moleculars, morfològiques i ecològiques entre altres factors, ha modificat classificacions prèvies fins al punt que estimacions filogenètiques d'institucions públiques internacionals com l'Integrated Taxonomic Information System (www.itis.gov), organisme creat l'any 1996 amb l'objectiu de subministrar informació consistent i fiable sobre la taxonomia de les espècies biològiques, difereixen de les que ofereixen bases de dades més especialitzades com la Reptile Database [7], base de dades mantinguda per Peter Uetz considerada de referència i actualitzada aproximadament cada 3 mesos.

Tanmateix, el clade *Colubroidea* inclou a més de dues terceres parts de les espècies de serps actuals, constituint un dels grups de vertebrats més divers i amb major èxit evolutiu que inclou, entre altres, les aproximadament 600 espècies de serps verinoses, la majoria de les quals pertanyen a les famílies *Elapidae* i *Viperidae*. Dins d'aquesta darrera família, cal destacar dues subfamílies com són els escurçons (de l'anglès *viper*) pertanyents a la subfamília *Viperinae* i els cròtals (de l'anglès *pit viper*), pertanyents a la subfamília *Crotalinae*. Els membres d'aquesta darrera subfamília constitueixen els únics membres de la família *Viperidae* que podem trobar al continent americà.

La herpetofauna de Costa Rica, a la qual pertanyen les espècies caracteritzades en aquest treball, inclou 138 espècies de serps

diferents, de les quals 22 són verinoses: 16 cròtals i 6 elàpids (5 terrestres pertanyents al gènere *Micrurus* i una marina, *Pelamis platura*) [8].

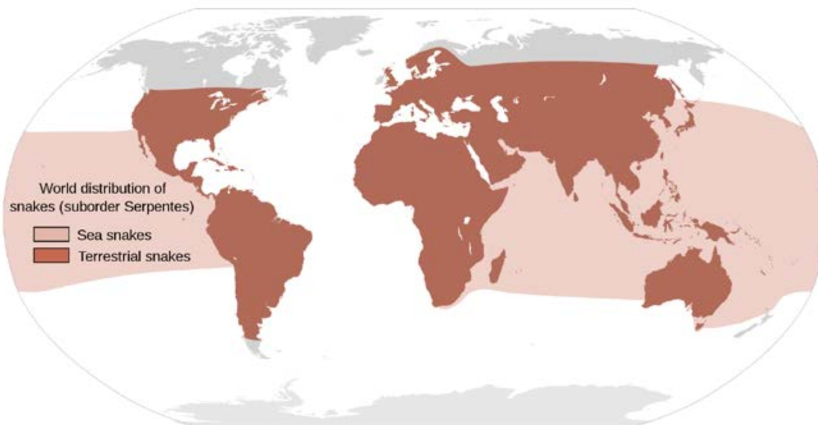


Figura 1. Distribució mundial del subordre *Serpentes*. Distribució d'espècies terrestres i marines de les serps. Veure el text per més informació. (Imatge obtinguda sota llicència Wikimedia Commons)

Des del punt de vista molecular la caracterització de les serps s'ha abordat, tal i com veurem més endavant, amb tècniques transcriptòmiques i proteòmiques, i només existeixen els esborranys dels genomes de dues espècies de serps, de les quals només una d'elles és verinosa [9]. A data de redacció d'aquesta tesi, l'abundància de dades referents a la família *Viperidae* al GenBank, del National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov), és de 16.700 seqüències de proteïnes, 58.127 seqüències de nucleòtids i 49 experiments amb dades provinents de Seqüenciació de Nova Generació depositades a la base de dades SRA (*Sequence Read Archive*) de 271 espècies de serps diferents.

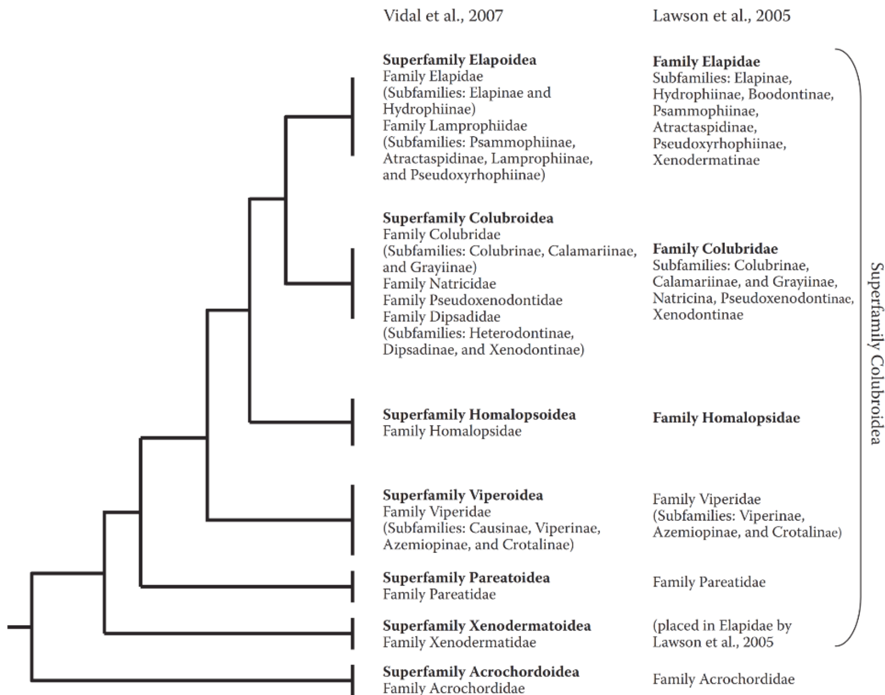


Figura 2. Filogènia de la superfamília Colubroidea. Modificacions successives del clade Colubroidea han posat de manifest la dificultat de classificació d'aquesta superfamília. (Imatge cedida per Stephen Mackessy des de [10])

Els verins

Estrictament, el verí pot ser definit com una secreció produïda en unes glàndules especialitzades i injectada a la presa mitjançant un aparell inoculador [11]. Tanmateix, la definició de "verí" està subjecta actualment a dues tendències diferenciades: una funcional segons la qual el verí és un compost tòxic injectat a la presa o al depredador que causa la seva mort o incapacitació ràpida, i una altra que busca acomodar o prioritzar evidències evolutives en la seva definició. Ambdues postures condueixen, inevitablement, a una categorització diferent dels animals verinosos i al diferent plantejament de qüestions

pel que fa la funció de les secrecions i la seva evolució, així com diferents plantejaments en la recerca farmacèutica [12].
Sigui com sigui, aquest verí suposa un complex còctel de compostos bioactius format en un 90% per unes poques famílies de proteïnes i pèptids, i un 10% de diferents components orgànics i inorgànics com sal, sucres, poliamines i neurotransmissors, que interfereix els processos fisiològics o bioquímics normals per tal de facilitar la captura de la presa o la defensa de l'animal productor. En aquest sentit, el verí actua com a element clau en la depredació per part de determinats organismes com llangardaixos, aranyes, escorpins, serps, insectes, cefalòpodes, gasteròpodes i cnidaris, mentre que suposa una adaptació defensiva en altres organismes com peixos, equinoderms i altres tipus d'insectes [13]. Aquesta diversitat taxonòmica al llarg de diferents fílums resalta la importància del verí com a una innovació evolutiva en el regne animal. Més concretament, la funció verinosa és considerada una innovació clau en el clade *Toxicofera* (del grec 'aquells que poseeixen verí'), que inclou les espècies pertanyents al subordre *Serpentes*, el subordre *Iguania* i l'infraordre *Anguimorpha*, i que va permetre la seva diversificació ecològica i el seu èxit evolutiu. L'origen de la funció verinosa en aquest clade apareix com un fenòmen únic fa aproximadament 170 milions anys [14, 15], cosa que va suposar una adaptació que va permetre la transició d'un mode mecànic (constricció) a un mode químic (verí) de matar i ingerir les preses en afectar als seus sistemes vitals amb l'alteració de funcions enzimàtiques crítiques i canals iònics.

Composició

L'evolució de les serps és, per tant, un procés íntimament lligat a l'evolució de la composició del verí, una marca de la història evolutiva. L'estudi del verí podria suposar, així, un valor taxonòmic [16] a més dels evidents beneficis clínics i de recerca, oferint una visió única de determinats camps de la biologia com la farmacologia (en el disseny de nous fàrmacs), la immunologia, la biologia estructural i l'ecologia, constituint models únics per a l'estudi de processos evolutius, tal i com veurem més endavant.

Els verins de les serps pertanyents a la família *Viperidae*, contenen proteïnes que interfereixen la cascada de coagulació, el sistema hemostàtic i la reparació de teixits [17, 18]. Aquestes proteïnes pertanyen a unes poques famílies de toxines multigèniques, que inclouen enzims i proteïnes sense activitat enzimàtica. L'abundància relativa d'aquestes proteïnes presenta una elevada variabilitat a nivell de gènere, d'espècie, poblacional i, inclús, una variabilitat relacionada amb l'estadi de desenvolupament de l'organisme (ontogenètica). Aquesta elevada variabilitat, juntament amb l'existència d'una gran varietat d'isoformes de cada família proteica, contribueix a la complexitat de la composició del verí i a la diversitat dels efectes biològics observats, descrits fa més de 20 anys per herpetòlegs i toxinòlegs [19]. Diferents treballs han intentat vincular aquesta variabilitat a la dieta [20–23] i també a mecanismes implicats en el control transcripcional i traduccional [24]. Sigui com sigui, diferents mecanismes evolutius participen en la creació d'aquesta diversitat de toxines [25]. Així, tot i que l'alteració en l'estructura de determinats dominis de gens de toxines o el seu empalmament alternatiu (de l'anglès *Alternative splicing*) poden participar en la generació de noves toxines, la duplicació gènica seguida d'una divergència funcional és la principal font de creació de diversitat molecular en el que es coneix com un procés de "naixement i mort" (de l'anglès *Birth and Death*) [26]. D'aquesta manera, la duplicació crea una redundància que permet a la còpia gènica ser expressada de forma selectiva a la glàndula del verí, escapant així de la selecció negativa i evolucionar cap a una nova funció, mantenint uns nivells elevats de variabilitat en les proteïnes del verí com una part essencial de l'adaptació de les serps a diferents preses i diferents ecosistemes [14, 27]. Aquesta evolució de les diferents famílies de toxines del verí va sovint acompanyada d'una evolució accelerada [28, 29] (principalment dels residus exposats a la superfície de la proteïna) a través d'una selecció positiva [30]. En aquest sentit, la selecció positiva sembla ser un mecanisme de neofuncionalització estès en la majoria d'organismes verinosos incloent serps, aranyes, escorpins i espècies del gènere *Conus*, on les substitucions no sinònimes d'aminoàcids són més nombroses que les substitucions sinònimes en els fragments codificants del gen, cosa que permet a diferents famílies

de toxines com les Fosfolipases A_2 (PLA2) ([31]) i les Lectines tipus C (CTL) ([32]) presentar diferents activitats biològiques com per exemple, miotoxicitat, cardiotoxicitat, neurotoxicitat i activitats anti- i procoagulants.
En els següents apartats intentarem descriure breument aquelles proteïnes del verí amb activitat enzimàtica i sense, que juguen un paper fonamental en els transcriptomes analitzats en el present treball.

Proteïnes amb activitat enzimàtica

Metaloproteases del verí de les serps

Les metaloproteases del verí de les serps (SVMPs, de l'anglès *Snake Venom MetalloProteinases*) són proteïnes multidomini depenents de Zn^{2+}, d'una mida aproximada d'entre 20 i 100 kDa que, juntament amb les ADAM cel·lulars (de l'anglès *A Disintegrin And Metalloprotease*), s'inclouen dins de la subfamília M12 (reprolisines) de les metaloproteases [33]. De fet, es considera que les SVMPs van evolucionar a partir de les ADAMs [34] com a resultat de la pressió adaptativa sobre les serps com a estratègia de divergència funcional tal i com s'ha comentat anteriorment [14].
L'existència de les SVMPs ja va ser descrita a la dècada dels 60 del segle passat [35], i actualment, es consideren unes toxines clau en la patogènesi (local i sistèmica) observada en l'enverinament causat per membres de la família *Viperidae*, arribant a constituir fins a un 32% de les proteïnes del verí [36].
La majoria de les activitats funcionals associades amb les SVMPs estan vinculades amb la disrupció de l'hemostàsia per degradació de diferents components de la matriu extracel·lular de les cèl·lules endotelials dels capil·lars, produint, en darrer terme, hemorràgia i danys a la regeneració muscular com a conseqüència dels efectes sobre el sistema vascular. També s'han descrit altres funcions relacionades amb les SVMPs que impliquen la inhibició de l'agregació plaquetària, apoptosi i inflamació.
Des del punt de vista molecular, les diferents SVMPs existents es diferencien, essencialment, en la seva mida, condicionada per l'existència o no de determinats dominis proteics. D'aquesta manera,

tal i com podem observar a la Figura 3, les SVMPs es classifiquen estructuralment en 3 tipus diferents: les PI, les PII i les PIII, podent aparèixer de forma simultània en el mateix verí, constituint un tret diferencial de cada verí. Tot i que la presència o absència de determinats dominis no és una característica de la capacitat hemorràgica del verí, en general, les PIII han demostrat ser la més potent de les 3 classes.

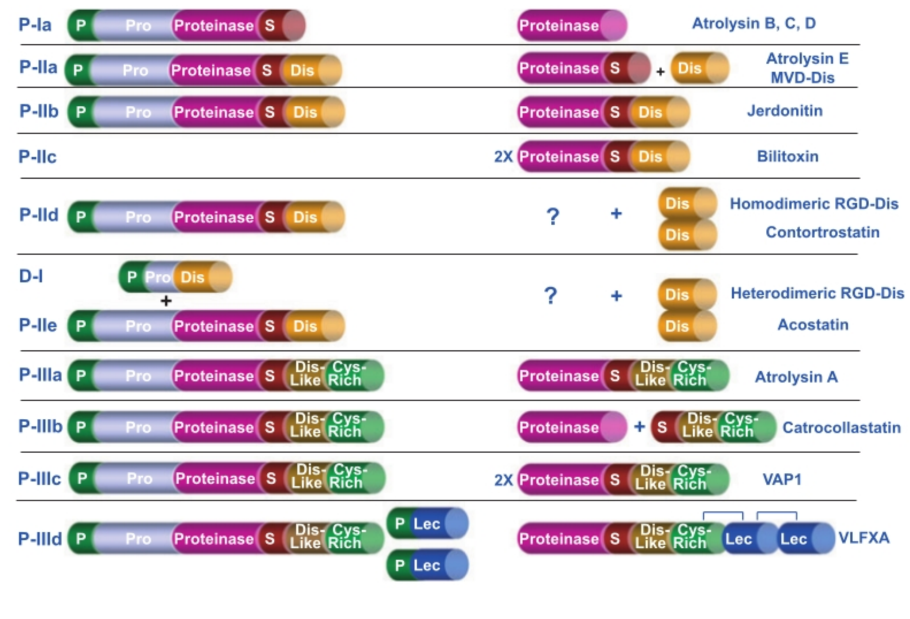


Figura 3. Classificació estructural de les Snake Venom Metalloproteinases. Dins de les 3 classes majoritàries, les diferents subclasses apareixen en funció del processament proteolític. (Imatge obtinguda de [36])

Fosfolipases A_2

Les Fosfolipases A_2 (PLA2, E.C.3.1.1.4) catalitzen la hidròlisi dels glicerofosfolípids donant lloc a lisofosfolípids i àcids grassos. Es classifiquen en 15 grups diferents en funció de diferentes paràmetres com estructura, seqüència d'aminoàcids, activitat catalítica i expressió, que s'agrupen en 4 categories principals com són les

fosfolipases secretades (sPLA2), les citosòliques (cPLA2), les Ca^{2+}-independents (iPLA2) i les acetil-hidrolases del factor activador de plaquetes (PAF-AH) [37].
En mamífers, les PLA2 juguen un paper fonamental en determinats processos cel·lulars com transducció del senyal o manteniment de la membrana entre d'altres. Tanmateix, també són unes de les toxines més potents dels verins de serps, escorpins, abelles i llangardaixos, on resulten essencials per a la immobilització i digestió de la víctima, amb múltiples efectes farmacològics com neurotoxicitat, miotoxicitat, efectes anticoagulants i activitat hemolítica entre d'altres. Aquesta diversitat d'activitats va evolucionar a partir d'un procés de neofuncionalització d'una PLA2 no tòxica, tal i com hem vist anteriorment en el cas de les SVMPs.
Segons la seqüència d'aminoàcids, estructura tridimensional i patró de ponts disulfur, les PLA2 del verí de les serps pertanyen als grups I (serps del Vell Món, majoritàriament de la família *Elapidae*) i al grup II (serps del Nou Món, majoritàriament de la família *Viperidae*). Els enzims pertanyents a aquest darrer grup estan formats per 120 – 125 aminoàcids amb 14 cisteïnes que formen 7 ponts disulfur i estan subdividits alhora en base a la seva expressió específica de teixit tal i com podem veure a la Taula 1.

II A	Trobada al líquid sinovial humà, al verí de les serps de cascavell i dels escurçons de la família *Viperidae*
II B	Trobada al verí de l'escurçó del Gabon (*Bitis gabonica*)
II C	Trobada als testicles de rata/ratolí
II D	Trobada al pàncrees i melsa humà/ratolí
II E	Trobada a cervell, cor i úter humà/ratolí
II F	Trobada a l'embrió i testicles de ratolí

Taula 1. Classificació de les PLA2 II. (Taula modificada de [38])

Les PLA2 que trobem al verí de les serps pertanyents a la família *Viperidae* (PLA2 IIA) es classifiquen en diferents subgrups segons l'aminoàcid existent a la posició 49. El residu existent en aquest centre actiu d'unió al Ca^{2+} determinarà la capacitat hidrolítica de la proteïna,

podent diferenciar entre PLA2 Asp49 i els anomenats homòlegs PLA2 (K49, S49,N49 i R49).

L-aminoàcid oxidases

Les L-aminoàcid oxidases (LAAO, E.C.1.4.3.2) són unes glicoproteïnes d'unió a FAD- (flavin adenina dinucleòtid) o FMN- (flavin mononucleòtid) que catalitzen la desaminació oxidativa d'un ampli ventall de L-aminoàcids per formar el corresponent $\alpha-$cetoàcid, amoni i peròxid d'hidrògen, responsable últim de la majoria d'efectes biològics atribuïbles a les LAAOs. Les flavines de les LAAOs són les responsables del color groc del verí de les serps i contribueixen a la toxicitat del verí a través de l'estrès oxidatiu derivat de la producció de peròxid d'hidrògen.
Les LAAOs estan àmpliament distribuïdes en bactèries, fongs i plantes, tot i que els verins de les serps són, probablement, una de les fonts més riques d'aquests enzims, i han estat objecte d'estudis biomèdics per la seva activitat antimicrobiana, anticoagulant, inductora / inhibidora de l'agregació plaquetària i inductora de l'apoptosi cel·lular [39].

Serines Proteases

Les serines proteases del verí de les serps (SVSPs, de l'anglès *Snake Venom Serine Proteinases*) comprenen un ampli grup d'enzims d'una mida aproximada de 30 kDa, que catalitzen reaccions implicades en la cascada de coagulació, el sistema del complement, el sistema fibrinolític i l'agregació plaquetària entre d'altres. Són proteïnes que pertanyen a la família S1 de les peptidases que es troben presents tant en procariotes com eucariotes, que poden arribar a representar fins el 20% de les proteïnes presents al verí de les famílies *Viperidae*, *Crotalidae*, *Elapidae* i *Colubridae*. Tal i com succeeix en el cas de les SVMPs i les PLA2, la diversitat de farmacologia associada a aquests enzims és el resultat d'una duplicació gènica seguida d'evolució accelerada.

Es caracteritzen per un mecanisme catalític comú en el qual participen un residu de serina altament reactiu estabilitzat per la presència d'un residu d'histidina i un aspàrtic en el centre actiu.
Els desordres de la coagulació poden tenir lloc en diferents nivells de la cascada de coagulació, però entre els verins dels rèptils, el fibrinògen n'és la diana principal. D'entre les mútiples funcions de les SVSPs, la depleció del fibrinògen per les anomenades TL-SVSPs (de l'anglès, *Thrombin-like Snake Venom Serine Proteinases*) n'és una de les més ben estudiades [40].

Nucleases, nucleotidases i fosfomonoesterases

Diferents enzims hidrolítics com DNAses, RNAses, fosfodiesterases i nucleotidases entre d'altres, estan presents en pràcticament tots els verins de serps, l'activitat farmacològica dels quals no ha estat encara ben caracteritzada, podent presentar activitats específiques solapants entre ells. L'aparent manca d'interès per part de la comunitat científica en aquests enzims rau en el fet que, probablement, participen només en processos de digestió i no són, per tant, tòxics.
Les nucleases són enzims que actuen sobre els àcids nuclèics. En el verí de les serps trobem tant endonucleases com exonucleases. El primer tipus, ja descrit per Delezenne i Morel l'any 1919, inclou les DNAses (E.C. 3.1.21.1) i les RNAses (E.C.3.1.21.-) en funció de l'especificitat per un tipus o altre de material genètic, mentre que les exonucleases inclouen les fosfodiesterases (E.C.3.1.4.1), enzims que catalitzen la hidròlisi d'RNA i de DNA indistintament.
Per la seva banda, les nucleotidases més àmpliament estudiades són les 5'-nucleotidases (E.C.3.1.3.5), enzims d'entre 73 i 100 kDa que catalitzen la hidròlisi del fosfat esterificat al carboni 5' de la ribosa i desoxiribosa dels nucleòtids. La seva presència al verí de les serps ja va ser descrita per Gulland i Jackson l'any 1938 [41] i des de llavors s'ha trobat en nombroses espècies, sent més abundant en verins de vipèrids que en elàpids. Existeix una evident mancança d'informació de l'activitat biològica de les 5'-nucleotidases en el verí de les serps, però determinats estudis semblen apuntar a la inhibició de l'agregació plaquetària.

De la mateixa manera, el verí conté enzims que poden degradar ràpidament l'ATP alliberat per determinades miotoxines a adenosina: són les ATPases i ADPases, enzims que participen en processos d'alliberament d'aquestes purines endògenes, potenciant la hipotensió i paràlisi induïda pel verí [42, 43]. L'avantatge evolutiva d'expressar en el mateix verí miotoxines i nucleotidases és, doncs, evident.
Finalment, les fosfomonoesterases són enzims que catalitzen la hidròlisi no específica dels èsters de fosfat i ja van ser descrits al verí de les serps l'any 1932 per Uzawa. Són enzims dividits en dues classes com són les fosfomonoesterases àcides i les fosfomonoesterases alcalines, sent aquesta darrera classe la més àmpliament distribuïda en els verins de les serps. Actualment no existeix informació de l'aïllament i caracterització de l'activitat biològica de les fosfomonoesterases del verí de les serps, tot i que el mateix enzim en el verí de les abelles és considerat un al·lèrgen i un potent alliberador d'histamina en basòfils humans.

Glutaminil ciclases

Les Glutaminil ciclases (GC, E.C.2.3.2.5) són aciltransferases de 33-40 kDa altament conservades des de llevats fins a humans que catalitzen la ciclació d'un residu de glutamina a l'extrem N-terminal en el processament de nombroses proteïnes com a mesura de prevenció davant l'acció d'exopeptidases [44]. Les GC han estat identificades en diferents proteomes de membres de la família *Viperidae* [45], on probablement juguen un paper fonamental en la biosíntesi de determinats pèptids hipotensius i inhibidors endògens de les metaloproteases. Així, les GC podrien no actuar com a toxines per elles mateixes, sinó participar en el processament post-traduccional de determinades toxines per tal d'aconseguir la conformació adequada o en la seva protecció de l'autodegradació.

Proteïnes sense activitat enzimàtica

Pèptids natriurètics i potenciadors de la bradiquinina

Els pèptids potenciadors de la bradiquinina (BPPs, de l'anglès *Bradykinin potentiating peptides*) són uns oligopèptids rics en prolina que inhibeixen l'enzim convertidor de l'angiotensina (ACE, E.C.3.4.15.1), una metalopeptidasa de la membrana de les cèl·lules endotelials que participa en la regulació de l'homeòstasi cardiovascular de l'organisme per la seva participació en dues rutes metabòliques diferents: per una banda, l'ACE catalitza la conversió d'angiotensina I en angiotensina II, un potent vasoconstrictor, i, per altra banda, metabolitza la bradiquinina inactivant una via metabòlica que podria conduir a un xoc hipotensiu. De fet, la importància de l'ACE en la regulació de la pressió sanguínia ja va descrita per Ferreira i col·laboradors, senyalant l'enzim com una potencial diana terapèutica [46]. D'aquesta manera, els BPP prevenen l'efecte hipertensiu de l'angiotensina II i potencien l'efecte hipotensiu de la bradiquinina circulant. Es troben àmpliament descrits en els verins de múltiples espècies de serps, i, tal i com veurem, la seva caracterització l'any 1965 va donar lloc al desenvolupament de fàrmacs antihipertensius com el Captopril® [47] i a la caracterització de determinades dianes terapèutiques involucrades en la disminució de la pressió arterial sense afectar l'activitat de l'ACE [48].
Els pèptids natriurètics (NPs) per la seva banda, són uns potents reguladors cardiovasculars i osmoreguladors en vertebrats, i han estat descrits en el verí de diferents espècies de serps amb funcions vasodilatadores.

Disintegrines

Les disintegrines representen una família de polipèptids d'una longitud aproximada de 40 – 100 aminoàcids, riques en cisteïnes que s'alliberen al verí per processament proteolític de les SVMPs de tipus PII [49] o sintetitzades a partir d'mRNAs mancats d'una regió codificant per una metaloproteasa [50], la complexitat estructural i funcional de les quals contrasta amb la seva petita mida molecular.

Descrites l'any 1987 per les investigacions d'Stefan Niewiarowski i Tur-Fu Huang [51], les disintegrines poden ser classificades en funció de la seva longitud i el nombre de ponts disulfur, donant lloc a disintegrines curtes, mitjanes, llargues i disintegrines dimèriques.
Constitueixen una família de proteïnes antagonistes de receptors de la família de les integrines, afectant així a diferents processos com angiogènesi, formació de trombos, integritat de la pell i sistema immunitari entre d'altres, amb nombroses aplicacions biomèdiques en l'estudi de diferents processos biològics en els quals les integrines juguen un paper fonamental. Tal i com veurem més endavant, s'han desenvolupat antagonistes sintètics de les integrines de baix pes molecular, actualment usats en el tractament i prevenció de processos tromboembolítics.

Proteïnes Semblants a Lectines tipus C

Tot i que semblants en la seva nomenclatura, les Lectines tipus C són diferents de les Proteïnes Semblants a les Lectines tipus C presents al verí de les serps (CLPs, de l'anglès, *C-type lectin-like proteins*), que Kenneth J. Clemetson anomena "snaclecs" (de l'anglès, *Snake Venom C-type lectins*) per tal d'evitar la confusió [52]. Així, les Lectines tipus C són homodímers depenents de Ca^{2+}, sense activitat enzimàtica, amb dominis d'unió a carbohidrats que trobem a molts organismes eucariotes. Per contra, les CLPs presents al verí de les serps són heterodímers mancats de la funció d'unió a carbohidrats tot i l'elevada homologia de seqüència (15 - 40%) amb aquests dominis, que mostren un ampli ventall d'activitats farmacològiques que afecten les plaquetes, el sistema hemostàtic i la trombosi [53].
Les CLPs juguen un paper important en l'efecte hemorràgic causat pels verins dels membres de la família *Viperidae* i contribueixen a la disrupció de l'hemostàsia amb l'activació o inhibició de determinats components cel·lulars.

Factors de creixement

L'existència de factors de creixement al verí de les serps ja va ser observada l'any 1956 [54], estudis que van ser mereixedors del premi

Nobel de Medicina l'any 1986 per la seva relació amb la regulació del creixement cel·lular. Tot i això, la purificació d'un d'aquests factors de creixement (NGF, de l'anglès *Nerve Growth Factor*) no es va produir fins a 20 anys després [55] i la seva cristal·lització va ser possible l'any 1991 [56]. A més, no ha estat fins recentment, que s'han trobat evidències d'un patró d'evolució accelerada, característic, tal i com hem vist, d'altres components del verí [57].
Actualment els factors de creixement s'han trobat com a constituents del verí dels membres de la família *Elapidae* (i en menor grau de la família *Viperidae*), amb funcions hipotensores i potenciadores de la permeabilitat vascular.

CRISP (Cysteine-Rich Secretory Proteins)

Les CRISPs són polipèptids de cadena senzilla d'un pes molecular aproximat de 20 – 30 KDa, conservats en diferents espècies del regne animal des de mamífers fins al verí de diferents rèptils, excercint diferents funcions biològiques [58]. Tanmateix, malgrat la caracterització de la proteïna fa més de 30 anys [59], existeix una evident mancança de coneixement de les funcions fisiològiques de les CRISP en els mamífers, on podrien participar en determinats processos involucrats amb l'aparell reproductor, i del paper que juguen aquestes proteïnes en el verí de les serps, on podrien participar en el bloqueig de determinats canals iònics i en la contracció de la musculatura llisa.

Aplicacions

Els animals verinosos han despertat la curiositat i fascinació de la humanitat al llarg dels segles. Els usos mèdics i bèl·lics dels verins estan documentats des de temps antics en diferents cultures com l'Antic Egipte, Grècia, la Xina o l' Islam. L'apiteràpia, és a dir, l'ús del verí de les abelles per al tractament de malalties que cursen amb processos inflamatoris, sembla que ja va ser usada per Hipòcrates o Galè. Més endavant, el mateix Carlemany, emperador d'Occident a l'Edat Mitjana, va ser tractat amb picades d'abelles per alleugerir el dolor a les seves articulacions. De la mateixa forma, l'ús terapèutic del

verí dels escorpins està ben documentat en remeis casolans o la medicina tradicional xinesa [60, 61].
Per la seva banda, la serp és el símbol de la medicina per la seva associació amb Asclepi, el Déu grec de la medicina, les observacions del qual van permetre descobrir les propietats curatives de substàncies verinoses i tòxiques administrades en dosis molt petites. El primer coneixement de l'ús mèdic dels efectes coagulants del verí de les serps data de l'any 67 a.C. : en el seu llibre titulat "The Poison King: The Life and Legend of Mithradates, Rome's Deadliest Enemy" [62], Adrienne Mayor descriu com el rei Mitridates VI va ser tractat de les hemorràgies causades per les ferides sofertes durant les guerres Mitridàtiques amb verí de la serp *Vipera ursini*. Aquest excitant descobriment va suposar l'anticipació de la toxinologia moderna més de 1500 anys.
Segles més tard, Francesco Redi (1626 – 1697) metge de la cort de Ferran II de Medici, va descriure com els escurçons produïen el verí i l'injectaven a les seves preses i els efectes coagulants d'aquest sobre la sang de la víctima. Les seves observacions, publicades l'any 1664, marquen el naixement de la toxinologia moderna [63]. Tanmateix, va fer falta un segle més per tal que Abbé Gasparo Ferdinando Felice Fontana (1730 – 1805), primer director del Museu d'Història Natural de Florència, portés a terme experiments amb el verí dels escurçons europeus *Vipera berus* i *Vipera aspis*. El seu text, *Traité sur le Vénin de la Vipère*, publicat l'any 1781 és considerat una obra clau de la toxinologia moderna [64].
Els verins representen, així, una sèrie de reservoris inexplorats de components bioactius que han estat seleccionats per la seva capacitat de ser secretats, per la seva estabilitat, plasticitat funcional i la capacitat d'establir interaccions amb les potencials dianes, i que només en les últimes dècades han estat estudiats, no només com a font natural per a la generació de productes farmacèutics, sinó també per tal d'entendre les bases moleculars de l'evolució.
Aquesta recerca biomèdica va donar lloc durant la dècada dels 70 a l'aparició del Captopril®, el primer fàrmac inhibidor de l'enzim convertidor de l'angiotensina, prescrit per al tractament de la hipertensió, obtingut a partir dels pèptids potenciadors de la bradiquinina trobats al verí de l'escurçó brasiler *Bothrops jararaca*

[47] i que va suposar el premi Albert Lasker en recerca biomèdica per a David Cushman i Miguel Ondetti l'any 1999 [65].
Un altre exemple de l'obtenció d'un fàrmac derivat d'una toxina és l'analgèsic Zicotonida (Prialt®), un pèptid bloquejant dels canals de calci (derivat sintètic d'una ω - conotoxina) provinent del verí del cargol marí (*Conus magus*) aprovat el 2004 per la FDA (US Food and Drug Administration) per al tractament del dolor crònic per a pacients refractaris a la morfina. D'altra banda, els fàrmacs Eptifibatidae (Integrilin®) i Tirofiban (Aggrastat®), fàrmacs derivats de components dels verins de les serps *Sistrurus miliarius* i *Echis carinatus* respectivament, són actualment usats en el prevenció de la síndrome coronària aguda en tractar-se d'inhibidors de la integrina $\alpha_{IIb}\beta_3$ i, per tant, antiagregants plaquetaris.
Altres toxines estan actualment en una fase avançada dels seus estudis clínics, estudis majoritàriament centrats en els camps del dolor, sistema vascular, infecció o càncer [66–68]. Així, la clorotoxina TM-601 és un pèptid bloquejant dels canals de clor de 36 aminoàcids de longitud provinent del verí de l'escorpí *Leiurus quinquestriatus* que podria resultar útil per al tractament i diagnòstic de diferents tipus de càncer com, per exemple, el glioma i que actualment ha superat la fase II dels estudis clínics. Per altra banda, les mambalgines, proteïnes obtingudes del verí de la mamba negra (*Dendroaspis polylepis polylepis*), mostren una potent funció analgèsica bloquejant els canals iònics de detecció d'àcid involucrats en diferents rutes metabòliques implicades en la sensació del dolor [69]. Aquestes proteïnes, la funció analgèsica de les quals pot ser tant efectiva com la morfina, suposen noves propostes contra el tractament del dolor amb un enorme potencial terapèutic.
Determinats components dels verins s'estan usant també amb finalitats diagnòstiques [70]: determinades proteïnes amb propietats anticoagulants del verí de les serps *Bothrops atrox*, *Echis carinatus*, *Oxyuranus scutellatus* i *Daboia russelii* són usades per al diagnòstic de trastorns en la coagulació així com per a fer el seguiment de determinats tractaments amb teràpies anticoagulants. És l'exemple de la botrocetina, una proteïna amb una forta capacitat d'agregació plaquetària trobada al verí de *Bothrops atrox* i usada per al diagnòstic de patologies vasculars d'origen genètic com per exemple la malaltia

de von Willebrand. Degut a la seva insensibilitat a l'heparina, Prefakit®Reptilase®Time, un assaig funcional que conté unes quantitats estandaritzades de l'enzim proteolític batroxobina (l'origen del qual es troba al verí de *Bothrops atrox*), és recomanat per a l'estudi de la darrera fase de la coagulació sanguínia.
Lògicament, la importància mèdica de l'estudi dels verins també rau en la producció dels antiverins. L'adequat tractament d'una mossegada de serp (la incidència de les quals veurem amb més detall al proper apartat) depèn de la capacitat dels antiverins per revertir la simptomatologia associada a un enverinament respecte la coagulopaties, hemorràgies o xoc hipotensiu. Els antiverins anomenats de primera generació, ja descrits fa més de 100 anys, provenien del sèrum no purificat d'animals hiperimmunitzats amb verí [71, 72]. Els antiverins actuals consisteixen en immunoglobulines purificades, la qual cosa ha reduït la incidència de reaccions adverses com, per exemple, xocs anafilàctics. Tanmateix, els protocols d'immunització no han canviat en aquests més 100 anys. Conseqüentment, els antiverins presenten anticossos redundants contra molècules no tòxiques del verí i poca capacitat neutralitzadora contra petites molècules altament tòxiques però molt poc immunogèniques. Això fa necessàries elevades quantitats d'antiverí per a un tractament efectiu amb el risc de les reaccions adverses ja comentades. D'altra banda, les mescles d'immunització usades actualment són específiques de cada país o regió degut a la variabilitat interespecífica del verí i al fet que diferents serps són responsables de la majoria dels accidents en diferents països [73]. De la mateixa manera, les manifestacions clíniques de la mossegada de serp d'una mateixa espècie en diferents àrees geogràfiques són heterogènies, donant lloc a una variabilitat intraespecífica del verí que s'observa també dels individus adults als juvenils (variabilitat ontogenètica). El bon coneixement de la composició del verí i dels seus perfils immunogènics, així com el seguiment d'unes directrius per a la producció, control i regulació [74] dels antiverins, seran fonamentals per a la seva millora pel que fa el seu disseny i la seva eficàcia [75].
Queda doncs clar, que l'aprofitament dels secrets ocults dels verins passa per una caracterització exhaustiva dels seus components. Actualment, aquesta caracterització s'està duent a terme en diferents

laboratoris d'arreu del món mitjançant l'estudi tant del transcriptoma com dels components finals existents al verí mitjançant tècniques proteòmiques. Un exemple de la importància d'aquest tipus de recerca és la participació de diferents entitats de 5 països europeus en el projecte *Venomics*, que en el marc del FP7-HEALTH, es proposa la identificació de nous productes terapèutics en diferents organismes verinosos, des d'insectes fins a organismes clàssicament considerats verinosos com serps, aranyes o escorpins, amb una estimació potencial de més de 40 milions de components bioactius (principalment pèptids i proteïnes), dels quals, actualment, se'n coneixen només 3.000.

Enverinament i símptomes

L'enverinament per mossegada de serp representa una emergència mèdica que afecta especialment a les comunitats rurals de països en desenvolupament a l'Àfrica sub-Sahariana, Àsia i Amèrica Llatina.
La seva incidència global, obtinguda mitjançant registres hospitalaris, s'estima en 421.000 casos d'enverinament anuals amb, almenys, 20.000 morts a tot el món [76]. Recentment, però, l'Organització Mundial de la Salut (OMS) ha publicat un estudi [77] on fixa aquesta taxa en 5 milions d'accidents que resulten en aproximadament 2,5 milions d'enverinaments i al voltant de 100.000 morts a tot el món (Figura 4). Tanmateix, la veritable incidència d'enverinament per mossegada de serp a nivell mundial, el seu impacte i les característiques en les diferents regions, són desconeguts degut a que en moltes zones rurals les víctimes no van a l'hospital i moren a casa seva sense ser registrats. El nombre d'amputacions permanents i d'incapacitats per les seqüeles i infeccions secundàries podria representar tres vegades més [78]. De fet, el nombre d'amputacions només a Àfrica com a conseqüència d'un accident per mossegada de serp s'estima en 8.000 [79], amb la conseqüent discriminació i exclosió social de les víctimes amb una deformitat o impediment físic permanent.
Per altra banda, pocs estudis han abordat les conseqüències econòmiques d'aquestes situacions [80], però l'impacte econòmic i

social contribueix significativament al cercle viciós que suposa el fet que siguin precissament les zones més empobrides del planeta les que pateixen de forma més important les conseqüències dels accidents per mossegada de serp. De fet, la relació entre pobressa i mortalitat per mossegada de serp està clarament demostrada, existint una correlació linial entre aquesta mortalitat i determinats marcadors socio-econòmics com el Producte Interior Brut (PIB) del país, la despesa governamental en salut i el percentatge d'ocupació en agricultura [81]. Tot i això, els accidents per mossegada de serp han rebut poca atenció per part de les autoritats de salut pública i la indústria farmacèutica, fins al punt de poder categoritzar-la com a una de les malalties més oblidades del segle XXI. Tant és així, que l'OMS va incorporar a l'Abril del 2009 [82] els accidents per mossegada de serp en la llista de condicions descuidades en el conjunt de les malaties tropicals descuidades (NTDs de l'anglès *Neglected Tropical Diseases*) juntament amb altres com per exemple el dengue o la malaltia de Chagas [83]. Les NTDs representen un grup de problemes de salut que comparteixen determinades característiques distintives des del punt de vista demogràfic, sociològic, epidemiològic i social, que afecten a més d'un bilió de persones a tot el món. Per tal de fer possible els objectius del Milennium Development Goals de les Nacions Unides [84], s'han dut a terme diverses estratègies en el marc del pla global de lluita contra les NTDs 2008 – 2015 [85], que inclouen, entre altres, la teràpia preventiva, el control dels vectors i la provisió d'aigua en bones condicions. Aquestes estratègies han permés reduir en més d'un 99% la incidència de determinades malaties infeccioses com per exemple la dracunculiasis. Tanmateix, els accidents per mossegada de serp han estat exclosos d'aquests esforços de les autoritats sanitàries i empreses farmacèutiques, probablement pel seu origen no infecciós. Aquesta interpretació resulta errada si considerem els trets principals que comparteixen les conegudes com a "clàssiques" NTDs i els accidents per mossegada de serp [86]. Amb la finalitat d'establir un marc global de col·laboració entre científics, empreses productores d'antiverins, diferents organitzacions civils, l'OMS i les autoritats públiques sanitàries, la Global Snakebite Organization [87] pretén millorar el coneixement i la prevenció dels accidents per mossegada

de serp, així com l'accés d'uns antiverins eficients a les comunitats més pobres del planeta [88].
A l'estat espanyol els accidents per mossegada de serp constitueixen un quadre poc freqüent amb una incidència d'aproximadament 100 accidents anuals amb una mortalitat anual d'entre 3 i 7 individus [89, 90]. A nivell europeu, una recent publicació [91], conseqüència d'una recerca exhaustiva de la literatura existent entre els anys 1970 i 2010 tant en revistes indexades com no indexades, conclou que no existeix variació entre la incidència de mossegada de serp entre el nord i el sud d'Europa, que es situa en aproximadament 8.000 accidents anuals amb aproximadament 4 morts a l'any.

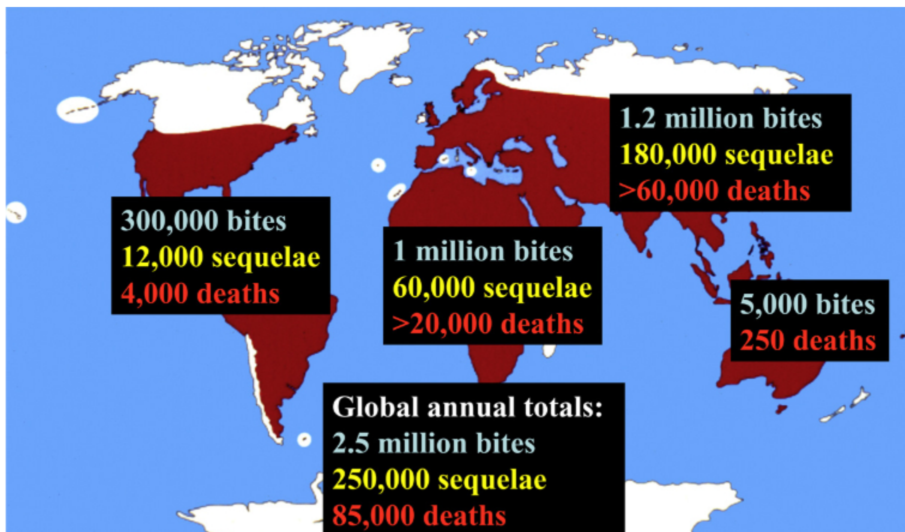


Figura 4. Incidència mundial dels accident per mossegada de serp. Estimació regional i global de la incidència i mortalitat de la mossegada de serp. (Imatge obtinguda de [92])

Pel que fa els símptomes, les víctimes per mossegada de serp de la família *Viperidae* pateixen danys del teixit local caracteritzats per dermomionecrosi, formació de butllofes, edema i hemorràgia local. En casos severs, les afectacions a nivell sistèmic impliquen defibrin(ogen)ació, trombocitopènia, hipoagregació plaquetària,

sagnat distant del lloc de la mossegada, coagulació intravascular disseminada, xoc cardiovascular i fallada aguda renal [93] .

Transcriptòmica i proteòmica

Les anomenades tècniques "òmiques" han revolucionat la recerca biològica. Les millores tant en la metodologia com en la tecnologia, han ampliat el rang dels estudis biològics des de simples anàlisis bioquímiques que implicaven unes poques molècules o uns pocs gens a la vegada, a un estudi sistemàtic de genomes, transcriptomes i proteomes, permetent una visió molt més global de l'organisme.
Per explorar els components del verí, diversos laboratoris han dut a terme l'anàlisi proteòmic i transcriptòmic del verí i de les glàndules productores respectivament. De forma individual o combinada, la proteòmica i la transcriptòmica han demostrat la seva fiabilitat per explorar en profunditat la composició dels verins. Un recull recent de les dades proteòmiques i transcriptòmiques actualment disponibles en més de 116 espècies diferents de serps podem trobar-lo a [94].
Així, l'ús de dades proteòmiques per estudiar els sis possibles marcs de lectura de les traduccions de dades transcriptòmiques, en el que hem anomenat "proteotranscriptòmica" [95] per contraposició amb el terme "proteogenòmica" [96], apareix com una combinació d'ambdues metodologies que ens permet estudiar la relació entre el proteoma i el transcriptoma per tal d'analitzar més profundament la complexitat del verí, en una evolució des del coneixement de la composició proteïca del verí, a la consideració de fenòmens biològics (evolutius i geogràfics entre d'altres) en el que s'ha anomenat vènomica de segona generació (de l'anglès, *Next-Generation Snake Venomics*).

Transcriptòmica

Els organismes han de canviar la seva expressió gènica d'acord a les condicions ambientals o en funció del seu creixement i desenvolupament. El transcriptoma, és a dir, el conjunt de totes les molècules d'RNA de la cèl·lula, apareix com a un component dinàmic

de la cèl·lula, en contraposició al genoma, més estàtic. El seu estudi esdevé, així, un mecanisme general de caracterització, no només de l'expressió cel·lular sinó de l'estructura i regulació gènica, que pot revelar molt sobre la dinàmica dels processos cel·lulars actius en determinades malalties.
La manera d'abordar aquesta caracterització del transcriptoma ha estat llargament condicionada per l'existència o no d'un genoma de referència i, tal i com veurem més endavant, ha estat lligada a l'evolució de la tecnologia de seqüenciació del DNA (Figura 5). En el cas d'organismes amb una seqüència genòmica coneguda, la hibridació de seqüències de DNA immobilitzades en xips de DNA (de l'anglès *DNA microarrays*) amb seqüències de cDNA marcades amb fluorescència, ha permès l'anàlisi de l'expressió gènica de milers de gens en un únic experiment al llarg de més de 15 anys. L'ús de xips de DNA per a la caracterització de l'expressió gènica va ser descrit per primera vegada l'any 1995 [97], i dos anys més tard va permetre l'anàlisi d'un genoma eucariota sencer [98]. Tanmateix, aquest tipus de tecnologia té determinades limitacions (Taula 2) com són la necessitat del coneixement de les seqüències genòmiques per al disseny de les sondes, un elevat soroll de fons degut a la hibridació creuada i un baix rang dinàmic de detecció degut a aquest soroll de fons i a la saturació del senyal.
L'estudi transcriptòmic de les glàndules del verí de les serps ha confiat per contra, des dels treballs pioners de Ho i col·laboradors l'any 1995 [100], en la seqüenciació Sanger de clons seleccionats aleatòriament d'una llibreria de cDNA construïda per transcripció reversa de les molècules d'RNA expressades al verí [101].

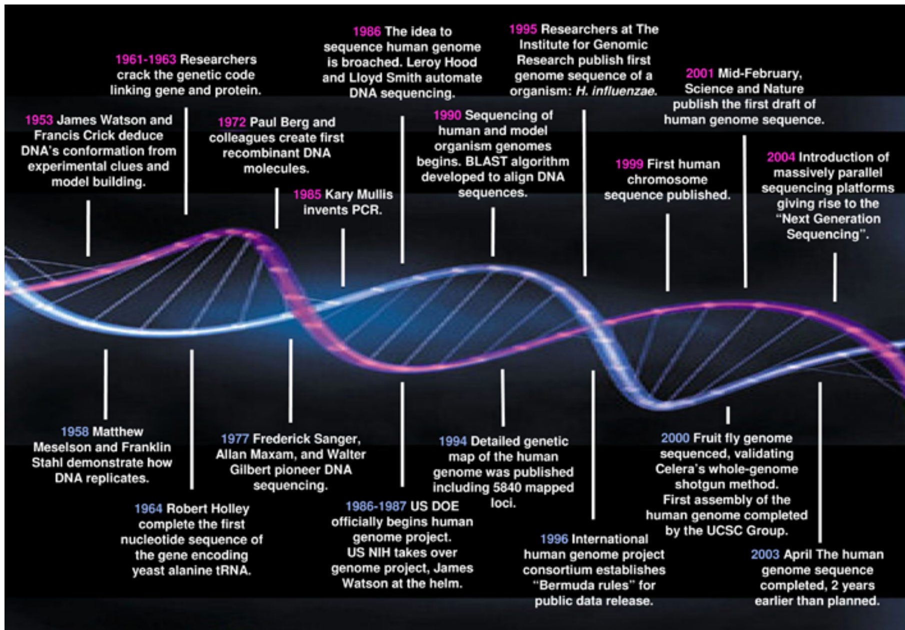


Figura 5. Evolució de la seqüenciació de DNA. Principals fites en la manipulació del DNA des del seu descobriment fins a la seva seqüenciació de forma massiva. (Imatge cedida per Valerio Costa des de [99])

En una estratègia orientada a la caracterització '*de novo*' del transcriptoma, les seqüències parcials de cDNA derivades de gens expressats, també anomenades EST (de l'anglès *Expressed Sequence Tags*), s'agrupaven en conjunts de seqüències solapants més llargues anomenades *contigs* les quals podien arribar a cobrir la molècula original d'RNA en tota la seva extensió, obtenint així una caracterització no només qualitativa sinó també quantitativa de l'expressió gènica, ja que el nombre d'EST agrupats en contigs resulta proporcional a la seva expressió [101]. Múltiples estudis transcriptòmics de les glàndules del verí de serp s'han dut a terme amb aquest tipus de tecnologia de baix rendiment [102–111]. Les limitacions d'aquest tipus de tècnica responen a possibles desviacions introduïdes pel procés de clonació en bactèries i l'esforç en termes de cost i mà d'obra necessària per construir una extensa col·lecció d'EST

(Taula 2). A més, l'estudi de l'expressió gènica amb aquest tipus de tecnologia queda restringida a aquelles molècules més expressades, perdent la informació que ofereixen aquells transcrits menys expressats que podrien aportar una informació molt valuosa respecte als mètodes usats per les cèl·lules de les glàndules per produir l'arsenal químic o per regular la seva expressió.
L'arribada de les tecnologies de Seqüenciació de Nova Generació (NGS, de l'anglès *Next Generation Sequencing*) ha canviat la forma d'abordar estudis moleculars, recordant, en paraules de Michael Metzker, els inicis de la PCR, on "l'única limitació al seu ús és la imaginació" [112]. La gran quantitat de dades produïdes en aquest tipus de projectes ha fet canviar la concepció de la biologia i de la recerca, des d'una investigació guiada per una determinada hipòtesi, a una investigació dirigida per la tecnologia, tal i com ja va avançar F.J Dyson l'any 1999 [113]. En aquest sentit, en paraules de S. Brenner, estaríem parlant de "caçadors" en el primer cas i de "recol·lectors" en el segon [114]. Tal i com veurem més endavant, l'estudi del transcriptoma s'ha beneficiat d'aquestes tecnologies d'alt rendiment donant lloc al que s'ha conegut com a RNA-Seq (Seqüenciació d'RNA) [99, 115, 116] o Seqüenciació Shotgun del Transcriptoma Total (WTSS, de l'anglès *Whole Transcriptome Shotgun Sequencing*).
L'ús de les tècniques NGS ha donat lloc a conclusions sorprenents, arribant al punt de qüestionar el concepte de gen tal i com era conegut al segle XX [117], sobretot arran de les conclusions del projecte ENCODE (ENCyclopedia Of DNA Elements) [118] i fins al punt d'afirmar, en paraules d'un investigador del projecte, que la unitat fonamental del genoma i la unitat bàsica d'herència hauria de ser el transcrit i no el gen. Aquest projecte públic del National Human Genome Research Institute (NHGRI), va començar l'any 2003 com a lògica evolució del desxiframent del genoma humà l'any 2000. Les seves més recents conclusions han atorgat una funció bioquímica al 80% del genoma humà, desterrant l'antic concepte de DNA "escombraries" [119]. Aquestes conclusions, tanmateix, no han estat exemptes de crítica per la definició de "funció" per part del Consorci i pel tractament estadístic del conjunt de les 1.640 dades recollides [120].

Sigui com sigui, la majoria d'aquest RNA no codificant està involucrat en la regulació de l'expressió gènica. Dins de tots els elements reguladors, els microRNA (miRNA) apareixen com a elements d'importància creixent en la regulació post-transcripcional. Aquests fragments curts (21 - 25 nucleòtids de longitud), de cadena senzilla d'RNA descrits per primera vegada per Lee i col·laboradors [121], actuen com a reguladors negatius de l'expressió gènica en organismes eucariotes [122] en un complex silenciador que s'uneix per aparellament de bases tipus Watson-Crick principalment a les regions 3' no traduïdes (3'-UTR) del transcrit. Tal i com veurem més endavant, l'existència de determinats miRNA podria estar lligada a l'expressió temporal de determinades toxines. Així, la variabilitat ontogenètica observada en el proteoma de determinades espècies de serps ([123, 124]) podria respondre a una regulació negativa post-transcripcional mediada per miRNAs.

	Xips DNA	**Seqüenciació EST**	**RNA-Seq**
Principi	Hibridació	Sanger	Seqüenciació massiva
Rendiment	Alt	Baix	Alt
Seqüència de referència	Necessària	No necessària	No necessària
Soroll de fons	Alt	Baix	Baix
Detecció isoformes	Limitada	Si	Si
Quantitat RNA	Alta	Alta	Baixa
Cost	Alt	Alt	Baix

Taula 2. Característiques dels principals mètodes d'anàlisi del transcriptoma. (Taula adaptada de [115])

Proteòmica

Diferents aproximacions proteòmiques s'han desenvolupat per a l'anàlisi de mostres complexes com els verins. Algunes d'elles han estat aplicades amb èxit per tal d'entendre els efectes patològics causats pel verí de les serps [125, 126], incloent, entre altres, l'ús de

l'electroforesi en gel de poliacrilamida, la separació cromatogràfica i l'espectrometria de masses. És evident que cap d'aquestes tècniques de forma individual és capaç de cobrir tots els components proteïcs / peptídics d'una mostra, cadascuna d'elles amb els seus particulars punts forts i punts dèbils. Tanmateix, en els últims anys, els protocols desenvolupats al nostre laboratori, coneguts com a "venomics" (Figura 6) [127] i "antivenomics" [128, 129], s'han demostrat adequats per a la caracterització, tant dels verins de les serps com de la seva neutralització per part del antiverins, respectivament. Així, l'acoblament de determinats protocols de preparació de mostres amb l'espectrometria de masses ha permés obtindre una aproximació qualitativa i quantitativa de la composició del verí. De la mateixa manera, els assajos funcionals de neutralització duts a terme mitjançant l'antivenòmica, han permés l'estudi preclínic de l'eficàcia dels antiverins i de la seva possible reactivitat creuada.

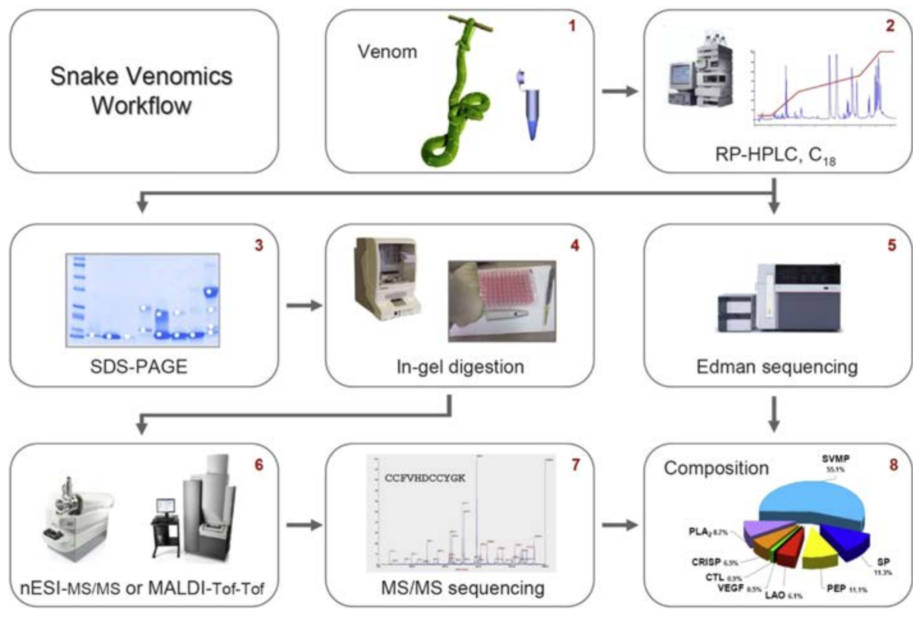


Figura 6. Representació esquemàtica de l'estratègia analítica "venomics" (Imatge obtinguda de [8])

A tall il·lustratiu, el proteoma de nombroses serps ha estat caracteritzat seguint aquest tipus de protocols (revisió a [8, 94]).
La relació existent entre les toxines trobades al verí (el proteoma) i els transcrits codificants d'aquestes (el transcriptoma) en el que anteriorment hem anomenat "proteotranscriptòmica", serà un punt important de discussió del nostre treball.

Seqüenciació de Nova Generació (NGS)

La tècnica de seqüenciació del DNA ha evolucionat de forma molt ràpida en els últims anys (Figura 7). El mètode descrit per Fred Sanger [130] l'any 1977 basat en l'ús de dideoxinucleòtids terminals ha estat durant més de 25 anys el mètode de seqüenciació de DNA dominant i va ser la tècnica usada per a la seqüenciació del genoma humà [131] en el que es va anomenar mètode de seqüenciació de "primera generació". Al llarg d'aquests 25 anys s'han produït successives millores i automatitzacions en el mètode descrit per Fred Sanger, des de l'eliminació dels nucleòtids marcats radioactivament per nucleòtids marcats amb un fluoròfor (1987) fins a l'ús de l'electroforesi capil·lar enlloc del gel de poliacrilamida (1999), donant lloc a una capacitat de seqüenciació d'aproximadament 3 Mbases per dia amb lectures de 900 nucleòtids de longitud i 384 mostres en paral·lel.
En els últims deu anys, però, l'escenari ha canviat radicalment gràcies a la introducció de les anomenades tècniques NGS [132–134]. El rendiment produït amb aquest tipus de tecnologies ha fet que sovint siguin anomenades tecnologies de seqüenciació massiva en paral·lel, on les milions de reaccions de seqüenciació simultànies estan separades espacialment.
Les diferents tècniques NGS comparteixen una estratègia comú com és la preparació de la llibreria, el procés de seqüenciació pròpiament dit, la captació de la imatge i l'anàlisi de les dades. Tal i com podem veure a la Taula 3, la combinació única de diferents estratègies en alguna d'aquestes fases distingeix un tipus de tecnologia de l'altra, i determina el tipus de dades produïdes per cadascuna de les diferents plataformes.

Així, la preparació de la llibreria implica la immobilització dels fragments del DNA sobre una determinada superfície, cosa que permetrà la paral·lelització del procés en milers de milions de reaccions de seqüenciació. Diferents estratègies d'immobilització han donat lloc als mètodes de "segona generació" (aquells en els quals té lloc una amplificació clonal del fragment de DNA), i als mètodes de "tercera generació" (o de molècula única). En el primer cas, dos mètodes són els més àmpliament usats per la majoria de plataformes de seqüenciació: la PCR en emulsió (emPCR) i l'amplificació en fase sòlida.

Pel que fa el mètode de seqüenciació usat, les diferents plataformes es poden classificar segons fagin ús de l'estratègia CRT (de l'anglès *Cyclic reversible termination*), piroseqüenciació, SBL (de l'anglès *Sequencing by ligation*), temps real (de l'anglès *Real Time*) i monitorització del canvi de pH. Una descripció de cadascuna d'aquestes estrategies podem trobar-la a [112]. La metodologia de seqüenciació usada en les dues primeres estratègies ha donat lloc al que s'ha anomenat seqüenciació per síntesi (SBS, de l'anglès *Sequencing By Synthesis*), on la seqüència és llegida a mesura que la polimerasa incorpora cada base de forma individual.

Plataforma	**Prep. Llibreria**	**Química NGS**
454 Roche	emPCR	Piroseqüenciació
Illumina	Fase sòlida	CRT
SOLiD	emPCR	SBL
PacBio	Molècula única	Real – time
Ion PGM	emPCR	Canvi pH

Taula 3. Característiques de les principals plataformes de seqüenciació. Combinació de característiques pel que fa a preparació de la llibreria i química de seqüenciació en diferents plataformes NGS.

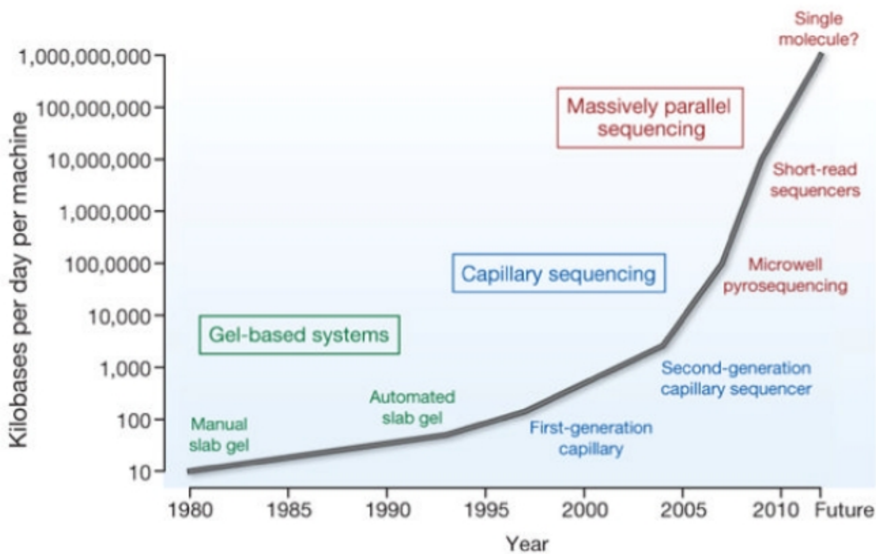


Figura 7. Evolució de les tècniques de seqüenciació. Rendiment (Kbases/dia) obtingut amb els mètodes de seqüenciació basats en gel de poliacrilamida, en electroforesi capil·lar i en la seqüenciació massiva en paral·lel. (Imatge obtinguda de [135])

Comparades amb la tecnologia desenvolupada per Sanger, les tècniques NGS han suposat una sèrie d'avantatges com són: i) la construcció de la llibreria *in vitro* seguida de l'amplificació clonal ha eliminat el coll d'ampolla que suposa la transformació en *Escherichia coli* i el creixement de les colònies, ii) les tècniques NGS permeten una elevada paral·lelització del procés i iii) disminució dels volums de reacció amb la conseqüent disminució del cost de seqüenciació.
De la mateixa manera, l'evolució de les diferents plataformes existents ha portat a una millora pel que fa la longitud total de la lectura obtinguda i al rendiment total del procés (mesurat com el nombre total de bases seqüenciades), donant lloc a lectures cada cop més llargues i un rendiment cada cop més elevat (Figura 8).

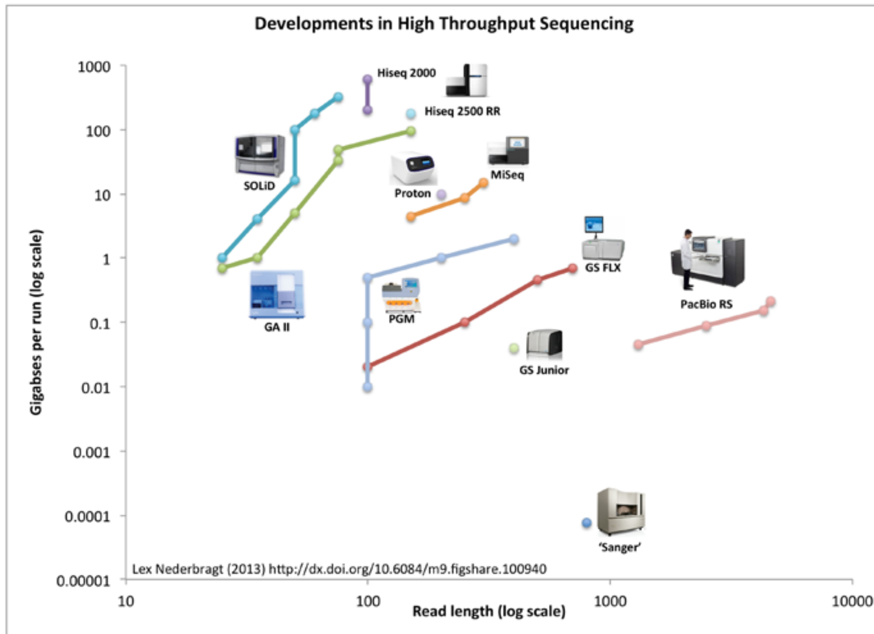


Figura 8. Evolució de les diferents plataformes NGS. Evolució pel que fa longitud de la lectura i rendiment total del procés de les diferents plataformes de seqüenciació existents al mercat. (Imatge obtinguda sota llicència Creative Commons sense restricció d'ús)

Tant és així, que recentment Illumina ha anunciat el seu seqüenciador HiSeq X [136], amb una capacitat de seqüenciació de 1.8 Tbases en menys de 3 dies, cosa que ha estat anunciada com el trencament de la 'barrera' de la genòmica humana, aconseguint el tant esperat genoma humà a un cost inferior a 1.000 dòlars. Com a conseqüència d'això, el cost de seqüenciació està disminuint a un ritme molt més ràpid que el que podríem esperar segons la Llei de Moore, llei de la tecnologia de la informació segons la qual la capacitat dels microprocessadors es dobla cada any i mig amb la conseqüent baixada de preus (Figura 9).
El gran volum de dades produït per totes aquestes plataformes està fent que el coll d'ampolla estigui actualment en l'emmagatzematge, manipulació i interpretació d'aquestes dades. En aquest punt la

bioinformàtica està emergint com una disciplina imprescindible per al treball diagnòstic i de recerca en l'ús d'aquest tipus de tecnologia. Les plataformes de seqüenciació actuals estan donant lloc a una gran quantitat de dades en formats diferents i amb determinats tipus d'error específics. D'aquesta forma, la implementació d'estratègies de processament de dades individualitzades combinades amb nous mètodes computacionals estandaritzats i una gran capacitat de càlcul sembla ser l'única manera de poder abordar la identificació, anàlisi i visualització d'aquesta enorme quantitat de dades biològiques.

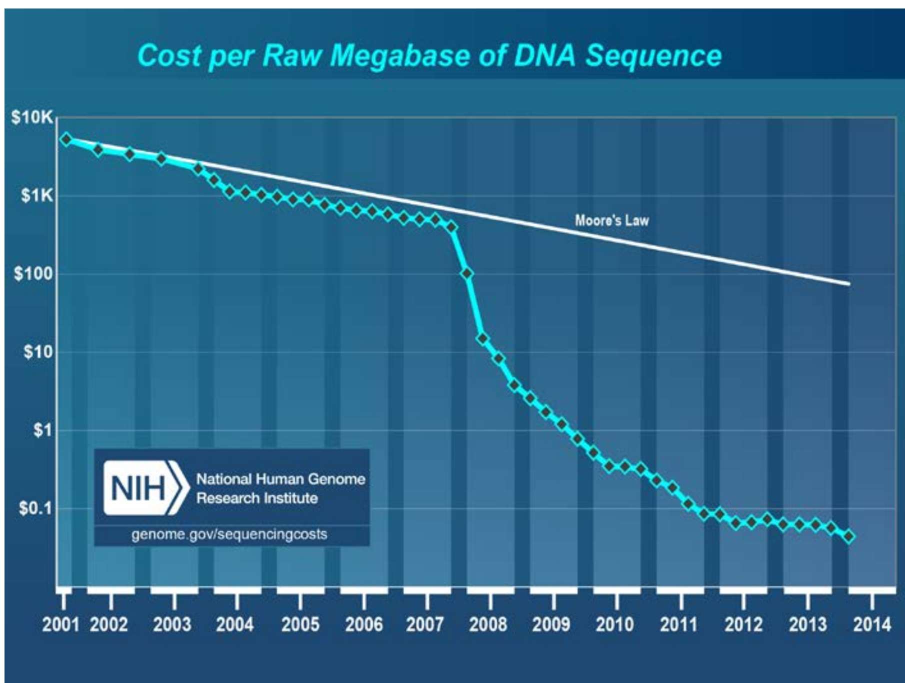


Figura 9. Cost en dòlars per Mbase seqüenciada. Les despeses considerades a la gràfica corresponen a: i) Reactius i manipulació, ii) Equips de seqüenciació, iii) Activitat informàtica relacionada, iv) Construcció de la llibreria, v) Enviament de dades a bases de dases públiques, vi) Despeses indirectes relacionades. (Imatge obtinguda de [137])

Plataformes actuals

A la Taula 4 trobem la comparativa pel que fa a rendiment, longitud de lectura i temps de les principals plataformes de seqüenciació descrites en el present treball. Tal i com podem veure, diferents empreses han diversificat la seva oferta de seqüenciadors oferint més capacitat de seqüenciació o més velocitat del procés total. D'aquesta manera, Roche 454 ofereix els aparells 454 GS FLX i 454 GS Junior com a versió de menor temps i menor rendiment, però més apropiada pel diagnòstic ràpid de determinades mutacions causants d'una patologia. Una estratègia similar ha adoptat Illumina amb els seus seqüenciadors HiSeq o MiSeq.

Instrument	Longitud lectura (nucleòtids)	Rendiment (Gb)	Lectures	Temps
454 GS FLX	700	0.7	1M	24 h
454 GS Junior	700	0.07	100000	12 h
Illumina MiSeq	300	15	25M	5 - 65 h
Illumina HiSeq 2000/2500 RR	125	1000	4G	11 dies
Illumina NextSeq 500	150	120	400M	< 30 h
Illumina HiSeq X	150	1800	3G	< 3 dies
SOLiD	75	320	3G	1 -2 setmanes
PacBio RS	8500	4.5	44000	30 min - 2 h
IonTorrent PGM	400	2	5M	2 h
IonTorrent Proton	200	10	50M	2 - 4 h

Taula 4. Comparativa de les principals plataformes de seqüenciació. Les principals plataformes de seqüenciació són comparades en longitud de lectura dels nucleòtids resultants, rendiment total del procés expressat en Gbases seqüenciades, nombre total de lectures obtingudes i temps total del procés de seqüenciació. M, milions de lectures; G, milers de milions de lectures.

Fora de l'abast del present treball està la descripció de totes les plataformes de seqüenciació existents. La tecnologia NGS està avançant a un ritme molt ràpid, com ho demostra el fet que al llarg de 2014, a la ja comentada introducció al mercat de la plataforma HiSeq X se li ha d'afegir la comercialització del dispositiu USB MinION

d'Oxford Nanopore. De forma similar, altres mètodes de seqüenciació han experimentat canvis recents com per exemple l'adquisició de l'empresa Complete Genomics (la metodologia de seqüenciació de la qual implica la detecció fluorescent provinent de la seqüenciació de nanoboles de DNA) per l'Institut Xinés de Genòmica (BGI, de l'anglès, *Beijing Genomic Institute*) o la fallida econòmica d'Helicos Biosciences, l'estratègia de seqüenciació de la qual implicava SBS de molècula única de DNA.

454 (Roche)

Aquesta plataforma de seqüenciació, va ser la primera introduïda comercialment l'any 2004 i amb la que es va seqüenciar el primer genoma usant tècniques NGS [138]. El seu mètode de seqüenciació fa ús de la piroseqüenciació, on es detecta el pirofosfat alliberat durant la incorporació nucleotídica mitjançant una reacció luminiscent [139]. Així, durant el procés de seqüenciació, fragments únics de cDNA són units a una única perla per a l'amplificació clonal del fragment un una PCR en emulsió. Aquest procés va ser definit com "Un fragment, una perla, una lectura" (de l'anglès, *"One fragment, one bead, one read").* D'aquesta manera, una vegada incorporat el nucleòtid corresponent per complementarietat de bases, s'allibera un pirofosfat que donarà lloc a una molècula d'ATP que serà usat per un enzim per a la producció de llum de forma proporcional a la quantitat de pirofosfats alliberats (Figura 10).

Les diferents evolucions del sistema han permès arribar a la plataforma a una longitud de lectura de 700 bases amb un rendiment total de 0,7 Gbases per procés de seqüenciació.

Donada l'elevada longitud de les seves lectures, 454 ha estat des dels seus inicis un molt bon sistema per a la seqüenciació '*de novo*', i és per això que va ser el sistema de seqüenciació triat en el nostre laboratori per a la seqüenciació dels 8 transcriptomes caracteritzats.

El seu elevat preu per base seqüenciada així com l'aposta de la companyia pels mètodes de seqüenciació de lectura única, ha fet que aquest tipus de plataforma no tingui suport més enllà de 2016.

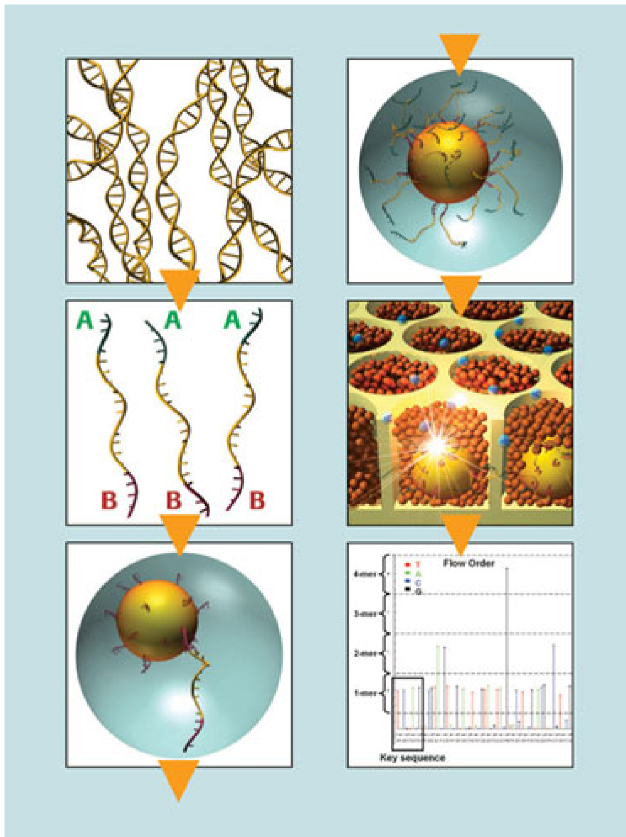


Figura 10. Estratègia seguida per Roche 454. En aquesta estratègia 4 fases principals donaran lloc a "Un fragment, una perla, una lectura": la generació de la cadena senzilla de DNA anirà seguida de l'amplificació en la PCR en emulsió, la generació de dades per SBS i l'anàlisi d'aquestes dades. (Imatge obtinguda de [140])

SOLiD (Life Technologies)

Sequencing by Oligo Ligation Detection (SOLiD) és un mètode de seqüenciació desenvolupat per Applied Biosystems (ara Life Technologies), comercialment disponible des de l'any 2006. El seu mètode de seqüenciació adopta l'anomenada tecnologia SBL, un mètode que usa la sensibilitat d'una DNA ligasa per identificar el nucleòtid present en una determinada posició d'una seqüència de DNA. Durant la construcció de la llibreria, aquesta seqüència de DNA

és flanquejada per, com a mínim, un extrem de seqüència coneguda, i serà amplificada en una emPCR. Així, la reacció de seqüenciació consta de successius cicles de ligació, detecció i trencament entre la molècula de DNA i diferents sondes marcades amb fluoròfors, on cada color representa un parell de bases, en el que es va anomenar codificació de 2 bases o *color space*.
Les contínues evolucions del sistema (5 des de 2007 a 2010) han donat lloc a la versió 5500XL W, amb un rendiment de 320 Gbases per procés de seqüenciació.
Life Technologies és també la propietària d'una altra plataforma de seqüenciació com és Ion Torrent, sobre la qual Life Technologies sembla haver concentrat tots els seus esforços. Això, juntament amb la manca de desenvolupament del *color space* per part de la comunitat científica ha fet que la plataforma SOLiD estigui actualment en retrocés respecte a les seves competidores.

Illumina

Sense cap mena de dubte, Illumina és l'empresa dominant actualment en el món de la seqüenciació. Inicialment desenvolupat per Solexa, el seqüenciador Genome Analyzer (GA) adopta la tecnologia de Terminació Cíclica Reversible (CRT) per a seqüenciar una llibreria de DNA construïda seguint el protocol d'amplificació clonal en fase sòlida, una estratègia ja presentada l'any 1998 per Pascal Mayer i col·laboradors al 5è congrés d'Automatització en el Mapeig i Seqüenciació de DNA [141]. Així, tal i com podem veure a la Figura 11, les molècules de DNA i els adaptadors són fixats en una superfície (la cel·la de fluxe, de l'anglès *flow cell*) i amplificades mitjançant una DNA polimerasa per donar lloc a colònies de DNA separades (*clusters*), cadascuna d'elles amb aproximadament 1.000 amplicons idèntics. Diversos milions de *clusters* poden ser amplificats a cadascun dels 8 carrils que existeixen en una única cel·la de fluxe.
Posteriorment, la seqüenciació tindrà lloc per repetits cicles d'incorporació d'un únic nucleòtid marcat amb un fluoròfor, adquisició de la imatge i trencament del grup bloquejant a 3' del nucleòtid i de l'etiqueta fluorescent per permetre un nou cicle d'amplificació.

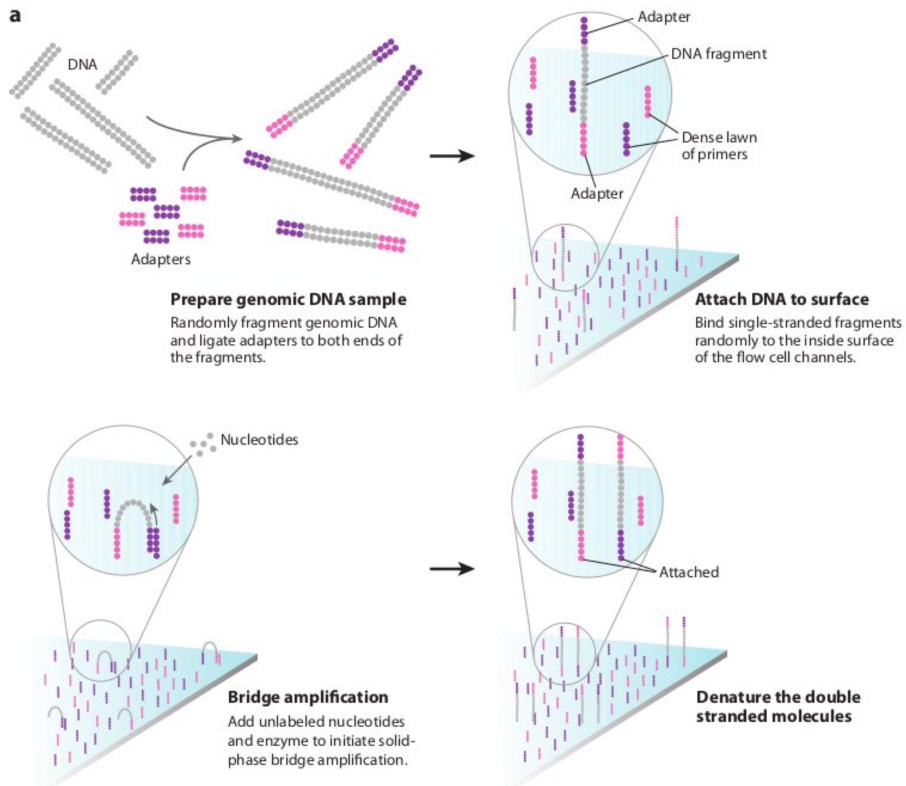


Figura 11. Estratègia SBS seguida per Illumina. Els clusters es creen per amplifiació clonal de cadenes de DNA unides a una superfície sòlida. L'addició de DNA polimerasa i nucleòtids modificats permetrà el procés de seqüenciació. (Imatge obtinguda de [133])

El seqüenciador GA de Solexa tenia un rendiment inicial d'una Gbase per procés de seqüenciació. Mitjançant millores en la polimerasa, tampons de reacció, la cel·la de fluxe i el software, l'any 2009 el seu rendiment va augmentar fins a 20 Gbases per procés de seqüenciació a l'agost, fins a 30 Gbases a l'octubre i fins a 50 Gbases al desembre d'aquell mateix any. A principis de 2010 Illumina va comercialitzar el seqüenciador HiSeq 2000, amb la mateixa estratègia de seqüenciació que el GA, i amb un rendiment de 200 Gbases per procés de seqüenciació, millorat fins als actuals 1000 Gbases.

Comparat amb els seus competidors, Illumina ofereix la seqüenciació més barata, amb un cost d'aproximadament 0.05 dòlars per milió de bases.

Ion Torrent (Life Tecnologies)

La tecnologia desenvolupada per Ion Torrent en els seus seqüenciadors Ion Torrent Personal Machine (PGM) i Ion Proton ha donat lloc als primers seqüenciadors que no necessiten fluorescència ni càmera CCD, cosa que s'ha traduït en una velocitat de seqüenciació més elevada, un cost més baix i una mida més reduïda. Desenvolupada per Jonathan M. Rothberg, un dels fundadors de l'empresa 454 Life Sciences, la seva tecnologia de seqüenciació fa ús de la seqüenciació per síntesi sobre una llibreria de DNA construïda en una emPCR [142]. A diferència de la resta de mètodes l'Ion Torrent detecta l'àtom d'H^+ alliberat en el procés d'incorporació d'un nucleòtid mitjançant un sensor de pH, cosa que dona lloc a un pols elèctric que serà traduït en la seqüència de DNA, evitant així l'ús de qualsevol sistema òptic. L'existència d'un xip semiconductor on té lloc la construcció de la llibreria i la reacció de seqüenciació ha fet que sovint la tecnologia es conegui com a seqüenciació Ion semiconductora.
D'entre les fites més importants d'aquest tipus de seqüenciador destaca la detecció del gen codificant d'una toxina a la soca O104:H4 d'*Escherichia coli*, que va causar 50 morts a principis de maig de 2011 [143].
Amb el seu seqüenciador Ion Proton, Life Technologies espera poder arribar a finals de 2014 a les 32 Gbases per procés de seqüenciació amb una longitud mitjana de lectura de 100 nucleòtids.

Pacific Biosciences

Pacific Biosciences és una companyia fundada l'any 2004 que ha desenvolupat una plataforma, el principi de seqüenciació de la qual es defineix com a seqüenciació en temps real d'una molècula única (SMRT, de l'anglès *Single Molecule Real Time Sequencing*) en el que s'ha anomenat mètode de "tercera generació". Així, la nova versió del seu seqüenciador, comercialitzat l'octubre de 2013 anomenat Pac Bio RS II, ha donat lloc a lectures amb una longitd mitjana de 8,500

nucleòtids amb un màxim de 30,000 nucleòtids en una única lectura. Les prediccions de l'empresa són aconseguir una longitud mitjana de lectura de 20,000 nucleòtids l'any 2015.

NGS i transcriptòmica

La seqüenciació d'RNA o RNA-Seq ha esdevingut en poc temps una pràctica estàndard per a la recerca en el camp de la transcriptòmica, beneficiada també pel continu descens del cost per base seqüenciada juntament amb la millora de rendiment i la reducció de temps de seqüenciació, tal i com hem vist anteriorment. Aquest tipus d'aproximació ha estat la tècnica usada en la recerca de determinats patrons d'expressió tant en organismes model com en organismes no model, i ha proveït informació sorprenent sobre la regulació transcripcional i post-transcripcional, revelant uns transcriptomes molt més complexos del que es podia anticipar. Així, l'ús de la tecnologia NGS ha permés l'estudi de RNAs petits no codificants la mida dels quals suposava una limitació per a ser capturats degut a la limitada resolució dels xips de DNA. El present treball, tal i com serà discutit més endavant, pretén determinar si l'expressió gènica al verí està controlada de manera transcripcional o si l'existència de determinats factors entre els quals podrien trobar-se aquest tipus d'RNA no codificants, està donant lloc a una regulació post-transcripcional.
De la mateixa manera, la tecnologia NGS ha permés la seqüenciació i la caracterització a màxima resolució i rang dinàmic del transcriptoma de múltiples organismes com l'ésser humà (amb finalitats, inclús diagnòstiques), el ratolí, el llevat, la mosca de la fruita, diferents plantes i organismes procariotes, fins al punt de poder ser aplicada a mostres més complexes amb l'existència de múltiples organismes en el que s'ha anomenat metatranscriptòmica [144].
El primer ús de les tècniques NGS per a l'estudi del transcriptoma va ser dut a terme per M. N. Bainbridge l'any 2006 [135], i la seva primera aplicació en organismes sense informació genòmica data d'un parell d'anys després [145]. Pel que fa l'estudi del transcriptoma de membres del subordre *Serpentes*, la primera aplicació de la tecnologia NGS data de l'any 2010, on el transcriptoma de l'espècie *Thamnophis elegans* (família *Colubridae*) va ser seqüenciat amb 454 GS-FLX

[146]. Posteriorment, i en especial al llarg de 2013, s'ha abordat l'estudi del transcriptoma de moltes espècies de serp amb tecnologia NGS (veure Discussió). El nostre treball, representa la meitat de les espècies de serp estudiades mitjançant NGS, tecnologia que ha servit per aprofundir en el coneixement de l'expressió gènica de determinades espècies que poden ajudar a revelar adaptacions i processos biològics que governen la composició del verí. A més, l'anàlisi del transcriptoma està proporcionant l'esquelet adequat sobre el qual investigar més profundament el proteoma, la composició final del verí.

OBJECTIUS

El treball desenvolupat durant la present tesi doctoral pretén aportar més evidències, no només en la caracterització de la variabilitat interespecífica observada en la composició del verí dels cròtals sud-americans *Crotalus simus*, *Atropoides picadoi*, *Atropoides mexicanus*, *Bothriechis lateralis*, *Bothriechis schlegelii*, *Cerrophidion sasai* (abans *Cerrophidion godmani*) i *Bothrops asper*, sinó també intentar explicar els fenòmens biològics responsables de la variabilitat intraespecífica observada en determinades espècies de serps tant a nivell geogràfic com ontogenètic. En aquest sentit, la feina desenvolupada durant la present tesi doctoral ha intentat estudiar de forma conjunta les dades transcriptòmiques i les dades proteòmiques per a organismes sense informació genòmica prèvia, en el que hem volgut anomenar "proteotranscriptòmica", per tal d'esbrinar si l'esmentada variabilitat podria respondre a una divergència genòmica o, per altra banda, determinats elements reguladors com els miRNA podrien jugar un paper clau en la diversitat de fenotips observats.
Amb això es pretén avançar en el coneixement dels mecanismes de regulació que fan possible aquesta variabilitat, la comprensió dels quals podrien obrir un nou camí en el tractament de les patologies associades en els accidents per mossegada de serp i a l'avaluació de les teràpies aplicades.
Per abordar aquests objectius, el desenvolupament de la tesi doctoral va necessitar de l'estudi de les relacions filogenètiques dels membres del subordre *Serpentes* així com del coneixement de les famílies de proteïnes existents als verins i de les tècniques de seqüenciació massiva de DNA, coneixements que intentaran ser explicats en els diferents apartats de la manera més entenedora possible. Els nostres resultats van possibilitar proposar potencials mecanismes de regulació post-transcripcional que podrien jugar un paper fins ara no descrit en la regulació de l'expressió de les toxines i van ajudar a la caracterització '*de novo*' de determinades proteïnes del verí no descrites anteriorment.

PUBLICACIONS

Capítol 1

Profiling the venom gland transcriptomes of Costa Rican snakes by 454 pyrosequencing

Durban *et al. BMC Genomics* 2011, **12**:259
http://www.biomedcentral.com/1471-2164/12/259

BMC Genomics

RESEARCH ARTICLE Open Access

Profiling the venom gland transcriptomes of Costa Rican snakes by 454 pyrosequencing

Jordi Durban[1], Paula Juárez[1], Yamileth Angulo[2], Bruno Lomonte[2], Marietta Flores-Diaz[2], Alberto Alape-Girón[2,3], Mahmood Sasa[2], Libia Sanz[1], José M Gutiérrez[2], Joaquín Dopazo[4], Ana Conesa[4*] and Juan J Calvete[1*]

Abstract

Background: A long term research goal of venomics, of applied importance for improving current antivenom therapy, but also for drug discovery, is to understand the pharmacological potential of venoms. Individually or combined, proteomic and transcriptomic studies have demonstrated their feasibility to explore in depth the molecular diversity of venoms. In the absence of genome sequence, transcriptomes represent also valuable searchable databases for proteomic projects.

Results: The venom gland transcriptomes of 8 Costa Rican taxa from 5 genera (*Crotalus*, *Bothrops*, *Atropoides*, *Cerrophidion*, and *Bothriechis*) of pitvipers were investigated using high-throughput 454 pyrosequencing. 100,394 out of 330,010 masked reads produced significant hits in the available databases. 5.165,220 nucleotides (8.27%) were masked by RepeatMasker, the vast majority of which corresponding to class I (retroelements) and class II (DNA transposons) mobile elements. BLAST hits included 79,991 matches to entries of the taxonomic suborder *Serpentes*, of which 62,433 displayed similarity to documented venom proteins. Strong discrepancies between the transcriptome-computed and the proteome-gathered toxin compositions were obvious at first sight. Although the reasons underlaying this discrepancy are elusive, since no clear trend within or between species is apparent, the data indicate that individual mRNA species may be translationally controlled in a species-dependent manner. The minimum number of genes from each toxin family transcribed into the venom gland transcriptome of each species was calculated from multiple alignments of reads matched to a full-length reference sequence of each toxin family. Reads encoding ORF regions of Kazal-type inhibitor-like proteins were uniquely found in *Bothriechis schlegelii* and *B. lateralis* transcriptomes, suggesting a genus-specific recruitment event during the early-Middle Miocene. A transcriptome-based cladogram supports the large divergence between *A. mexicanus* and *A. picadoi*, and a closer kinship between *A. mexicanus* and *C. godmani*.

Conclusions: Our comparative next-generation sequencing (NGS) analysis reveals taxon-specific trends governing the formulation of the venom arsenal. Knowledge of the venom proteome provides hints on the translation efficiency of toxin-coding transcripts, contributing thereby to a more accurate interpretation of the transcriptome. The application of NGS to the analysis of snake venom transcriptomes, may represent the tool for opening the door to systems venomics.

Keywords: Snake venom gland transcriptomics, next generation high-throughput DNA sequencing, 454 pyrosequencing, bioinformatic analysis, Costa Rican snakes, *Bothrops asper*, Bothriechis, Atropoides, Crotalus, Cerrophidion

* Correspondence: aconesa@cipf.es; jcalvete@ibv.csic.es
[1]Consejo Superior de Investigaciones Científicas, Jaime Roig 11, 46010 Valencia, Spain
[4]Centro de Investigación Príncipe Felipe, Valencia, Spain
Full list of author information is available at the end of the article

© 2011 Durban et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Durban et al. BMC Genomics 2011, 12:259
http://www.biomedcentral.com/1471-2164/12/259

Background

Venomous snakes of the families Viperidae and Elapidae possess paired specialized venom glands located in the upper jaw ventral and posterior to the eyes [1] that produce an arsenal of toxins [2,3], which they inject into prey tissues through high-pressure delivery fangs [4]. Within the reptile clade Toxicofera, venom was a single ancient innovation [5]. Snake venom toxins are the result of recruitment events by which ordinary genes were duplicated, and the new genes selectively expressed in the venom gland and amplified to multigene families with extensive neofunctionalization throughout ~100 million years of evolution [5,6]. Given the central role that diet has played in the adaptive radiation of snakes [7], venom thus represents a key adaptation that has played an important role in the diversification of snakes.

Envenoming by snakebites constitutes a highly relevant, though neglected, public health issue on a global basis [8], as there are venomous organisms in every continent and almost every country. However, venomous animals are particularly abundant in tropical regions, the kitchen of evolution. Arthropod stings constitute the most common cause of envenoming by animals, although around 80% of the more than 150.000 yearly deaths by envenomings worldwide are caused by snakebites, followed by scorpion stings, which cause 15% [9,10]. The venoms of extant snakes comprise complex cocktails of proteins tailored by Natural Selection to act on vital systems of the prey [11]. Medical uses of venoms are well documented in folk remedies and in Western and Chinese traditional medicine [12]. However, despite their remarkable potency and high degree of target specificity, only in the last decades have toxins been increasingly used as pharmacological tools, and it has been realized that venoms represent a vast and essentially untapped resource of preoptimized lead molecules for the medicinal chemist [12-17].

Adequate treatment of snakebites is critically dependent on the ability of antivenoms to neutralize the lethal and tissue-damaging toxins, reversing thereby the signs of envenoming [18,19]. A long term research goal of venomics, of applied importance for improving current antivenom therapy, but also for drug discovery, is to understand the molecular mechanisms and evolutionary forces that underlie the enormous pharmacological potential of venoms [12]. Individually or combined, proteomic and transcriptomic studies have demonstrated their feasibility to explore in depth the molecular diversity of venoms [[20-28], and references therein]. In the absence of genome sequence, transcriptomes represent also valuable searchable databases for proteomic projects.

Since the pioneer report by Ho and co-workers in 1995 [29], snake venom transcriptomic studies have relied on sequencing DNA clones randomly picked from a cDNA library constructed by reverse transcription of the RNA molecules expressed in the venom gland [25]. The partial cDNA sequences derived from expressed genes, also known as Expressed Sequence Tags (ESTs) [30], cluster into groups of contiguous sequences (contigs), which eventually cover the entire extension of the original RNA molecule. In addition, the number of ESTs clustered into a contig is proportional to the transcriptional level of the parent RNA in the venom gland [25]. However, in the few instances in which transcriptomics and proteomics databases have been compared [26,27,31,32], a low degree of concordance has been reported. The occurrence of non-venom-secreted toxin transcripts might indicate that these messengers exhibit an individual or a temporal expression pattern over the life time of the snake [33], or may encode very low-abundance venom proteins. On the other hand, the presence in the venom of toxins not represented in the transcriptome clearly indicates that construction of the cDNA library was biased, i.e. due to the necessary fractionating steps to avoid interfering substances like short, partial length 3'-end cDNAs and adapter sequences [34]. A second bias of cDNA libraries is the potential of the mRNA transcript in plasmids to be partially expressed in their bacterial cells with lethal effects [35]. Moreover, smaller cDNA fragments are over represented compared to larger ones, due to the higher transformation efficiency of smaller plasmids [35].

The high demand for low-cost sequencing has driven the development of high-throughput next-generation sequencing (NGS) technologies such as 454 Roche, Illumina's Solexa, and Applied Biosystems' SOLiD, and and most recently released Helicos HeliScope platforms as alternatives to the classical chain-termination Sanger method of DNA sequencing for the qualitative and quantitative analyses of transcriptomes [36,37]. NGS technologies are revolutionizing the field of transcriptomics by rapidly reducing the time and cost per base sequenced [38]. For example, snake venom gland transcriptomes reported are typically arranged from few hundreds to few thousands ESTs [25]. The largest transcriptome database was assembled from 8696 ESTs (mean read length of 398 bp) from *Deinagkistrodon acutus* venom gland [39]. Only very recently Rokyta and colleagues [40] reported a high-throughput venom gland transcriptome of the Eastern Diamond Rattlesnake (*Crotalus adamanteus*) using Roche 454 sequencing technology. 82,621 reads were singletons, and the remaining 552,863 reads were assembled into 24,773 contigs of average length 513 nucleotides [40]. NGS technologies applied to the transcriptomic analysis of non-model species has the advantage of providing a genome-wide,

unbiased insight into the transcriptome [41]. However, NGS techniques applied to non-model species, which like snakes lack a suitable reference genome sequence, are not devoid of limitations. NGS technologies provide shorter and more error-prone reads than Sanger sequencing, making transcript assembly a challenging bioinformatic task, which frequently yields a large set of contigs but a fragmented transcriptome [38,41]. Here, we report the application of the 454 platform to infer the venom gland transcriptomes of Costa Rican snakes, *Bothrops asper* (from Caribbean (car) and Pacific (pac) populations), *Bothriechis lateralis*, *Bothriechis schlegelii*, *Atropoides picadoi*, *Atropoides mexicanus*, *Crotalus simus*, and *Cerrophidion godmani*, with special emphasis on the strategy used to assemble and analyze the gathered DNA sequences. Although the average length of singletons (174 bp) and contigs (208.2 bp), and the low coverage of reads per contig (6), prevented the generation of definitive and reliable full-length gene sequences, our results provide a deep comparison of the transcriptional activity of the venom glands of these medically relevant species in Central America [42,43].

Results and Discussion

454 sequencing statistics and annotation of transcripts

The eight venom gland single-strand cDNA libraries were sequenced using a multiplex strategy. To this end, each cDNA library was barcoded with a unique 10-base sequence (MID, Multiplex IDentifier) that is recognized by the sequencing analysis software, allowing for automated sorting of MID-containing reads. A total of 334,540 reads (amounting ~ 62 Mb of 8 snake venom gland transcriptomes) were simultaneously sequenced in two runs of the Genome Sequencer FLX System. Raw sequencing data are archived under accession number SRP003780 in the NCBI Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/Traces/sra) [44]. Accession codes by species are SRS117169.3 (*B. asper* (car)); SRS117211.3 (*B. asper* (pac)); SRS117214.2 (*C. simus*); SRS117215.1 (*A. picadoi*); SRS117216.1 (*A. mexicanus*); SRS117217.1 (*C. godmani*); SRS117218.1 (*B. lateralis*); and SRS117219.2 (*B. schlegelii*). The first run included only cDNA from *B. asper* from the Caribbean versant of Costa Rica, and was performed as a test run. The second run was done using cDNA from all the species investigated. The average length per read was 186.6 bases (max. 645 bp, and only 3.27% of reads < 50 bp), and this figure is in keeping with other 454 transcriptomic reports conducted in non-model species [45]. 4530 reads could not be assigned due to sequencing errors in the 3' 10-mer label. This figure corresponds to a sequencing error rate of 1.35%, which is higher than that (~ 0.04%) reported in other studies[35,44]. In addition, 5,165,220 nucleotides (8.27%) were masked with N's characters by RepeatMasker. The vast majority of sequence elements masked are class I (retroelements) and class II (DNA transposons) mobile elements [46-48]. Retroelements may have a profound impact on the plasticity of the host genomes [49], i.e. modulating transcription of immediately downstream host genes [50,51]. The bulk (64%) of mobile elements identified in the snake venom gland transcriptomes investigated here are retrotransposons (Additional file 1: Table S1). Retrotransposable elements have been previously reported in the transcriptomes of *Bothrops insularis* (4.1% of ESTs) [52], *Lachesis muta* (0.3%) [53], and *Philodryas olfersii* (4.1%) [54], and in PLA2 genes from the venom gland of *Vipera ammodytes* [55,56] and *Protobothrops flavoviridis* [57]. In the context of multigene toxins, which like the snake venom PLA_2s are evolving under strong positive adaptive selection [58-60], it is worth mentioning that transposable elements are overrepresented in the mRNAs of rapidly evolving genes [61], suggesting that they have played a role in the diversification and expansion of these gene families [61,62].

The 454 sequencing run yielded a total of 330,010 masked reads, which were distributed among the 8 venom gland transcriptomes as displayed in Additional file 1: Table S2. In the absence of any reference genome to guide the assembly, the sets of reads of each species were separately processed with program Newbler, the *de novo* assembler tool of the 454 Sequencing platform. However, only 58.4% of all reads clustered into 31,025 contigs (average length of the contigs was 208.2 bp; average number of reads per contig = 6), of which 43% comprised only 2 reads. The program also returned 103,357 singletons (mean length, 174.4 bp). Employing other assembler programs, such as MIRA (http://www.chevreux.org/projects_mira.html) or Velvet [63], and using different settings (i.e. Velvet: hash-length 21 or 31; MIRA: job normal or accurate), did not improve the assembly performance. The transcriptome assembly problem has been documented [64,65], particularly for organisms without a reference genome database. Because of the low data compression gained in the assembly step and the small difference between contigs and reads mean length, bioinformatic processing of the 454 sequence data was performed on whole sets of unassembled reads. The set of 330,010 reads was searched against non-redundant GenBank databases using BLASTX and BLASTN algorithms to identify similar sequences with an e-value cutoff $<10^{-3}$. 100,394 (30.4%) produced significant hits (Additional file 1: Table S2). The high percentage of reads without significant similarity to any known sequence is in line with previous transcriptomic studies. Hence, the 8696 high quality ESTs from a non-normalized cDNA library of *D. acutus* were assembled into 2855 clusters, of which

only 45.60% matched known sequences and 54.40% had no match to any known sequence in Genbank [39,66].

BLAST hits included 79,991 matches to entries of the taxonomic suborder *Serpentes*, of which 62,433 (62% of BLAST hits) displayed similarity to documented venom proteins (Additional file 1: Table S2). The set of reads lacking similarity to *Serpentes* entries was searched for the presence of cysteine-rich domains (eg. stretches of 50-100 amino acids containing ≥10% cysteine residues), as this feature is commonly shared by many toxin sequences [67]. The survey proved fruitless. Further attempts to enlarge the toxin dataset by searching specific databases, such as the Animal Toxin Database [68] or MEROPS [69], were also unsuccessful. The venom protein families identified, and their relative abundance, in the whole 454 read sequence dataset are listed in Table 1. The relative distribution of these venom protein families among the eight taxa investigated is shown in Table 2.

Among the proteins listed in Table 1, glutaminyl cyclase (GC) belongs to the group of venom proteins without demonstrated toxic activity. Glutaminyl cyclase has been been identified in the venom proteomes of

Table 1 Identity and relative abundance of venom protein entries identified in the whole 454 read sequence dataset of the 8 Costa Rican snake venom gland transcriptomes

	Number of reads	% of total venom protein entries
Bradykinin potentiating peptide (BPP)	9231	14.8
Cysteine-Rich Secretoy Peptide (CRISP)	1066	1.7
C-type lectin-like protein (CTL)	1039	1.6
Growth factor (GF)	789	1.2
L-amino acid oxidase (LAO)	2535	4.0
Phospholipase A_2 (PLA_2)	7065	11.3
Metalloproteinase (SVMP)	26646	42.7
Serine Proteinase (SP)	10019	16.0
5'-nucleotidase (5'-NTase)	374	0.6
Phosphodiesterase (PDE)	119	0.2
Glutaminyl cyclase (GC)	170	0.3
Cobra Venom Factor (CVF)	8	0.01
Crotamine (CRO)	22	0.04
Sarafotoxin (SARA)	3	0.005
Waprin (WAP)	26	0.04
Kunitz-type inhibitor (KUN)	21	0.03
Kazal-type inhibitor (KAZ)	21	0.03
Hyaluronidase (HYA)	24	0.04
Ohanin (OHA)	2412	3.9
Three-Finger Toxin (3FTx)	845	1.3

Their relative abundance in each transcriptome is displayed in Table 2.

B. jararaca, *C. atrox*, and *C. durissus terrificus* [70-72]. N-terminal pyrrolidone carboxylic acid (pyroglutamate, pGlu) formation from its glutaminyl (or glutamyl) precursor is required in the maturation of numerous bioactive peptides. Snake venom GC is likely involved in the biosynthesis of pyroglutamyl peptides such as hypotensive BPPs [73,74], and endogenous inhibitors of metalloproteinases, pEQW, pENW, and pEKW [75,76]. Accumulation of peptide inhibitors in venoms provides a basis for attenuating the proteolytic activity of venom gland-stored SVMPs, preventing thereby autodigestion [77]. Mature PIII-SVMPs secreted into the venom proteome usually contain an N-terminal pyroglutaminyl residue (unpublished results), suggesting the action of the glutaminyl cyclase downstream of the proteolytic processing of the metalloproteinase precursor. However, the structural/functional consequences of N-terminal cyclation are unknown.

To estimate the number of toxin transcript sequences expressed in each transcriptome, multiple alignments of all reads clustered in the same protein family were generated, using the most similar full-length reference sequence as template. It was then realized that a large number of "*Serpentes* venom protein" reads did not align with translated ORFs. Instead, these reads appeared to represent 5'UTR, 3'UTR, and microsatellite loci. Particularly, all reads matching "cobra venom factor", "crotamine", "crotasin", and "sarafotoxin" entries corresponded to non-translated, mostly (87-100%) microsatellite DNA. In addition, 2397 out of 2412 reads for ohanin, and 840/845 3FTx reads aligned with microsatellite DNA. Microsatellite sequences accounted also for 66% GF, 49% SP, 36% PLA_2, 27% CRISP, 15% LAO, and 8% CTL, but represented less than 5% of the reads of the rest of venom protein classes listed in Table 2 and Additional file 1: Table S3. On the other hand, the bulk (> 99%) of non-microsatellite untranslated sequences corresponded to 3' UTRs. Additional file 1: Table S3 summarizes the number of reads aligned to translated regions of reference snake venom toxin sequences. The occurrence of a large number of microsatellites in the venomous snake *Agkistrodon contortrix* has been recently reported by Castoe and colleagues [78], who used the 454 Genome Sequencer FLX next-generation sequencing platform to sample randomly ~27 Mbp (128,773 reads) of the genome of this species. These authors identified microsatellite loci in 11.3% of all reads obtained, with 14612 microsatellite loci identified in total.

The presence of mRNA coding for 3'-untranslated regions of toxins points to a) a bias due to the first-strand synthesis method used, which produced cDNA libraries enriched in 3'-end-transcripts, b) incompletely sequenced transcripts or c) to transcription of nonfunctional gene

Durban *et al. BMC Genomics* 2011, **12**:259
http://www.biomedcentral.com/1471-2164/12/259

Table 2 Relative contribution of the different venom protein family hits in each of the Costa Rican snake venom gland transcriptome

	C. simus		*B. asper* (Car)		*B. asper* (Pac)		*C. godmani*		*A. picadoi*		*A. mexicanus*		*B. schlegelii*		*B. lateralis*	
	Reads	**%**	**Reads**	**%**	**Reads**	**%**	**Reads**	**%**	**Reads**	**%**	**Reads**	**%**	**Reads**	**%**	**Reads**	**%**
BPP	178	13.4	4232	15.0	169	12.2	818	8.3	1634	16.4	406	13.0	444	15.3	1350	23.6
CRISP	0	0	245	0.8	3	0.2	253	2.6	261	2.6	17	0.5	105	3.6	182	3.2
CTL	32	2.4	287	1.0	19	1.3	23	0.2	476	4.8	15	0.5	7	0.2	180	3.1
GF	60	4.5	275	0.9	38	2.7	64	0.6	56	0.6	50	1.6	142	4.9	103	1.8
LAO	50	3.8	1197	4.2	33	2.4	537	5.5	308	3.1	121	3.9	92	3.2	197	3.4
PLA_2	161	12.1	5026	17.8	195	14.2	777	7.9	228	2.3	224	7.2	278	9.6	176	3.0
SVMP	86	6.5	11733	41.6	559	40.6	4144	42.2	5583	56.1	1204	38.7	762	26.3	2575	44.9
SP	373	28.1	3790	13.4	114	8.3	2770	28.1	1089	10.9	692	22.2	594	20.5	597	10.4
5'-NTase	3	0.2	214	0.7	3	0.2	67	0.7	13	0.1	27	0.8	5	0.17	42	0.7
PDE	3	0.2	56	0.2	3	0.2	13	0.1	8	0.08	0	0	2	0.06	33	0.6
GC	5	0.4	108	0.4	4	0.3	32	0.3	13	0.1	7	0.2	0	0	1	0.02
CVF	2	0.15	2	0.006	0	0	1	0.01	0	0	2	0.06	1	0.03	0	0
CRO	0	0	10	0.03	4	0.3	2	0.02	4	0.04	0	0	1	0.03	1	0.02
SARA	1	0.07	2	0.006	0	0	0	0	0	0	0	0	0	0	0	0
WAP	0	0	26	0.09	0	0	0	0	0	0	0	0	0	0	0	0
KUN	2	0.15	10	0.03	3	0.2	0	0	1	0.01	0	0	4	0.12	1	0.02
KAZ	0	0	0	0	0	0	0	0	0	0	0	0	9	0.3	12	0.2
HYA	3	0.2	7	0.02	1	0.07	4	0.04	1	0.01	1	0.03	5	0.17	2	0.03
OHA	199	15.0	779	2.8	165	11.9	256	2.	233	2.3	254	8.2	338	11.6	188	3.3
3FTx	169	12.7	221	0.8	65	4.7	63	0.6	43	0.4	89	2.8	104	3.6	91	1.6

Protein family names are abbreviated as in Table 1.

copies. Relevant to the latter possibility, excepting *A. mexicanus* venom, which contains a small amount (<0.1%) of a 3Ftx [50], CVF, SARA, OHA, WAP, 5'NTase, PDE, GC, KUN, HYA and 3FTx have not been detected in the venom proteomes of the Costa Rican snakes sampled here [49-52] (Additional file 1: Table S4). Fry and colleagues [79] have shown that the venom system is a basal characteristic of the advanced snakes, and have investigated the timing of toxin recruitment events and patterns of toxin diversification across the full range of the ~100 million-year-old advanced snake clade [2,3,5,6,79]. These studies revealed single early recruitment events for each toxin type, including those identified here (Table 1), indicating that the venomous function arose once in squamate reptile evolution, at about 200 Myr ago. Structural and functional diversification of the venom system is best described by the birth-and-death model of protein evolution [80,81]. Pseudogenes in Costa Rican pitviper venom transcriptomes may thus represent relics of the evolution of their venom arsenal.

The 37,961 reads comprising non-venom-protein BLAST hits were classified based on the presumed biological process to which they may contribute (Figure 1A) and on their putative molecular function (Figure 1B), according to the Gene Ontology database [48]. Their relative abundance, biological processes (general metabolism, response to external stimuli, cell differentiation, proliferation, and communication, cell cycle...), and molecular functions (transcription and translation, protein binding, catalysis, etc.) identified in this work generally agree with the broad categories reported for other viperid transcriptomes [27,32,52,53,82-84], and will not be described here in detail again. The most abundant transcripts are related to DNA transcription, mRNA translation, and post-translational processing of proteins, reflecting the high specialization of this tissue for expressing and secreting toxins to the lumen of the venom gland. Furthermore, many venom toxins bear a high number of cysteinyl residues, which are engaged in extensive intra- and intermolecular disulphide crosslinking [20]. Venom proteins such as disintegrins, C-type lectin-like proteins, serine proteinases, PLA_2s, 3FTxs, and SVMPs occur in different oligomerization states [85-88]. The large structural impact at low energy cost of engineering disulphide bonds represents an opportunity for structural (and functional) diversification of proteins during evolution. Not surprisingly, protein disulphide isomerase (PDI), an enzyme and chaperone involved in disulphide bond formation in the endoplasmic reticulum [89,90], represents a highly expressed gene transcript (1859 reads; 5.1% of non-venom-protein reads) in all venom gland transcriptomes, ranging from 2.3% in *C. simus* to 6.9% in *B. lateralis*.

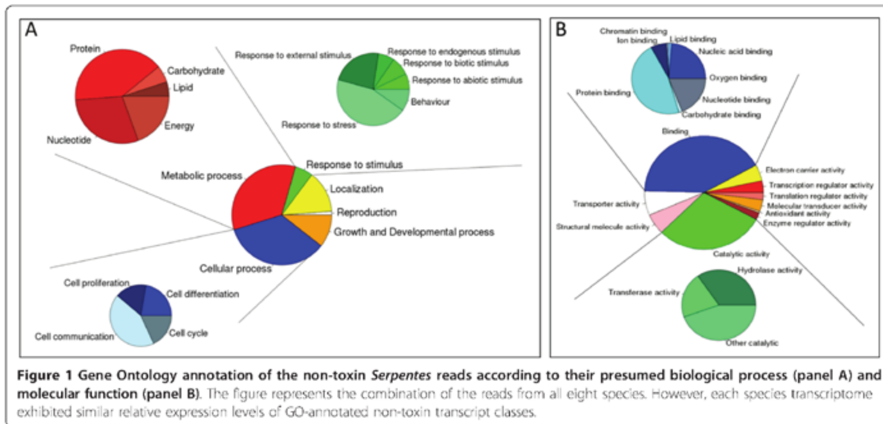


Figure 1 Gene Ontology annotation of the non-toxin *Serpentes* reads according to their presumed biological process (panel A) and molecular function (panel B). The figure represents the combination of the reads from all eight species. However, each species transcriptome exhibited similar relative expression levels of GO-annotated non-toxin transcript classes.

It is also worth mentioning the finding of reads for ribosomal 12S and 16S RNAs. This finding suggests that either internal mRNA A-rich tracts may have acted as priming sites in the cDNA synthesis, or that these messengers contained poly(A) tails. The possibility that rRNAs represent some residual contamination in the mRNA preparation should also be taken into account. Although polyadenylation is a distinctive feature of mRNA, polyadenylation of rRNA has been reported to occur in mammals and several unicellular organisms (*Candida albicans, Saccharomyces cerevisiae, Leishmania braziliensis* and *L. donovani*), and it may have a quality control role in rRNA degradation [91-94]. Polyadenylated ribosomal RNA has been also reported in the venom gland transcriptome of the Desert Massasauga Rattlesnake (*Sistrurus catenatus edwardsii*) [95].

Calculation of the minimum number of gene copies from each toxin family

An estimation of the minimum number of genes from each toxin family transcribed into the venom gland transcriptome of each species was calculated from the multiple alignments of reads matched to a full-length reference sequence of each toxin family (Figure 2). To this end, the nucleotide sequences of the ORF-coding reads of each venom protein family were assembled into contigs using MIRA and an iterative multiple-pass reference-guided protocol. MIRA is recommended for analysis of a normalized dataset or a filtered set of reads that did not have extreme coverage [64]. Each line of the multiple alignment (Figure 2) contained a distinct set of contigs spanning the maximum possible number of nucleotide positions of the reference sequence. Since the short average length of singletons and contigs (174-208 bp), and the low coverage of reads per contig (6), prevented the generation of full-length gene sequences, each line of the alignment corresponds to one or more synthetic gene. We considered two contigs as different if their nucleotide sequences depart in more number of positions than expected from a sequencing error rate of 1.35%, and the same mutated residues were found in at least two other reads. For each toxin family from each venom gland transcriptome a representation of the "number of reads per contig" vs. "number of contigs" was plotted, and only contigs accounting for ~95% of all assembled reads were considered. The rationale for

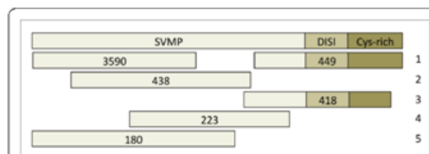


Figure 2 Calculation of the minimum number of gene copies. Multiple alignment of the top six SVMP transcripts of *B. asper* (Car) (Additional file 1: Table S6) using the sequence (top) of the most similar database-annotated toxin sequence as template. Each line of the multiple sequence alignment displays a distinct set of contig(s), comprised by a unique set of reads indicated in parentheses (see also Additional file 1: Table S5). Since the short average length of the reads and the low coverage of reads per contig prevented the assemblage of reliable gene sequences, each line of the alignment corresponds to at least a distinct gene of the SVMP multigene family translated into the venom gland transcriptome of *B. asper* (car).

introducing this quality trimming is because in this way only contigs in which the observed sequence differences were validated in a significant number of reads were taken into account, eliminating thus potential false positives due to sequencing errors, which generate "orphan" reads. The number of topologically equivalent homologous multiply-aligned reads corresponds thus to the minimum number of genes from a given toxin family transcribed into the venom gland transcriptome (Figure 2). The outcome of this analysis is displayed in Table 3. The estimated number of toxin-coding genes is in line with the number of different proteins identified in the respective venom proteomes: *C. simus* (27 reverse-phase HPLC fractions, ~20 proteins) [52]; *B.asper* (Car) (31 HPLC fractions, ~30 proteins) [51]; *B. asper* (Pac) (30 HPLC fractions, ~27 proteins) [51]; *A. mexicanus* (41 HPLC fractions) and *A. picadoi* (30 HPLC fractions), each containing bradykinin-potentiating peptides and around 25-27 proteins [50]; and *B. schlegelii* (34 HPLC fractions) and *B. lateralis* (34 HPLC fractions) matched to ~29 and 27 proteins, respectively [49]. Moreover, in most cases the overwhelming majority of reads of the large multigene toxin families (i.e., SVMP, PLA_2, and SP) cluster into a small subset of contigs (Additional file 1: Tables S5, S6 and S7). The uneven distribution of SVMPs of *B. asper* (Car) (Figure 3) clearly illustrates this point: 3590 out of 6746 reads clustered into a single contig, and only 6 other contigs were assembled from 449-173 reads. The remaining 22 transcripts comprised each 95-12 reads. The low number of venom proteins inferred from our 454 transcriptomic analysis is also in concordance with a recent high-throughput venom pland transcriptomic analysis for the Eastern Diamond Rattlesnake (*C. adamanteus*), which identified 40 unique toxin transcript [40]. The most diverse and highly expressed toxin classes were the SVMPs (11 isoforms), serine proteinases and C-type lectin-like proteins (9 different protein species each).

The insight provided by our present transcriptomic data, supported by previous proteomic studies, indicate that the venoms of the Costa Rican snakes investigated are comprised by toxins belonging to a few major protein families. In addition, our data suggest that different genes of a multigene family are subjected to very distinct transcription (and translation) yields, i.e. as the result of distinct stability and translational rates of the messengers.

Comparison of transcriptomes and proteomes

The relative abundance of the different toxin families in each transcriptome was calculated as the percentage of toxin family-specific reads relative to all BLAST hits (Table 2) or to the set of reads aligned to translated (ORF) regions of a reference sequences (Additional file 1: Table S3). To estimate the relative contribution of each toxin family, the total number of nucleotides of the ORF-coding reads was normalized for the full-length nucleotide sequence of a canonical member of the protein family. When available, the obtained figures were compared with the percentages of toxin families reported for the venom proteome of the same species. The outcome of this comparative analysis is compiled in

Table 3 Estimation of the minimum number of toxin family gene copies translated in the venom gland transcriptomes of Costa Rican snakes

	C. simus	*B. asper* (Car)	*B. asper* (Pac)	*C. godmani*	*A. picadoi*	*A. mexicanus*	*B. schlegelii*	*B. lateralis*
BPP	1	1	1	2	1	1	4	2
CRISP	0	2	1	2	4	1	2	1
CTL	2	5	2	3	9	1	0	5
GF	2	5	1	3	3	1	1	1
LAO	3	2	2	4	5	2	3	3
PLA_2	3	9	4	4	2	2	1	3
SVMP	9	29	5	19	15	4	14	20
SP	6	15	1	13	7	6	8	11
5'-NTase	1	3	1	2	2	2	1	3
PDE	1	1	1	2	2	0	1	2
GC	1	2	1	1	1	1	0	1
WAP	0	2	0	0	0	0	0	0
HYA	2	2	0	1	1	1	0	1
OHA	0	0	0	0	1	1	0	0
3FTx	1	0	0	0	0	0	0	0
KUN	0	0	0	0	0	0	1	0
KAZ	0	0	0	0	0	0	1	1

Protein family names are abbreviated as in Table 1.

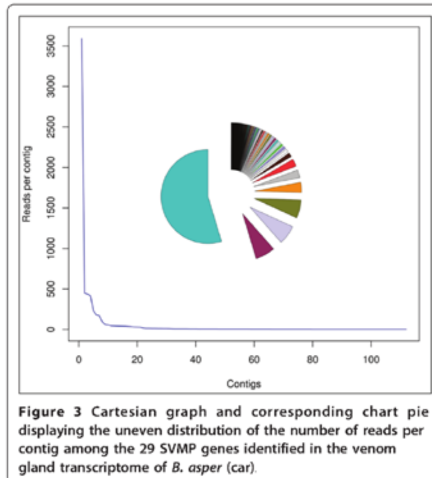


Figure 3 Cartesian graph and corresponding chart pie displaying the uneven distribution of the number of reads per contig among the 29 SVMP genes identified in the venom gland transcriptome of *B. asper* (car).

Additional file 1: Tables S3 and S4. Strong discrepancies between the transcriptome-computed and the proteome-gathered toxin compositions are obvious at first sight. The best, although still far from perfect, agreement between proteomic and transcriptomic data occured when the relative abundance of transcripts was computed using all the reads (ORFs + UTRs) belonging only to toxin classes detected in the venom proteome (Figure 4). This would support the view that the majority of reads matching UTRs may indeed form part of parent translatable mRNAs. However, particularly *B. asper* (Car) (Figure 4A), *B. schlegelii* (Figure 4B), *A. mexicanus* (Figure 4C), and *C. simus* (Figure 4D) strongly depart from this picture. The reasons underlying this discrepancy are elusive, since no clear trend within or between species is apparent, but both intrinsic (methodological) and extrinsic (biological) factors may be involved. Hence, besides the difficulty of deciding between bias due to cDNA libraries enriched in 3'-end-transcripts, and the presence of transcripts of pseudogenes in the transcriptome, we hypothesize that the distinct stability and translational rates of the messengers might also contribute to the observed differences between transcriptome and proteome. Thus, a high abundant messenger subjected to a higher degradation rate may produce the same concentration as a low abundance but more stable mRNA or exhibiting a higher translational rate. The observation that *B. asper* (Car) and *B. asper* (Pac), as well as *A. mexicanus* and *A. picadoi*, exhibit highly similar transcriptomes but strongly depart in the relative toxin composition of their venom proteomes (Figure 4), indicates that individual mRNA species are translationally controlled in a species-dependent manner. The same conclusion can be drawn by comparing the proteome and transcriptome of *C. simus* (Figure 4D). In this respect, mounting evidences in yeast, indicate that the loading of ribosomes onto individual mRNA species varies broadly across a cellular transcriptome, and this finding is consistent with each transcript having a uniquely defined efficiency of translation [96-99].

A rough comparison of the transcriptomes (Table 2) shows that some toxin families are relatively constantly expressed among snakes while others exhibit greater variability. Principal Component Analysis (PCA) revealed that the abscissa (PC1) and the ordinate (PC2) each explained 32% of the observed variability (Figure 5A). PCA discriminated the eight transcriptomes into four groups (Figure 5B). Referring to the average value of the toxin family, the transcriptomes of *A. picadoi* and *B. lateralis* contain higher content (⇑) of SVMP reads and lower number (⇓) of PLA_2 and SP reads; the two *B. asper* taxa express ⇑SVMPs and PLA_2s and ⇓ SPs; *A. mexicanus* and *C. godmani* contain ⇑SVMPs, ⇑SPs, and slightly ⇓PLA_2s; and *B. schlegelii* and *C. simus* contain ⇓SVMPs, ⇑SPs, and average number (⇔) of PLA_2 reads. These data point to convergent and divergent evolutionary trends among pitvipers. PCA of the venom proteomes listed in Additional file 1: Table S4 revealed different clustering of taxa (Figure 5, panels C and D), and unequal contributions of PC1 (48%) and PC2 (22%) to venom variability. These results illustrate the versatility of snake venoms as a system to achieve the purpose of subduing prey through different strategies [100]. On the other hand, the lack of any apparent correlation between the PCA score plots for transcriptome (Figure 5B) and proteome (Figure 5D) data, further highlights the existence of variable translational patterns across species. Clearly, our results emphasize the relevance of combining detailed proteomic and transcriptomic studies for a thorough characterization of the venom toxin repertoire and the factors regulating transcription and translation.

In a previous proteomic study, we identified two RP-HPLC fractions of *B. schlegelii* venom as Kazal-type proteinase inhibitor-like proteins (family PD000417 in the ProDom database, http://prodom.prabi.fr/prodom/current/html/home.php) [49]. Kazal-type inhibitor-like proteins (KAZ) have not been found in any other snake venom reported to date, casting doubts on their venom gland origin, on the one hand, or pointing to a recruitment event of these proteins along the speciation of the Neotropical pitviper clade [49]. Now we report the finding of 9 and 12 reads encoding ORF regions of KAZ in

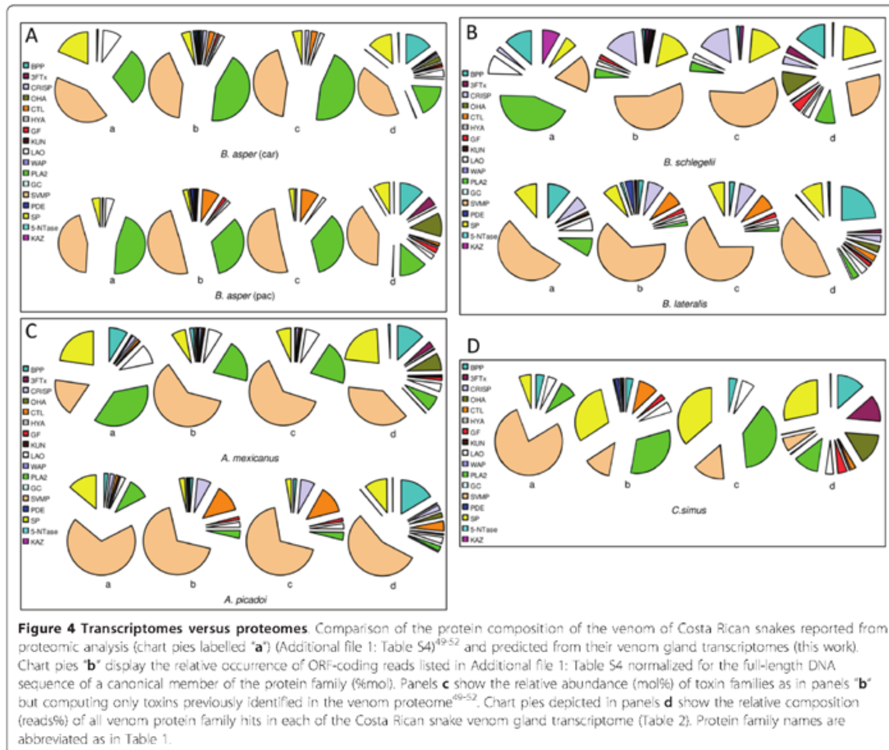


Figure 4 Transcriptomes versus proteomes. Comparison of the protein composition of the venom of Costa Rican snakes reported from proteomic analysis (chart pies labelled "**a**") (Additional file 1: Table S4)[49-52] and predicted from their venom gland transcriptomes (this work). Chart pies "**b**" display the relative occurrence of ORF-coding reads listed in Additional file 1: Table S4 normalized for the full-length DNA sequence of a canonical member of the protein family (%mol). Panels **c** show the relative abundance (mol%) of toxin families as in panels "**b**" but computing only toxins previously identified in the venom proteome[49-52]. Chart pies depicted in panels **d** show the relative composition (reads%) of all venom protein family hits in each of the Costa Rican snake venom gland transcriptome (Table 2). Protein family names are abbreviated as in Table 1.

B. schlegelii and *B. lateralis* transcriptomes, respectively (Table 2 and Additional file 1: Table S4). In each species, all KAZ reads assembled into a single transcript (Table 3), suggesting a monogenic origin. The occurrence of Kazal-type inhibitor-like proteins only in the venom gland transcriptomes of the two Bothriechis taxa (Additional file 1: Tables S4 and S5) supports the view of a genus-specific recruitment event during the early-Middle Miocene ~14 Mya, the estimated divergence time for Bothriechis in a model of Middle American highland speciation [101]. On the other hand, the presence of Kazal-type proteins in the venom proteome of *B. schlegelii*, the basal species of the Bothriechis clade [102], suggests a species-specific expression of this class of protein. Though a number of Kazal-like domains harbor serine proteinase inhibitor activity, these protein scaffolds are also present in the extracellular part of a number of proteins, which are not known to be proteinase inhibitors. Clearly, further investigations are needed to assess the biological activity of the Kazal-type proteins, and the role that these proteins may have played in the early adaptive radiation of the *Bothriechis* clade.

Transcriptome-based cladistic analysis of the Costa Rican snake venom gland transcriptomes

To assess structural relationships between the transcriptomes, consensus sequences were constructed for those major toxin families shared by all snake venom gland transcriptomes, i.e. BPP, LAO, PLA2, SVMP, and SP. To assess the degree of kinship between the Costa Rican snake venom gland transcriptomes, species-specific synthetic sequences were generated by the concatenation of the 5 toxin-family consensus sequences (in the order described above), and using these synthetic sequences as input, a cladogram was built using the suite of web-

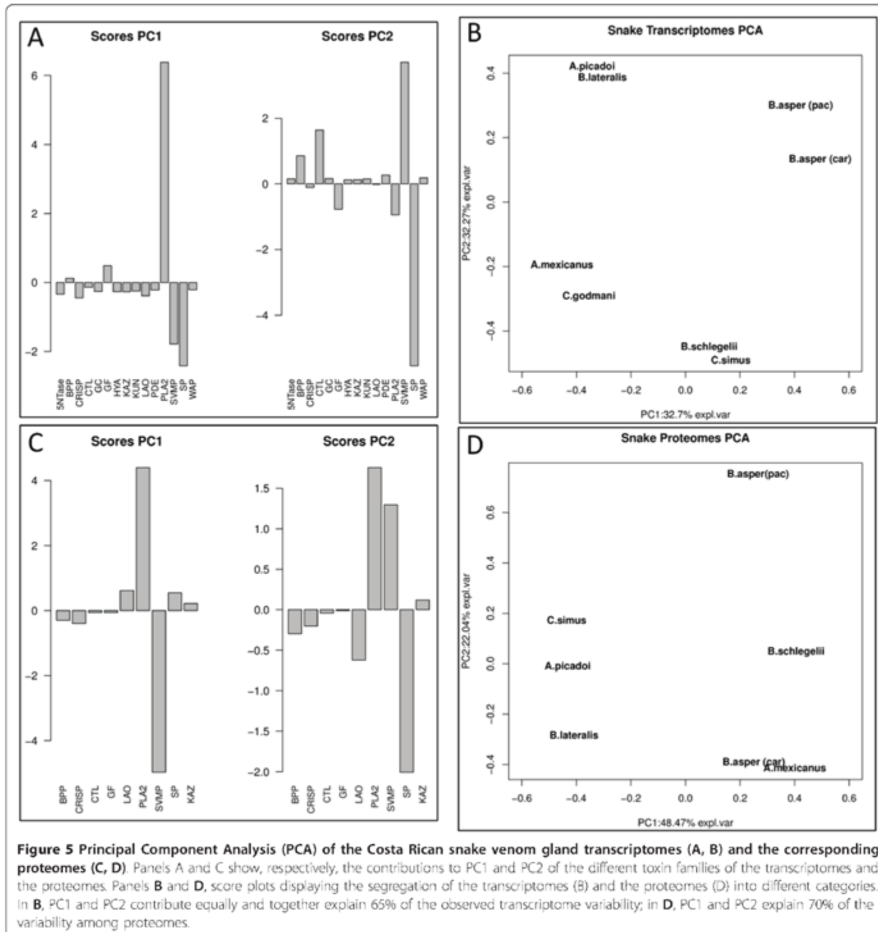


Figure 5 Principal Component Analysis (PCA) of the Costa Rican snake venom gland transcriptomes (A, B) and the corresponding proteomes (C, D). Panels A and C show, respectively, the contributions to PC1 and PC2 of the different toxin families of the transcriptomes and the proteomes. Panels **B** and **D**, score plots displaying the segregation of the transcriptomes (B) and the proteomes (D) into different categories. In **B**, PC1 and PC2 contribute equally and together explain 65% of the observed transcriptome variability; in **D**, PC1 and PC2 explain 70% of the variability among proteomes.

tools Phylemon (http://phylemon.bioinfo.cipf.es) (Figure 6). Mutation is the driving force of evolution, but inferring evolutionary distances from multiple sequence alignments can yield misleading results if the mutation rates of the genes being compared are unequal across species. Given knowledge of the degree of mutation rate heterogeneity, appropriate algorithms can be applied to correct unbiasness and inaccuracy of the phylogenetic reconstruction [103-105]. Globally, translated assembled sequences displayed mean variability levels between 0 and 7.4% (computed as number of variable residues divided by sequence length), being the SVMPs and SPs the toxin families which accumulate more amino acid substitutions. Although the cladogram depicted in Figure 6 should not be regarded as an evolutionary hypothesis, the divergence of the Atropoides taxa, and the clustering of *A. mexicanus* and *C. godmani* deserves discussion.

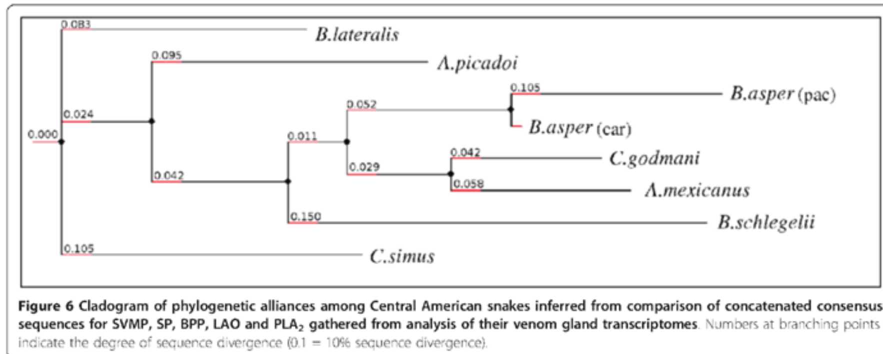


Figure 6 Cladogram of phylogenetic alliances among Central American snakes inferred from comparison of concatenated consensus sequences for SVMP, SP, BPP, LAO and PLA_2 gathered from analysis of their venom gland transcriptomes. Numbers at branching points indicate the degree of sequence divergence (0.1 = 10% sequence divergence).

Despite the efforts of numerous authors, phylogenetic relationships within the subfamily Crotalinae remain controversial, particularly at the intergeneric level [100,105]. In particular, several analyses, even from the same research group, support different phylogenetic models. Thus, Bayesian Markov chain Monte-Carlo results suggested the monophyly of the three genera of the Porthidium group (*Atropoides*, *Cerrophidion*, and *Porthidium*) and indicated that *Cerrophidion* and *Porthidium* form a clade that is the sister taxon to *Atropoides* [106]. On the other hand, genus *Atropoides* has been also inferred through Bayesian phylogenetic methods to be paraphyletic with respect to *Cerrophidion* and *Porthidium*, due to *Atropoides picadoi* being distantly related to other *Atropoides* species [107,108]. Although resolving the phylogenetic relationships among the Neotropical pitvipers of the Porthidium group requires a detailed genomic study of species occupying geographically close ecological niches, i.e., *Porthidium nasutum*, *Porthidium ophryomegas*, *Porthidium volcanicum*, and *Cerrophidion godmani*, both a previous proteomic study by Angulo and co-workers [50], who estimated that the similarity of venom proteins between the two Atropoides taxa may be around 14-16%, and the present transcriptomic analysis, support a large divergence between *A. mexicanus* and *A. picadoi*, and the closer kinship between *A. mexicanus* and *C. godmani*.

Conclusions

The snake venom gland is a highly specialized and sophisticated organ, which harbors the cellular machinery that transformed throughout > 200 million years of evolution genes coding for ordinary proteins of disparate scaffolds, diverse ancestral bioactivities, and recruited from a wide range of tissues, into lethal toxins [2]. Although the details of recruitment and neofunctionalization of these proteins remain elusive, gene duplication events, followed by the accelerated evolution of some copies and degradation of others to pseudogenes, underlay the emergence of venom gland multigene toxin families. Comparative analysis of complete genome sequences of squamate reptiles would be extremely valuable for tracking the evolution of the venom system in lizards and snakes [5,6]. On the other hand, a deep understanding of the toxin gene expression pattern revealed by high-throughput transcriptomic may reveal taxon-specific adaptations and the underlying biological processes governing the formulation of the venom arsenal. In this respect, invoking a reverse venomics approach, knowledge of the end product of transcriptome translation, the venom proteome, may provide hints on the translation efficiency of toxin-coding transcripts. The high-throughput capability of next-generation sequencing technologies offer the opportunity to generate large transcriptome databases relatively rapidly, which may help to speed up the tedious, often *de novo* [20], assignment of proteomic-gathered data. Furthermore, analysis of the venom gland transcriptome enhances the comprehensibility of the venom proteome, and this in turn contributes to a more accurate interpretation of the transcriptome. The application of NGS to the analysis of snake venom transcriptomes, may represent the tool for opening the door to systems venomics.

Methods

Snake venom gland cDNA synthesis and sequencing

Venom glands were removed 3 days after venom milking, when transcription is maximal [109], from anesthetized snakes using fine forceps and immediately placed in RNAlater™ solution (Qiagen). 30 mg of tissue were disrupted and homogenized by a rotor-stator homogenizer, and total RNA was isolated using RNeasy Mini kit (Qiagen), quantified in a spectrophotometer, and

quality-checked on an agarose gel discerning the 28S and 18S bands of ribosomal RNA. First strand cDNA was synthesized using RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas), which selectively transcribes full-length polyadenylated mRNA. The manufacturer's recommendations were followed except where specified. Approximately 5 μg of total RNA was used as starting material. In order to avoid polymerase slippage, a modified 3' 54-mer adaptor (5' GAGC-TAGTTCTGGAG$(T)_{16}$VN), which includes a type IIs enzyme (GsuI) site (underlined), was used for first-strand synthesis. This modified oligonucleotide effectively converts the long run of adenosine residues at the polyA tail into a sequence that causes fewer problems for dideoxy sequencing chemistry, and thus the resulting cDNA libraries were enriched in 3'-end-transcripts. To avoid internal cuts, the cDNA was hemimethylated by adding 5-methyl-dCTP to the dNTPs mix. The first strand cDNA was used as template for second strand synthesis by *E. coli* DNA Polymerase I and RNase H. Double strand (ds) cDNA was precipitated with ethanol and the pellet was resuspended in 70 μL of nuclease-free water and subjected to enzymatic digestion with GsuI for 4 hours at 30°C. The enzyme was then inactivated at 65°C for 20 minutes and the digested cDNA was stored at -20°C. For 454 pyrosequencing, the GS FLX General Library Preparation Method Manual workflow (Roche Diagnostics) was followed. To this end, 3 μg of final non-normalized cDNA library were sheared by nebulization into small fragments. The fragment ends were polished and short A/B adaptors were ligated onto both ends, providing priming regions to support both emulsion amplification and the pyrosequencing process. A biotin tag on the B adaptor allowed immobilization of the dscDNA library fragments onto streptavidin-conjugated magnetic beads and the subsequent isolation of the library of single strand cDNA sequencing templates. Each of the eight cDNA libraries was tagged with a unique 10-base sequence (MID, Multiplex IDentifier) that is recognized by the sequencing analysis software, allowing for automated sorting of MID-containing reads. Barcoded libraries were simultaneously sequenced in a Genome sequencing FLX System (Roche Applied Science) at Life Sequencing S.L. (Parc Científic Universitat de Valencia, Paterna, Valencia, Spain; http://www.lifesequencing.com) using the method developed by Margulies et al. [110].,

Bioinformatic processing of the 454 reads, identification of toxin transcripts, and quantitation of the expression levels of toxin families

Additional file 2: Figure S1 displays a scheme of the data analysis pipeline developed to identify sequences of toxin molecules by similarity search against nucleotide databases, and elaborate a reference-guide estimation of transcripts, which included available NGS algorithms and in-house scripts. To this end, interspersed repeats and low complexity DNA sequences were masked from the transcript reads using RepeatMasker (version 3.2.7) [111]. RepeatMasker is available from the Institute for Systems Biology (http://www.systemsbiology.org) addressing http://repeatmasker.org. The program makes use of Repbase, which is a service of the Genetic Information Research Institute (http://www.girinst.org). Repbase is a comprehensive database of repetitive element consensus sequences (update of 20 January 2009). Data and computational resources for the Pre-Masked Genomes page is provided courtesy of the UCSC Genome Bioinformatics group (http://genome.ucsc.edu). Masked reads were then searched against the non-redundant NCBI database (http://blast.ncbi.nlm.nih.gov, release of March 2009) and the UniProtKB/Swiss-Prot Toxin Annotation Program database (http://us.expasy.org/sprot/tox-prot), using BlastX and BlastN [112] algorithms, specifying a cut-off value of e-03 and BLOSUM62 as scoring matrix. Snake venom gland-specific transcripts were selected from best BLAST-hit descriptions identifying GenBank entries belonging to the taxonomic suborder *Serpentes*. This taxonomic group is represented by 37,070 records comprising entries from 357 different genera. The subset of reads exhibiting similarity to *Serpentes* sequences were further filtered using a list of keywords (including the acronyms of all known toxin protein families described so far [20,25]) to distinguish putative snake venom toxins from non-toxin (ribosomal, mitochondrial, nuclear, etc.), ordinary proteins. In a second round of filtering, non-matched sequences were searched for structural features (eg. high cysteine content) expected for a putative toxin molecule. Non-toxin-assigned transcripts were functionally annotated using the Blast2GO software [113] and classified using GO-terms [114]. The relative expression of a given toxin protein family (mol %) was calculated as the number of reads assigned to this protein family (R_i) normalized by the length (in nucleotides) of the reference transcript sequence (ntREF) and expressed as the % of total reads in the snake transcriptome (ΣReads): mol% toxin family "i" = $\%[(R_i/\text{ntREF})/\Sigma\text{Reads})$. The relationship between the expression levels of the different protein families in the different species sampled was analyzed by Principal Component Analysis (PCA). When possible, toxin transcriptome profiles were compared with available proteomic-based venom toxin profiles [115-118].

Assessing molecular diversity within toxin families and cladistic analysis

To assess the molecular diversity within each toxin family, the phylogenetically nearest top-hit sequence was

designated as the reference sequence for aligning all the toxin family-specific reads. To this end, each toxin family read was translated into the 6 possible reading frames and blasted against the reference protein sequence using tBlastN. Matched frames exhibiting e-value thresholds better than e-03 were aligned onto the reference sequence to create a multiple alignment using COBALT [119]. The multiple alignment was then parsed to create an assembled (consensus) toxin sequence in which each amino acid position is supported by at least four reads, representing at least 30% of the total number of reads at that position. Non-sequenced positions or those that did not meet the minimum coverage condition, are depicted in lower case in the assembled sequences. Positions where two or more amino-acids fulfilled the selection criteria were annotated as variable residues suggesting the occurrence of different alleles (isoforms) of the protein. Multiple sequence alignments based on transcriptomic data were computed with program MUSCLE [120] (version 3.52) using species-specific synthetic sequences constructed by the concatenation of the consensus sequences of the major toxin families shared by all snake venom gland transcriptomes, i.e. BPP, LAO, PLA2, SVMP, and SP, in this order. Phylogenetic tree reconstruction was done using the suite of web-tools Phylemon (http://phylemon.bioinfo.cipf.es/cgi-bin/tools.cgi) using the MUSCLE, ProtDist and Neighbor options of Phylip (version 3.65).

Additional material

Additional file 1: Table S1: RepeatMasker usage results and features of the sequence elements masked in the 8 Costa Rican venom gland transcriptomes analyzed **Table S2**: Summary of the 454 sequencing statistics and annotation of transcripts in the 8 venom gland transcriptomes. **Table S3**: Number of reads aligned to translated (ORF) regions of reference snake venom toxin sequences. **Table S4**: Relative occurrence (in %) of the ORF-coding reads listed in Table S3. **Table S5**: Distribution of reads per contig among the SVMP genes. **Table S6**: Distribution of reads per contig among the PLA_2 genes. **Table S7**: Distribution of reads per contig among the serine proteinase genes.

Additional file 2: Figure S1. Summary of the strategy employed to assembly and analyze the 454 pyrosequencing reads from the venom gland transcriptomes of the Costa Rican snakes *Bothrops asper* (from Caribbean and Pacific populations), *Bothriechis lateralis*, *Bothriechis schlegelii*, *Atropoides picadoi*, *Atropoides mexicanus*, *Crotalus simus*, and *Cerrophidion godmani*.

Acknowledgements

This work has been financed by grants BFU2007-61563 and BFU2010-17373 (to JJC), and BIO2009-10799 (to AC), from the Ministerio de Ciencia e Innovación, Madrid, Spain; PROMETEO/2010/005 from the Generalitat Valenciana; projects 741-A7-611 from the Vicerrectoría de Investigación, Universidad de Costa Rica, CRUSA-CSIC (2007CR0004 and 2009CR0021), and CYTED (206AC0281). This work was also partly supported by the National Institute of Bioinformatics (http://www.inab.org), an initiative of the Instituto de Salud Carlos III (ISCIII), Madrid (Spain). Traveling between Spain and Costa Rica was financed by Acciones Integradas 2006CR0010 between CSIC and the University of Costa Rica (UCR).

Author details

[1]Consejo Superior de Investigaciones Científicas, Jaime Roig 11, 46010 Valencia. Spain. [2]Instituto Clodomiro Picado, Universidad de Costa Rica, San José, Costa Rica. [3]Centro de Investigaciones en Estructuras Microscópicas, Universidad de Costa Rica, San José, Costa Rica. [4]Centro de Investigación Príncipe Felipe, Valencia. Spain.

Authors' contributions

PJ, MF-D, AA-G, and MS dissected the venom glands and construct the cDNA libraries. JD, along with JMG, AC, LS, and JJC, analyzed the transcriptomic data and participated in data interpretation and discussion, as well as in revising the article drafted by JJC. JD provided bioinformatic tools and lab space for conducting this study. All authors have and approved the final manuscript.

Received: 2 January 2011 Accepted: 23 May 2011
Published: 23 May 2011

References

1. Mackessy SP: **Morphology and ultrastructure of the venom glands of the northern Pacific rattlesnake *Crotalus viridis oreganus*.** *J Morphol* 1991, **208**:109-128.
2. Fry BG, Wüster W: **Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences.** *Mol Biol Evol* 2004, **21**:870-883.
3. Fry BG, Scheib H, van der Weerd L, Young B, McNaughtan J, Ramjan SF, Vidal N, Poelmann RE, Norman JA: **Evolution of an arsenal: structural and functional diversification of the venom system in the advanced snakes (Caenophidia).** *Mol Cell Proteomics* 2008, **7**:215-246.
4. 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 A, Richardson MK: **Evolutionary origin and development of snake fangs.** *Nature* 2008, **454**:630-633.
5. Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SF, Kuruppu S, Fung K, Hedges SB, Richardson MK, Hodgson WC, Ignjatovic V, Summerhayes R, Kochva E: **Early evolution of the venom system in lizards and snakes.** *Nature* 2006, **439**:584-588.
6. Fry BG, Vidal N, van der Weerd L, Kochva E, Renjifo C: **Evolution and diversification of the Toxicofera reptile venom system.** *J Proteomics* 2009, **72**:127-136.
7. Greene HW: **Dietary correlates of the origin and radiation of snakes.** *Am Zool* 1983, **23**:431-441.
8. World Health Organization: **Rabies and envenomings. A neglected public health issue: Report of a consultative meeting.** Geneva: WHO; 2007 [http://www.who.int/bloodproducts/animal_sera/Rabies.pdf].
9. Kasturiratne A, Wickremasinghe AR, de Silva N, Gunawardena NK, Pathmeswaran A, Premaratna R, Savioli L, Lalloo DG, de Silva HJ: **The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths.** *PLoS Med* 2008, **5**:e218.
10. Chippaux JP, Goyffon M: **Epidemiology of scorpionism: a global appraisal.** *Acta Trop* 2008, **107**:71-79.
11. Koh DCI, Armugam A, Jeyaseelan K: **Snake venom components and their applications in biomedicine.** *Cell Mol Life Sci* 2006, **63**:3030-3041.
12. Calvete JJ: **Venomics: digging into the evolution of venomous systems and learning to twist nature to fight pathology.** *J Proteomics* 2009, **72**:121-126.
13. Harvey AL, Bradley KN, Cochran SA, Rowan EG, Pratt JA, Quillfeldt JA, Jerusalinsky DA: **What can toxins tell us for drug discovery?** *Toxicon* 1998, **36**:1635-1640.
14. Ménez A: **Functional architectures of animal toxins: a clue to drug design?** *Toxicon* 1998, **36**:1557-1572.
15. Ménez A, Stöcklin R, Mebs D: **"Venomics" or: the venomous systems genome project.** *Toxicon* 2006, **47**:255-259.
16. Escoubas P, King GF: **Venomics as a drug discovery platform.** *Expert Rev Proteomics* 2009, **6**:221-224.
17. Vetter I, Davis JL, Rash LD, Anangi R, Mobli M, Alewood PF, Lewis RJ, King GF: **Venomics: a new paradigm for natural products-based drug discovery.** *Amino Acids* 2011, **17**:15-28.

18. Gutiérrez JM, León G: **Snake antivenoms: Technological, clinical and public health issues.** In *Animal Toxins: State of the Art. Perspectives in Health and Biotechnology.* Edited by: de Lima ME, Pimenta AMC, Martin-Euclaire MF, Zingali RB, Rochat H. Belo Horizonte: Editora UFMG; 2009:393-421.
19. Espino-Solis GP, Riaño-Umbarila L, Becerril B, Possani LD: **Antidotes against venomous animals: state of the art and prospectives.** *J Proteomics* 2009, **72**:183-199.
20. Calvete JJ, Juárez P, Sanz L: **Snake venomics. Strategy and applications.** *J Mass Spectrom* 2007, **42**:1405-1414.
21. Fox JW, Serrano SM: **Exploring snake venom proteomes: multifaceted analyses for complex toxin mixtures.** *Proteomics* 2008, **8**:909-920.
22. Escoubas P, Quinton L, Nicholson GM: **Venomics: unravelling the complexity of animal venoms with mass spectrometry.** *J Mass Spectrom* 2008, **43**:279-295.
23. Dos Santos LD, Santos KS, Pinto JR, Dias NB, de Souza BM, dos Santos MF, Perales J, Domont GB, Castro FM, Kalil JE, Palma MS: **Profiling the proteome of the venom from the social wasp *Polybia paulista*: a clue to understand the envenoming mechanism.** *J Proteome Res* 2010, **9**:3867-3877.
24. Rodríguez de la Vega RC, Schwartz EF, Possani LD: **Mining on scorpion venom biodiversity.** *Toxicon* 2009, **56**:1155-1161.
25. Junqueira de Azevedo ILM, Diniz MRV, Ho PL: **Venom gland transcriptomic analysis.** In *Animal Toxins: State of the Art. Perspectives in Health and Biotechnology.* Edited by: de Lima ME, Pimenta AMC, Martin-Euclaire MF, Zingali RB, Rochat H. Belo Horizonte: Editora UFMG; 2009:693-713.
26. Sanz L, Escolano J, Ferretti M, Biscoglio MJ, Rivera E, Crescenti EJ, Angulo Y, Lomonte B, Gutiérrez JM, Calvete JJ: **Snake venomics of the South and Central American Bushmasters. Comparison of the toxin composition of *Lachesis muta* gathered from proteomic versus transcriptomic analysis.** *J Proteomics* 2008, **71**:46-60.
27. 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**:609-623.
28. Fasoli E, Sanz L, Wagstaff S, Harrison RA, Righetti PG, Calvete JJ: **Exploring the venom proteome of the African puff adder, *Bitis arietans*, using a combinatorial peptide ligand library approach at different pHs.** *J Proteomics* 2010, **73**:932-942.
29. Ho PL, Soares MB, Yamane T, Raw I: **Reverse biology applied to *Micrurus corallinus*, a south american coral snake.** *J Toxicol Toxin Reviews* 1995, **14**:327-337.
30. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC: **Complementary DNA sequencing: expressed sequence tags and human genome project.** *Science* 1991, **252**:1651-1656.
31. Calvete JJ, Marcinkiewicz C, Sanz L: **Snake venomics of *Bitis gabonica gabonica*. Protein family, composition, subunit organization of venom toxins, and characterization of dimeric disintegrins Bitisgabonin-1 and Bitisgabonin-2.** *J Proteome Res* 2007, **6**:326-336.
32. Valente RH, Guimarães PR, Junqueira M, Neves-Ferreira AG, Soares MR, Chapeaurouge A, Trugilho MR, León IR, Rocha SL, Oliveira-Carvalho AL, Wermelinger LS, Dutra DL, Leão LI, Junqueira-de-Azevedo IL, Ho PL, Zingali RB, Perales J, Domont GB: ***Bothrops insularis* venomics: a proteomic analysis supported by transcriptomic-generated sequence data.** *J Proteomics* 2009, **72**:241-255.
33. Chippaux JP, Williams V, White J: **Snake venom variability: methods of study, results and interpretation.** *Toxicon* 1991, **29**:1279-1303.
34. 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.
35. Wheat CW: **Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing.** *Genetica* 2010, **138**:433-451.
36. Linnarsson S: **Recent advances in DNA sequencing methods. General principles of sample preparation.** *Exp Cell Res* 2010, **316**:1339-1343.
37. Morozova O, Hirst M, Marra MA: **Applications of new sequencing technologies for transcriptome analysis.** *Annu Rev Genomics Hum Genet* 2009, **10**:135-151.
38. Mardis ER: **The impact of next-generation sequencing technology on genetics.** *Trends Genet* 2008, **24**:133-141.
39. Zhang B, Liu Q, Yin W, Zhang X, Huang Y, Luo Y, Qiu P, Su X, Yu J, Hu S, Yan G: **Transcriptome analysis of *Deinagkistrodon acutus* venomous gland focusing on cellular structure and functional aspects using expressed sequence tags.** *BMC Genomics* 2006, **7**:152.
40. Rokyta DR, Wray KP, Lemmon AR, Lemmon EM, Caudle SB: **A high-throughput venom-gland transcriptome for the Eastern Diamond Rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes.** *Toxicon* 2011, **57**:657-671.
41. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet* 2009, **10**:57-63.
42. Gutiérrez JM: **Snakebite envenomation in Central America.** In *Handbook of Venoms and Toxins of Reptiles.* Edited by: Mackessy SP. CRC Press: Boca Raton, Florida; 2009:491-507.
43. Campbell JA, Lamar WW: *The Venomous Reptiles of the Western Hemisphere* Cornell University Press: Ithaca; 2004.
44. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Edgar R, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Ostell J, Pruitt KD, Schuler GD, Shumway M, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, Tatusova TA, Wagner L, Yaschenko E: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res* 2008, , **36 Database**: D13-21.
45. Wang W, Wang Y, Zhang Q, Qi Y, Guo D: **Global characterization of *Artemisia annua* glandular trichome transcriptome using 454 pyrosequencing.** *BMC Genomics* 2009, **10**:465.
46. McClintock B: **Controlling elements and the gene.** *Cold Spring Harb Symp Quant Biol* 1956, **21**:197-216.
47. Bannert N, Kurth R: **Retroelements and the human genome: new perspectives on an old relation.** *Proc Natl Acad Sci USA* 2004, **101**:14572-14579.
48. Deininger PL, Batzer MA: **Mammalian retroelements.** *Genome Res* 2002, **12**:1455-1465.
49. Hughes JF, Coffin JM: **Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution.** *Nat Genet* 2001, **29**:487-489.
50. Jern P, Coffin JM: **Effects of retroviruses on host genome function.** *Annu Rev Genet* 2008, **42**:709-732.
51. Pi W, Zhu X, Wu M, Wang Y, Fulzele S, Eroglu A, Ling J, Tuan D: **Long-range function of an intergenic retrotransposon.** *Proc Natl Acad Sci USA* 2010, **107**:12992-12997.
52. 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.
53. Junqueira-de-Azevedo ILM, Ching ATC, Carvalho E, Faria F, Nishiyama MY Jr, 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, **172**:877-889.
54. Ching AT, Rocha MM, Leme AFP, Pimenta DC, Furtado MFD, Serrano SM, Ho PL, Junqueira de Azevedo IL: **Some aspects of the venom proteome of the Colubridae snake *Philodryas olfersii* revealed from a Duvernoy's (venom) gland transcriptome.** *FEBS Lett* 2006, **580**:4417-4422.
55. Kordis D, Gubenšek F: **Bov-B long interspersed repeated DNA (LINE) sequences are present in *Vipera ammodytes* phospholipase A2 genes and in genomes of Viperidae snakes.** *Eur J Biochem* 1997, **246**:772-779.
56. Kordis D, Gubenšek F: **The Bov-B lines found in *Vipera ammodytes* toxic PLA2 genes are widespread in snake genomes.** *Toxicon* 1998, **36**:1585-1590.
57. Ikeda N, Chijiwa T, Matsubara K, Oda-Ueda N, Hattori S, Matsuda Y, Ohno M: **Unique structural characteristics and evolution of a cluster of venom phospholipase A_2 isozyme genes of *Protobothrops flavoviridis* snake.** *Gene* 2010, **461**:15-25.
58. Ohno M: **Evolution by gene duplication.** Springer Verlag: Berlin; 1970.
59. Lynch VJ: **Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A_2 genes.** *BMC Evol Biol* 2007, **7**:2.
60. Gibbs HL, Rossiter W: **Rapid evolution by positive selection and gene gain and loss: PLA_2 venom genes in closely related *Sistrurus* rattlesnakes with divergent diets.** *J Mol Evol* 2008, **66**:151-166.

61. van de Lagemaat LN, Landry JR, Mager DL, Medstrand P: **Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions.** *Trends Genet* 2003, **19**:530-536.
62. Medstrand P, van de Lagemaat LN, Dunn CA, Landry JR, Svenback D, Mager DL: **Impact of transposable elements on the evolution of mammalian gene regulation.** *Cytogenet Genome Res* 2005, **110**:342-345.
63. Zerbino DR, Birney E: **Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs.** *Genome Res* 2008, **18**:821-829.
64. Kumar S, Blaxter ML: **Comparing *de novo* assemblers for 454 transcriptome data.** *BMC Genomics* 2010, **11**:571.
65. Flicek P, Birney E: **Sense from sequence reads: methods for alignment and assembly.** *Nature Methods* 2009, **6**:S6-S12.
66. Liu Q, Zhang X, Yin W, Li C, Huang Y, Qiu P, Su X, Hu S, Yan G: **A catalog for transcripts in the venom gland of the *Agkistrodon acutus*: Identification of the toxins potentially involved in coagulopathy.** *Biochem Biophys Res Commun* 2006, **341**:522-531.
67. Naamati G, Askenazi M, Linial M: **ClanTox: a classifier of short animal toxins.** *Nucleic Acid Res* 2009, **37**:W363-W368.
68. He Q, He Q, Deng X, Yao L, Meng E, Liu Z, Liang S: **ATDB: a uni-database platform for animal toxins.** *Nucleic Acids Res* 2008, **36**:D293-297.
69. Rawlings ND, Barrett AJ, Bateman A: **MEROPS: the peptidase database.** *Nucleic Acids Res* 2010, **38**:D227-233.
70. Fox JW, Ma L, Nelson K, Sherman NE, Serrano SMT: **Comparison of indirect and direct approaches using ion-trap and Fourier transform ion cyclotron resonance mass spectrometry for exploring viperid venom proteomes.** *Toxicon* 2006, **47**:700-714.
71. Georgieva D, Öhler M, Seifert J, von Bergen M, Arni RK, Genov N, Betzel C: **Snake venomic of *Crotalus durissus terrificus*. Correlation with pharmacological activities.** *J Proteome Res* 2010, **9**:2302-2316.
72. Calvete JJ, Fasoli E, Sanz L, Boschetti E, Righetti PG: **Exploring the venom proteome of the western diamondback rattlesnake, *Crotalus atrox*, via snake venomics and combinatorial peptide ligand library approaches.** *J Proteome Res* 2009, **8**:3055-3067.
73. Wermelinger LS, Dutra DLS, Oliveira-Carvalho AL, Soares MR, Bloch JrC, Zingali RB: **Fast analysis of low molecular mass compounds present in snake venom: identification of ten new pyroglutamate-containing peptides.** *Rapid Commun Mass Spectrom* 2005, **19**:1703-1708.
74. Menin L, Perchuć A, Favreau P, Perret F, Michalet S, Schöni R, Wilmer M, Stöcklin R: **High throughput screening of bradykinin-potentiating peptides in *Bothrops moojeni* snake venom using precursor ion mass spectrometry.** *Toxicon* 2008, **51**:1288-1302.
75. Munekiyo SM, Mackessy SP: **Presence of peptide inhibitors in rattlesnake venoms and their effects on endogenous metalloproteases.** *Toxicon* 2005, **45**:255-263.
76. Wagstaff SC, Favreau P, Cheneval O, Laing GD, Wilkinson MC, Miller RL, Stöckin R, Harrison RA: **Molecular characterisation of endogenous snake venom metalloproteinase inhibitors.** *Biochem Biophys Res Commun* 2008, **365**:650-656.
77. Huang KF, Chiou SH, Ko TP, Wang AH-J: **Determinants of the inhibition of a Taiwan habu venom metalloprotease by its endogenous inhibitors revealed by X-ray crystallography and synthetic inhibitor analogues.** *Eur J Biochem* 2002, **269**:3047-3056.
78. Castoe TA, Poole AW, Gu W, Koning J, Daza JM, Smith EN, Pollock DD: **Rapid identification of thousands of copperhead snake (*Agkistrodon contortrix*) microsatellite loci from modest amounts of 454 shotgun genome sequence.** *Mol Ecol Res* 2010, **10**:341-347.
79. Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDA, King GF, Nevalainen TJ, Norman JA, Lewis RJ, Norton RS, Renjifo C, Rodriguez de la Vega RC: **The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms.** *Annu Rev Genom Human Genet* 2009, **10**:483-511.
80. Nei M, Gu X, Sitnikova T: **Evolution by the birth-and-death process in multigene families of the vertebrate immune system.** *Proc Natl Acad Sci USA* 1997, **94**:7799-7806.
81. Fry BG: **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* 2005, **15**:403-420.
82. Kashima S, Roberto PG, Soares AM, Astolfi-Filho S, Pereira JO, Giuliati S, Faria JrM, Xavier MAS, Fontes MRM, Giglio JR, França SC: **Analysis of *Bothrops jararacussu* venomous gland transcriptome focusing on structural and functional aspects: gene expression profile of highly expressed phospholipases A_2.** *Biochimie* 2004, **86**:211-239.
83. 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.
84. Neiva M, Arraes FB, de Souza JV, Rádis-Baptista G, Prieto da Silva AR, Walter ME, Brigido MM, Yamane T, López-Lozano JL, Astolfi-Filho S: **Transcriptome analysis of the Amazonian viper *Bothrops atrox* venom gland using expressed sequence tags (ESTs).** *Toxicon* 2009, **53**:427-4.
85. Doley R, Kini RM: **Protein complexes in snake venom.** *Cell Mol Life Sci* 2009, **66**:2851-2871.
86. Juárez P, Comas I, González-Candelas F, Calvete JJ: **Evolution of snake venom disintegrins by positive Darwinian selection.** *Mol Biol Evol* 2008, **25**:2391-2407.
87. Lu Q, Navdaev A, Clemetson JM, Clemetson KJ: **Snake venom C-type lectins interacting with platelet receptors. Structure-function relationships and effects on haemostasis.** *Toxicon* 2005, **45**:1089-1098.
88. Walker JR, Nagar B, Young NM, Hirama Y, Rini JM: **X-ray crystal structure of a galactose-specific C-type lectin possessing a novel decameric quaternary structure.** *Biochemistry* 2004, **43**:3783-3792.
89. Freedman RB, Hirst TR, Tuite MF: **Protein disulphide isomerase: building bridges in protein folding.** *Trends Biochem Sci* 1994, **19**:331-336.
90. Yao Y, Zhou Y, Wang C: **Both the isomerase and chaperone activities of protein disulfide isomerase are required for the reactivation of reduced and denatured acidic phospholipase A_2.** *EMBO J* 1997, **16**:651-658.
91. Fleischmann J, Liu H, Wu CP: **Polyadenylation of ribosomal RNA by *Candida albicans* also involves the small subunit.** *BMC Mol Biol* 2004, **5**:17.
92. Kuai L, Fang F, Butler JS, Sherman F: **Polyadenylation of rRNA in *Saccharomyces cerevisiae*.** *Proc Natl Acad Sci USA* 2004, **101**:8581-8586.
93. Decuypere S, Vandesompele J, Yardley V, De Donckeri S, Laurent T, Rijal S, Llanos-Cuentas A, Chappuis F, Arevalo J, Dujardin JC: **Differential polyadenylation of ribosomal RNA during post-transcriptional processing in Leishmania.** *Parasitology* 2005, **131**:121-329.
94. Slomovic S, Laufer D, Geiger D, Schuster G: **Polyadenylation and degradation of human mitochondrial RNA: the prokaryotic past leaves its mark.** *Mol Cell Biol* 2005, **25**:6427-6435.
95. 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.
96. Arava Y, Wang Y, Storey JD, Liu CL, Brown PO, Herschlag D: **Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*.** *Proc Natl Acad Sci USA* 2003, **100**:3889-3894.
97. Preiss T, Baron-Benhamou J, Ansorge W, Hentze MW: **Homodirectional changes in transcriptome composition and mRNA translation induced by rapamycin and heat shock.** *Nat Struct Mol Biol* 2003, **10**:1039-1047.
98. MacKay VL, Li X, Flory MR, Turcott E, Law GL, Serikawa KA, Xu XL, Lee H, Goodlett DR, Aebersold R, Zhao LP, Morris DR: **Gene expression in yeast responding to mating pheromone: Analysis by high-resolution translation state analysis and quantitative proteomics.** *Mol Cell Proteomics* 2004, **3**:478-489.
99. Law GL, Bickel KS, MacKay VL, Morris DR: **The undertranslated transcriptome reveals widespread translational silencing by alternative 5' transcript leaders.** *Genome Biol* 2005, **6**:R111.
100. Fernández J, Lomonte B, Sanz L, Angulo Y, Gutiérrez JM, Calvete JJ: **Snake venomics of *Bothriechis nigroviridis* reveals extreme variability among palm pitviper venoms: different evolutionary solutions for the same trophic purpose.** *J Proteome Res* 2010, **9**:4234-4241.
101. Castoe TA, Daza JM, Smith EN, Sasa MM, Kuch U, Campbell JA, Chippindale PT, Parkinson CL: **Comparative phylogeography of pitvipers suggests a consensus of ancient Middle American highland biogeography.** *J Biogeogr* 2009, **36**:88-103.
102. Crother BI, Campbell JA, Hillis DM: **Phylogeny and historical biogeography of the palm pitvipers, genus Bothriechis: biochemical and morphological evidence.** In *Biology of the Pitvipers.* Edited by: Campbell JA, Brodie EDJr. Selva: Tyler, TX; 1992:1-19.
103. Kuhner MK, Felsenstein J: **A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates.** *Mol Biol Evol* 1994, **11**:459-468.

104. Deng HW, Fu YX: **The effects of variable mutation rates across sites on the phylogenetic estimation of effective population size or mutation rate of DNA sequences.** *Genetics* 1996, **144**:1271-1281.
105. Yang Z: **Among-site rate variation and its impact on phylogenetic analyses.** *Tree* 1996, **11**:367-372.
106. Castoe TA, Sasa MM, Parkinson CL: **Modeling nucleotide evolution at the mesoscale: the phylogeny of the Neotropical pitvipers of the *Porthidium* group (Viperidae: Crotalinae).** *Mol Phylogenet Evol* 2005, **37**:881-898.
107. Castoe TA, Parkinson CL: **Bayesian mixed models and the phylogeny of pitvipers (Viperidae: Serpentes).** *Mol Phylogenet Evol* 2006, **39**:91-110.
108. Castoe TA, Chippindale PT, Campbell JA, Ammermann LK, Parkinson CL: **Molecular systematics of the Middle American jumping pitvipers (genus Atropoides) and phylogeography of the *Atropoides nummifer* complex.** *Herpetologica* 2003, **59**:420-431.
109. Paine MJ, Desmond HP, Theakston RDG, Crampton JM: **Gene expression in *Echis carinatus* (carpet viper) venom glands following milking.** *Toxicon* 1992, **30**:379-386.
110. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437**:376-380.
111. Smit AFA, Hubley R, Green P: **RepeatMasker (release of January 2009).** [http://repeatmasker.org].
112. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
113. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M: **Blast2GO: a universal tool for annotation visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**:3674-3676.
114. The Gene Ontology Consortium: **Gene ontology: tool for the unification of biology.** *Nat Genet* 2000, **25**:25-29.
115. Lomonte B, Escolano J, Fernández J, Sanz L, Angulo Y, Gutiérrez JM, Calvete JJ: **Snake venomics and antivenomics of the arboreal neotropical pitvipers *Bothriechis lateralis* and *Bothriechis schlegelii*.** *J Proteome Res* 2008, **7**:2445-2457.
116. Angulo Y, Escolano J, Lomonte B, Gutiérrez JM, Sanz L, Calvete JJ: **Snake venomics of Central American pitvipers: clues for rationalizing the distinct envenomation profiles of *Atropoides nummifer* and *Atropoides picadoi*.** *J Proteome Res* 2008, **7**:708-719.
117. Alape-Girón A, Sanz L, Escolano J, Flores-Díaz M, Madrigal M, Sasa M, Calvete JJ: **Snake venomics of the lancehead pitviper *Bothrops asper*: geographic individual and ontogenetic variations.** *J Proteome Res* 2008, **7**:3556-3571.
118. Calvete JJ, Sanz L, Cid P, de la Torre P, Flores-Díaz M, dos Santos MC, Borges A, Bremo A, Angulo Y, Lomonte B, Alape-Girón A, Gutiérrez JM: **Snake venomics of the Central American rattlesnake *Crotalus simus* and the South American *Crotalus durissus* complex points to neurotoxicity as an adaptive paedomorphic trend along *Crotalus* dispersal in South America.** *J Proteome Res* 2010, **9**:528-544.
119. Papadopoulos JS, Agarwala R: **COBALT: constraint-based alignment tool for multiple protein sequences.** *Bioinformatics* 2007, **23**:1073-1079.
120. Edgar RC: **MUSCLE: a multiple sequence alignment method with reduced time and space complexity.** *BMC Bioinformatics* 2004, **5**:113.

doi:10.1186/1471-2164-12-259
Cite this article as: Durban *et al.*: Profiling the venom gland transcriptomes of Costa Rican snakes by 454 pyrosequencing. *BMC Genomics* 2011 **12**:259.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

Capítol 2

Integrated “omics” profiling indicates that miRNAs are modulators of the ontogenetic venom composition shift in the Central American rattlesnake, *Crotalus simus simus*

Durban *et al. BMC Genomics* 2013, **14**:234
http://www.biomedcentral.com/1471-2164/14/234

BMC Genomics

RESEARCH ARTICLE Open Access

Integrated "omics" profiling indicates that miRNAs are modulators of the ontogenetic venom composition shift in the Central American rattlesnake, *Crotalus simus simus*

Jordi Durban[1], Alicia Pérez[1], Libia Sanz[1], Aarón Gómez[2], Fabián Bonilla[2], Santos Rodríguez[2], Danilo Chacón[2], Mahmood Sasa[2], Yamileth Angulo[2], José M Gutiérrez[2] and Juan J Calvete[1*]

Abstract

Background: Understanding the processes that drive the evolution of snake venom is a topic of great research interest in molecular and evolutionary toxinology. Recent studies suggest that ontogenetic changes in venom composition are genetically controlled rather than environmentally induced. However, the molecular mechanisms underlying these changes remain elusive. Here we have explored the basis and level of regulation of the ontogenetic shift in the venom composition of the Central American rattlesnake, *Crotalus s. simus* using a combined proteomics and transcriptomics approach.

Results: Proteomic analysis showed that the ontogenetic shift in the venom composition of *C. s. simus* is essentially characterized by a gradual reduction in the expression of serine proteinases and PLA_2 molecules, particularly crotoxin, a β-neurotoxic heterodimeric PLA_2, concominantly with an increment of PI and PIII metalloproteinases at age 9–18 months. Comparison of the transcriptional activity of the venom glands of neonate and adult *C. s. simus* specimens indicated that their transcriptomes exhibit indistinguisable toxin family profiles, suggesting that the elusive mechanism by which shared transcriptomes generate divergent venom phenotypes may operate post-transcriptionally. Specifically, miRNAs with frequency count of 1000 or greater exhibited an uneven distribution between the newborn and adult datasets. Of note, 590 copies of a miRNA targeting crotoxin B-subunit was exclusively found in the transcriptome of the adult snake, whereas 1185 copies of a miRNA complementary to a PIII-SVMP mRNA was uniquely present in the newborn dataset. These results support the view that age-dependent changes in the concentration of miRNA modulating the transition from a crotoxin-rich to a SVMP-rich venom from birth through adulthood can potentially explain what is observed in the proteomic analysis of the ontogenetic changes in the venom composition of *C. s. simus*.
(Continued on next page)

* Correspondence: jcalvete@ibv.csic.es
[1]Instituto de Biomedicina de Valencia, Consejo Superior de Investigaciones Científicas, Jaime Roig 11, Valencia 46010, Spain
Full list of author information is available at the end of the article

© 2013 Durban et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

(Continued from previous page)

Conclusions: Existing snake venom toxins are the result of early recruitment events in the Toxicofera clade of reptiles by which ordinary genes were duplicated, and the new genes selectively expressed in the venom gland and amplified to multigene families with extensive neofunctionalization throughout the approximately 112–125 million years of ophidian evolution. Our findings support the view that understanding the phenotypic diversity of snake venoms requires a deep knowledge of the mechanisms regulating the transcriptional and translational activity of the venom gland. Our results suggest a functional role for miRNAs. The impact of specific miRNAs in the modulation of venom composition, and the integration of the mechanisms responsible for the generation of these miRNAs in the evolutionary landscape of the snake's venom gland, are further challenges for future research.

Keywords: Ontogenetic venom shift, Venomics, Snake venom gland transcriptomics, 454 pyrosequencing, Ion-Torrent microRNA profiling, *Crotalus simus simus*

Background

The presence of a venom system is a shared derived character of the advanced snakes [1-3]. Venom represents an adaptive trophic trait in snake evolution [4-7] that has played a central role in the origin of the advanced snakes during the Cenozoic era [8,9]. The diversity of modern snakes appeared during the Paleocene period of the Cenozoic Era, approximately 54–64 Mya, following the split of the Pareatidae from the remaining Caenophidians [6]. In Occidental culture, Francesco Redi (1626–1697), court physician to Ferdinando II de' Medici, Grand Duke of Tuscany and his successor, Cosimo III, is credited for discovering how vipers produce venom and inject it into their prey [10]. One century latter, Abbé Gasparo Ferdinando Felice Fontana (1730–1805), first director of the Museum of Physics and Natural History in Florence, performed experiments on the venom of the European viper. His classic text published in 1781 [11] is regarded the Opera Prima of modern toxinology [12].

Research on venoms has been continuously advanced by technological developments [13]. Particularly for the past decade, and fueled by the application of "omic" technologies [14-18], the field of molecular toxinology has experienced a sustained exponential growth. In-depth venom proteomics and transcriptomic analyses have recently become available for a number of venomous snake lineages, revealing a vast unexplored source of evolutionary innovation [2,19,20]. Snake venoms are well documented as having different venom compositions and toxicity based on taxonomic or geographical locations [21]. Inter- and intraspecific individual, gender-specific, regional, and seasonal variations in venom toxin composition may reflect local adaptations conferring fitness advantages to individuals specializing on different prey, or phylogenetic carry on. Understanding the processes that drive the evolution of snake venom variability is a topic of intense research interest in molecular and evolutionary toxinology. Recent studies of the molecular basis of adaptations have sought to understand the relative importance of gene regulation effects as determinants of venom phenotype [6,22-24]. Most significantly, a number of snakes show age-related (ontogenetic) changes in venom composition [25-32]. Surprisingly, despite previous suggestions that ontogenetic changes in venom are prey-related [33], juvenile Dusky Pigmy rattlesnakes, *Sistrurus miliarius barbouri*, raised on different diets showed similar albeit highly-variable venom compositions by the end of a 26-months study, suggesting little effect of diet on the overall make-up of venom in snakes this age or younger [23]. Over the same period shifts in venom composition occurred in females raised on different prey in all diet treatments with the magnitude of those changes strongly related to diet. This work provided evidence that venom composition is somewhat plastic in both juvenile and adult *S. m. barbouri* and that, at least in adults, prey consumed may influence the relative abundance of possibly functionally-distinct classes of venom proteins [23]. However, the molecular mechanisms that underly age-related changes in venom remain elusive.

In this work we have explored the basis and level of regulation of the ontogenetic shift in the venom composition of the Central American rattlesnake, *Crotalus s. simus*. Biogeographical data indicate a basal cladogenesis in the Central American *C. simus* clade, dating back to the late Miocene/early Pliocene (6.4-6.7 Mya) [34]. Neonate and juvenile *C. s. simus* venoms contain large relative amounts of crotoxin, a heterodimeric PLA_2 molecule exhibiting presynaptic β-neurotoxicity, along with low concentration of hemorrhagic snake venom metalloproteinases (SVMPs) [29]. By contrast, adult *C. s. simus* venom presents a high content of SVMP and is largely devoid of neurotoxic activity [29,35,36]. Juvenile and adult *C. s. simus* venom phenotypes broadly correspond, respectively, to type II (low metalloproteinase activity and high toxicity, LD_{50} <1 μg/g mouse body weight) and type I (high levels of SVMPs and low toxicity, LD_{50} >1 μg/g mouse body weight) venoms defined by Mackessy [37,38]. Here we investigate the transition from type II to type I venom phenotype in *C. s. simus* through an integrated "omics" approach involving proteomic survey of

time-course venom composition variation, from birth through adulthood, and Next Generation sequencing of the venom gland transcriptomes and microRNAs of neonate and adult specimens. With this experimental design we intended to determine whether venom gene expression is transcriptionally controlled by developmental stage-dependent factors, or whether the ontogenetic changes in venom composition involve regulation at the post-transcriptional level.

Results and discussion

Time-resolved proteomic analysis of the ontogenetic changes in the venom composition of *C. s. simus*

A previous proteomic survey of the venom of the Central American rattlesnake, *C. s. simus*, laid the foundation for understanding the changes in toxin composition and overall pharmacological features between adult (predominantly hemorrhagic) and newborn (mainly neurotoxic) snakes [29]. This study revealed prominent stage-dependent protein expression changes between the pooled venoms from newborn and adult specimens from Costa Rica, characterized by a shift from a type II to a type I venom phenotype [29]. Such a conspicuous venom composition change has been hypothesized to be related to variations in the size of prey consumed by snakes of different ages and the variable requirements for immobilizing and digesting them [37,38]. It has also been suggested that the significance of venom adaptation to specific diets represents a trade-off between the metabolic cost of venom production and increasing foraging efficiency [7,39]. In the present work we have sought to establish the molecular mechanism responsible for the reported ontogenetic shift in venom protein composition.

Figure 1 presents reverse-phase HPLC profiles showing changes in the composition of venom pooled from specimens from age 8-weeks to adulthood (≥ 36 months). The ontogenetic shift is essentially characterized by a gradual reduction in the expression of serine proteinases and PLA_2 molecules, particularly crotoxin, a β-neurotoxic heterodimeric PLA_2 [40-42], concomitantly with an increased secretion of PI and PIII SVMPs at age 9–18 months. Of particular note, whereas venoms from individual 9-month-old *C. s. simus* specimens showed indistinguishable reverse-phase chromatographic profiles, venoms from 18-month-old snakes exhibited large individual variation in their crotoxin and SVMP contents (Figure 2), suggesting a key point in shifts in venom composition at this developmental stage. Consistent with this view, juvenile specimens around this age exhibit a range of venom phenotypes, between predominantly type I, type II, and combinations of the two. Systemic effects involving hemostatic disturbances, i.e. coagulopathy, have not been documented in Central American rattlesnake envenomings [35,43]. However, there is little information on the clinical observations of envenomings induced by specimens of *C. simus* of various ages in Central America. Envenomings by adults are characterized by local tissue damage, i.e. edema and hemorrhage and systemic manifestations associated with cardiovascular effects and coagulopathy [43]. Our proteomic data suggest that bites by juvenile specimens might be characterized by a combination of SVMP-induced hemorrhage and crotoxin-induced neurotoxicity, in addition to serine proteinase-induce coagulopathy. Thus variable clinical manifestations might occur in accidents by *C. simus* of different ages. A similar trend regarding geographical differences in venom composition and toxicity has been described in the literature for the North American Mojave rattlesnake, *C. s. scutulatus*. Venom populations of *C. s. scutulatus* exhibit an intergradation pattern from SVMP-rich (type B) to Mojave toxin-rich (type A) phenotypes, from south central to southeastern Arizona [44]. Type A venom has large amounts of Mojave (crotoxin-like) toxin and shows neurotoxic effects. Type B venom has large amounts of SVMPs and shows hemorrhagic effects, and type A+B venom is a combination of the two and induces both neurotoxic and hemorrhagic effects [45,46]. Geographic venom variation throughout the *C. s. scutulatus* range correlated with clinical severity outcomes [44]. Hence, besides ecological and taxonomical implications, knowledge of the natural history and toxin composition of venoms is of fundamental importance for the treatment of bite victims and for the selection of specimens for the preparation of venom pools for antivenom production [47,48].

Analysis of the venom gland transcriptomes of neonate and adult *C. s. simus*. Retroelements

To investigate the basis of the ontogenetic venom shift revealed by the time-resolved proteomic analysis, we compared the transcriptional activity of the venom glands of a neonate and an adult *C. s. simus* specimens, using 454 pyrosequencing and the bioinformatic processing strategy outlined in Durban *et al.* [15]. 408,505 and 349,170 raw 454 reads from adult and newborn venom gland transcriptomes, respectively, were quality trimmed using the Prinseq software and only reads having a Quality Value (QV) greater than 20 [49] (355,140 (adult) and 320,907 (newborn) Table 1) were considered for assembling with Newbler 2.6 software. The analysis yielded 33,408 (adult) and 24,136 (newborn) singletons, and among the resulting 6,484 (adult) and 6,047 (newborn) total contigs, 1.43% and 10.74% comprised only 2 reads. Table 1 displays a summary of the 454 sequencing statistics.

Contig sequences were inspected for repetitive elements using Repeat Masker. 97,204 bases (1.93% of adult *C. s. simus* venom gland transcriptome) and 65,742 nucleotides (1.94% of the newborn venom gland transcriptome) were masked with N characters, a large part of them

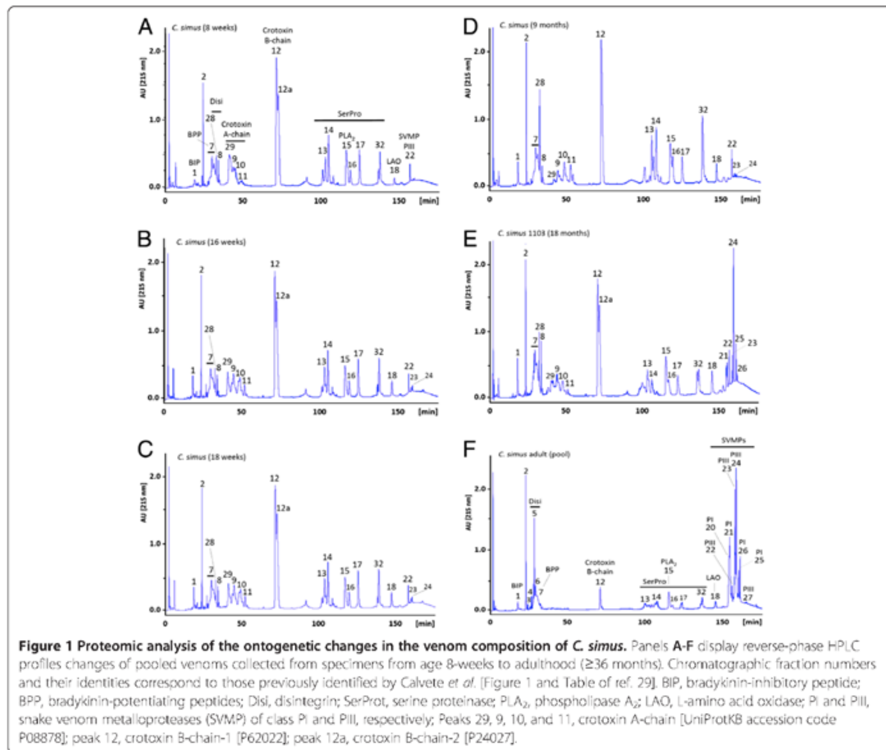


Figure 1 Proteomic analysis of the ontogenetic changes in the venom composition of *C. simus*. Panels **A-F** display reverse-phase HPLC profiles changes of pooled venoms collected from specimens from age 8-weeks to adulthood (≥36 months). Chromatographic fraction numbers and their identities correspond to those previously identified by Calvete *et al.* [Figure 1 and Table of ref. 29]. BIP, bradykinin-inhibitory peptide; BPP, bradykinin-potentiating peptides; Disi, disintegrin; SerProt, serine proteinase; PLA_2, phospholipase A_2; LAO, L-amino acid oxidase; PI and PIII, snake venom metalloproteases (SVMP) of class PI and PIII, respectively; Peaks 29, 9, 10, and 11, crotoxin A-chain [UniProtKB accession code P08878]; peak 12, crotoxin B-chain-1 [P62022]; peak 12a, crotoxin B-chain-2 [P24027].

comprising Short and Long INterspersed repetitive Elements (SINEs and LINEs) retroelements (Additional file 1: Tables S1 and S2). Transposable elements (TE) that propagate within the host genome via RNA intermediates occupy a large fraction of eukaryotic genomes. Their mobility and amplification represent a major source of genomic variation [50,51]. Retrotransposable elements have been reported in the transcriptomes of *Bothrops insularis* (4.1% of ESTs) [52], *Lachesis muta* (0.3%) [53], and *Philodryas olfersii* (4.1%) [54], in PLA_2 genes from the venom gland of *Vipera ammodytes* [55,56] and *Protobothrops flavoviridis* [57,58], and in an *E. ocellatus* PIII-SVMP gene [59]. Although their functional relevance in the venom gland remains unknown, transposable elements appear to be overrepresented in UTRs of mRNAs of rapidly evolving genes [60], suggesting that they have played a role in the diversification and expansion of these gene families [61,62].

Sauria SINE have been characterized in all major lineages of squamate reptiles [62,63], and phylogenetic analysis of *E. ocellatus* Sauria SINEs [57] indicated that their origin correlates with the time frame of the evolution of the snake venom system. Sauria SINEs may have arisen by a fusion of a tRNA-related sequence with the 3' end of a LINE [64] more than 200 million years ago and are uniquely widespread in lepidosaurian genomes [62]. SINEs have no protein coding capacity, and their retrotransposition depends on a "target-primed reverse transcription" by autonomous partner LINEs, that encode an endonuclease for cleaving the genomic integration site and a reverse transcriptase to copy the RNA to DNA [65,66]. Since Sauria SINEs share an identical 3' sequence with Bov-B LINEs, it has been proposed [63] that they utilize *in trans* the enzymatic machinery of Bov-B LINEs for their mobility and dispersal throughout the genome. In squamates

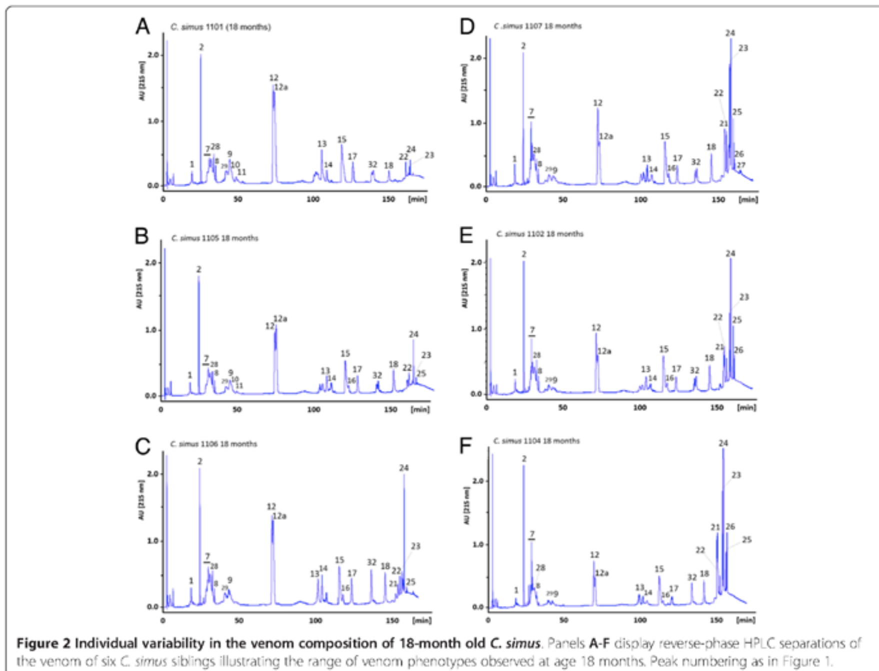


Figure 2 Individual variability in the venom composition of 18-month old *C. simus*. Panels **A-F** display reverse-phase HPLC separations of the venom of six *C. simus* siblings illustrating the range of venom phenotypes observed at age 18 months. Peak numbering as in Figure 1.

and turtles, CR1 and L2 LINES are also partners of diverse SINEs [63,67].

An overview of the landscape of retrotransposable elements reported in the sauropsida taxon, a sister group of mammals that includes all extant reptiles and birds, has recently emerged from analysis of the draft genomes of the red jungle fowl, *Gallus gallus* [68], and the green anole, *Anolis carolinensis* [69]. Whereas a single type of LINE, CR1, comprises over 80% of all interspersed repeats in the chicken genome (200,000 copies; 9% of the chicken genome), approximately 30% of the *A. carolinensis* genome is composed of a wide diversity of LINE and SINE mobile element families. Since SINEs are among the largest multigene families in reptilian genomes, they may act as sites for homologous recombination events between dispersed SINEs, resulting in a variety of genetic rearrangements, including duplication, deletion and translocation, that likely represent mechanisms that generates genetic diversity in reptilian genomes [70]. The prevalence of transposable elements in untranslated regions of mRNAs of recently expanded gene classes suggested that TEs could affect gene expression through the donation of transcriptional regulatory signals [60]. The indistinguishable distribution of TEs in the venom gland transcriptomes of neonate and adult *C. s. simus* (Additional file 1: Tables S1 and S2) argues against this type of transcriptional regulation to explain the ontogenetic shift in venom composition.

Table 1 454 sequencing statistics

	Total reads	Mean length read (nt)	Max. length read (nt)	Assembled reads (%)	Number of contigs	Mean length contig (nt)	Max. length contig (nt)	Singletons
Newborn	320,907	567.5	1,193	276,729 (86.23)	6,047	560.87	5,857	24,136
Adult	355,140	535.4	1,320	296,470 (83.48)	6,484	777.88	5,947	33,408

Raw reads, Roche Signal Processing Application reads, and Newbler 2.6-assembled contigs were inspected for poly-N's, the presence of library construction tags, and their Quality Values (QV). Only reads showing a QV > 20 were assembled.

Durban *et al. BMC Genomics* 2013, **14**:234
http://www.biomedcentral.com/1471-2164/14/234

Comparison of toxin-coding transcript distribution in newborn and adult *C. s. simus* transcriptomes provides clues for streamlining their divergent venom phenotypes

The sets of 6,484 (adult) and 6,047 (newborn) masked contigs were searched against the NCBI nucleotide sequence database using the BLASTN algorithm to identify similar sequences. 4,141 hits representing 63.9% of the total contigs of the adult snake transcriptome were retrieved, 431 of which (10.4% of matched hits) displayed similarity (e-value cutoff $<10^{-3}$) to documented venom proteins of the taxonomic suborder *Serpentes*. In addition, 21,460 singleton sequences (64.23% of the total adult venom gland singletons, Table 1) produced significant BLASTN hits, and of these only 431 (2%) corresponded to documented snake venom entries.

On the other hand, 3,022 (49.9% of the newborn venom gland transcriptome) found a BLAST hit, including 658 (10.9%) matches to snake venom proteins. Also, 526 sequences out of the 15,052 singleton BLAST hits (3.5%) corresponded to snake venom toxins. Additional file 2: Table S3 lists the relative abundances of the different venom protein family hits in the non-normalized venom gland transcriptomes of newborn and adult *C. s. simus*. The venom protein families identified in the newborn and adult venom gland transcriptomes, and the relative abundances of the length-normalized ORFs, simulated with NoiSeq for five technical replicates (nss=5), are displayed, are listed in Table 2. An estimation of the minimum number of genes from each toxin family transcribed into the venom gland transcriptome was calculated from the multiple alignments of reads matched to a full-length reference sequence [15] (Table 3). These newborn and adult synthetic gene sequences were clustered with CD-HIT into shared (identity threshold > 0.8) and unique clusters (Table 3). Figure 3 shows the distribution of clusters from the major toxin families between newborn and adult transcriptomes. These results indicated that both newborn and adult *C. s. simus* venom glands expressed non-overlapping gene sets. In particular, C-type lectin-like (CTL), (nerve growth factor) NGF, phospholipase A_2 (PLA_2), and snake venom metalloproteinase (SVMP) toxin families exhibited high NOISeq probabilities (*prob* value = 0.98, 0.93, 0.96, and 0.91, respectively) of being differentially expressed between the adult and the newborn transcriptomes. CTL, NGF, and SVMP toxin families were down-regulated in newborn *versus* adult database (-2.01, -1.69, and 0-77, respectively), whereas the PLA_2 family appeared to be up-regulated (1.49).

Figure 4 displays chart pies comparing the relative protein family compositions computed from the adult and newborn transcriptomes (Table 2) and the venom proteomes [15]. The amino acid sequences of transcript-deduced amino acid sequences of PLA_2 molecules, serine proteinases, and SVMPs are shown in Additional file 3: Figures S1-S3). In line with previous transcriptomic

Table 2 RPKM (Reads per Kilobase per Million mapped reads)-normalized contigs and singletons from the venom gland transcriptomes of newborn and adult *C. s. simus* aligned to a reference snake venom toxin ORF

	Newborn					Adult				
	Contigs	Reads	Singletons	RPKM	%	Contigs	Reads	Singletons	RPKM	%
5'-NTase	1	395	1	212.20	0.06	2	334	0	532.54	0.08
BPP	2	42	1	74.66	0.02	1	11	0	56.83	0.01
CRISP	3	62	0	81.26	0.02	0	0	0	0	0
CTL	6	483	5	1,035.42	0.31	13	1,462	4	9,255.5	1.5
GC	1	143	2	124.08	0.04	3	348	2	891.07	0.14
HYA	2	346	4	245.52	0.07	4	369	1	772.30	0.12
KUN	2	15	0	18.70	0.01	2	115	3	437.76	0.07
LAO	1	1,300	0	793.68	0.24	1	1,440	1	2,617.77	0.42
NGF	1	384	0	501.22	0.15	1	917	4	3,577.02	0.58
OHA	0	0	0	0	0	0	0	0	0	0
PDE	1	308	0	383.50	0.11	1	77	0	285.28	0.04
PLA_2	27	41,700	37	94,942.21	28.37	18	11,015	9	74,618.46	12.13
SVMP	48	54,601	79	28,306.34	8.46	62	69,198	62	106,685.8	17.34
SP	191	156,291	194	207,689.54	62.07	83	105,104	62	415,323.2	67.51
VEGF	2	132	2	219.77	0.07	4	19	0	92.72	0.01

Abbreviations: *5'-NTase* 5'-nucleotidase, *BPP* bradykinin potentiating peptide, *CRISP* cysteine-rich secretory protein, *CTL* C-type lectin-like protein, *GC* glutaminyl cyclase, *HYA* hyaluronidase, *KUN* Kunitz-type inhibitor, *LAO* L-amino acid oxidase, *NGF* nerve growth factor, *OHA* ohanin, *PDE* phosphodiesterase, *PLA_2* phospholipase A_2, *SVMP* snake venom metalloproteinase, *SP* serine proteinase, *VEGF* vascular endothelial growth factor. Protein families found in *C. s. simus* proteome [15] are underlined.

Table 3 Estimation of minimum number of toxin genes transcribed into the venom gland transcriptomes of newborn (N) and adult (A) *C. s. simus*

Toxin family	Newborn Unique	Adult	N + A Shared
5'-NTase			1
BPP	2	1	
CRISP	3		
CTL	6	2	
GC			1
HYA	1		1
KUN	2	2	
LAO			1
NGF			1
PDE			1
PLA_2	2	1	5
SVMP	7	7	7
SP	15	0	50
VEGF	2	1	

Toxin family acronyms as in Table 2.
Protein families found in the respective venom proteomes [15] are underlined.

surveys, the overall composition of neonate and adult transcriptomes (Figure 4, pie charts "a" and "b") are more similar to each other than their respective proteomes (Figure 4, pie charts "c"), indicating that the venom transcriptome may be subjected to stage-dependent translational control. In particular, newborn and adult venom glands expressed similar amounts of a transcript encoding a protein sequence 100% identical to crotoxin B-chain [P62022], whereas the concentration of this protein markedly differ in their respective venom proteomes (compare peak 12,12b in between panels of Figure 1). Moreover, shared clusters 1, 2, and 4 encode precursor crotoxin A-chain [P08878]-like sequences, although this protein (peaks 29, 9–11 in Figure 1), which is necessary for generating the heterodimeric presynaptic β-neurotoxic PLA_2 molecule [40,41] responsible for the neurotoxicity of newborn and juvenile Central and adult South American rattlesnakes [42,43], is absent from the venom of adult *C. s. simus* venom [29]. Similarly, cluster 5, shared by newborn and adult venom gland transcriptomes (Figure 3B), encodes the PI-SVMP 20–21 exclusively found in venoms of juvenile (18-month) and adult specimens (Figure 1). Newborn and adult transcriptomes also share a number of clusters encoding PIII-SVMPs (0, 1, 2, 4, 6, and 16) although only a PIII-SVMP (peak 22, Figure 1) is present in the venom proteome of snakes aged 0–9 month. Moreover, PIII-SVMPs 8, 11, 14 and 18 and 9, 12, 15, 17 and 19 exhibited exclusive transcription in the newborn and the adult, respectively. On the other hand, although fragmentary, the protein sequences encoded by clusters 2, 12 and 15 match the MS/MS-derived peptide sequence information derived from the PIII-SVMPs 23–24 isolated from venom of adult snakes [29]. The higher diversity of serine proteinase transcripts characterized in the newborn *versus* the adult venom gland also mirrors the proteomic data (Figure 1). Shared cluster 0 (Figure 3C) encodes serine proteinase (SP) 14, an abundant enzyme in newborn and juvenile venom proteomes whose expression is significantly down-regulated in adult snakes (Figure 1). A precise matching of other venom serine proteinases characterized in *C. s. simus* is impeded by the lack of MS/MS data discriminating between the different transcript sequences (Additional file 3: Figure S3). Nevertheless, our results indicate that the age-dependent expression of both SP 14 and the major PIII-SVMPs found in adult *C. s. simus* venom might be due to a transcriptional regulatory mechanism. In addition, although it cannot be excluded that some shared transcripts correspond to pseudogenes, and also taking into consideration that the sampled transcriptomes were from single specimens, another main conclusion from our findings is that the molecular mechanism by which shared transcriptomes generate divergent venom phenotypes may operate post-transcriptionally.

Distinct venom gland miRNA sets in newborn and adult snakes: modulators of the Central American rattlesnake's ontogenetic venom composition variation?

Phenotypic diversity generated through the regulation of RNA transcripts has been proposed as an explanation for the discrepancy between organismal complexity and the relatively limited number of primary coding transcripts [72,73]. Regulation by micro RNAs is one such mechanism [74,75]. MicroRNAs (miRNAs) are a class of small (~ 22 nucleotides in length), single-stranded, non-coding endogenous RNAs that are recently found to be negative post-transcriptional regulators of gene expression in eukaryotic organisms [76]. MicroRNAs act as adaptors that employ a silencing ribonucleoprotein complex to target mRNAs by selective Watson-Crick base-pairing, primarily in the 3'-UTR region. miRNAs anchored into specific binding pockets guide members of the Argonaute (Ago) protein family to target mRNA molecules for silencing or destruction [77]. The evolutionary dynamics of miRNAs across metazoan phylogeny and through deep evolutionary time suggests that metazoan morphological complexity might have its roots in miRNA innovation [78]. To explore the possible involvement of miRNAs in the post-transcriptional regulation of *C. s. simus* venom gland transcriptome, miRNA libraries from neonate and adult specimens were sequenced on an Ion Torrent Personal Genome Machine. Table 4A displays a summary of candidate miRNA

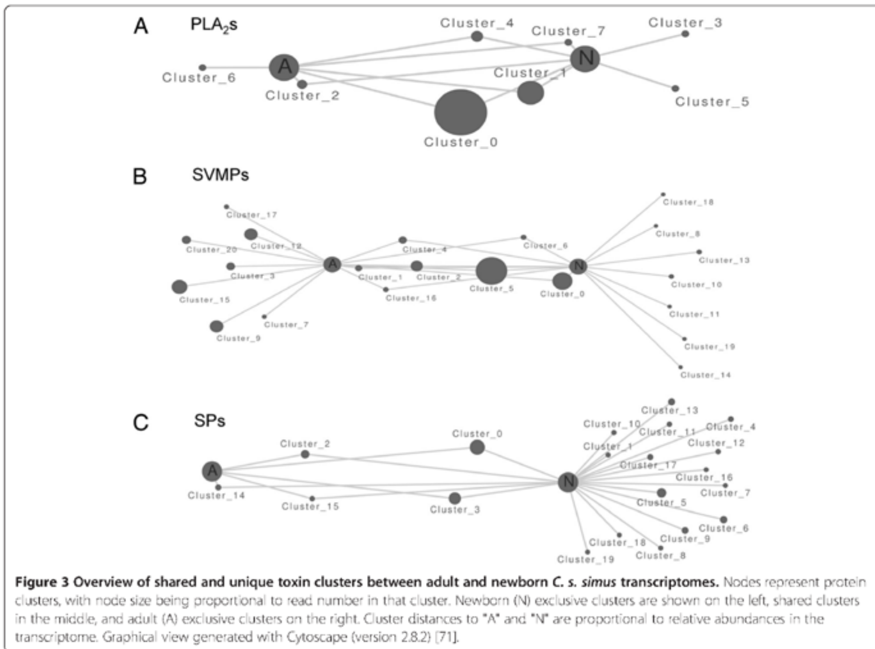


Figure 3 Overview of shared and unique toxin clusters between adult and newborn *C. s. simus* transcriptomes. Nodes represent protein clusters, with node size being proportional to read number in that cluster. Newborn (N) exclusive clusters are shown on the left, shared clusters in the middle, and adult (A) exclusive clusters on the right. Cluster distances to "A" and "N" are proportional to relative abundances in the transcriptome. Graphical view generated with Cytoscape (version 2.8.2) [71].

sequencing statistics. Candidate miRNAs were filtered out according to nucleotide length and the presence of either the 30-mer IonA or the 23-mer P1 adapter sequences. This quality processing step yielded 38,738 (newborn) and 64,493 (adult) clusters, of which, respectively, 238 and 419 comprised at least 100 reads (Table 4B). These newborn and adult datasets contained 132 and 268 unique miRNAs, respectively, and 151 were shared between them (Table 4B). Although to date no snake miRNA has been reported in miRBase (http://www.mirbase.org), which includes 21,643 mature miRNA products from 168 species [79], BLAST analysis of *C. s. simus*

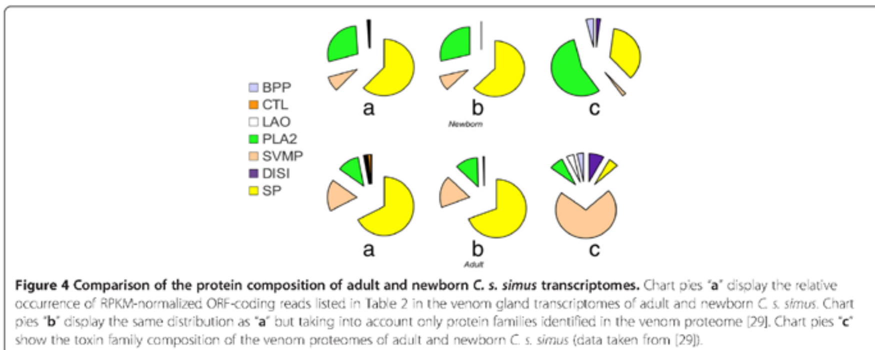


Figure 4 Comparison of the protein composition of adult and newborn *C. s. simus* transcriptomes. Chart pies "**a**" display the relative occurrence of RPKM-normalized ORF-coding reads listed in Table 2 in the venom gland transcriptomes of adult and newborn *C. s. simus*. Chart pies "**b**" display the same distribution as "**a**" but taking into account only protein families identified in the venom proteome [29]. Chart pies "**c**" show the toxin family composition of the venom proteomes of adult and newborn *C. s. simus* (data taken from [29]).

Table 4 A. Summary of candidate miRNA sequencing statistics. B. Clustering and miRBase annotation statistics

A	Total raw reads	Prinseq quality filtered reads	FastX too-short reads	FastX adapter-only reads		Processed reads	Mean length read (nt)
Newborn	314,592	263,517	34,579	8,390		220,548	24.58
Adult	515,040	450,902	55,882	5,956		389,064	23.68
B	**Total clusters**	**≥100 reads**	**Unique clusters**	**Shared clusters**	**BLAST miRBase hits**	**Anolis hits**	**% Anolis hits**
Newborn	38,738	283	132	151	48	22	45.8
Adult	64,493	419	268	151	70	30	42.8

Total clusters were obtained after the CD-HIT usage. Clusters were compared between newborn and adult datasets and blasted against miRBase.

the adult and newborn miRNA datasets against the mature miRBase retrieved 118 hits matching such diverse taxa as mammals (gray short-tailed opossum, *Monodelphis domestica;* platypus, *Ornithorhynchus anatinus;* Tasmanian devil, *Sarcophilus harrisii*; bull, *Bos taurus;* horse, *Equus caballus*; Sumatran orangutan, *Pongo abelii*; wild boar, *Sus scrofa*; mouse, *Mus musculus*,), birds (zebra finch, *Taeniopygia guttata*), fishes (sea lamprey, *Petromyzon marinus*; Japanese ricefish, *Oryzias latipes*; zebra fish, D*anio rerio*), reptile (Anole lizard, *Anolis carolinensis*), and the solitary sea squirt, *Ciona intestinalis* (Ascidian). However, 52 miRNA sequences (44%) corresponded to miRNAs reported in *Anolis carolinensis*, the only available squamate genome [69].

MicroRNAs with frequency count of 1000 or greater exhibited a distinct distribution between the newborn and adult datasets (Figure 5). Noteworthy, the most expressed miRNAs in the newborn venom gland showed a significantly lower abundance in adults and *visa versa*, miRNAs abundantly expressed in adults have a lower expression in neonates (Figure 5, right panels; Tables S4 and S5 in Additional file 2). This uneven distribution of miRNAs suggests a regulatory mechanism by which a single transcriptome may result in different proteomes. Prediction of putative target genes was assessed by miRanda using the parameter specified in the Methods section. MiRanda is a bioinformatic tool for finding genomic targets for microRNAs that incorporates current biological knowledge on target rules and computes optimal sequence complementarity between a set of mature microRNAs and a given mRNA using a weighted dynamic algorithm [80-82]. MiRanda predicted 10 miRNAs complementary of 3'-UTR loci of *C. s. simus* SVMP 454-transcripts (5 shared between newborn and adult; 5 newborn-exclusive) and 3 miRNAs from adult snakes targeting PLA_2 mRNAs (Figure 6; but see also Figure S4 in Additional file 3). When these matches were filtered through MapMi using thermodynamics, sequence conservation, and pairwise alignment criteria [83], positive hits were restricted to the three miRNA sequences mapping to PLA_2 loci (clusters New299/Ad368, New1849/Ad1078, and Ad2166) and a pair of miRNAs against SVMP mRNAs (clusters New4393/Ad3416 and New2578) (Figure 6). miRNA counts in the newborn and adult venom gland transcriptomes, and best BLAST hits of their putative target genes are displayed in Figure 7. In addition, Table S6 (Additional file 2) lists the miRanda-only predicted miRNAs and their best BLAST hit, and Figure S4 (Additional file 3 shows their Watson-Crick pairing to target 3'-UTR loci of PLA2 and SVMP 454 transcripts, and the corresponding binding energy calculated by MapMi. Noteworthy, 590 copies of miRNA 2166, targeting crotoxin B-subunit, was exclusively found in the transcriptome of the adult snake whereas 1185 copies of miRNA 2578, complementary to a PIII-SVMP mRNA, was uniquely present in the newborn dataset (Figure 6).

Animal miRNAs guide proteins to repress the translation of target mRNAs via imperfect base pairing between the miRNA and the target [75]. Although the precise rules for pairing between a miRNA and its mRNA target sites are not known, an obvious requirement for miRNA regulation is the concurrent expression of both a miRNA and its target genes, and requiring conserved Watson-Crick pairing to the 5' region of the miRNA centered on nucleotides 2–7 (the so-called miRNA "seed") markedly reduces the occurrence of false-positive predictions [84,85]. However, a modest role for 3'-supplementary in targeting specificity, and the rare occurrence of 3'-compensatory sites that can compensate for a single-nucleotide bulge or mismatch in the seed region, both centered on miRNA nucleotides 13-16/17, have been reported [86]. In addition, mismatch at position 1 is supported by biochemical and crystallographic studies, indicating that the 5'-most nucleotide of an Argonaute-bound guide RNA is not paired to the target strand [87,88]. Figure 8 and S6 (Additional file 2) display the complementarity between the dataset-exclusive miRNAs and their (miRanda + MapMi)-predicted PLA_2 and SVMP target mRNA loci listed in Figure 7.

Concluding remarks and perspectives

The most important concept that emerges from our results is the possibility that miRNAs play a role in both the ontogenetic shift observed in certain venoms and in the plasticity of venoms from adult snakes underlying adaptations to changing environments. An important

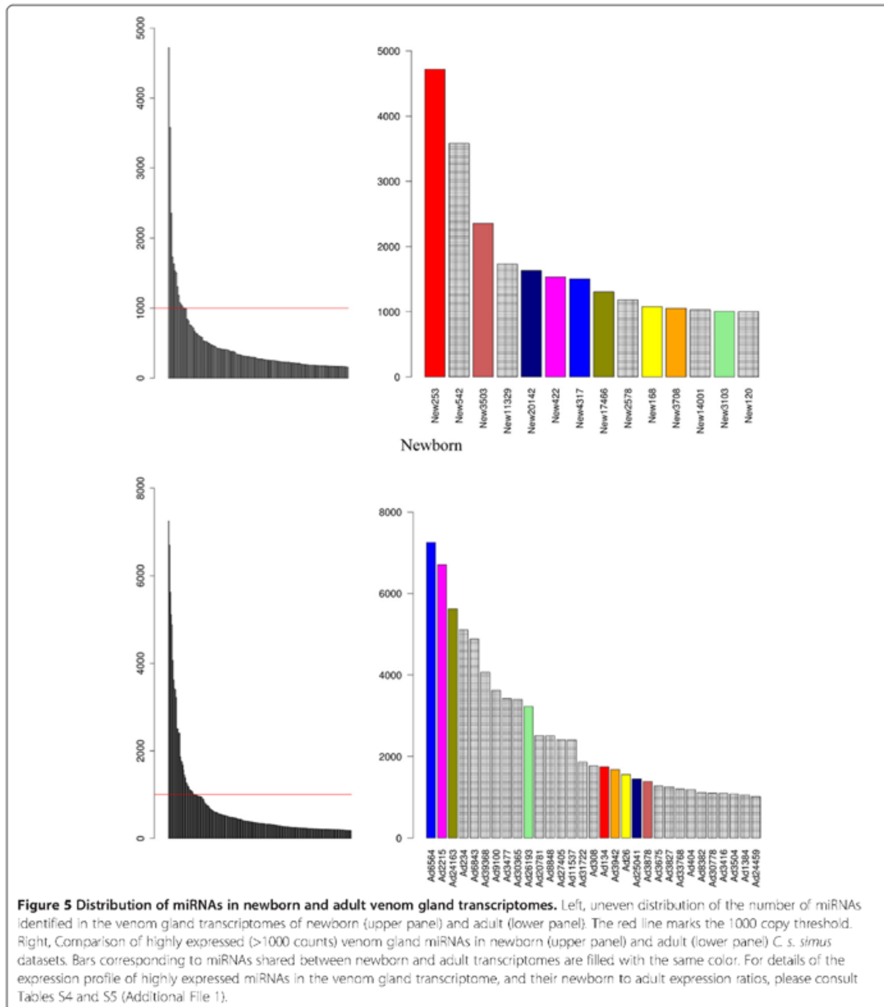


Figure 5 Distribution of miRNAs in newborn and adult venom gland transcriptomes. Left, uneven distribution of the number of miRNAs identified in the venom gland transcriptomes of newborn (upper panel) and adult (lower panel). The red line marks the 1000 copy threshold. Right, Comparison of highly expressed (>1000 counts) venom gland miRNAs in newborn (upper panel) and adult (lower panel) *C. s. simus* datasets. Bars corresponding to miRNAs shared between newborn and adult transcriptomes are filled with the same color. For details of the expression profile of highly expressed miRNAs in the venom gland transcriptome, and their newborn to adult expression ratios, please consult Tables S4 and S5 (Additional File 1).

caveat of our current understanding of miRNA target recognition and post-transcriptional venom gland transcriptome regulation is the absence of mapeable snake genomes. Nonetheless, almost half of the most abundant miRNAs sequenced from the venom gland of both, newborn and adult, *C. s. simus* were orthologous to sequences in *A. carolinensis*. More significant is the fact that a *C. s. simus* newborn venom gland-exclusive miRNA and a venom gland miRNA uniquely found in adult snakes target, respectively, mRNAs encoding a PIII-SVMP and the B-subunit of crotoxin/Mojave toxin. Relevant to this point, Mojave toxin is a neurotoxic heterodimeric PLA_2, whose translation into the venom proteome is ontogenetically regulated [29] (Figure 1).

Durban *et al. BMC Genomics* 2013, **14**:234
http://www.biomedcentral.com/1471-2164/14/234

Cluster ID		miRNA Sequence	Adult	Newborn	Adult	Newborn
			SVMP		PLA$_2$	
New 752		AGCTGAAGAGGAAAACAACACTTTCCTTAGTGACA	X^1	X^1		
New 253		AGCTGAAGAGGAAAACAACACTTTCCTTAGTGACCA	X^1	X^1		
New 3503		GCATTGGTGGTTCAGTGGTAGAATTCTCGCC		X^1		
New 2681		AGCAGCTGACTTAGAACTGGCGCGGA	X^1	X^1		
Ad 368	**New299**	**ATGGCCTCCGTCGCCCTCAGCCGATCGA**			$X^{1,2}$ (181)	$X^{1,2}$ (186)
Ad 1078	**New1849**	**ACGTGCAAATCGGTCGTCCGACCTGGGTA**			$X^{1,2}$ (113)	$X^{1,2}$ (105)
New 20		TGATCCTTCGATGTCGGCTCT		X^1		
New		3708 GCGTTGGTGGTATAGTGGTGAGCATAGC		X^1		
New 87		GAGGTAAACGGGTGGGGTCCGCGCA	X^1	X^1		
New 4393	**Ad3416**	**GCGTTGGTGGTATAGTGGTTAGCATAGC**	$X^{1,2}$ (1092)	$X^{1,2}$ (496)		
New 166		AAGAGGTAAACGGGTGGGGTCCGCGC	X^1	X^1		
New 2578		**GGATGGGAGACCGCCTGGGAATACCGGGTGCTGT**		$X^{1,2}$ (1185)		
Ad 2166		**CGAGGCGTCCAGTGCGGTAACGCGAC**			$X^{1,2}$ (590)	

Figure 6 miRNA sequences predicted by miRanda (1) and MapMi (2) to target 3'-UTR loci of SVMP (blue) and PLA2 (red) transcripts. miRNA counts in their respective transcriptome are indicated in parentheses. Best BLAST hits of the clusters mapped by the miRanda- and MapMi-predicted miRNAs are listed in Figure 7. Best BLAST hits for miRanda-only-predicted miRNA sequences are displayed in Additional file 2: Table S6.

The noncovalent association of two dissimilar subunits, the small acidic, nonenzymatic, and nontoxic A-subunit with the basic and weakly toxic PLA$_2$ B-subunit increases the lethal potency of the uncomplexed crotoxin B-subunit by enabling the toxin to reach its specific site of action at the neuromuscular junction [40,41,89-91]. Hence, miRNAs targeting a crotoxin subunit messenger would serve the goal of eliminating the generation of the neurotoxic heterodimer. Age-dependent changes in the concentration of miRNA modulating the transition from a crotoxin-rich to a SVMP-rich venom from birth through adulthood can potentially explain what is observed in the age-dependent proteomic analysis of the ontogenetic changes in the venom composition of *C. s. simus* illustrated in Figure 1.

Large-scale proteomic analysis, performed for one miRNA (miR-223) in only one cell type (murine neutrophils), revealed that although some proteins were repressed by 50-80%, miR-223 typically had more modest effects [92], suggesting that perhaps other miRNAs in their endogenous context have targets for which protein output is more dramatically repressed. Clearly, a challenge ahead in molecular toxinology is to design laboratory experiments to uncover the impact, molecular details, and evolution of the regulatory miRNA-target interactions that shape venom phenotype during snake development. In this regard, it has been proposed that a single origin of venom in Squamata, the order of reptiles including lizards and snakes, dates back roughly 200 Mya to the Late Triassic/Early Jurassic period [1,2,93]. Existing snake venom toxins are the result of recruitment events by which ordinary genes were duplicated, and the new genes selectively expressed in the venom gland and amplified to multigene families with extensive neofunctionalization throughout the approximately 112–125 Mya of snake evolution. Given the central role that diet has played in the adaptive radiation of snakes [4], venom thus represents a key adaptation that has played an important role in the diversification of snakes. Our findings here reported support the view that understanding the basis for the phenotypic diversity of snake venoms requires a deep understanding of the mechanisms regulating the transcriptional and translational activity of the venom gland. Our results, though restricted

Cluster ID	Contig map	Best blast hit description
Ad 368/New 299	contig02859	X12603 | Crotalus durissus terrificus RNA for crotoxin B
	contig02858	X12603 | Crotalus durissus terrificus RNA for crotoxin B
Ad 1078/New 1849	contig01024	U01027 | Crotalus scutulatus scutulatus Mojave Toxin subunit gene
	contig00975	U01026 | Crotalus scutulatus scutulatus Mojave Toxin subunit gene
	contig00986	AF269131 | Crotalus atrox acidic phospholipase A2 precursor mRNA
	contig00974	U01026 | Crotalus scutulatus scutulatus Mojave Toxin subunit gene
Ad 2166	contig02859	X12603 | Crotalus durissus terrificus RNA for crotoxin B
	contig02858	X12603 | Crotalus durissus terrificus RNA for crotoxin B
	contig01024	U01027 | Crotalus scutulatus scutulatus Mojave Toxin subunit gene
	contig01004	U01027 | Crotalus scutulatus scutulatus Mojave Toxin subunit gene
New 4393/Ad 3416	contig03089	GQ451440 | Crotalus viridis viridis metalloproteinase VMP-II precursor
	contig03090	D28871 | AGKAHB Gloydius halys gene for prepro-halystatin 2 and 3
	contig03078	HM443642 | Bothrops neuwiedi MP_IIa SVMP precursor
New 2578	contig03130	U01234 | CRLPREHTA Crotalus atrox hemorrhagic toxin a, atrolysin a (Ht-a)
	contig03228	GQ451441 | Agkistrodon piscivorus leucostoma metalloproteinase VMP-III
	contig03079	GQ451440 | Crotalus viridis viridis metalloproteinase VMP-II precursor
	contig03128	GQ451443 | Agkistrodon piscivorus leucostoma disintegrin precursor

Figure 7 Best BLAST hit descriptions of the 454 contigs mapped by the miRNA sequences highlighted in Figure 6. Predicted MapMi-calculated Watson-Crick pairing and binding energy to their target 3'-UTR loci are displayed in Figure S4 (Additional file 3).

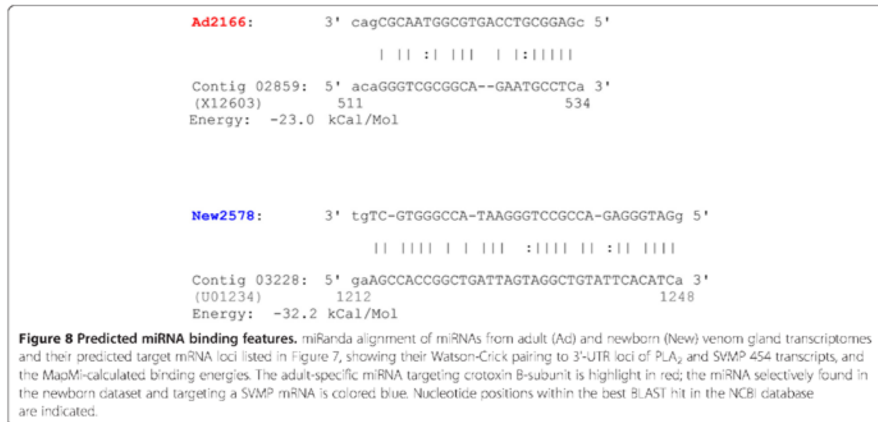


Figure 8 Predicted miRNA binding features. miRanda alignment of miRNAs from adult (Ad) and newborn (New) venom gland transcriptomes and their predicted target mRNA loci listed in Figure 7, showing their Watson-Crick pairing to 3'-UTR loci of PLA_2 and SVMP 454 transcripts, and the MapMi-calculated binding energies. The adult-specific miRNA targeting crotoxin B-subunit is highlight in red; the miRNA selectively found in the newborn dataset and targeting a SVMP mRNA is colored blue. Nucleotide positions within the best BLAST hit in the NCBI database are indicated.

to individual specimens from a single species, suggest a functional role for miRNAs. The impact of specific miRNAs in the modulation of venom composition, and the integration of the mechanisms responsible for the generation and impact of these miRNAs on patterns of expression in the snake's venom gland, are further challenges for future research.

Methods

Proteomics assessment of the ontogenetic shift of venom composition in C. simus

Venoms from a large number (>25) of adult *Crotalus s. simus* specimens and from 11 captive-born siblings (from an age of 8-weeks to 21-months) kept at the serpentarium of the Instituto Clodomiro Picado, University of Costa Rica in San José were collected by snake biting on a parafilm-wrapped jar (adults and juveniles) or by aspiration from the fangs of neonates using an Eppendorf pipette. Crude venoms were centrifuged at low speed to remove cells and debris, lyophilized, weighed on a microbalance, and stored at −20°C until used. Venom proteins were separated by reverse-phase HPLC using a Teknokroma Europa C_{18} (0.4 cm × 25 cm, 5 mm particle size, 300 Å pore size) column and an Agilent LC 1100 High Pressure Gradient System equipped with DAD detector and micro-Auto-sampler. The flow-rate was set to 1 ml/min and the column was developed with a linear gradient of 0.1% TFA in water (solution A) and acetonitrile (solution B), isocratically (5% B) for 10 min, followed by 5–25 % B for 20 min, 25-45% B for 120 min, and 45-70% for 40 min. Isolated proteins were characterized as described [29].

Snake venom gland cDNA synthesis and 454 sequencing

Venom glands of an 8-week-old and an adult *C. s. simus* specimens were removed 3 days after venom milking, when transcription is maximal [94], from anesthetized snakes using fine forceps and immediately placed in RNAlater™ solution (Qiagen). About 50 mg of tissue were disrupted and homogenized by a rotor-stator homogenizer, and total RNA was isolated using RNeasy Mini kit (Qiagen), quantified in a NanoDrop ND_spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) through the A260/A280 ratio, and quality-checked on an agarose gel discerning the 28S and 18S bands of ribosomal RNA. First strand cDNA was synthesized using RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas), which selectively transcribes full-length polyadenylated mRNA. The manufacturer's recommendations were followed except where specified. Approximately 5 μg of total RNA was used as starting material. In order to avoid polymerase slippage, a modified 3' 54-mer adaptor (5' GAGCTAGT TCTGGAG$(T)_{16}$VN), which includes a type IIs enzyme (GsuI) site (underlined), was used for first-strand synthesis. This modified oligonucleotide effectively converts the long run of adenosine residues at the polyA tail into a sequence that causes fewer problems for dideoxy sequencing chemistry, and thus the resulting cDNA libraries were enriched in 3'-end-transcripts. To avoid internal cuts, the cDNA was hemimethylated by adding 5-methyl-dCTP to the dNTPs mix. The first strand cDNA was used as template for second strand synthesis by *E. coli* DNA Polymerase I and RNase H. Double strand (ds) cDNA was precipitated with ethanol and the pellet was resuspended in 70 μL of nuclease-free water

and subjected to enzymatic digestion with GsuI for 4 hours at 30°C. The enzyme was then inactivated at 65°C for 20 minutes and the digested cDNA was stored at −20°C. For 454 pyrosequencing, the GS FLX General DNA Library Preparation Method Manual workflow (Roche Diagnostics) was followed. To this end, 3 μg of final non-normalized cDNA library were sheared by nebulization into small fragments. The fragment ends were polished and short A/B adaptors were ligated onto both ends, providing priming regions to support both emulsion amplification and the pyrosequencing process. A biotin tag on the B adaptor allowed immobilization of the dscDNA library fragments onto streptavidin-conjugated magnetic beads and the subsequent isolation of the library of single strand cDNA sequencing templates. Each of the eight cDNA libraries was tagged with a unique 10-base sequence (MID, Multiplex IDentifier) that is recognized by the sequencing analysis software, allowing for automated sorting of MID-containing reads. Barcoded libraries were simultaneously sequenced in a Genome sequencing FLX Titanium System (Roche Applied Science) at Life Sequencing S.L. (Parc Científic Universitat de Valencia, Paterna, Valencia, Spain; http://www.lifesequencing.com) using the method developed by Margulies et al. [95].

Bioinformatic analysis of the 454 reads

454 data were analyzed using the workflow developed in [15] to identify sequences of toxin molecules by similarity search against nucleotide databases, which includes available NGS algorithms and in-house scripts. An initial quality test step of both the raw reads provided by the Genome Sequencer FLX System prior to the assembly process and the longer contig sequences obtained after running the 454 Newbler assembler program (version 2.6) (Titanium chemistry) was run using the Prinseq program (standalone version 0.17.4) [96]. Interspersed repeats and low complexity DNA sequences were masked from the transcript reads using RepeatMasker (version 3.2.9) [97]. RepeatMasker is available from the Institute for Systems Biology (http://www.systemsbiology.org) addressing http://repeatmasker.org. The program screens DNA sequences for interspersed repeats and low complexity DNA sequences included in the Repbase database (http://www.girinst.org). Repbase, a comprehensive database of repetitive element consensus sequences (update of 20 September 2011), operates as a service of the Genetic Information Research Institute (http://www.girinst.org). Data and computational resources for the Pre-Masked Genomes page is provided courtesy of the UCSC Genome Bioinformatics group (http://genome.ucsc.edu). Masked contig sequences were translated into the 6 possible reading frames and blasted against the non-redundant NCBI database (http://blast.ncbi.nlm.nih.gov, release of February 2012) and the UniProtKB/Swiss-Prot Toxin Annotation Program database (http://us.expasy.org/sprot/tox-prot), using BlastX and BlastN [98] algorithms (version 2.2.24), specifying a cut-off value of e-03 and BLOSUM62 as scoring matrix. Snake venom gland-specific transcripts were selected from best BLAST-hit descriptions identifying GenBank entries belonging to the taxonomic suborder *Serpentes*. This taxonomic group is represented by 44,141 nucleotide records comprising entries from 2,396 different species. A second filtering round was carried out using a list of keywords (including the acronyms of all known toxin protein families described so far to distinguish putative snake venom toxins from non-toxin (ribosomal, mitochondrial, nuclear, etc.) ordinary proteins [15]. The phylogenetically nearest top-hit full-length sequence was designated as the reference sequence onto which all toxin family-specific reads were aligned to create a multiple alignment using COBALT [99]. The multiple alignment was then parsed to create an assembled (consensus) toxin sequence in which each amino acid position is supported by at least four reads. We considered two contigs as different if their nucleotide sequences depart in more number of positions than expected from a sequencing error rate of 1.5%, and the same mutated residues were found in at least two other reads. Positions where two or more amino-acids fulfilled this criterium were annotated as variable residues suggesting the occurrence of different alleles (isoforms) of the protein.

The relative expression of a given toxin protein family was calculated according to the RPKM (Reads per Kilobase of exon per Million mapped reads) standard procedure described by Mortazavi and coworkers [100]. This normalization procedure provides an analog of the relative molar concentrations of transcripts. To this end, all the reads contributing to the contigs (regardless whether that read uniquely maps to that contig or not) and the length of the phylogenetically nearest coding sequence were taken into account to calculate the RPKM normalized figure:

$$\mathrm{RPKM} = \frac{\text{ORF reads} + \text{ORF Singletons}}{\frac{\text{Total reads}}{10^6} \times \frac{\text{Length}}{10^3}}$$

Possible differential expression of venom proteins between adult and the newborn individuals was assessed with the non-parametric NOISeq-sim algorithm [101] using the following parameters recommended for counts without replicates in the NOISeq-sim manual were used: k (counts equal to zero) = 0.5; nss (number of samples to be simulated) ≥ 5; pnr (percentage of the total

sequencing depth) = 0.2; v (variability in the total sequencing depth of the simulated sample) = 0.02. Reads were normalized by the length of the phylogenetically nearest sequence and a threshold of 0.9 was allowed.

Comparison of newborn and adult *C. s. simus* venom transcriptomes

To assess the degree of similarity between transcripts synthesized by newborn an adult venom glands, a Perl script was written that aligned singletons and Newbler-assembled contigs onto the open reading frame (ORF) of the phylogenetically nearest protein sequence used as reference. The aligned nucleotide sequences were re-assembled with MIRA (http://www.chevreux.org/projects_mira.html) to infer the minimum number of different assemblies. These resulting sequences from newborn and adult individuals were compiled into a single FASTA file, translated into protein sequences, and manually inspected to discard possible mispaired BLAST annotations of local regions due to incorrect frame translations. Protein sequences were then clustered with CD-HIT (standalone version 4.5.7 built on 27th February 2012) [102,103] to identify protein sequences shared between newborn and adult transcriptomes.

Snake venom gland microRNA profiling

Total RNA from neonate and adult venom glands was used to isolate small RNA libraries using a RNeasy Mini Kit following the manufacture's (Qiagen) instructions. Samples were quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) through the 260/280 absorbance relation. 5 microliter of neonate (258.6 ng/microliter) and adult (490.7 ng/microliter) size (<200 nt)-enriched microRNA (miRNA) library were sequenced on an Ion Torrent Personal Genome Machine (PGM™) Sequencing platform at Life Sequencing S.L.

Bioinformatic processing of the Ion-Torrent miRNA reads

Using Prinseq [96], the Ion-Torrent miRNA reads were filtered out according to nucleotide length (min_length 15 nt; max_length 40 nt) and the presence of either the 30-mer IonA or the 23-mer P1 adapter sequences. Adapter sequences were removed using the fasxt_clipper tool from the Fastx-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit). The resulting reads were clustered using CD-HIT-454 [102] setting an identity threshold of 0.98 and the accurate mode option g1. HIT-454 was designed to identify artificial duplicates from raw 454 sequencing reads, including exact duplicates and near identical duplicates. Script cdhit-cluster-consensus (v.13) of the CD-HIT suite of programs was run to derive consensus sequences for each of the miRNA clusters. This program is currently maintained by Dr. Li's group (http://weizhong-lab.ucsd.edu). The relative abundances of miRNAs with frequency count ≥ 100 were normalized by scaling the number of reads clustered by CD-HIT-454 to the total number of processed reads (220,548 for newborn + 389,064 for adult). To identify unique and shared sequences between the newborn and adult datasets, these newborn and adult miRNA clusters were compared between themselves using CD-HIT. Differential expression between miRNAs in these two datasets was assessed with NOISeq-sim [101] allowing a threshold of 0.9. miRNAs comprising ≥ 100 reads were subjected to BLAST analysis against miRBase (release 18, November 2011, which included 21643 mature miRNA products from 168 species) (http://www.mirbase.org) [79].

The search for miRNA targets

The precise rules and energetics for pairing between a miRNA and its mRNA target sites are not completely understood [104,105]. A variety of computational algorithms aimed at predicting miRNA target genes incorporate rules for miRNA-mRNA interactions such as base pairing complementarity and favourable miRNA-mRNA duplex thermodynamics. Current prediction methods are diverse, both in approach and performance [106]. We have employed the freely available target prediction, position-weighted local alignment miRanda algorithm (standalone version 3.3a) [80,81] and the MapMi webserver (version 1.5.0-build 01, release March 2012) (http://www.ebi.ac.uk/enright-srv/MapMi) [83] to identify candidate miRNA target sites in 3'-UTR regions of 454 reads by base complementarity, and putative miRNA loci, respectively. MapMi is a tool designed to locate miRNA precursor sequences in existing genomic sequences, using potential mature miRNA sequences as input. miRanda is an algorithm for finding genomic targets for microRNAs that incorporates current biological knowledge on target rules and computes optimal sequence complementarity between a set of mature microRNAs and a given mRNA using a weighted dynamic programming algorithm. In addition, miRanda uses an estimate of the free energy of formation of the miRNA:mRNA duplex as a secondary filter [82]. The following parameters were set to reduce the estimated false positives to ≤ 9% [80,81]: total Score >100; ΔG of the intermediate duplex < −19 Kcal/mol; and output of 2 or more 454-contig targets.

Database accession

The raw 454 GS FLX Titanium reads of *C. s. simus* adult and neonate venom gland transcriptomes, and the Ion Torrent PGM reads (miRNA sequences) of adult and neonate *C. s. simus* have been archived as Standard Flowgram Format (sff) files with the NCBI Sequence

Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra?term=SRA051956) under accession codes SRX143982 and SRX143985, SRX143983 and SRX143984, respectively.

Animal ethics

Procedures for snake maintenance, sacrifice, and gland extraction in this study followed the Quality Management Protocols from the Instituto Clodomiro Picado, University of Costa Rica, and comply the Animal Welfare Law #7451, chapters II and III, and the Biodiversity Law, Decree 7788, Republic of Costa Rica. Research permission was allowed under Resolution ACG- SINAC- PI-012-2010 (Ministry of Environment of Costa Rica).

Additional files

Additional file 1: Table S1. RepeatMasker usage results and features of the sequence elements masked in the adult *C. s. simus* venom gland transcriptomes analyzed. **Table S2.** RepeatMasker usage results and features of the sequence elements masked in the adult *C. s. simus* venom gland transcriptomes analyzed.

Additional file 2: Table S3. Relative abundances of the different venom protein family hits in the venom gland transcriptomes of newborn and adult *C. s. simus*. **Table S4.** NoiSeq computed expression profile of highly expressed (>100 counts) miRNAs in the venom gland transcriptome listed by decreasing newborn (N) to adult (A) expression ratio. **Table S5.** NoiSeq computed expression profile of highly expressed (>100 counts) miRNAs in the venom gland transcriptome listed by decreasing adult (A) to newborn (N) expression ratio. **Table S6.** MiRanda predicted miRNAs complementary of 3'-UTR loci of 454 *C. s. simus* venom transcripts.

Additional file 3: Figure S1. Multiple alignment of transcript-deduced amino acid sequences of PLA_2 molecules. **Figure S2.** Multiple alignment of transcript-deduced serine proteinase amino acid sequences. **Figure S3.** Multiple alignment of transcript-deduced amino acid sequences of snake venom metalloproteinases. **Figure S4.** Predicted targets for the miRNAs displayed in Figure 7, showing their Watson-Crick pairing to target 3'-UTR loci of PLA_2 and SVMP 454 transcripts, and the corresponding binding energy calculated by MapMi.

Competing interests

The authors declare that they have no competing interests

Authors' contributions

AG, FB, SR and DC were responsible for all stages of animal care and venom extraction. FB, MS and AG dissected the venom glands. AP and LS prepared the RNA samples for 454 and Ion Torrent sequencing. All co-authors (JD, AP, LS, AG, FB, SR, DC, MS, YA, JMG, JJC) analyzed the data and participated in data interpretation and discussion of the results, as well as in revising the article drafted by JJC. All co-authors have seen, reviewed, and approved the final version of the manuscript.

Acknowledgements

This work has been financed by grant BFU2010-17373 from the Ministerio de Economía y Competitividad, Madrid, Spain; PROMETEO/2010/005 from the Generalitat Valenciana; and projects 741-B2-093 from the Vicerrectoría de Investigación, Universidad de Costa Rica, CRUSA-CSIC (2009CR0021), and CYTED (206AC0281). We also acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI). JD is grateful to Eva Alloza (CIPF) and José Afonso Guerra-Assunção (EBI) for useful comments and advice regarding miRNA sequence analysis.

Author details

[1]Instituto de Biomedicina de Valencia, Consejo Superior de Investigaciones Científicas, Jaime Roig 11, Valencia 46010, Spain. [2]Instituto Clodomiro Picado, Facultad de Microbiología, Universidad de Costa Rica, San José, Costa Rica.

Received: 12 November 2012 Accepted: 14 March 2013
Published: 10 April 2013

References

1. Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SF, Kuruppu S, Fung K, Hedges SB, Richardson MK, Hodgson WC, Ignjatovic V, Summerhayes R, Kochva E: **Early evolution of the venom system in lizards and snakes.** *Nature* 2006, **439**:584–588.
2. Fry BG, Casewell NR, Wüster W, Vidal N, Young B, Jackson TN: **The structural and functional diversification of the Toxicofera reptile venom system.** *Toxicon* 2012, **60**:434–448.
3. Vonk FJ, Admiraal JF, Jackson K, Reshef R, de Bakker MA, Vanderschoot K, van den Berge I, van Atten M, Burgerhout E, Beck A, Mirtschin PJ, Kochva E, Witte F, Fry BG, Woods AE, Richardson MK: **Evolutionary origin and development of snake fangs.** *Nature* 2008, **454**:630–633.
4. Greene HW: **Dietary correlates of the origin and radiation of snakes.** *Am Zool* 1983, **23**:431–441.
5. Daltry JC, Wüster W, Thorpe RS: **Diet and snake venom evolution.** *Nature* 1996, **379**:537–540.
6. Gibbs HL, Mackessy SP: **Functional basis of a molecular adaptation: prey-specific toxic effects of venom from Sistrurus rattlesnakes.** *Toxicon* 2008, **53**:672–679.
7. Barlow A, Pook CE, Harrison RA, Wüster W: **Coevolution of diet and prey-specific venom activity supports the role of selection in snake venom evolution.** *Proc Biol Sci* 2009, **276**:2443–2449.
8. Vidal N, Hedges SB: **The molecular evolutionary tree of lizards, snakes, and amphisbaenians.** *C Rendus Biol* 2009, **332**:129–139.
9. Vidal N, Rage J-C, Couloux A, Hedges SB: **Snakes (Serpentes)**. In *The Timetree of Life*. Edited by Hedges SB, Kumar S. Oxford Univ. Press; 2009:390–397.
10. Redi F: In *Osservazioni intorno alle Vipere*. Edited by All'Insegna Della Stella F.; 1664. a digitalized version is freely accessible in http://archive.org/details/osservazioniint00redigoog.
11. Fontana F: *Traité sur le vénin de la vipere, sur les poisons americains, sur le laurier-cerise et sur quelques autres poisons végetaux*. Florence: Gibelin; 1781.
12. Hawgood BJ: **Abbé Felice Fontana (1730–1805): founder of modern toxinology.** *Toxicon* 1995, **33**:591–601.
13. Calvete JJ: **Venomics, what else?** *Toxicon* 2012, **60**:427–433.
14. de Azevedo ILM J, Diniz MRV, Ho PL: **Venom gland transcriptomic analysis.** In *Animal Toxins: State of the Art. Perspectives in Health and Biotechnology*. Edited by De Lima ME, Pimenta AMC, Martin-Euclaire MF, Zingali RB, Rochat H. Belo Horizonte: Editora UFMG; 2009:693–713.
15. Durban J, Juárez P, Angulo Y, Lomonte B, Flores-Diaz M, Alape-Girón A, Sasa M, Sanz L, Gutiérrez JM, Dopazo J, Conesa A, Calvete JJ: **Profiling the venom gland transcriptomes of Costa Rican snakes by 454 pyrosequencing.** *BMC Genomics* 2011, **12**:259.
16. Rokyta DR, Wray KP, Lemmon AR, Lemmon EM, Caudle SB: **A high-throughput venom-gland transcriptome for the Eastern Diamondback Rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes.** *Toxicon* 2011, **57**:657–671.
17. Calvete JJ: **Proteomic tools against the neglected pathology of snake bite envenoming.** *Exp Rev Proteomics* 2011, **8**:739–758.
18. Calvete JJ: **Snake venomics: from the inventory of toxins to biology.** *Toxicon* 2013. in press.
19. Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JD, King GF, Nevalainen TJ, Norman JA, Lewis RJ, Norton RS, Renjifo C, de la Vega RC: **The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms.** *Annu Rev Genomics Hum Genet* 2009, **10**:483–511.
20. Fry BG, Scheib H, De LM, Junqueira De Azevedo I, Silva DA, Casewell NR: **Novel transcripts in the maxillary venom glands of advanced snakes.** *Toxicon* 2012, **59**:696–708.
21. Chippaux JP, Williams V, White J: **Snake venom variability: methods of study, results, and interpretation.** *Toxicon* 1991, **29**:1279–1303.

22. Gibbs HL, Sanz L, Calvete JJ: **Snake population venomics: proteomics-based analyses of individual variation reveals significant gene regulation effects on venom protein expression in Sistrurus rattlesnakes.** *J Mol Evol* 2009, **68**:113–125.
23. Gibbs HL, Sanz L, Chiucchi JE, Farrell TM, Calvete JJ: **Proteomic analysis of ontogenetic and diet-related changes in venom composition of juvenile and adult Dusky Pigmy rattlesnakes (*Sistrurus miliarius barbouri*).** *J Proteomics* 2011, **74**:2169–2179.
24. Chiucchi JE, Gibbs HL: **Similarity of contemporary and historical gene flow among highly fragmented populations of an endangered rattlesnake.** *Mol Ecol* 2010, **19**:5345–5358.
25. Mackessy SP: **Venom ontogeny in the Pacific rattlesnakes, Crotalus viridis helleri and C. viridis oreganus.** *Copeia* 1988, **1988**:92–101.
26. Guércio RAP, Shevchenko A, Shevchenko A, López-Lozano JL, Paba J, Sousa MV, Ricart CAO: **Ontogenetic variations in the venom proteome of the Amazonian snake *Bothrops atrox*.** *Proteome Sci* 2006, **4**:11.
27. Mackessy SP, Sixberry NM, Heyborne WH, Fritts T: **Venom of the Brown Treesnake, *Boiga irregularis*: ontogenetic shifts and taxa-specific toxicity.** *Toxicon* 2006, **47**:537–548.
28. Alape-Girón A, Sanz L, Escolano J, Flores-Díaz M, Madrigal M, Sasa M, Calvete JJ: **Snake venomics of the lancehead pitviper *Bothrops asper*: geographic, individual, and ontogenetic variations.** *J Proteome Res* 2008, **7**:3556–3571.
29. Calvete JJ, Sanz L, Cid P, De La Torre P, Flores-Diaz M, Dos Santos MC, Borges A, Bremo A, Angulo Y, Lomonte B, Alape-Girón A, Gutiérrez JM: **Snake venomics of the Central American rattlesnake *Crotalus simus* and the South American *Crotalus durissus* complex points to neurotoxicity as an adaptive paedomorphic trend along Crotalus dispersal in South America.** *J Proteome Res* 2010, **9**:528–544.
30. Zelanis A, Travaglia-Cardoso SR, Furtado MFD: **Ontogenetic changes in *Bothrops insularis* (Serpentes: Viperidae) snake venom and its biological implications.** *South Am J Herpetol* 2008, **3**:45–50.
31. Zelanis A, Tashima AK, Pinto AF, Leme AF, Stuginski DR, Furtado MF, Sherman NE, Ho PL, Fox JW, Serrano SM: ***Bothrops jararaca* venom proteome rearrangement upon neonate to adult transition.** *Proteomics* 2011, **11**:4218–4228.
32. Madrigal M, Sanz L, Flores-Diaz M, Sasa M, Núñez V, Alape-Girón A, Calvete JJ: **Snake venomics across genus Lachesis. Ontogenetic changes in the venom composition of *L. stenophrys* and comparative proteomics of the venoms of adult *L. melanocephala* and *L. acrochorda*.** *J Proteomics*. in press.
33. Andrade DV, Abe AS: **Relationship of venom ontogeny and diet in *Bothrops*.** *Herpetologica* 1999, **55**:200–204.
34. Wüster W, Ferguson JE, Quijada-Mascareñas JA, Pook CE, Salomão MG, Thorpe RS: **Tracing an invasion: landbridges, refugia and the phylogeography of the Neotropical rattlesnake (Serpentes: Viperidae: *Crotalus durissus*).** *Mol Ecol* 2005, **14**:1095–1108.
35. Gutiérrez JM, Dos Santos MC, Furtado MF, Rojas G: **Biochemical and pharmacological similarities between the venoms of newborn *Crotalus durissus durissus* and adult *Crotalus durissus terrificus* rattlesnakes.** *Toxicon* 1991, **29**:1273–1277.
36. Saravia P, Rojas E, Arce V, Guevara C, López JC, Chaves E, Velásquez R, Rojas G, Gutiérrez JM: **Geographic and ontogenic variability in the venom of the neotropical rattlesnake *Crotalus durissus*: Pathophysiological and therapeutic implications.** *Rev Biol Trop* 2002, **50**:337–346.
37. Mackessy SP: **Venom composition in rattlesnakes: trends and biological significance.** In *The Biology of Rattlesnakes*. Edited by Hayes WK, Beaman KR, Cardwell MD, Bush SP. Loma Linda, California: Loma Linda University Press; 2008:495–510.
38. Mackessy SP: **Evolutionary trends in venom composition in the Western Rattlesnakes (*Crotalus viridis* sensu lato): Toxicity vs. tenderizers.** *Toxicon* 2010, **55**:1463–1474.
39. McCue MD: **Cost of producing venom in three North American pitviper species.** *Copeia* 2006, **2006**:818–825.
40. Bon C: **Multicomponent neurotoxic phospholipases A_2.** In *Venom phospholipase A2 enzymes: structure, function and mechanism*. Edited by Kini RM. Chichester: Wiley; 1997:269–286.
41. Faure G, Xu H, Saul FA: **Crystal structure of crotoxin reveals key residues involved in the stability and toxicity of this potent heterodimeric β-neurotoxin.** *J Mol Biol* 2011, **412**:176–191.
42. Faure G, Saul F: **Crystallographic characterization of functional sites of crotoxin and ammodytoxin, potent β-neurotoxins from Viperidae venom.** *Toxicon* 2012, **60**:531–538.
43. Gutiérrez JM: **Clinical toxicology of snakebite in Central America**. In *Handbook of clinical toxicology of animal venoms and poisons*. Edited by Meier J, White J. Boca Raton, Florida: CRC; 1995:645–665.
44. Massey DJ, Calvete JJ, Sánchez EE, Sanz L, Richards K, Curtis R, Boesen K: **Venom variability and envenoming severity outcomes of the *Crotalus scutulatus scutulatus* (Mojave rattlesnake) from Southern Arizona.** *J Proteomics* 2012, **75**:2576–2587.
45. Glenn JL, Straight RC, Wolfe MC, Hardy DL: **Geographical variation in *Crotalus scutulatus scutulatus* (Mojave rattlesnake) venom properties.** *Toxicon* 1983, **21**:119–130.
46. Glenn JL, Straight RC: **Intergradation of two different venom populations of the Mojave rattlesnake (*Crotalus scutulatus scutulatus*) in Arizona.** *Toxicon* 1989, **27**:411–418.
47. Gutiérrez JM, Lomonte B, León G, Alape-Girón 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.
48. Williams DJ, Gutiérrez JM, Calvete JJ, Wüster W, Ratanabanangkoon K, Paiva O, Brown NI, Casewell NR, Harrison RA, Rowley PD, O'Shea M, Jensen SD, Winkel KD, Warrell DA: **Ending the drought: new strategies for improving the flow of affordable, effective antivenoms in Asia and Africa.** *J Proteomics* 2011, **74**:1735–1767.
49. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM: **Accuracy and quality of massively parallel DNA pyrosequencing.** *Genome Biol* 2007, **8**:R143.
50. Biémont C, Vieira C: **Junk DNA as an evolutionary force.** *Nature* 2006, **443**:521–524.
51. Castoe TA, Hall KT, Guibotsy Mboulas ML, Gu W, de Koning AP, Fox SE, Poole AW, Vemulapalli V, Daza JM, Mockler T, Smith EN, Feschotte C, Pollock DD: **Discovery of highly divergent repeat landscapes in snake genomes using high-throughput sequencing.** *Genome Biol Evol* 2011, **3**:641–653.
52. 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.
53. Junqueira-de-Azevedo ILM, Ching ATC, Carvalho E, Faria F, Nishiyama MY Jr, 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, **172**:877–889.
54. Ching AT, Rocha MM, Leme AFP, Pimenta DC, Furtado MFD, Serrano SM, Ho PL, de Azevedo IL J: **Some aspects of the venom proteome of the Colubridae snake *Philodryas olfersii* revealed from a Duvernoy's (venom) gland transcriptome.** *FEBS Lett* 2006, **580**:4417–4422.
55. Kordis D, Gubenšek F: **Bov-B long interspersed repeated DNA (LINE) sequences are present in *Vipera ammodytes* phospholipase A2 genes and in genomes of Viperidae snakes.** *Eur J Biochem* 1997, **246**:772–779.
56. Kordis D, Gubenšek F: **The Bov-B lines found in *Vipera ammodytes* toxic PLA2 genes are widespread in snake genomes.** *Toxicon* 1998, **36**:1585–1590.
57. Ikeda N, Chijiwa T, Matsubara K, Oda-Ueda N, Hattori S, Matsuda Y, Ohno M: **Unique structural characteristics and evolution of a cluster of venom phospholipase A_2 isozyme genes of *Protobothrops flavoviridis* snake.** *Gene* 2010, **461**:15–25.
58. Nobuhisa I, Ogawa T, Deshimaru M, Chijiwa T, Nakashima K-I, Chuman Y, Shimohigashi Y, Fukumaki Y, Hattori S, Ohno M: **Retrotransposable CR1-like elements in Crotalinae snake genomes.** *Toxicon* 1998, **36**:915–920.
59. Sanz L, Harrison RA, Calvete JJ: **First draft of the genomic organization of a PIII-SVMP gene.** *Toxicon* 2012, **60**:455–469.
60. van de Lagemaat LN, Landry JR, Mager DL, Medstrand P: **Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions.** *Trends Genet* 2003, **19**:530–536.
61. Medstrand P, Van de Lagemaat LN, Dunn CA, Landry JR, Svenback D, Mager DL: **Impact of transposable elements on the evolution of mammalian gene regulation.** *Cytogenet Genome Res* 2005, **110**:342–345.

62. Piskurek O, Austin CC, Okada N: **Sauria SINEs: novel short interspersed transposable elements that are widespread in reptile genomes.** *J Mol Evol* 2006, **62:**630–644.
63. Piskurek O, Okada N: **Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals.** *Proc Natl Acad Sci USA* 2007, **104:**12046–12051.
64. Oshima K, Okada N: **SINEs and LINEs: symbionts of eukaryotic genomes with a common tail.** *Cytogenet Genome Res* 2005, **110:**475–490.
65. Luan DD, Korman MH, Jakubczak JL, Eickbush TH: **Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition.** *Cell* 1993, **72:**595–605.
66. Deininger PL, Batzer MA: **Mammalian retroelements.** *Genome Res* 2002, **12:**1455–1465.
67. Piskurek O, Nishihara H, Okada N: **The evolution of two partner LINE/SINE families and a full-length chromodomain-containing *Ty3/Gypsy* LTR element in the first reptilian genome of *Anolis carolinensis*.** *Gene* 2009, **441:**111–118.
68. International Chicken Genome Sequencing Consortium: **Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.** *Nature* 2004, **432:**695–716.
69. Alföldi J, Di Palma F, Grabherr M, Williams C, Kong L, Mauceli E, Russell P, Lowe CB, Glor RE, Jaffe JD, Ray DA, Boissinot S, Shedlock AM, Botka C, Castoe TA, Colbourne JK, Fujita MK, Moreno RG, ten Hallers BF, Haussler D, Heger A, Heiman D, Janes DE, Johnson J, de Jong PJ, Koriabine MY, Lara M, Novick PA, Organ CL, Peach SE, *et al*: **The genome of the green anole lizard and a comparative analysis with birds and mammals.** *Nature* 2011, **477:**587–591.
70. Kazazian HHJ: **Mobile elements: drivers of genome evolution.** *Science* 2004, **303:**1626–1632.
71. Cline M, Smooth M, Cerami E, Kushinski A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD: **Integration of biological networks and gene expression data using Cytoscape.** *Nat Protoc* 2007, **2:**2366–2382.
72. Sharp PA: **The centrality of RNA.** *Cell* 2009, **136:**577–580.
73. Licatalosi DD, Darnell RB: **RNA processing and its regulation: global insights into biological networks.** *Nat Rev Genet* 2010, **11:**75–87.
74. Ambros V: **The functions of animal microRNAs.** *Nature* 2004, **431:**350–355.
75. He L, Hannon GJ: **MicroRNAs: small RNAs with a big role in gene regulation.** *Nat Rev Genet* 2004, **5:**522–531.
76. Bartel DP: **MicroRNAs: target recognition and regulatory functions.** *Cell* 2009, **136:**215–233.
77. Höck H, Meister G: **The Argonaute protein family.** *Genome Biol* 2008, **9:**210.
78. Wheeler BM, Heimberg AM, Moy VN, Sperling EA, Holstein TW, Heber S, Peterson KJ: **The deep evolution of metazoan microRNAs.** *Evol Dev* 2009, **11:**145–150.
79. Kozomara A, Griffiths-Jones S: **miRBase: integrating microRNA annotation and deep-sequencing data.** *Nucleic Acid Res* 2011, **39:**D152–D157.
80. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS: **MicroRNA targets in Drosophila.** *Genome Biol* 2003, **5:**R1.
81. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS: **Human MicroRNA targets.** *PLoS Biol* 2004, **2:**e363.
82. Wuchty S, Fontana W, Hofacker IL, Schuster P: **Complete suboptimal folding of RNA and the stability of secondary structures.** *Biopolymers* 1999, **49:**145–165.
83. Guerra-Assunção JA, Enright AJ: **MapMi: automated mapping of microRNA loci.** *BMC Bioinformatics* 2012, **11:**133.
84. Chen K, Rajewsky N: **Deep conservation of microRNA-target relationships and 3' UTR motifs in vertebrates, flies, and nematodes.** *Cold Spring Harb Symp Quant Biol* 2006, **71:**149–156.
85. Lewis BP, Burge CB, Bartel DP: **Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.** *Cell* 2005, **120:**15–20.
86. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP: **MicroRNA targeting specificity in mammals: determinants beyond seed pairing.** *Mol Cell* 2007, **27:**91–105.
87. Ma JB, Yuan YR, Meister G, Pei Y, Tuschl T, Patel J: **Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein.** *Nature* 2005, **434:**666–670.
88. Parker JS, Roe SM, Barford D: **Structural insight into mRNA recognition from Piwi domain-siRNA guide complex.** *Nature* 2005, **434:**663–666.
89. Bon C, Changeux JP, Jeng TW, Fraenkel-Conrat H: **Postsynaptic effects of crotoxin and its isolated subunits.** *Eur J Biochem* 1979, **99:**471–481.
90. Faure G, Copic A, Le Porrier S, Gubensek F, Bon C, Krizaj I: **Crotoxin acceptor protein isolated from Torpedo electric organ: binding properties to crotoxin by surface plasmon resonance.** *Toxicon* 2003, **41:**509–517.
91. Pereañez JA, Gómez ID, Patiño AC: **Relationship between the structure and the enzymatic activity of crotoxin complex and its phospholipase A_2 subunit: An in silico approach.** *J Mol Graph Model* 2012, **35:**36–42.
92. Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP: **The impact of microRNAs on protein output.** *Nature* 2008, **455:**64–71.
93. Vidal N, Hedges SB: **The phylogeny of squamate reptiles (lizards, snakes, and amphisbaenians) inferred from nine nuclear protein-coding genes.** *Comptes Rendus Biologies* 2005, **328:**1000–1008.
94. Paine MJ, Desmond HP, Theakston RDG, Crampton JM: **Gene expression in *Echis carinatus* (carpet viper) venom glands following milking.** *Toxicon* 1992, **30:**379–386.
95. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, *et al*: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437:**376–380.
96. Schmieder R, Edwards R: **Quality control and preprocessing of metagenomic datasets.** *Bioinformatics* 2011, **27:**863–864.
97. Smit AFA, Hubley R, Green P: *RepeatMasker Open-3.0*. http://repeatmasker.org.
98. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403–410.
99. Papadopoulos JS, Agarwala R: **COBALT: constraint-based alignment tool for multiple protein sequences.** *Bioinformatics* 2007, **23:**1073–1079.
100. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008, **5:**621–628.
101. Tarazona S, García-Alcalde F, Dopazo J, Ferrer A, Conesa A: **Differential expression in RNA-seq: a matter of depth.** *Genome Res* 2011, **21:**2213–2223.
102. Li W, Godzik A: **Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences.** *Bioinformatics* 2006, **22:**1658–1659.
103. Niu BF, Fu LM, Sun SL, Li WZ: **Artificial and natural duplicates in pyrosequencing reads of metagenomic data.** *BMC Bioinformatics* 2010, **11:**187.
104. Stefani G, Slack FJ: **A 'pivotal' new rule for microRNA-mRNA interactions.** *Nature Struct & Mol Biol* 2012, **19:**265–266.
105. Chi SW, Hannon GJ, Darnell RB: **An alternative mode of microRNA target recognition.** *Nature Struct & Mol Biol* 2012, **19:**321–328.
106. Seelbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N: **Widespread changes in protein synthesis induced by microRNAs.** *Nature* 2008, **455:**58–63.

doi:10.1186/1471-2164-14-234
Cite this article as: Durban *et al.*: Integrated "omics" profiling indicates that miRNAs are modulators of the ontogenetic venom composition shift in the Central American rattlesnake, *Crotalus simus simus*. *BMC Genomics* 2013 **14**:234.

Capítol 3

Amino acid sequence and biological characterization of BlatPLA2, a non-toxic acidic phospholipase A2 from the venom of the arboreal snake *Bothriechis lateralis* from Costa Rica

30 de Junio de 2014

A quien interese,

Por este medio me permito hacer constar que en la publicación "Van der Laat M, Fernández J, **Durban J**, Villalobos E, Camacho E, Calvete JJ, Lomonte B (2013) Amino acid sequence and biological characterization of BlatPLA$_2$, a non-toxic acidic phospholipase A$_2$ from the venom of the arboreal snake *Bothriechis lateralis* from Costa Rica. *Toxicon* **73**, 71-80", hemos contado con la valiosa participación y co-autoría de Jordi Durban, del Instituto de Biomedicina de Valencia.

En dicho trabajo se aisló una nueva fosfolipasa A$_2$ acídica a partir del veneno de la serpiente *Bothriechis lateralis* de Costa Rica, y se determinó su secuencia completa de aminoácidos, sus propiedades biológicas, y sus relaciones filogenéticas con respecto a proteínas homólogas descritas en venenos de serpientes, a través de la colaboración entre el grupo de Costa Rica y el grupo de Valencia. El aporte central de Jordi Durban consistió en la realización de los análisis filogenéticos de la secuencia obtenida a través de métodos bioinformáticos.

La distribución general de los aportes de los distintos autores se pueden resumir de la siguiente manera: Marco Van der Laat, Julián Fernández y Eva Villalobos purificaron y caracterizaron los aspectos bioquímicos generales e inmunoquímicos de la nueva proteína, Jordi Durban realizó los análisis de bioinformática y filogenia, Erika Camacho realizó algunas pruebas funcionales, y Juan J. Calvete y Bruno Lomonte realizaron la secuenciación y la supervisión de la redacción del manuscrito.

Quedando a su disposición para cualquier información adicional que se requiera, se suscribe

Atentamente,

Dr Bruno Lomonte
Catedrático
Instituto Clodomiro Picado
Universidad de Costa Rica
bruno.lomonte@ucr.ac.cr

Toxicon 73 (2013) 71–80

Contents lists available at SciVerse ScienceDirect

Toxicon

journal homepage: www.elsevier.com/locate/toxicon

Amino acid sequence and biological characterization of BlatPLA$_2$, a non-toxic acidic phospholipase A$_2$ from the venom of the arboreal snake *Bothriechis lateralis* from Costa Rica

Marco Van der Laat[a], Julián Fernández[a], Jordi Durban[b], Eva Villalobos[a], Erika Camacho[a], Juan J. Calvete[b], Bruno Lomonte[a,*]

[a] *Instituto Clodomiro Picado, Facultad de Microbiología, Universidad de Costa Rica, San José 11501, Costa Rica*
[b] *Instituto de Biomedicina de Valencia, CSIC, Jaume Roig 11, 46010 Valencia, Spain*

A R T I C L E   I N F O

Article history:
Received 13 May 2013
Received in revised form 28 June 2013
Accepted 2 July 2013
Available online 16 July 2013

Keywords:
Snake venom
Viperidae
Phospholipase A$_2$
Arboreal snake
Bothriechis lateralis

A B S T R A C T

Bothriechis is considered a monophyletic, basal genus of arboreal Neotropical pitvipers distributed across Middle America. The four species found in Costa Rica (*B. lateralis, B. schlegeli, B. nigroviridis, B. supraciliaris*) differ in their venom proteomic profiles, suggesting that different *Bothriechis* taxa have evolved diverse trophic strategies. In this study, we isolated a phospholipase A$_2$ (PLA$_2$) from *B. lateralis* venom, aiming at increasing our knowledge on the structural and functional characteristics of group II acidic PLA$_2$s, whose toxic actions are generally more restricted than those displayed by basic PLA$_2$s. The new acidic enzyme, BlatPLA$_2$, occurs as a monomer of 13,917 Da, in contrast to many basic group II PLA$_2$s which associate into dimers and often display myotoxicity and/or neurotoxicity. Its amino acid sequence of 122 residues predicts an isoelectric point of 4.7, and displays significant differences with previously characterized acidic PLA$_2$s, with which it shows a maximum sequence identity of 78%. BlatPLA$_2$ is catalytically active but appears to be devoid of major toxic activities, lacking intravenous or intracerebroventricular lethality, myotoxicity, *in vitro* anticoagulant activity, and platelet aggregation or inhibition effects. Phylogenetic relationships with similar group II enzymes suggest that BlatPLA$_2$ may represent a basal sequence to other acidic PLA$_2$s. Due to the metabolic cost of venom protein synthesis, the presence of a relatively abundant (9%) but non-toxic component is somewhat puzzling. Nevertheless, we hypothesize that BlatPLA$_2$ could have a role in the pre-digestion of prey, possibly having retained characteristics of ancestral PLA$_2$s without evolving towards potent toxicity.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Phospholipases A$_2$ (PLA$_2$s; EC 3.1.1.4) are widely distributed in animal venoms, where they play major roles in toxicity. In snake venoms, these enzymes often display potent neurotoxic and/or myotoxic effects *in vivo*, therefore being relevant for subduing and pre-digestion of prey (Rosenberg, 1990; Kini, 1997; Montecucco et al., 2008). During ophidian evolution, at least two independent recruitments of ancestral PLA$_2$ genes have been proposed to explain the presence of group I enzymes in elapid and colubrid venoms and group II enzymes in the venom of viperids (Fry and Wüster, 2004). The evident structural conservation between these two groups of PLA$_2$ toxins and their corresponding non-toxic counterparts of mammals, proves the versatility of the recruited scaffolds to evolve into successful 'chemical weapons' in venom glands (Casewell et al., 2013). Moreover, the fact that neofunctionalization towards neurotoxicity and myotoxicity

* Corresponding author.
E-mail address: bruno.lomonte@ucr.ac.cr (B. Lomonte).

0041-0101/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.toxicon.2013.07.008

emerged in both lineages (groups I and II) of ancestral PLA_2 genes has been pointed out as an interesting case of convergent evolution (Lomonte and Gutiérrez, 2011; Lomonte and Rangel, 2012).

Despite important achievements in understanding the structural bases of toxic activities exerted by venom PLA_2s, many fundamental aspects of their actions are still open questions (reviewed by Gutiérrez and Lomonte, 2013). It has long been recognized that the ability to hydrolyze phospholipids *per se* is insufficient to generate toxicity (Rosenberg, 1990; Kini, 2003; Petan et al., 2013). On the other hand, some effects of PLA_2s involve interactions with proteins independently of catalysis (Lambeau and Lazdunski, 1999; Fujisawa et al., 2008; Šribar and Krizaj, 2011), and, moreover, a divergent subgroup of these proteins known as 'Lys49 PLA_2 homologues' display toxicity in spite of their lack of catalytic activity (Lomonte and Rangel, 2012). Thus, correlating the structures of snake venom PLA_2s to toxic functions remains a challenging area of research.

As a general trend, basic venom PLA_2s appear to be more toxic than their acidic counterparts. In the case of group II PLA_2s, an increasing number of acidic enzymes have been found to lack major toxic actions (de Araújo et al., 1994; Daniele et al., 1995; Fernández et al., 2010a; Garcia Denegri et al., 2010; Silveira et al., 2013) or to display moderate myotoxicity or pharmacological activities (Serrano et al., 1999; Fuly et al., 2002; Andrião-Escarso et al., 2002; Ketelhut et al., 2003; Roberto et al., 2004; Cogo et al., 2006; de Albuquerque Modesto et al., 2006; Rodrigues et al., 2007; Santos-Filho et al., 2008). On these grounds, the isolation and characterization of new acidic PLA_2s may provide a valuable comparative framework to disclose the structural determinants responsible for the remarkable toxicity of basic PLA_2s. With this aim, we focused in this study on a PLA_2 present in the venom of the arboreal snake *Bothriechis lateralis*. Previous work on the proteomic profiling of venoms from four *Bothriechis* species that inhabit Costa Rica, *B. lateralis*, *B. schlegeli* (Lomonte et al., 2008), *B. nigroviridis* (Fernández et al., 2010b), and *B. supraciliaris* (Lomonte et al., 2012), revealed strikingly different compositions, in spite of their common adaptation to arboreal habitats, diet, and close geographical distributions within this small country. In contrast to the other three species of *Bothriechis*, the venom of *B. lateralis* does not contain basic proteins, but presents a relatively abundant (~9%) acidic PLA_2, which was isolated, sequenced, and characterized in the present study.

2. Materials and methods

2.1. Isolation of BlatPLA$_2$

Venom was collected from ten specimens of *B. lateralis* from Costa Rica, kept at the serpentarium of Instituto Clodomiro Picado (ICP). The pooled venom was freeze-dried and stored at −20 °C. Batches of 2–3 mg were dissolved in 200 μl of 0.1% trifluoroacetic acid (TFA) in water, centrifuged at 15,000 × *g* for 5 min, and fractionated by RP-HPLC on a C_{18} column (4.6 × 250 mm; 5 μm particle; Teknokroma) using an Agilent 1200 chromatograph. Elution, monitored at 215 nm, was performed at 1 ml/min by applying a gradient towards acetonitrile containing 0.1% TFA, as follows: 0–33% in 19 min, 33–40% in 26 min, and 40–70% in 5 min. This gradient was modified on the basis of previous studies on the venom proteome of *B. lateralis* (Lomonte et al., 2008) in order to optimize the obtention of the PLA_2 peak. Fractions were collected manually and dried by vacuum centrifugation (Savant). The peak eluting at 28.6 min yielded a protein with PLA_2 activity, which was analyzed by SDS-PAGE (12%) under reducing and non-reducing conditions, and by analytical RP-HPLC on a C_4 column (4.6 × 150 mm). This protein, named BlatPLA$_2$, was further characterized as described below.

2.2. Molecular mass determination

BlatPLA$_2$ was analyzed by MALDI-TOF mass spectrometry (MS) on an Applied Biosystems 4800 Plus instrument. Various dilutions of the protein were mixed with an equal volume of saturated sinapinic acid in 50% acetonitrile containing 0.1% TFA, spotted (1 μl) onto an Opti-TOF 384 plate, and allowed to dry. Spectra were acquired in positive linear mode, in the *m/z* range 4000–30,000, using 1500 laser shots at an intensity level of 4000. CalMix-5 standards (AB Sciex) spotted on the same plate were used for external calibration.

2.3. Amino acid sequencing

BlatPLA$_2$ was dissolved in ammonium bicarbonate (50 mM) and subjected to reduction with dithiothreitol (10 mM) and alkylation with iodoacetamide (50 mM). Its N-terminal sequence was obtained by Edman degradation on a Procise instrument (Applied Biosystems). Additionally, the reduced and alkylated protein samples were digested with sequencing grade trypsin, chymotrypsin, or Asp-N, and the resulting peptides were analyzed by MALDI-TOF-TOF MS and/or nano-ESI MS/MS, in order to obtain *de novo* amino acid sequences. For MALDI-TOF-TOF MS, proteolytic peptides were mixed with saturated α-cyano-4-hydroxycinnamic acid in 50% acetonitrile containing 0.1% TFA, spotted, and analyzed in positive reflector mode (1625 shots, laser intensity 3000). Up to ten precursor ions from each MS spectrum were selected for automated collision-induced dissociation (CID) at 2 kV (500 shots, laser intensity 3000). Amino acid sequences were deduced with the aid of ProteinPilot v.4.0.8 (ABSciex) or interpreted manually. For nano-ESI-MS/MS, peptides were analyzed by direct infusion in a QTrap 3200 (Applied Biosystems) mass spectrometer. Selected doubly- or triply-charged ions from spectra obtained in Enhanced Resolution mode (250 amu/s) were subjected to fragmentation using the Enhanced Product Ion option with Q_0 trapping. Settings were: Q_1, unit resolution; collision energy, 25–40 eV; linear ion trap Q_3 fill time, 250 ms; and Q_3 scan rate, 1000 amu/s (Calvete et al., 2007). CID spectra were interpreted manually to derive *de novo* amino acid sequences.

2.4. Phylogenetic relationships

The phylogenetic relationships of BlatPLA$_2$ with 26 snake venom PLA_2s showing highest similarity in a BLAST search (http://blast.ncbi.nlm.nih.gov) were inferred using the

neighbor-joining method (Saitou and Nei, 1987) implemented in MEGA5 (Tamura et al., 2011), after a multiple sequence alignment performed with MUSCLE. The bootstrap consensus tree was inferred from 1000 replicates, and evolutionary distances were computed using the p-distance method (Felsenstein, 1985; Nei and Kumar, 2000).

2.5. Phospholipase A_2 activity

Enzymatic activity of BlatPLA$_2$ was determined by the phenol red colorimetric method described by (de Araújo and Radvanyi, 1987). Twenty microliters of various enzyme solutions in water, containing 2000, 1000, 500, 250, or 125 ng, were added to 1 ml of the micellar substrate (0.25% w/v *sn*-3-phosphatidylcholine, 0.4% v/v Triton X-100, 0.004% w/v phenol red) in a thermoregulated cuvette at 30 °C. After a stabilization period of 20 s, the decrease in absorbance at 558 nm was monitored continuously for 1 min. One unit of PLA$_2$ activity was defined as the change of 0.001 in absorbance per min. Assays were performed in triplicate.

2.6. Lethal activity

A group of five CD-1 mice (16–18 g body weight) received an intravenous injection of 100 µg of BlatPLA$_2$ (equivalent to 5.9 µg/g), dissolved in 100 µl of phosphate-buffered saline (0.12 M NaCl, 0.04 M sodium phosphate, pH 7.2; PBS). Three control mice received the same volume of PBS alone. Animals were observed to record deaths for up to 48 h. All procedures were approved by the Institutional Committee for the Care and Use of Laboratory Animals (CICUA) of Universidad de Costa Rica (#82-08).

2.7. Intracerebroventricular lethality

A group of three CD-1 mice (16–18 g) received an intracerebroventricular injection of 10 µg of BlatPLA$_2$, dissolved in 5 µl of PBS. A negative control group received 5 µl of PBS alone, whereas a positive control group received 5 µg of myotoxin III, a basic Asp49 PLA$_2$ isolated from *Bothrops asper* venom (Kaiser et al., 1990). Animals were observed for up to 48 h to record deaths.

2.8. Myotoxic activity

A group of five CD-1 mice (18–20 g) received an intramuscular injection of 50 µg of BlatPLA$_2$, dissolved in 50 µl of PBS, in their right gastrocnemius. A control group of mice received an identical injection of PBS alone. For comparison, 50 µg of crude *B. lateralis* venom was injected in another group. After 3 h, a blood sample was collected from the tail of all animals into heparinized capillaries, centrifuged, and plasma aliquots of 4 µl were utilized to determine the activity of creatine kinase (CK; E.C. 2.7.3.2) using a kinetic assay (CK-Nac, Biocon Diagnostik) (Gutiérrez et al., 1986). Enzyme activity was expressed in U/mL.

2.9. Edema-forming activity

Edema was determined in a group of four CD-1 mice (18–20 g) by measuring footpad thickness with a low-pressure spring caliper (Oditest) before and at various intervals (30, 60, 120, 180, 240, 300, and 360 min) after the subcutaneous intraplantar injection of 10 µg of BlatPLA$_2$, dissolved in 50 µl of PBS (Lomonte et al., 1993). A control group of mice received an identical injection of PBS alone. Edema was expressed as the percentage increase in footpad thickness relative to the pre-injection measurement.

2.10. Anticoagulant activity

Different amounts of BlatPLA$_2$ (5, 10, or 20 µg), dissolved in 50 µl of PBS were added to 200 µl of citrated human plasma (pre-warmed at 37 °C) and incubated for 10 min at 37 °C. As a control, 50 µl of PBS alone was added to 200 µl of plasma. Then, 50 µl of 0.25 M $CaCl_2$ was added to each tube and the clotting time was recorded (Gutiérrez et al., 1986). Assays were performed in triplicate.

2.11. Effects on platelet aggregation

Platelet-rich plasma (PRP) was prepared from citrated blood of healthy donors, by centrifugation at 60 g for 15 min. Different amounts of BlatPLA$_2$ (2.5, 5, and 10 µg), in 25 µl of PBS, were added to 225 µl of PRP, and incubated at 37 °C for 5 min. Then, platelet aggregation was induced by the addition of 25 µl of ADP (200 µM final concentration; Rucavado et al., 2001), and continuously monitored by measurement of percentage transmitted light in an AggRAM analyzer (Helena Laboratories) at a constant spin rate of 600 rpm. Experiments were performed in duplicates.

2.12. Immunological cross-recognition

The antigenic cross-reactivity between BlatPLA$_2$ and a previously isolated acidic enzyme from *B. asper* venom, BaspPLA$_2$-II (Fernández et al., 2010a), was tested by ELISA. Proteins were adsorbed onto 96-well microplates (Nunc Maxisorp) at 0.2 µg/100 µl/well, by overnight incubation in 0.1 M Tris, 0.15 M NaCl, pH 9.0 buffer. After washing with PBS, free sites were blocked with PBS containing 1% bovine serum albumin (BSA) for 1 h. Then, serial dilutions of rabbit serum against BaspPLA$_2$-II (Fernández et al., 2010a), or normal rabbit serum as a negative control, were added and incubated for 1 h. Plates were then washed with 0.05 M Tris, 0.15 M NaCl, 20 µM $ZnCl_2$, 1 mM $MgCl_2$ (pH 7.4), and bound antibodies were detected by incubation with an anti-rabbit IgG-alkaline phosphatase conjugate (1:5000) for 1 h, followed by washing and color development with *p*-nitrophenylphosphate (1 mg/mg) in diethanolamine buffer (pH 9.8). Absorbances were recorded on a microplate reader (Labsystems) at 405 nm. Assays were performed in triplicate wells.

2.13. Statistical analyses

Results are presented as mean ± SD. Comparisons between means of two groups were performed with the Student's *t*-test, where *p* values lower than 0.05 were considered significant.

3. Results

The RP-HPLC method originally used for the proteomic profiling of *B. lateralis* venom was optimized to obtain a prominent peak, eluting at ~28.6 min (Fig. 1A), which contained a protein with PLA_2 activity. Collection of this peak was followed by the bulk elution of the remaining venom components and column regeneration. The homogeneity of the protein obtained, here named BlatPLA_2, was evaluated by analytical RP-HPLC (Fig. 1B) and SDS-PAGE (Fig. 1C), which in both cases showed a single component. By the latter technique, BlatPLA_2 migrated at ~15 kDa in either reducing or non-reducing conditions, evidencing its monomeric nature and lack of tendency to form aggregates. Further indication of the homogeneity of this protein was provided by MALDI-TOF mass spectrometry analysis, which showed sharp peaks at m/z 6959.4 [$(M + 2H)^{2+}$] and 13,917.6 [$(M + H)^{+}$] (Fig. 2A). The PLA_2 activity of the isolated protein upon micellar phosphatidylcholine/TX-100 substrate was preserved, as shown in Fig. 2A.

BlatPLA_2 was subjected to N-terminal amino acid sequencing, which yielded its first 24 residues, and to proteolytic digestions. Analysis of the resulting peptides by tandem mass spectrometry allowed completion of its amino acid sequence (Fig. 3). The protein sequence data reported in this paper will appear in the UniProt Knowledgebase under the accession number C0HJC1. The calculated isotope-averaged molecular mass (M_{av}) for this sequence, including seven disulphide bonds, is 13,877.6 Da, departing from the experimental value of 13,916.7 ± 0.4 Da. Since the full-length sequence of BlatPLA_2 shown in Fig. 3 was deduced by combination of Edman degradation and *de novo* sequencing of overlapping peptide ions, whose monoisotopic masses exactly matched the deduced amino acid sequence, the mass difference of 39.1 Da may indicate that the MALDI-TOF m/z values corresponded to a potassium adduct of the native protein ion (calculated $(M_{av} + H + K)^{+}$ = 13,917.7 Da).

As expected for a group II PLA_2, BlatPLA_2 presents 14 Cys residues which adhere to the strictly conserved pattern that defines this group. Results of similarity search and multiple alignment are presented in Fig. 4. The highest amino acid sequence identity obtained in this analysis was 78% with an acidic PLA_2 isolated from the venom of *Lachesis stenophrys* from Costa Rica (P84651; de Assis et al., 2008), followed by 69–76% with a group of viperid PLA_2s. The acidic nature of BlatPLA_2 is supported by its predicted pI value of 4.7 (Fig. 4). However, interestingly, snake venom PLA_2s showing high identity values to BlatPLA_2 include not only acidic enzymes but also several basic ones, with theoretical pI values between 8.2 and 8.8 (Fig. 4). The multiple sequence alignment of the newly described

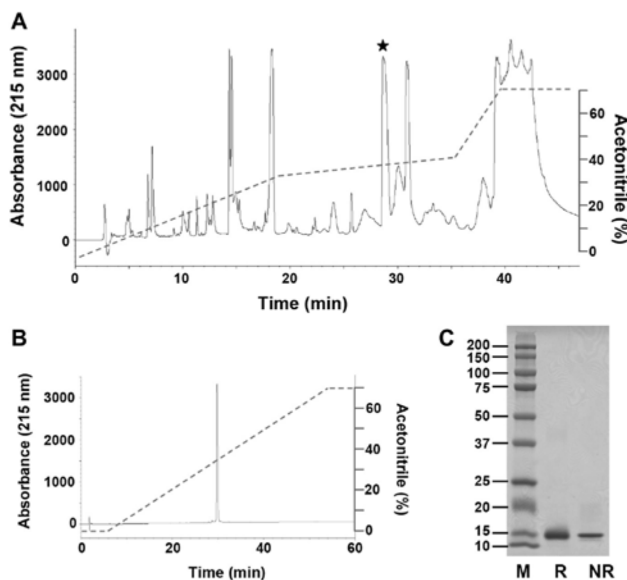


Fig. 1. **(A)** Reverse-phase HPLC of *Bothriechis lateralis* venom (2.6 mg) on a C18 column (250 × 4.6 mm) eluted at 1 ml/min with a gradient from water to acetonitrile, containing 0.1% trifluoroacetic acid, as follows: 0–33% in 19 min, 33–40% in 26 min, and 40–70% in 5 min. BlatPLA_2 was collected in the peak eluting at ~28.6 min, indicated with the symbol "★". **(B)** Analytical RP-HPLC of the purified BlatPLA_2 in a C4 column (150 × 4.6 mm) eluted at 1 ml/min with a linear gradient from water to 70% acetonitrile, containing 0.1% trifluoroacetic acid. **(C)** SDS-PAGE analysis of BlatPLA_2 under reducing (R) or non-reducing (NR) conditions. Molecular weight standards are shown in the left lane (M) with values indicated in kDa.

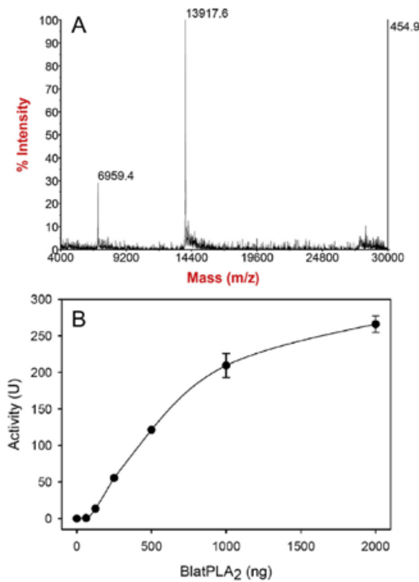


Fig. 2. **(A)** MALDI-TOF mass spectrum of BlatPLA$_2$, analyzed in positive linear mode in the presence of sinapinic acid as matrix. The spectrum was acquired in an Applied Biosystems 4800 Plus instrument, in the *m/z* range 4000–30,000, using 1500 laser shots at an intensity of 4000. The left and right ordinates indicate relative and absolute signal intensities, respectively. **(B)** Phospholipase A$_2$ activity of BlatPLA$_2$ on phosphatidylcholine-Triton X-100 micelles, as determined by the phenol red-based method of de Araújo and Radvanyi (1987). One unit of PLA$_2$ activity was defined as the change of 0.001 in the absorbance at 558 nm per min. Points represent mean ± SD of triplicates.

enzyme with similar proteins shows that 48 out of its 122 amino acid positions are absolutely conserved (shaded gray in Fig. 4). Overall, these comparisons indicate that BlatPLA$_2$ presents significant differences from previously described group II PLA$_2$s from snake venoms, as supported by the maximum identity value of 78%. The phylogenetic position of BlatPLA$_2$ within the group of 27 proteins aligned in Fig. 4 was inferred by the neighbor-joining tree presented in Fig. 5. The tree clearly discerned the acidic from the basic enzymes, and placed BlatPLA$_2$ in a relatively basal position within the former group, as a sister branch to all of the acidic PLA$_2$s analyzed, except for two enzymes from *Agkistrodon halys* venom (O42190 and O42189) which formed an independent clade (Fig. 5).

An evaluation of *in vivo* and *in vitro* activities of BlatPLA$_2$ was performed. The intravenous injection of 100 µg of this enzyme did induce neither lethality, nor caused any evident alterations in mice. Similarly, the protein was devoid of lethal effect when injected by the intracerebroventricular route at a dose of 10 µg, whereas the outcome in a control group of mice receiving 5 µg of a basic PLA$_2$ instead (*B. asper* myotoxin III) was of 100% lethality. Also, BlatPLA$_2$ was devoid of myotoxic activity when administered by the intramuscular route (Fig. 6), since a dose of 50 µg induced a negligible and statistically non-significant ($p > 0.05$) increase in plasma creatine kinase levels in comparison to control animals that received vehicle alone. In the mouse footpad assay, subcutaneous injection of 10 µg of the enzyme caused a mild, although statistically significant edema of rapid onset, which vanished by 5–6 h (Fig. 7). *In vitro* assays showed that BlatPLA$_2$ was devoid of anticoagulant effect on human plasma (data not shown), and did neither promote human platelet aggregation, nor inhibited ADP-induced platelet aggregation (Fig. 8).

Since the profile of bioactivities of BlatPLA$_2$ closely resembled that of a previously isolated acidic enzyme of *B. asper*, BaspPLA$_2$-II (Fernández et al., 2010a), the immunological cross-recognition of the former enzyme by rabbit

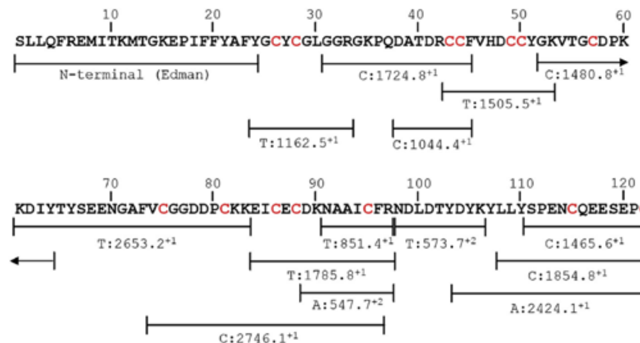


Fig. 3. Amino acid sequence of BlatPLA$_2$. Overlapping peptides, obtained after reduction of disulphide bonds and alkylation of cysteines followed by digestion with proteases (T: trypsin; C: chymotrypsin; A: Asp-N), were *de novo* sequenced by MALDI-TOF/TOF MS or nano-ESI-MS/MS, as described in Materials and methods. The first 24 amino acids were determined by N-terminal Edman degradation. Cysteines are indicated in red (for color descriptions the reader is referred to the web version of the figure).

```
BlatPLA2|B.lateralis          SLLQFREMITKMTGKEPIFFYAFYGCYCGLGGRGKPQDATDRCCFVHDCCYGKVTGCDPK
P84651| L.stenophrys          HLLQFGDLIDKIAGRSGFWYYGFYGCYCGLGGRGRPQDATDRCCFVHDCCYGKVTGCDPK
O42190| A.halys               NLLQFEKMIKKMTGKEPVVSYAFYGCYCGSGGQGKPKDATDRCCFVHDCCYGKVTGCDPK
A8E2V4| B.schlegeli           DLLQFEGMIMTIAGRSGIWYYGSYGCYCGAGGQGKPQDASDRCCFVHDCCYGKVTECNPK
O42189| A.halys               SLIQFEKMIKKMTGKEPVVSYAFYGCYCGSGGQGKPKDATDRCCFVHDCCYEKVTGCDPK
O42187| A.halys               SLLQFRKMIKKMTGKEPVVSYAFYGCYCGSGGRGKPKDATDRCCFVHDCCYEKLTGCDPK
K7TC65| A.halys               HLLQFRKMIKKMTGKEPVVSYAFYGCYCGSGGRGKPKDATDRCCFVHDCCYEKVTGCDPK
P04417| A.halys blomhoffi     HLLQFRKMIKKMTGKEPVISYAFYGCYCGSGGRGKPKDATDRCCFVHDCCYEKVTGCKPK
Q02517| P.flavoviridis        HLLQFRKMIKKMTGKEPIVSYAFYGCYCGKGGRGKPKDATDRCCFVHDCCYEKVTGCDPK
Q8JIG0| P.flavoviridis        HLLQFRKMIKKMTGKEPIVSYAFYGCYCGKGGRGKPKDATDRCCFVHDCCYGKVTGCDPK
Q6H3D0| V.stejnegeri          HLMQFETLIMKVAGRSGVWYYGSYGCFCGAGGQGRPQDASDRCCFVHDCCYGKVNGCDPK
P20249| A.halys blomhoffi     SLMQFETLIMKIAGRSGIWYYGSYGCYCGAGGQGRPQDASDRCCFVHDCCYGKVTGCDPK
Q6T5K9| A.shedaoensis         SLVQFETLIMKIAGRSGIWYYGSYGCYCGAGGQGRPQDASDRCCFVHDCCYGKVTGCDPK
O42192| A.halys               SLMQFKTLIMKIAGRSGIWYYGSYGCYCGGGGQGRPQDASDRCCFVHDCCYGKVTGCNPK
P81480| T.gramineus           NLMQFETLIMKVAGRSGVWYYGSYGCFCGAGGQGRPQDASDRCCFVHDCCYGKVNGCDPK
P81479| T.gramineus           HLMQFETLIMKVAGRSGVWYYGSYGCFCGAGGQGRPQDASDRCCFVHDCCYGKVNGCDPK
O42188| A.halys               NLIQFKKMIKKMTGKEPVVSYAFYGCYCGSGGRGKPKDATDRCCFVHNCCYEKVTGCDPK
I2DAL5| B.diporus             NLVQFETLIMKIAGRSGVWYYGSYGCYCGSGGQGRPQDASDRCCFVHDCCYGKVTGCNPK
C9DPL5| B.pirajai             NLWQFGKLIMKIAGESGVFKYLSYGCYCGLGGQGQPTDATDRCCFVHDCCYGKVTGCDPK
A8E2V8| T.gracilis            SLMQFEMLIMKLAKSSGMFWYSAYGCYCGWGGQGRPQDATDRCCFVHDCCYGKATGCDPK
B0LSG7| S.miliarius           HLIQFETLIMKIAGRSGVFWYSAYGCYCGWGGYGRPQDATDRCCFVHDCCYGKVTGCNPK
Q2HZ28| B.erythromelas        SLVQFETLIMKIAGRSGVWYYGSYGCYCGSGGQGRPQDASDRCCFVHDCCYGKVTDCDPK
Q9I8F8| B.pictus              SLVQFETLIMKIAKRSGVWFYGSYGCFCGSGGQGRPQDASDRCCFVHDCCYGKVTDCDPK
A8E2V5| C.godmani             NLMQFERLIMKVAKRSGMFWYSAYGCYCGWGGQGKPQDPTDRCCFVHDCCYGKVTDCDPK
P86389| B.asper               NLWQFGQMMSDVMRKNVVFKYLSYGCYCGWGGIGQPKDATDRCCFVHDCCYGKVTGCDPK
B0LSG0| S.catenatus           HLIQFETLIMKIAGRSGVFWYSAYGCYCGWGGYGRPQDPTDRCCFVHDCCYGKVTGCNPK
P81243| B.jararaca            DLWQFGQMMNDVMREYVVFNYLYYGCYCGWGGIGKPRDATDRCCFVHDCCYGKVTGCNPK

                                                                                Id%   n   pI
BlatPLA2  KD--IYTYSEENGAFVCGGDDPCKKEICECDKNAAICFRDNLDTYDYK-YLLYSPENCQ-EESEPC  --  122 4.7
P84651|   KD--IYTYSEENGAIVCGGDNPCKKEICECDRDAAICFRDNLDTYDNK-YWLFPNKYCK-EESEPC  78% 122 5.0
O42190|   MD--VYSFSEENGDIVCGGDDPCKKEICECDRAAAICFRDNLNTYNDKKYWAFGAKNCPQEESEAC  76% 124 5.0
A8E2V4|   KD--SYTYTEENGAVVCGGDDPCMKEVCECDKNAAICFRDNLDTYDNK-YWFYPAKYCQ-EESEPC  75% 122 4.4
O42189|   WD--DYTYSWKDGDIVCGGDDPCKKEICECDRAAAICFRDNLNTYNDRKYWAFGAKNCPQEESEPC  75% 124 5.0
O42187|   WD--DYTYSWKNGTIVCGGDDPCKKEVCECDKAAAICFRDNLKTYKKR-YMTYPNILCS-SKSEKC  74% 122 8.7
K7TC65|   WD--DYTYSWKDGDIVCGGDDPCKKEVCECDKAAAICFRDNLKTYKKR-YMAYPDILCS-SKSEKC  74% 122 8.3
P04417|   WD--DYTYSWKNGDIVCGGDDPCKKEICECDRAAAICFRDNLKTYKKR-YMAYPDILCS-SKSEKC  74% 122 8.7
Q02517|   WD--YYTYSSENGDIVCGGDNPCTKEVCECDKAAAICFRDNLKTYKKR-YMTFPDIFCT-DPTEKC  73% 122 8.2
Q8JIG0|   WD--YYTYSSENGDIVCEGDNPCTKEVCECDKAAAICFRDNLKTYKKR-YMTFPDIFCT-DPTEKC  73% 122 8.2
Q6H3D0|   KD--FYTYSEENGAIVCGGDDPCKKEICECDKDAAICFRDNKDTYDNK-YWFFPAKNCQ-EESEPC  73% 122 4.8
P20249|   LD--VYTYTEENGAIVCGGDDPCKKQICECDKDAAICFRDNIDTYDNK-YWFFPAKNCQ-EESEPC  73% 122 4.6
Q6T5K9|   MD--VYTYTEENGAIVCGGDDPCKKQICECDKDAAICFRDNIDTYDNK-YWFFPAKNCQ-EESEPC  73% 122 4.6
O42192|   MD--VYTYTEENGAIVCGGDDPCKKQICECDKDAAICFRDNIDTYDNK-YWFFPAKNCQ-EESEPC  72% 122 4.9
P81480|   KD--FYTYSEENGAIVCGGDDPCKKEICECDKDAAICFRDNKDTYDNK-YWFFPAKNCQ-EESEPC  72% 122 4.7
P81479|   KD--FYTYSEENGDIVCGGDDPCKKEICECDKDAAICFRDNKDTYDNK-YWFFPAKNCQ-EESEPC  72% 122 4.7
O42188|   WD--DYTYSWKNGTIVCGGDDPCKKEVCECDKAAAICFRDNLKTYKKR-YMTYPNILCS-SKSEKC  72% 122 8.8
I2DAL5|   AD--TYTYSEENGVVVCGGDDPCKKQICECDRVAATCFRDNKDTYDNK-YWFFPAKNCQ-EESEPC  72% 122 4.9
C9DPL5|   ID--SYTYSKENGDVVCGGDDPCKKQICECDRVAATCFRDNKDTYDIK-YWFYGAKNCQ-EESEPC  72% 122 4.9
A8E2V8|   KD--VYTYSEENGDIVCGGDDPCRKEVCECDKAAAICFRDNMDTYNSKTYWMFPAKNCQ-EESEPC  71% 123 4.7
B0LSG7|   KD--FYSYTEENGEIVCGGDDPCKKQICECDRVAAICFRDNIPTYDDK-YWRFPTENCQ-EEPEPC  71% 122 4.8
Q2HZ28|   AD--VYTYSEENGVVVCGGDDPCKKQICECDRVAATCFRDNKDTYDNK-YWFFPAKNCQ-EESEPC  70% 122 4.7
Q9I8F8|   TD--IYTYSEENGVVVCGGDDPCKKQICECDRVAAVCFRDNKDTYDNK-YWFFPANNCQ-EESEPC  70% 122 4.7
A8E2V5|   LD--VYTYNEENGEVVCGGDDPCKKQICECDKAAAICFRDNKDTYDWKTYWRFPPEKCR-EESEPC  70% 123 5.0
P86389|   MDIYTYTYSKENGDVVCGGDDPCKKQICECDRVAAICFRDNKDTYDSK-YWFYGAKNCQ-EDSEPC  70% 124 5.0
B0LSG0|   KD--FYSYSEENGEIVCGGEDPCKNQICECDRVAAICFRDNIPTYDDK-YWRFPTENCQ-EEPEPC  69% 122 4.7
P81243|   TDSYTYTYSEENGDVVCGGDDLCKKQICECDRVAATCFRDNKDTYDTK-YWLYGAKNCQ-EESEPC  69% 124 4.6
```

Fig. 4. Multiple alignment of the amino acid sequence of BlatPLA$_2$ with the most similar proteins after a BLAST search and processing with MUSCLE in MEGA5, as described in Materials and methods. Cysteines are highlighted in bold, and identical residues among the 27 proteins are shaded in gray. Accession codes in the UniProt database are indicated at the left of each sequence. At the right, percentage of identity (Id%), total number of amino acids (n), and predicted isoelectric points (pI) are indicated. The theoretical pI values were obtained using the "Compute pI/MW" tool at ExPASy (http://http://web.expasy.org/compute_pi/).

antibodies generated against the latter was evaluated. ELISA results (Fig. 9) evidenced significant cross-reactivity between the two antigens when confronted to these antibodies, where the immunorecognition of BlatPLA$_2$ corresponded2 to nearly 65% of the recognition of the homologous antigen.

4. Discussion

The relatively small size (~15 kDa; ~120 amino acids) and highly conserved three-dimensional fold of snake venom PLA$_2$s is in contrast with their wide range of bioactivities and pharmacological effects (Kini, 2003). Subtle

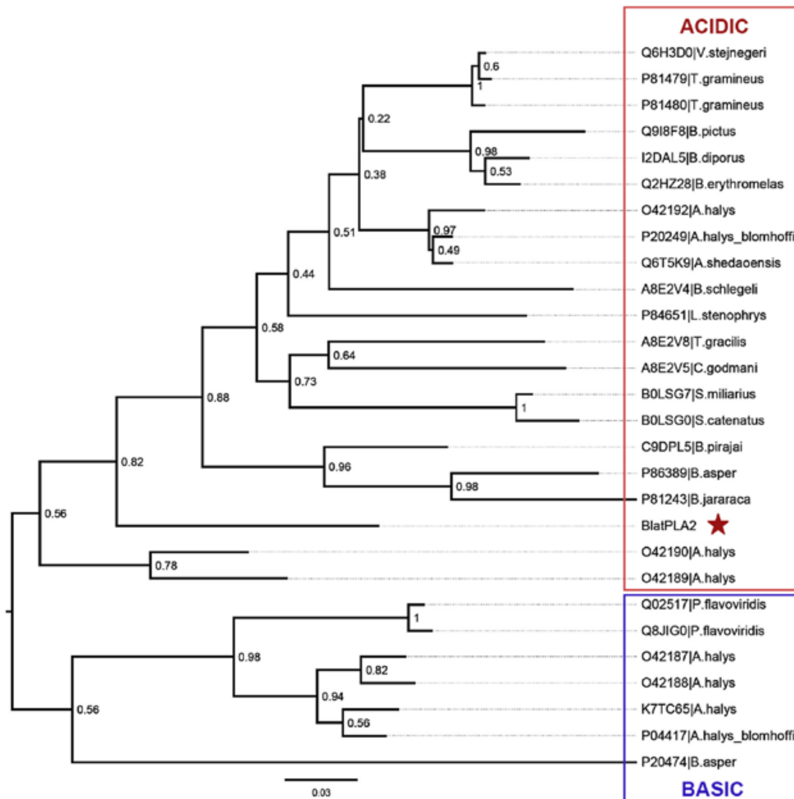


Fig. 5. Phylogenetic relationships of the PLA_2 proteins aligned in Fig. 4, inferred by the neighbor-joining algorithm implemented in MEGA5, as described in Materials and methods. The bootstrap consensus tree was inferred from 1000 replicates, and evolutionary distances were computed using the p-distance method. The position of $BlatPLA_2$ is marked with the symbol "★". Acidic and basic PLA_2s are enclosed in red and blue boxes, respectively (for color descriptions the reader is referred to the web version of the figure).

changes in amino acids located in exposed regions of these interfacial enzymes, seemingly played a significant role in their evolution of diverse toxic actions (Ogawa et al., 1995; Kini and Chan, 1999). However, the molecular basis of toxicity of snake venom PLA_2s is only partially understood. In group II PLA_2s, a correlation between basic character and high toxicity is generally observed. Acidic enzymes, on the other hand, tend to be of lower, if any, toxicity, although a series of pharmacological actions such as *in vitro* anticoagulant activity, platelet aggregation or its inhibition, and a mild hypotensive action *in vivo* or moderate myotoxicity have been reported (Serrano et al., 1999; Fuly et al., 2002; Andrião-Escarso et al., 2002; Ketelhut et al., 2003; Roberto et al., 2004; Cogo et al., 2006; de Albuquerque Modesto et al., 2006; Rodrigues et al., 2007; Santos-Filho et al., 2008). In some cases, acidic group II PLA_2 enzymes appear to be devoid of toxicity or pharmacological actions (de Araújo et al., 1994; Daniele et al., 1995; Fernández et al., 2010a; Garcia Denegri et al., 2010; Silveira et al., 2013), their presence in snake venoms being therefore somewhat puzzling. It has been speculatively proposed that the role of these apparently non-toxic enzymes could be limited to contributing in the pre-digestion of prey, rather than in its immobilization or killing (Fernández et al., 2010a). On the basis of its lack of lethal, myotoxic, anticoagulant, platelet aggregating or anti-aggregating activities, the new enzyme here characterized, $BlatPLA_2$, would fit into this latter group of non-toxic, acidic PLA_2s. Only a minor, transient edema of

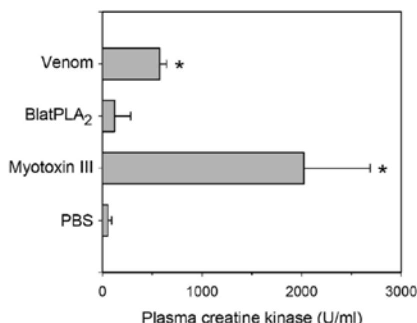


Fig. 6. Lack of myotoxic activity of BlatPLA$_2$. Mice received an intramuscular injection of BlatPLA$_2$ (50 μg) in their right gastrocnemius, and plasma creatine kinase levels were determined after 3 h, as described in Materials and methods. The effects of similar injections containing, respectively, crude *B. lateralis* venom (50 μg) or purified *B. asper* myotoxin III (50 μg), a basic PLA$_2$, or vehicle alone (PBS), are shown for comparison. Bars represent mean ± SD of five mice per group. Asterisks indicate a statistically significant ($p < 0.05$) difference of values in comparison to the PBS control group.

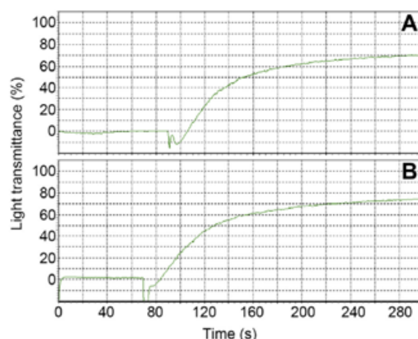


Fig. 8. Lack of effects of BlatPLA$_2$ on human platelets *in vitro*. BlatPLA$_2$ **(A)** or PBS vehicle alone **(B)** were added to platelet-rich plasma obtained from healthy donors, as described in Materials and methods. After 80–100 s, when the lack of platelet aggregation was confirmed, aggregation was induced by the addition of ADP (200 μM) and monitored by the change in light transmittance in an AggRAM analyzer. Experiments were performed in duplicates, and representative traces are shown in **(A)** and **(B)**.

rapid onset was recorded for this enzyme, in common with many PLA$_2$s of diverse origins (Cirino et al., 1989; Texeira et al., 2003). Antibodies to BaspPLA$_2$-II, another non-toxic enzyme from *B. asper* venom (Fernández et al., 2010a), significantly cross-recognized BlatPLA$_2$ by ELISA. Previous studies have shown that acidic and basic PLA$_2$s from *B. asper* form two antigenically distinct groups, which do not cross-react (Fernández et al., 2010a). The present ELISA results are in agreement with the acidic nature of both proteins, which share 70% sequence identity, as well as with their resemblance in lacking major toxic activities.

BlatPLA$_2$ represents nearly 9% of the venom proteome of *B. lateralis*, which, interestingly, does not contain any basic PLA$_2$, often acting as myotoxins in viperids. The moderate myotoxic action of this venom (Fig. 6) can be attributed to its high content of hemorrhagic metalloproteinases (Lomonte et al., 2008; Durban et al., 2011), which may indirectly damage muscle tissue by promoting ischemia (Gutiérrez et al., 2009). Nevertheless, a significant metabolic cost of protein synthesis should be considered when rationalizing the presence of a highly abundant non-toxic

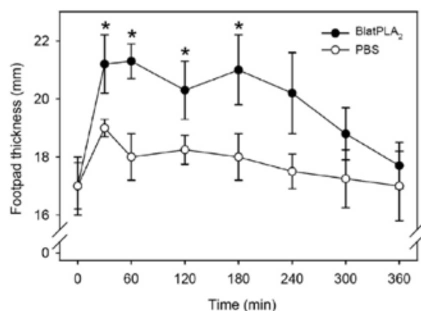


Fig. 7. Edema-inducing activity of BlatPLA$_2$. Mice received a subcutaneous, intraplantar injection containing 10 μg of BlatPLA$_2$ (●) or PBS vehicle alone (○), and their footpad thickness was measured at the indicated time intervals, as described in Materials and methods. Points represent mean ± SD of four mice per group. Asterisks indicate a statistically significant ($p < 0.05$) difference of values in comparison to the PBS control group.

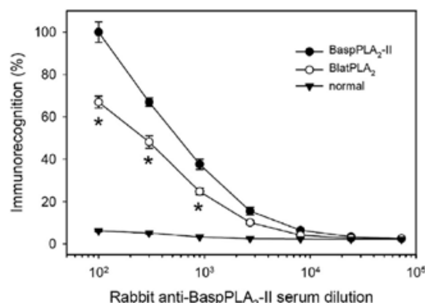


Fig. 9. Immunological cross-reactivity between BlatPLA$_2$ and BaspPLA$_2$-II, evaluated by ELISA. BaspPLA$_2$-II is an acidic, non-toxic enzyme that was previously isolated from *B. asper* venom (Fernández et al., 2010a). Both proteins were adsorbed onto 96-well microplates and the binding of anti-BaspPLA$_2$-II rabbit antibodies to each enzyme was detected as described in Materials and methods. Normal rabbit serum (◆) was included as a negative control. Absorbance values for BaspPLA$_2$-II (●) and BlatPLA$_2$ (○) were normalized by considering the immunorecognition of the homologous antigen as 100%. Points represent mean ± SD of triplicate wells. Asterisks indicate a statistically significant ($p < 0.05$) difference of values in comparison to the normal rabbit serum controls.

component such as BlatPLA$_2$ in the venom (McCue, 2006; Morgenstern and King, 2013). One possibility would be that this enzyme has a hitherto unknown toxic effect on a very specific type of prey, relevant in the diet of *B. lateralis*. However, rodents are known to constitute the main component of its diet, at least in adults (Solórzano, 2004), and still, injection of BlatPLA$_2$ in mice did not evidence toxicity. A second possibility to consider relates to the possible synergisms that acidic PLA$_2$s may exert in combination with other venom components, as recently described by Bustillo et al. (2012). These authors demonstrated that an acidic PLA$_2$ from *Bothrops alternatus* which *per se* had no effects upon cultured myogenic cells, enhanced the detaching action exerted by a purified hemorrhagic metalloproteinase from the same venom (Garcia Denegri et al., 2010; Gay et al., 2005; Bustillo et al., 2012). A further possibility would be that enzymes with a general digestive action, but devoid of major toxic effects such as BlatPLA$_2$, may represent persisting remnants of the evolutionary process that transformed such proteins into potent toxins. We hypothesize that the catalytically active BlatPLA$_2$ could have a role in the pre-digestion of prey, possibly having retained characteristics of ancestral PLA$_2$s without evolving towards potent toxicity.

The amino acid sequence of BlatPLA$_2$ presents similarities with a number of acidic, and a few basic, group II PLA$_2$s, although identity values obtained in pairwise alignments are not higher than 78%. A reconstruction of phylogenetic relationships among these proteins suggests that BlatPLA$_2$ occupies a basal position to a group of acidic enzymes from Old- and New World viperid snakes, representing a sister branch to them. This observation would be in agreement with the possibility that BlatPLA$_2$ might represent a remnant of ancestral, non-toxic acidic enzymes with a digestive role, as discussed above. As more amino acid sequences of group II PLA$_2$ become available, together with information on their toxic/pharmacological actions, future comparative structural analyses or mutagenesis experiments may provide key clues to reveal the topology of the sites that define functional properties of these enzymes. BlatPLA$_2$ might prove to be a useful tool for such purposes.

Acknowledgments

The authors thank the staff of the serpentarium at Instituto Clodomiro Picado and Dr. Alexandra Rucavado for contributing valuable materials for this study, as well as Dr José María Gutiérrez for fruitful discussions. The financial support from Vicerrectoría de Investigación, UCR (741-A9-513), ICGEB (CRP Program COS-08-03), grant BFU2010-17373 from the Ministerios de Ciencia e Innovación, and Economía y Competitividad (Madrid, Spain), PROMETEO/2010/005 from the Generalitat Valenciana (Valencia, Spain), CRUSA-CSIC (2009CR0021), and CYTED (project BIOTOX P211RT0412) are gratefully acknowledged.

Conflict of interest statement

The authors declare that no competing interests exist regarding this manuscript.

References

Andrião-Escarso, S.H., Soares, A.M., Fontes, M.R., Fuly, A.L., Corrêa, F.M., Rosa, J.C., Greene, L.J., Giglio, J.R., 2002. Structural and functional characterization of an acidic platelet aggregation inhibitor and hypotensive phospholipase A$_2$ from *Bothrops jararacussu* snake venom. Biochem. Pharmacol. 64, 723–732.

Bustillo, S., Gay, C.C., García Denegri, M.E., Ponce-Soto, L.A., Bal de Kier Joffe, E., Acosta, O., Leiva, L.C., 2012. Synergism between baltergin metalloproteinase and Ba SPII RP4 PLA$_2$ from *Bothrops alternatus* venom on skeletal muscle (C2C12) cells. Toxicon 59, 338–343.

Calvete, J.J., Escolano, J., 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.

Casewell, N.R., Wüster, W., Vonk, F.J., Harrison, R.A., Fry, B.G., 2013. Complex cocktails: the evolutionary novelty of venoms. Trends Ecol. Evol. 28, 219–229.

Cirino, G., Peers, S.H., Wallace, J.L., Flower, R.J., 1989. A study of phospholipase A$_2$-induced edema in rat paw. Eur. J. Pharmacol. 166, 505–510.

Cogo, J.C., Lilla, S., Souza, G.H., Hyslop, S., de Nucci, G., 2006. Purification, sequencing and structural analysis of two acidic phospholipases A$_2$ from the venom of *Bothrops insularis* (jararaca ilhoa). Biochimie 12, 1947–1959.

Daniele, J.J., Bianco, I.D., Fidelio, G.D., 1995. Kinetic and pharmacologic characterization of phospholipases A$_2$ from *Bothrops neuwiedii* venom. Arch. Biochem. Biophys. 318, 65–70.

de Albuquerque Modesto, J.C., Spencer, P.J., Fritzen, M., Valença, R.C., Oliva, M.L., da Silva, M.B., Chudzinski-Tavassi, A.M., Guarnieri, M.C., 2006. BE-I-PLA$_2$, a novel acidic phospholipase A$_2$ from *Bothrops erythromelas* venom: isolation, cloning and characterization as potent anti-platelet and inductor of prostaglandin I$_2$ release by endothelial cells. Biochem. Pharmacol. 72, 377–384.

de Araújo, A.L., Radvanyi, F., 1987. Determination of phospholipase A$_2$ activity by a colorimetric assay using a pH indicator. Toxicon 25, 1181–1188.

de Araújo, A.L., Radvanyi, F., Bon, C., 1994. Purification of an acidic phospholipase A$_2$ from *Bothrops lanceolatus* (fer de lance) venom: molecular and enzymatic properties. Toxicon 32, 1069–1081.

de Assis, E.B., Estevao-Costa, M.I., do Carmo Valentim, A., Silva-Neto, A., Agostini Cotta, G., Mudado, M.A., Richardson, M., Fortes-Dias, C.L., 2008. Purification and complete primary structure of the first PLA$_2$ from *Lachesis stenophrys* (the Central American Bushmaster) snake venom. Protein J. 27, 327–333.

Durban, J., Juárez, P., Angulo, Y., Lomonte, B., Flores-Díaz, M., Alape-Girón, A., Sasa, M., Sanz, L., Gutiérrez, J.M., Dopazo, J., Conesa, A., Calvete, J.J., 2011. Profiling the venom gland transcriptomes of Costa Rican snakes by 454 pyrosequencing. BMC Genomics 12, 259.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Fernández, J., Gutiérrez, J.M., Angulo, Y., Sanz, L., Juárez, P., Calvete, J.J., Lomonte, B., 2010a. Isolation of an acidic phospholipase A$_2$ from the venom of the snake *Bothrops asper* of Costa Rica: biochemical and toxicological characterization. Biochimie 92, 273–283.

Fernández, J., Lomonte, B., Sanz, L., Angulo, Y., Gutiérrez, J.M., Calvete, J.J., 2010b. Snake venomics of *Bothriechis nigroviridis* reveals extreme variability among palm viper venoms: different evolutionary solutions for the same trophic purpose. J. Proteome Res. 9, 4234–4241.

Fry, B.G., Wüster, W., 2004. Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. Mol. Biol. Evol. 21, 870–883.

Fujisawa, D., Yamazaki, Y., Lomonte, B., Morita, T., 2008. Catalytically inactive phospholipase A$_2$ homologue binds to vascular endothelial growth factor receptor-2 via C-terminal loop region. Biochem. J. 411, 515–522.

Fuly, A.L., de Miranda, A.L., Zingali, R.B., Guimarães, J.A., 2002. Purification and characterization of a phospholipase A$_2$ isoenzyme isolated from *Lachesis muta* snake venom. Biochem. Pharmacol. 63, 1589–1597.

Garcia Denegri, M.E., Acosta, O.C., Huancahuire-Vega, S., Martins-de-Souza, D., Marangoni, S., Marunak, S.L., Teibler, G.P., Leiva, L.C., Ponce-Soto, L.A., 2010. Isolation and functional characterization of a new acidic PLA$_2$ Ba SpII RP4 of the *Bothrops alternatus* snake venom from Argentina. Toxicon 56, 64–74.

Gay, C.C., Leiva, L.C., Marunak, S., Teibler, P., Acosta de Perez, O., 2005. Proteolytic, edematogenic and myotoxic activities of a hemorrhagic metalloproteinase isolated from *Bothrops alternatus* venom. Toxicon 46, 546–554.

Gutiérrez, J.M., Lomonte, B., 2013. Phospholipases A$_2$: unveiling the secrets of a functionally versatile group of snake venom toxins. Toxicon 62, 27–39.

Gutiérrez, J.M., Lomonte, B., Chaves, F., Moreno, E., Cerdas, L., 1986. Pharmacological activities of a toxic phospholipase A isolated from the venom of the snake *Bothrops asper*. Comp. Biochem. Physiol. 84C, 159–164.

Gutiérrez, J.M., Rucavado, A., Chaves, F., Díaz, C., Escalante, T., 2009. Experimental pathology of local tissue damage induced by *Bothrops asper* venom. Toxicon 54, 958–975.

Kaiser, I.I., Gutiérrez, J.M., Plummer, D., Aird, S.D., Odell, G.V., 1990. The amino acid sequence of a myotoxic phospholipase from the venom of *Bothrops asper*. Arch. Biochem. Biophys. 278, 319–325.

Ketelhut, D.F., de Mello, M.H., Veronese, E.L., Esmeraldino, L.E., Murakami, M.T., Arni, R.K., Giglio, J.R., Cintra, A.C., Sampaio, S.V., 2003. Isolation, characterization and biological activity of acidic phospholipase A_2 isoforms from *Bothrops jararacussu* snake venom. Biochimie 10, 983–991.

Kini, R.M., 1997. Phospholipase A_2 – a complex multifunctional protein puzzle. In: Kini, R.M. (Ed.), Venom Phospholipase A_2 Enzymes: Structure, Function and Mechanism. John Wiley & Sons, England, pp. 1–28.

Kini, R.M., 2003. Excitement ahead: structure, function and mechanism of snake venom phospholipase A_2 enzymes. Toxicon 42, 827–840.

Kini, R.M., Chan, Y.M., 1999. Accelerated evolution and molecular surface of venom phospholipase A_2 enzymes. J. Mol. Evol. 48, 125–132.

Lambeau, G., Lazdunski, M., 1999. Receptors for a growing family of secreted phospholipases A_2. TIPS 20, 162–170.

Lomonte, B., Gutiérrez, J.M., 2011. Phospholipases A_2 from Viperidae snake venoms: how do they induce skeletal muscle damage? Acta Chim. Slov. 58, 647–659.

Lomonte, B., Rangel, J., 2012. Snake venom Lys49 myotoxins: from phospholipases A_2 to non-enzymatic membrane disruptors. Toxicon 60, 520–530.

Lomonte, B., Tarkowski, A., Hanson, L.Å., 1993. Host response to *Bothrops asper* snake venom: analysis of edema formation, inflammatory cells, and cytokine release in a mouse model. Inflammation 17, 93–105.

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

Lomonte, B., Tsai, W.C., Bonilla, F., Solórzano, A., Solano, G., Angulo, Y., Gutiérrez, J.M., Calvete, J.J., 2012. Snake venomics and toxicological profiling of the arboreal pitviper *Bothriechis supraciliaris* from Costa Rica. Toxicon 59, 592–599.

McCue, M.D., 2006. Cost of producing venom in three North American pitviper species. Copeia 2006, 818–825.

Montecucco, C., Gutiérrez, J.M., Lomonte, B., 2008. Cellular pathology induced by snake venom phospholipase A_2 myotoxins and neurotoxins: common aspects of their mechanisms of action. Cell. Mol. Life Sci. 65, 2897–2912.

Morgenstern, D., King, G.F., 2013. The venom optimization hypothesis revisited. Toxicon 63, 120–128.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Ogawa, T., Kitakima, M., Nakashima, K., Sakaki, Y., Ohno, M., 1995. Molecular evolution of group II phospholipases A_2. J. Mol. Evol. 41, 867–877.

Petan, T., Prijatelj, P., Pungerčar, J., 2013. Protein engineering in structure-function studies of viper's venom secreted phospholipases A_2. In: Figurski, D. (Ed.), Genetic Manipulation of DNA and Protein. Examples from Current Research. InTech, New York, pp. 107–132.

Roberto, P.G., Kashima, S., Marcussi, S., Pereira, J.O., Astolfi-Filho, S., Nomizo, A., Giglio, J.R., Fontes, M.R., Soares, A.M., França, S.C., 2004. Cloning and identification of a complete cDNA coding for a bactericidal and antitumoral acidic phospholipase A_2 from *Bothrops jararacussu* venom. Protein J. 23, 273–285.

Rodrigues, R.S., Izidoro, L.F., Teixeira, S.S., Silveira, L.B., Hamaguchi, A., Homsi-Brandeburgo, M.I., Selistre-de-Araújo, H.S., Giglio, J.R., Fuly, A.L., Soares, A.M., Rodrigues, V.M., 2007. Isolation and functional characterization of a new myotoxic acidic phospholipase A_2 from *Bothrops pauloensis* snake venom. Toxicon 50, 153–165.

Rosenberg, P., 1990. Phospholipases. In: Shier, W.T., Mebs, D. (Eds.), Handbook of Toxinology. Marcel Dekker, New York, pp. 67–277.

Rucavado, A., Soto, M., Kamiguti, A., Theakston, R.D.G., Fox, J.W., Escalante, T., Gutiérrez, J.M., 2001. Characterization of aspercetin, a platelet aggregating component from the venom of the snake *Bothrops asper* which induces thrombocytopenia and potentiates metalloproteinase-induced hemorrhage. Thromb. Hemost. 85, 710–715.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Santos-Filho, N.A., Silveira, L.B., Oliveira, C.Z., Bernardes, C.P., Menaldo, D.L., Fuly, A.L., Arantes, E.C., Sampaio, S.V., Mamede, C.C., Beletti, M.E., de Oliveira, F., Soares, A.M., 2008. A new acidic myotoxic, anti-platelet and prostaglandin I_2 inductor phospholipase A_2 isolated from *Bothrops moojeni* snake venom. Toxicon 52, 908–917.

Serrano, S.M., Reichl, A.P., Mentele, R., Auerswald, E.A., Santoro, M.L., Sampaio, C.A., Camargo, A.C., Assakura, M.T., 1999. A novel phospholipase A_2, BJ-PLA_2, from the venom of the snake *Bothrops jararaca*: purification, primary structure analysis, and its characterization as a platelet-aggregation-inhibiting factor. Arch. Biochem. Biophys. 367, 26–32.

Silveira, L.B., Marchi-Salvador, D.P., Santos-Filho, N.A., Silva Jr., F.P., Marcussi, S., Fuly, A.L., Nomizo, A., da Silva, S.L., Stábeli, R.G., Arantes, E.C., Soares, A.M., 2013. Isolation and expression of a hypotensive and anti-platelet acidic phospholipase A2 from *Bothrops moojeni* snake venom. J. Pharm. Biomed. Anal. 73, 35–43.

Solórzano, A., 2004. Serpientes de Costa Rica. Editorial INBio, San José, Costa Rica.

Šribar, J., Krizaj, I., 2011. Secreted phospholipases A_2 – not just enzymes. Acta Chim. Slov. 58, 678–688.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.

Texeira, C.F.P., Landucci, E.C.T., Antunes, E., Chacur, M., Cury, Y., 2003. Inflammatory effects of snake venom myotoxic phospholipases A_2. Toxicon 42, 947–962.

RESULTATS

Capítol 1

J. Durban, P. Juárez, Y. Angulo, B. Lomonte, M. Flores-Diaz, A. Alape-Girón, M. Sasa, L. Sanz, J. M. Gutiérrez, J. Dopazo, A. Conesa, i J. J. Calvete, «**Profiling the venom gland transcriptomes of Costa Rican snakes by 454 pyrosequencing**», *BMC Genomics*, vol. 12, p. 259, 2011.

Aquest treball ha abordat la caracterització de l'activitat transcripcional amb tècniques NGS de 8 transcriptomes de 7 espècies de serp costa-riquenyes mèdicament rellevants, per tal d'aprofundir en el coneixement de les adaptacions i processos biològics que governen la formulació del verí. L'estudi dels mecanismes moleculars i de les forces d'evolució implicades en la producció de les toxines esdevé, tal i com hem explicat a la Introducció, de vital importància per a millorar l'actual teràpia usada amb els antiverins i per al descobriment de potencials nous fàrmacs. L'extensa caracterització del proteoma de nombroses serps es veu així complementada, en absència d'un genoma de referència, d'unes valuoses dades transcriptòmiques donant lloc al que hem anomenat "proteotranscriptòmica" i que ens permetrà tratar d'esbrinar potencials mecanismes de regulació entre la transcripció i la traducció.

Seqüenciació i ensamblatge

El procés de seqüenciació es va produir simultàniament per al conjunt dels 8 transcriptomes en un Genome Sequencer FLX (454 Roche) seguint una estratègia multiplex, cosa que va donar lloc a un total de 334,540 lectures (de l'anglès, *read*) per al conjunt dels 8 transcriptomes, cadascuna d'elles teòricament identificada mitjançant una etiqueta de 10 nucleòtids específica de mostra. Les 62 Mbases resultants d'aquest procés de seqüenciació han estat incloses en el repositori SRA de l'NCBI amb l'identificador de l'anàlisi SRP003780. La longitud mitjana de les lectures va ser de 186.6 nucleòtids, amb un màxim de 645 nucleòtids en una única lectura i només el 3.27% d'elles va ser inferior a una longitud de 50 nucleòtids. Tal i com està detallat a [147], el processament de les lectures pel que

fa la seva qualitat de seqüenciació, separació segons l'etiqueta identificadora i la identificació i substitució per caràcters neutres (N) d'elements repetitius, va donar lloc, finalment, a un total de 330,010 lectures útils. En absència d'un genoma de referència sobre el qual localitzar cadascuna d'aquestes lectures, l'estratègia habitual és ensamblar (de l'anglès *assembly*) aquestes mateixes lectures per superposició amb la finalitat d'intentar reconstruir la seqüència original, en el que s'ha anomenat ensamblatge '*de novo*'. Així, l'ensamblatge '*de novo*' és un procés bioinformàtic que dona lloc a fragments més llargs (*contigs*) producte de l'ensamblatge de múltiples lectures solapants, i a lectures úniques no ensamblades (*singletons*). Tal i com serà discutit més endavant, el nombre de contigs i la seva longitud, així com el nombre de singletons i determinats paràmetres específics, són utilitzats per avaluar la qualitat d'un ensamblatge, procés altament condicionat pel tipus de plataforma usada i el mètode de construcció de la llibreria de DNA.

En el present treball es van utilitzar diferents algoritmes d'ensamblatge, el millor resultat dels quals va ser, finalment, l'obtingut amb el programa proveït per la plataforma amb la qual es va dur a terme la seqüenciació (Newbler v1.1, Roche). Tot i això, els resultats van estar lluny de ser òptims, amb només el 58.4% de les lectures ensamblades en 31,025 contigs, la longitud mitjana dels quals va ser de 208.2 nucleòtids. A més, el 43% d'aquests contigs eren resultat de l'ensamblatge de només dues lectures. D'altra banda, el procés d'ensamblatge '*de novo*' va donar lloc també a 103,357 singletons, la longitud mitjana dels quals va ser de 174.4 nucleòtids.

En conjunt, el resultat del procés d'ensamblatge no va suposar un guany pel que fa la compressió de les dades com ho demostra la petita diferència entre la longitud mitjana dels contigs i de les lectures resultants del procés de seqüenciació, amb la qual cosa la caracterització dels transcriptomes en termes d'anotació funcional i quantificació va tenir lloc considerant les lectures individuals i no el resultat de l'ensamblatge.

Anotació funcional

El conjunt de les 330,010 lectures útils va ser comparat per homologia de seqüència amb la base de dades no redundant del NCBI mitjançant blast [148], obtenint un total de 100,394 homologies segons un determinat llindar de significació. Aquesta baixa proporció de resultats positius resultava similar a la obtinguda en la caracterització d'altres organismes no model dutes a terme mitjançant tecnologies NGS, tal i com es discutirà més endavant.

Aproximadament el 80% dels resultats positius obtinguts per al global dels 8 organismes van donar lloc a coincidències pertanyents al subordre *Serpentes* (Taula 5b), de les quals aproximadament el 78% van poder ser classificades com a proteïnes pertanyents al verí d'acord a la nostra acurada llista de paraules clau, i van poder ser assignades a una de les famílies de proteïnes del verí.

a)

Espècie de serp	**Lectures totals**	**Longitud mitjana (nt)**
C.simus	22,389	193.6
B.asper (Car)	123,485	185.6
B.asper (Pac)	16,076	184.8
C.sasai	44,843	182.1
A.picadoi	31,027	188.8
A.mexicanus	27,080	192.1
B.schlegelii	33,276	191.5
B.lateralis	31,833	186.1
TOTAL	330,010	187.3

b)

Espècie de serp	Blast		*Serpentes*		Verí	
	Lectures	% total	Lectures	% blast	Lectures	% *Serpentes*
C.simus	3,608	16.1	2,320	64.3	1,327	57.2
B.asper (Car)	43,818	35.5	35,655	81.4	28,220	79.1
B.asper (Pac)	2,848	17.7	2,078	72.9	1,378	66.3
C.sasai	13,743	30.7	11,252	81.9	9,824	87.3
A.picadoi	13,295	42.8	11,350	85.4	9,951	87.7
A.mexicanus	6,418	23.7	4,070	63.4	3,109	76.4
B.schlegelii	6,826	20.5	4,525	66.2	2,893	63.8
B.lateralis	9,838	30.9	7,970	81	5,731	71.9
TOTAL	100,394	30.4	79,991	79.7	62,443	78

Taula 5. Resum de les estadístiques obtingudes en els processos de seqüenciació i anotació dels 8 transcriptomes. Per a cada organisme, es mostren: a) Les estadístiques del procés de seqüenciació pel que fa al nombre total de lectures obtingudes i la seva longitud mitjana, i b) Estadístiques del procés d'anotació pel que fa lectures amb un resultat blast positiu, un resultat vinculat al subordre *Serpentes* i un resultat vinculat a una proteïna pertanyent al verí.

Segons aquests resultats, es va calcular l'abundància relativa de cadascuna d'aquestes famílies en cada organisme, normalitzant pel nombre total de lectures obtingudes en el procés de seqüenciació per cada organisme i per la longitud estimada del transcrit, amb la qual cosa es van poder obtenir uns perfils d'expressió característics de cada organisme (Figura 12). Aquests perfils transcriptòmics van mostrar que unes determinades famílies de toxines eren abundants i relativament expressades de forma constant al llarg de la majoria de les espècies en estudi, mentre que altres van mostrar una major variabilitat.

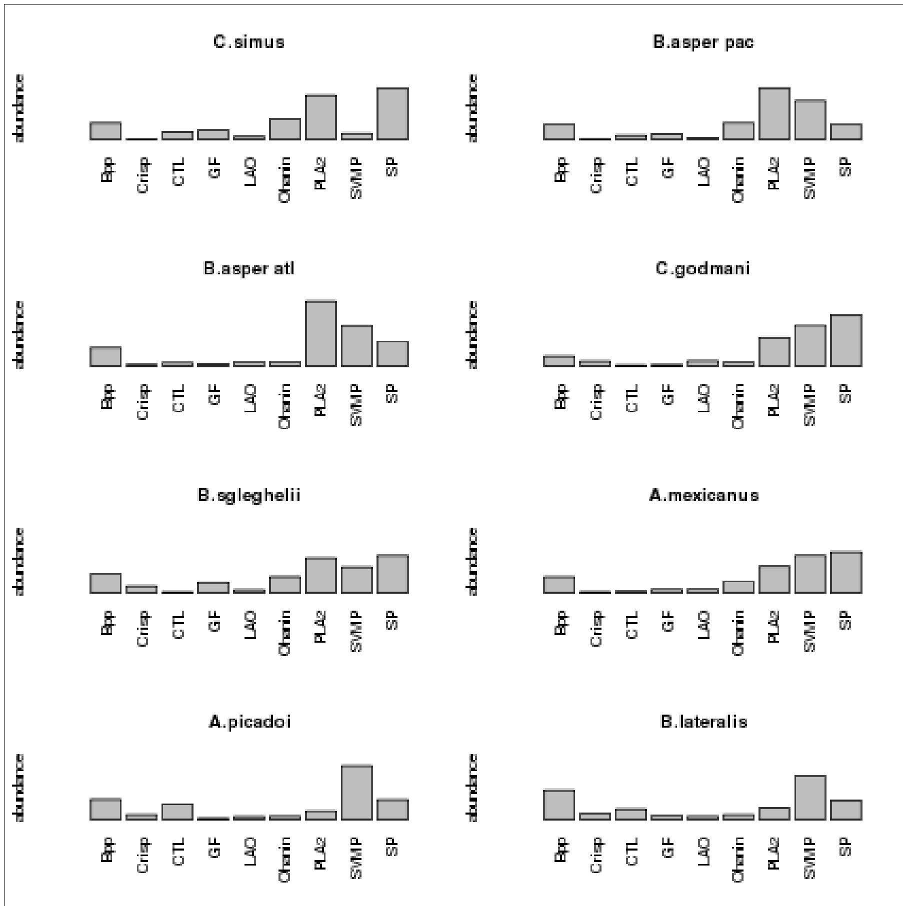


Figura 12. Perfils d'expressió de toxines. El nombre de lectures normalitzat per nombre de lectures totals i per la longitud estimada del transcrit va donar lloc a uns patrons d'expressió proteïca per a cada organisme.

Per tal d'investigar les relacions en la composició del verí entre les espècies, les dades transcriptòmiques van ser sotmeses a una Anàlisi de Components Principals (PCA). El PCA va revelar que les 8 serps en estudi es podien classificar en 4 grups principals d'acord a l'expressió relativa d'SVMP, SP i PLA2 (Figura 5 de [147], veure pàgina 76).

Estimació del nombre de còpies gèniques

Finalment, en un intent per estimar el nombre mínim de gens per a cada toxina, les lectures pertanyents a la mateixa família proteïca van poder ser classificades segons si es localitzaven en un marc obert de lectura (ORF, de l'anglès *Open Reading Frame*) o no, mitjançant un alineament múltiple d'aquestes lectures sobre la seqüència de referència més semblant segons una extensa recerca bibliogràfica. Amb això es van aconseguir dues coses: i) Obtenir aquelles lectures que realment no corresponien a ORFs sinó a regions no traduïdes (5'UTR, 3'UTR i microsatèl·lits), i ii) Obtenir totes aquelles lectures pertanyents a ORFs l'ensamblatge de les quals (tal i s'ha descrit anteriorment el cas d'ensamblatges '*de novo*') va donar lloc a una estimació del nombre de gens per a cada família de toxines (Taula 3 de [147], veure pàgina 73).

Comparativa transcriptoma - proteoma

Pocs estudis han intentat estudiar els factors que governen la transcripció dels gens de les toxines secretades al verí. Tot i que determinats estudis han detectat concordància entre el transcriptoma i el proteoma de determinades espècies de serp [149, 150], en el nostre estudi es van trobar fortes discrepàncies entre les dades transcriptòmiques i la composició final del verí. Tal i com podem veure a la Figura 4 de [147] (veure pàgina 75), la comparativa entre el proteoma i les dades obtingudes del transcriptoma no mostrava una clara tendència intra o interespecífica, i les raons d'aquestes discrepàncies són difícils de trobar, però, tal i com serà discutit més endavant, determinats factors que afecten l'estabilitat del mRNA i la seva traducció podrien estar darrera de les diferències observades entre el transcriptoma i el proteoma.

Clarament, els nostes resultats recolzen la importància de combinar amplis estudis proteòmics i transcriptòmics per a una caracterització profunda del repertori de toxines del verí i dels factors que regulen la transcripció i la traducció.

Capítol 2

J. Durban, A. Pérez, L. Sanz, A. Gómez, F. Bonilla, S. Rodríguez, D. Chacón, M. Sasa, Y. Angulo, J.M. Gutiérrez, i J.J. Calvete. «**Integrated "omics" profiling indicates that miRNAs are modulators of the ontogenetic venom composition shift in the Central American rattlesnake, *Crotalus simus simus***», *BMC Genomics*, vol. 14, p. 234, 2013.

Seqüenciació i ensamblatge

Es van seqüenciar els transcriptomes d'un individu adult i d'un individu juvenil del cròtal centreamericà *Crotalus simus simus*. Estudis proteòmics previs d'aquest organisme van revelar un patró d'expressió de toxines depenent de l'estadi de desenvolupament [151]. En concret, els individus juvenils mostren una elevada expressió d'una PLA2 amb activitat neurotòxica mentre que els individus adults mostren una elevada expressió d'una SVMPs, amb activitat hemorràgica. El present treball va voler investigar la transició d'un tipus de verí a l'altre i els seus mecanismes de regulació.
Els processos de seqüenciació d'aquests transcriptomes van tenir lloc en una plataforma 454 GS FLX amb química Titanium i van donar lloc a 408,505 lectures per a l'individu adult i a 349,170 lectures per a l'individu juvenil. Les dades resultants d'aquests processos de seqüenciació han estat incloses en el repositori SRA del NCBI amb l'identificador de l'anàlisi SRP012408.
En una estratègia similar a la definida a [147] (veure Capítol 1), les lectures van ser processades segons un determinats valors de qualitat donant lloc a 355,140 lectures útils per a l'individu adult i a 320,907 per a l'individu juvenil. Tal i com ja hem detallat al capítol anterior, en absència d'un genoma de referència, l'ensamblatge '*de novo*' és l'estratègia habitual per tal d'intentar reconstruir potencials transcrits. Així, l'ensamblatge '*de novo*' d'aquestes lectures amb el programa recomanat per l'empresa responsable del seqüenciador (Newbler, v2.6, Roche) va donar lloc a 6,484 contigs i 33,408 singletons per l'individu adult, i 6,047 contigs i 24,136 singletons per l'individu

juvenil. Les estadístiques d'ambdós processos de seqüenciació estan detallades a la Taula 1 de [152] (veure pàgina 89). Cal destacar que, tal i com serà discutit més endavant, aquest ensamblatge va posar de manifest la ràpida evolució de les eines bioinformàtiques, ja que, comparat amb l'ensamblatge de [147], en aquest cas l'ensamblatge '*de novo*' sí va donar lloc a uns bons resultats de compressió, on entre el 86% i el 83% de les lectures van ser ensamblades i només un 1.43% dels contigs de l'individu adult i un 10.74% dels contigs de l'individu juvenil estaven formats només per dues lectures.
D'altra banda, la regulació sobre els transcrits ha estat proposat com un possible mecanisme de generació de diversitat fenotípica. La regulació amb miRNAs podria ser un d'aquests mecanismes. Per tal d'esbrinar el paper dels miRNAs en la variació ontogenètica observada en *Crotalus simus simus*, dues llibreries de RNA total enriquides en RNAs de mida inferior a 200 nucleòtids (una per a l'individu adult i una per a l'individu juvenil) van ser sequenciades amb la plataforma Ion Torrent Personal Genome Machine. Els resultats d'aquest procés de seqüenciació i del seu processament estan detallats a la Taula 4A de [152] (veure pàgina 93) i han estat depositats a la base de dades SRA del NCBI amb d'identificador SRP012408. Així, es van obtenir 220,548 lectures de longitud mitjana de 24.58 nucleòtids per a l'individu juvenil i 389,064 lectures de longitud mitjana 23.68 nucleòtids per a l'individu adult, que van poder ser agrupades per homologia de seqüència en 283 i 419 grups amb més de 100 lectures respectivament. Tal i com serà discutit, el mode d'acció additiu en la inhibició dels miRNA va ser la raó principal per considerar aquest llindar.

Anotació funcional

La Taula 6 mostra els resultats pel que fa la recerca d'homologies mitjançant blast en la base de dades no redundant del NCBI dels contigs i singletons resultants del procés d'ensamblatge, tant per l'individu juvenil com per l'individu adult.

	***C.simus* Juvenil**	***C.simus* Adult**
Contigs blast	3,022	4,141
%contigs totals	49.97	63.86
Contigs *Serpentes*	946	902
% blast	31.3	21.78
Proteïnes verí	658	431
% *Serpentes*	69.55	47.78
Singleton blast	15,052	21,460
% totals	62.36	64.23
Singletons *Serpentes*	2,306	3,272
% blast	15.3	15.24
Singletons verí	526	431
% *Serpentes*	22.8	13.17

Taula 6. Resum de les estadístiques d'anotació per als contigs i singletons dels transcriptomes dels individus juvenil i adult. Estadístiques del procés d'anotació pel que fa lectures amb un resultat blast positiu, un resultat vinculat al subordre *Serpentes* i un resultat vinculat a una proteïna pertanyent al verí.

Aproximadament el 50% dels contigs obtinguts per a l'individu juvenil i el 63.86% de l'individu adult van donar alguna homologia significativa a la base de dades. També van donar resultats positius el 62.36% dels singletons per a l'individu juvenil i el 64.23% per a l'individu adult. Aquests resultats positius van ser classificats segons pertanyien al subordre *Serpentes* i a alguna de les famílies de proteïnes del verí en una segona ronda de filtratge segons un llistat de paraules clau producte d'una extensa recerca bibliogràfica. La Taula 2 de [152] (veure pàgina 90) mostra l'abundància relativa de les diferents famílies de proteïnes que van ser identificades en ambdós transcriptomes. Per a cadascuna d'elles es va estudiar la seva expressió diferencial, trobant que 4 toxines del verí estaven expressades de forma diferent entre l'organisme adult i el juvenil. A més, en un intent per determinar el grau de similitud de seqüència entre aquests transcrits, es va obtenir el nombre de seqüències

compartides entre ambdós transcriptomes i aquelles expressades de forma única per cada organisme. Aquests resultats apareixen a la Taula 3 i la Figura 3 de [152] (veure pàgines 91 i 92 respectivament).

Els miRNA

Es va dur a terme la seqüenciació de dues llibreries de RNA de molècules de mida inferior a 200 nucleòtids amb un Ion Torrent PGM, una per a l'individu adult i una per a l'individu juvenil. Aquest procés va donar lloc a 515,040 lectures (aproximadament 12 Mbases) i 314,592 (aproximadament 7.5 Mbases) respectivament, que van ser processades per valors de qualitat de la seqüenciació i eliminant possibles adaptadors. Aquestes lectures van ser agrupades per homologia de seqüència, i aquell conjunt de seqüències amb un mínim de 100 lectures cadascun (veure Taula 4b [152], veure pàgina 93) va ser classificat segons si estava expressat de forma única en un dels dos transcriptomes o de forma compartida per ambdós. En aquest darrer cas es van obtenir 151 grups de seqüències, l'anàlisi dels quals va mostrar una distribució desigual entre les dades de l'individu juvenil i l'adult. És a dir, d'entre els miRNA amb una freqüència de 1000 còpies o superior, els miRNA més expressats a l'individu juvenil mostraven una baixa abundància a l'individu adult i viceversa (Figura 5 de [152], veure pàgina 94). Aquesta distribució desigual suggeria un mecanisme de regulació pel qual un transcriptoma pot donar lloc a diferents composicions del verí.

La caracterització d'aquest potencial mecanisme de regulació va implicar l'ús de les eines bioinformàtiques adients per a determinar: i) Predicció de potencials dianes dels miRNA en els fragments 3'-UTR dels transcrits identificats com a PLA2 i SVMP dels transcriptomes adult i juvenil, mitjançant miRanda [153] i ii) Predicció de seqüències codificants de miRNA per criteris de conservació de seqüència i d'alineament amb seqüències miRNA descrites, mitjançant MapMi [154]. En el primer cas miRanda va predir 10 miRNAs complementaris a la regió 3'-UTR de transcrits de *Crotalus simus simus* identificats com a SVMP (5 compartits entre l'adult i el juvenil i 5 exclusius de l'individu juvenil) i 3 miRNAs provinents de l'individu adult amb complementarietat a seqüències de les PLA2 (Figura 6 de

[152], veure pàgina 95). L'ús de MapMi per a discernir potencials gens ortòlegs de miRNA sobre aquests resultats, va donar lloc, finalment, a 3 miRNA provinents de l'individu adult inhibidors de transcrits codificants de PLA2 i a 2 miRNA provinents de l'individu juvenil inhibidors de seqüències SVMP. Cal destacar, tal i com apareix a la Figura 6 de [152], que 590 còpies del miRNA 2166, inhibidor de la subunitat B de la crotoxina (PLA2), va ser trobat exclusivament al transcriptoma de l'individu adult, mentre que 1,185 còpies del miRNA 2578, complementari al mRNA d'una SVMP tipus PIII, va ser trobat de forma exclusiva al transcriptoma de l'individu juvenil.

Comparativa transcriptoma - proteoma

La Figura 4 de [152] (veure pàgina 92) mostra la comparativa de la composició relativa de famílies proteïques del transcriptoma de l'individu adult i juvenil. Tal i com serà discutit més endavant, de forma similar a com succeïa a [147], la composició general dels transcriptomes de l'individu adult i juvenil són molt més similars entre ells que els respectius proteomes, indicant que el verí podria estar sotmès a un control ontogenètic post-transcripcional pel qual transcriptomes similars donen lloc a fenotips del verí diferents en el qual podrien intervenir els miRNA tal i com hem vist anteriorment.

Capítol 3

M. Van der Laat, J. Fernández, J. Durban, E. Villalobos, E. Camacho, J.J. Calvete, i B. Lomonte, «**Amino acid sequence and biological characterization of BlatPLA2, a non-toxic acidic phospholipase A2 from the venom of the arboreal snake Bothriechis lateralis from Costa Rica**», *Toxicon*, vol. 73, p. 71-80, 2013

En aquest estudi es va aïllar i caracteritzar una nova PLA2 II acídica (BlatPLA2, UniProt C0HJC1) del verí de *Bothriechis lateralis* que mostra diferències significatives amb les PLA2 prèviament descrites, cosa que pot aportar un marc comparatiu important per tal de determinar aquelles bases moleculars responsables de l'elevada toxicitat observada en les PLA2 bàsiques, una toxicitat moderada en les PLA2 àcides descrites fins ara. Així, el paper que juguen en el verí aquestes PLA2 àcides mancades d'activitat farmacològica és quelcom desconegut, i determinades hipòtesis apunten a un paper digestiu de la presa en un caràcter ancestral retingut, més que no pas d'immobilització o tòxic.
Aquesta proteïna, que pot arribar a representar el 9% del proteoma de *Bothriechis lateralis*, va ser caracteritzada per tècniques de biologia molecular per valorar la seva activitat letal, el reconeixement immunològic, la formació d'edema i l'activitat anticoagulant entre altres. La determinació de la seva seqüència d'aminoàcids va permetre un alineament múltiple entre aquesta PLA2 i les 26 seqüències de PLA2 que van mostrar major homologia de seqüència en una recerca a la base de dades del NCBI usant blast, obtenint el percentatge més elevat de similitud amb una PLA2 aïllada del verí de *Lachesis stenophrys* (78%) i conservant les 14 cisteïnes que conformen el patró que defineix aquest grup de proteïnes (Figura 4 de [155], veure pàgina 112). Aquest alineament múltiple va permetre una estimació de la relació filogenètica entre aquestes proteïnes (Figura 5 de [155], veure pàgina 113). Aquesta estimació va ser capaç de separar clarament els enzims àcids dels bàsics, i va situar la BlatPLA2 en una posició basal dins del grup dels enzims àcids dels vipèrids del Vell i del Nou Món, com a una branca germana de totes les PLA2 àcides analitzades amb

excepció de dos enzims pertanyents a *Agkistrodon halys* que van donar lloc a un clade independent.

DISCUSSIÓ

Capítols 1 i 2

Seqüenciació i ensamblatge

En el moment de redacció d'aquest treball, les tècniques NGS han servit per caracteritzar els transcriptomes de 11 espècies de cròtals (*Bothrops asper, Cerrophidion sasai, Atropoides mexicanus, Atropoides picadoi, Bothriechis lateralis, Bothriechis schlegelii, Crotalus simus, Crotalus horridus, Crotalus adamanteus, Ovophis okinavensis* i *Protobothrops flavoviridis*) dos colúbrids (*Atractaspis aterrima* i *Thamnophis elegans*) i un elàpid (*Micrurus fulvius*) en 9 estudis diferents [146, 147, 152, 156–161]. En aquests estudis s'han empleat dos tipus diferents de tecnologia NGS: per a 5 d'ells es van utilitzar les lectures més llargues obtingudes amb 454, mentre que per a 4 estudis es va usar Illumina, 3 d'ells realitzats al llarg de 2013, en una clara tendència de l'evolució del mercat de la tecnologia de seqüenciació d'alt rendiment tal i com ja ha sigut explicat anteriorment (veure Introducció). Així, segons Rokyta i col·laboradors [156], la diferència en longitud de lectura entre 454 i Illumina era compensada per l'augment, en dos ordres de magnitud, del nombre total de lectures resultants del procés de seqüenciació d'aquest últim. D'aquesta manera, mentre el nombre de lectures obtingudes amb 454 no va superar en cap cas les 800.000 lectures per procés de seqüenciació, el nombre de lectures totals per als treballs realitzats amb Illumina va oscil·lar des dels aproximadament 13 milions de lectures per a la caracterització dels cròtals originaris de Japó (*Ovophis okinavensis* i *Protobothrops flavoviridis*), fins a les més de 96 milions de lectures usades per a la caracterització de *Crotalus adamanteus*. La quantitat de nucleòtids seqüenciats en el nostre cas va ser de 62 Mbases per al total del 8 transcriptomes caracteritzats a [147] i de 190 Mbases per a l'estudi de *Crotalus simus* on es va usar 454 amb química Titanium. En conjunt, els treballs realitzats amb 454 van assolir un màxim de 249 Mbases en la caracterització d' *Atractaspis aterrima* mentre que aquest mateix valor en els estudis on es va usar Illumina va arribar a les més de 19 Gbases seqüenciades de *Crotalus adamanteus*.

En tots els casos excepte [147], les lectures obtingudes van ser ensamblades mitjançant diferents algoritmes bioinformàtics. Diversos factors ens van empènyer a ometre el procés d'ensamblatge en els transcriptomes caracteritzats a [147]: en primer lloc, només el 58.4% de les lectures van ensamblar per donar lloc a contigs, dels quals aproximadament el 43% estava format per només dues lectures. En aquest punt cal destacar que els estudis realitzats amb 454 sobre els transcriptomes de *Crotalus simus* a [152], *Thamnophis elegans* i *Crotalus adamanteus* van donar lloc a un nombre de lectures ensamblades superior al 80% i el nombre de contigs amb dues lectures en la caracterització del primer anava des de l'1.43% fins al 10% per a l'individu adult i juvenil respectivament, mentre que més de 24,000 contigs d'un total 24,773 estaven formats per dues o més lectures en [157]; i en segon lloc, l'ensamblatge no suposava una millora de la longitud dels transcrits, obtenint una diferència molt petita entre la longitud mitjana de les lectures i dels contigs (186.6 i 208.2 respectivament). En els ensamblatges realitzats per als transcriptomes de *Crotalus adamanteus* i *Thamnophis elegans* el guany en la longitud mitjana dels transcrits va ser de 213 i 215 nucleòtids respectivament.
En conjunt, el mal resultat del nostre millor ensamblatge ens va conduïr a la decisió d'anotar les lectures de forma individual a [147], cosa que va estar recolçada pel fet que estudis transcriptòmics previs en organismes no model mitjançant NGS incloïen l'anotació dels singletons que, juntament amb els contigs, eren tractats com a potencials unigens [162].
Així, en el nostre cas, el percentatge de lectures ensamblades i la longitud mitjana dels contigs obtinguts van resultar paràmetres determinants per a la decisió d'anotar les lectures de forma individual d'una manera tan eficient com la tradicional seqüenciació Sanger [163]. En aquest sentit, en absència d'un genoma de referència, l'avaluació de la validessa i utilitat d'un ensamblatge ha fet ús de diferents tipus de paràmetres com són el nombre de contigs obtinguts, la seva longitud màxima i longitud mitjana, el nombre de lectures sense ensamblar i l'estadístic N50 (la longitud dels contigs per la qual usant contigs d'aquest valor o superior obtenim el 50% de les bases ensamblades), entre d'altres. Fins a la data, però, no està clar quins

paràmetres poden realment valorar la qualitat d'un ensamblatge d'una forma acurada [164].

Anotació funcional

Una vegada obtingudes les lectures i contigs dels diferents processos de seqüenciació que van ser duts a terme durant el desenvolupament d'aquest treball, l'estratègia definida implicava la recerca d'homologies en diferents bases de dades. En concret, tal i com ja s'ha detallat en els Resultats, es van anotar 10 transcriptomes pertanyents a 7 espècies diferents. D'aquesta manera, tal i com podem veure a les Taules 5 i 6 (pàgines 121 i 127 respectivament), *Crotalus simus* va ser seqüenciat en el conjunt dels 8 transcriptomes amb 454 donant lloc a un 16,1% de les lectures amb un resultat blast positiu (un 30,4% per al conjunt dels 8 transcriptomes), mentre que el transcriptoma d'un individu juvenil i d'un individu adult seqüenciats amb 454 Titanium van donar lloc al 63,86% i 49.97% dels contigs amb resultat positiu respectivament. A la llum d'aquests resultats, tant el guany en longitud de transcrit que va suposar l'ensamblatge (193.6 nucleòtids per a les lectures individuals anotades a [147], i 777.8 i 560.87 per als contigs obtinguts per a l'individu adult i juvenil respectivament a [152]) com la possible millora en la tecnologia de seqüenciació en la química Titanium, podrien ser la causa de la millora de resultats blast obtinguts. Tanmateix, l'alt percentatge de lectures sense una homologia significativa a una seqüència coneguda obtingut a [147] no s'allunya dels resultats obtinguts en altres estudis similars. Així, el 87% de lectures ensamblades en 96,379 contigs a [146] va donar lloc a un 34% d'homologies. De la mateixa manera, només el 45.6% dels 2,855 contigs obtinguts a [107] van mostrar una homologia a una seqüència del GenBank. Finalment, el nostre percentatge de resultats positius per al conjunt dels 8 transcriptomes va ser molt superior a l'obtingut a [165], estudi en el qual es va obtindre el 15.5 %.

Una vegada obtingudes homologies de les nostres seqüències amb seqüències depositades en bases de dades, es van identificar aquelles pertanyents al subordre *Serpentes* i aquelles que formaven part de les proteïnes de verí. El 67% de les lectures identificades com a pertanyents al subordre *Serpentes* a [147] i el 47.78% i el 69.55% dels

contigs obtinguts per a l'individu adult i juvenil respectivament a [152], van poder ser identificades mitjançant una acurada llista de paraules clau. Així, els resultats obtinguts podrien reflectir la pobresa de seqüències de membres del subordre *Serpentes* no pertanyents al verí en les diferents bases de dades o respondre a una alta especialització del tipus cel·lular existent a les glàndules per a la producció de proteïnes específiques del verí. En aquesta darrera hipòtesi cal destacar els resultats obtinguts per Rokyta i col·laboradors en la seqüenciació del transcriptoma de *Crotalus adamanteus* amb 454 [157]. En l'anotació de toxines l'estudi va identificar un 11% dels contigs i un 46% de les lectures individuals, un motiu indicador més, segons els autors, de la sobreexpressió de toxines en la glàndula del verí. En el mateix sentit, la caracterització del transcriptoma de *Micrurus fulvius* [160] va donar lloc a un 45.8% de lectures pertanyents a toxines.

Composició del transcriptoma

L'anotació funcional dels trancriptomes ens va permetre la caracterització d'aquests pel que fa les principals proteïnes del verí de les serps, descrites a la Introducció. A més, també es van trobar altres proteïnes com les hialuronidases, els inhibidors de proteases tipus Kazal, les disulfit isomerases i determinats elements mòbils com SINEs (de l'anglès *Short Interspersed repetitive Element*) i LINEs (de l'anglès *Long Interspersed repetitive Element*).
L'existència de les primeres en el verí de les serps està més relacionada amb l'expansió de les toxines a la sang i als teixits que amb una activitat tòxica pròpiament dita. És considerada, així, un component no tòxic del verí. Tanmateix, la degradació de l'àcid hialurònic de la matriu extracel·lular per part de la hialuronidasa esdevé un element essencial per a l'expansió sistèmica de les toxines d'una manera molt més ràpida a com ho farien per simple difusió [166].
Respecte als inhibidors de proteases tipus Kazal, aquestes proteïnes havien estat prèviament identificades només en el proteoma de *Bothriechis schlegelii* [167], amb dubtes sobre el seu origen glandular. En la caracterització dels transcriptomes feta a [147], es van trobar

lectures codificants per a aquesta proteïna en les dues serps pertanyents al gènere *Bothriechis*, recolçant la idea d'un reclutament específic de gènere [168].
La majoria de proteïnes del verí estan caracteritzades per un elevat nombre de residus de cisteïnes que donen lloc a ponts disulfur intra- i intermoleculars. Així, no va resultar sorprenent observar que aproximadament un 5% les lectures dels transcriptomes caracteritzats a [147] van donar una homologia de seqüència amb una disulfit isomerasa, un enzim involucrat en la formació de ponts disulfur al reticle endoplasmàtic [169].
Cal destacar també la presència de toxines de tres dits (3FTX, de l'anglès "*Three finger toxins*") i d'Ohanina en tots els transcriptomes caracteritzats a [147] i en el transcriptoma adult i juvenil caracteritzats a [152]. Les 3FTXs són proteïnes no enzimàtiques del verí que contenen entre 60 i 74 aminoàcids amb una estructura de 3 fulles β que s'extenen des d'un nucli central conservat, globular i hidrofòbic [170]. Han estat llargament descrites en els verins dels elàpids, i la seva presència en membres de la subfamília *Crotalinae* només s'ha pogut comprovar en el verí d'*Atropoides mexicanus* [171]. La Ohanina, per la seva banda, és una proteïna de 107 aminoàcids, amb un únic residu de cisteïna. Com les 3FTX, és una proteïna característica en membres de la família *Elapidae*, i en vipèrids només s'ha descrit un transcrit similar a *Lachesis muta* [109] i en els treballs desenvolupats per Rokyta i col·laboradors sobre *Crotalus adamanteus*, on tant 454 com Illumina van poder detectar un transcrit que codificava per aquesta proteïna, causant de dolor i hipolocomoció en ratolins segons els estudis realitzats per Pung i col.·laboradors [172].
En el nostre cas, però, ambdós tipus de proteïnes no semblaven donar lloc a productes gènics funcionals, ja que les lectures i contigs identificats no van poder ser localitzats en un ORF, amb excepció de dues lectures en les dues espècies del gènere *Atropoides* caracteritzades a [147] (Taula 3 de [147], veure pàgina 73). En cap cas la seqüenciació del transcriptoma de *Crotalus simus* va donar lloc a lectures dins del CDS (de l'anglès, "*Coding sequences*"). Així, la presència de mRNA corresponents a fragments UTR podria respondre a transcrits seqüenciats de forma incompleta, a un esbiaix causat pel

mètode de síntesi del cDNA (que produeix llibreries de cDNA enriquides en fragments 3'-UTR) o a la presència de pseudogens.
És rellevant també la presència d'elements mòbils en les nostres caracteritzacions, que, d'altra banda, també va ser observada en l'estudi amb NGS del colúbrid *Atractaspis aterrima* i en altres estudis previs a partir dels EST resultants de la seqüenciació de determinats cròtals com *Bothrops insularis, Lachesis muta,* del colúbrid *Philodryas olfersii* i en els gens de les PLA2 de *Vipera ammodytes* i *Protobothrops flavoviridis*. Tot i que la seva rellevància funcional a les glàndules continua sent desconeguda, aquests elements apareixen sobreexpressats en gens d'evolució ràpida, suggerint la seva possible implicació en la diversificació i expansió d'aquestes famílies gèniques.
Finalment, els nostres estudis dels transcriptomes de *Crotalus simus* van coincidir amb els de Rokyta i col·laboradors en l'anotació de les toxines més representades en els transcriptomes del gènere *Crotalus*. Així, les serines proteases semblen ser la familia més abundant de toxines, indicant que la disrupció del sistema hemostàtic de la presa és un mecanisme primari d'incapacitació d'aquest gènere.

Regulació post-transcripcional

Un dels objectius del treball proposat en aquesta tesi doctoral va ser intentar esbrinar l'existència de mecanismes de regulació per tal d'explicar la varibilitat observada en la composició final del verí.
Per això, es van seqüenciar individus de la costa atlàntica i de la costa pacífica de *Bothrops asper* (veure Annex I, pàgina 151) i es va seqüenciar un individu adult i un individu juvenil de *Crotalus simus*. En ambdós casos, la composició del transcriptoma va resultar ser molt més similar entre els individus de la mateixa espècie que la composició final del verí (Figura 4 de [147] i Figura 4 de [152], pàgines 75 i 92 respectivament), suggerint un mecanisme de regulació post-transcripcional.
Tal i com ja havia estat descrit anteriorment [173], diferències observades en la concentració de proteïnes són només atribuïbles en un 20 - 40% a la diferència de concentració de mRNA, posant de manifest la importància de la regulació post-transcripcional. La

caracterització computacional dels miRNAs de *Crotalus simus*, com a complement de les clàssiques aproximacions genètiques, va voler donar un pas més en aquesta direcció Els miRNA són una classe d'RNA petit (aproximadament 22 nucleòtids), de cadena única, no codificants, que han estat descrits com a reguladors negatius de l'expressió gènica en organismes eucariotes [122]. Aquestes seqüències, conservades entre diferents organismes [174], presenten característiques específiques com són la multiplicitat (un miRNA inhibeix l'expressió de diferents gens) i la cooperativitat (un gen és inhibit per diferents miRNAs) [153], característiques que van ser considerades en el nostre treball per tal de reduïr al màxim el resultat de falsos positius.
L'estudi de la potencial participació dels miRNA en l'expressió diferencial ontogenètica observada en *Crotalus simus* va ser dut a terme mitjançant dues aproximacions: a nivell de la pròpia seqüència del miRNA i de la seva conservació en altres organismes, i a nivell de la seqüència diana en les regions UTR dels transcrits de les SVMP i les PLA2 identificats en la caracterització dels transcriptomes de l'individu adult i de l'individu juvenil. D'aquesta manera, en absència d'un genoma de referència, l'estratègia basada en la recerca d'homologies és l'aproximació més adient per a la identificació de gens codificants de miRNA [175, 176]. Així, el nostre treball va implicar la recerca d'homologies a la miRBase [177], un repositori de precurssors i seqüències madures de miRNA principalment d'organismes model, que, a data de redacció d'aquesta tesi, compta amb 28,645 seqüències diferents (http://www.mirbase.org, versió 21). Els resultats positius d'aquesta recerca d'homologia ens va permetre un anàlisi comparatiu de conservació de seqüències entre organismes, com un tret distintiu dels miRNA tal i com ja s'ha comentat.
Per altra banda, les prediccions computacionals de dianes dels miRNAs han estat usades en la identificació de seqüències en diferents organismes [153, 178]. La majoria d'aquestes aproximacions computacionals estan basades en determinades característiques empíriques [179]: i) Un patró d'aparellament de bases entre el miRNA i el mRNA, ii) Determinades propietats termodinàmiques d'aquest aparellament de bases, i iii) Múltiples dianes per un mateix miRNA, segons la multiplicitat comentada anteriorment.

Paral·lelament, la variabilitat geogràfica i ontogenètica de *Bothrops asper* ha estat descrita prèviament a partir d'aproximacions proteòmiques [180]. El grau de similitud en la composició del verí entre les poblacions del Pacífic i del Carib (Figura 13) semblava estar proper al 52%. Tanmateix, tal i com discutirem a l'Annex I, aquestes diferències poden respondre, o bé a una regulació post-transcripcional que doni lloc a una expressió diferencial de gens similars, o bé que l'aïllament de les dues poblacions fa 5-8 milions d'anys ha augmentat la complexitat genòmica d'aquests organismes en un fenòmen normalment associat amb el reclutament de noves funcions o fenòmens d'especiació. Dos resultats més del nostre treball exposats, a l'Annex I, apunten a aquesta regulació post-transcripcional: per un cantó, l'existència de mRNA codificant en la població atlàntica de *Bothrops asper* per a pèptids trobats de forma exclusiva mitjançant tècniques proteòmiques en el verí de la població pacífica de *Bothrops asper*, recolça a nivell de seqüència la similitud de perfils transcriptòmics observats entre ambdues poblacions (veure Figura 4 [147]). En segon lloc, la majoria (9 sobre 12) de fragments peptídics d'una SVMP trobats de forma exclusiva en la variant atlàntica tenien una seqüència nucleotídica codificant en la variant del Pacífic sense observar el corresponent pèptid en el seu proteoma.
Cal fer esment, però, que la menor quantitat de dades de seqüenciació del transcriptoma obtingudes per a la variant del Pacífic, podria explicar el fet que determinats pèptids observats en ambdós proteomes només van poder ser trobats en lectures codificants de la variant del Carib.
Sigui com sigui, la predicció computacional de seqüències de miRNA i de les seves dianes suposa un repte per a la comunitat científica que necessitarà ser validada per mètodes experimentals com, per exemple, Northern Blot.

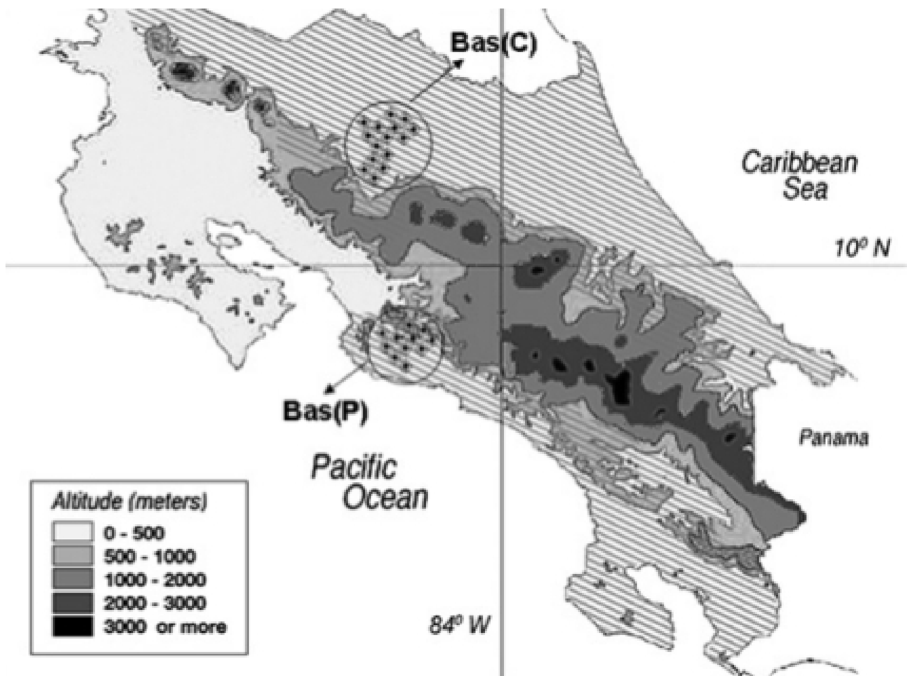


Figura 13. Distribució de les poblacions de *Bothrops asper* a Costa Rica. Mapa físic de Costa Rica mostrant la distribució geogràfica de *Bothrops asper* (àrea ratllada). (Imatge obtinguda de [180])

Capítol 3

Les bases moleculars de la toxicitat de les PLA2 només són parcialment conegudes i la correlació de l'estructura tridimensional amb determinades funcions tòxiques és una àrea intensa de recerca [181]. De forma general, però, les PLA2 bàsiques semblen ser més tòxiques que les seves parents àcides, descrites com a mancades de qualsevol activitat tòxica [182] o causants d'una moderada miotoxicitat [183]. L'existència d'una PLA2 àcida monomèrica a *Bothriechis lateralis* (el verí de la qual, d'altra banda, no conté cap PLA2 bàsica), podria respondre a diferents motius. En primer lloc, aquesta PLA2 podria resultar tòxica per a determinats organismes i no a d'altres. Tanmateix, el principal component de la dieta de *Bothriechis lateralis* són els ratolins, que no manifesten símptomes

d'activitat tòxica després de la injecció de la proteïna. En segon lloc, l'existència de possibles sinergismes entre diferents toxines tal i com ja s'ha demostrat a *Bothrops alternatus* [184] podrien atribuir a la BlatPLA2 un paper potenciador de l'efecte tòxic d'altres proteïnes del verí com les metaloproteases. Finalment, i en consonància amb la nostra estimació filogenètica, la BlatPLA2 podria suposar un remanent evolutiu en el procés que va transformar les PLA2 fisiològiques en toxines, adquirint una funció pre-digestiva, retenint carcterístiques d'una PLA2 ancestral sense evolucionar cap a la seva potent toxicitat.

CONCLUSIONS

1. L'anàlisi dels transcriptomes de diferents espècies de serp costa-riquenyes mitjançant NGS, ha revelat tendències específiques per a cada tàxon que governen la formulació del verí.
2. La caracterització de l'activitat transcripcional de les glàndules del verí d'un individu adult i d'un individu juvenil de *Crotalus simus* ha revelat uns perfils transcriptòmics molt similars entre ells, suggerint que el mecanisme pel qual es genera la divergència ontogenètica observada en la composició final del verí podria actuar de manera post-transcripcional, de tal forma que el silenciament de l'expressió d'mRNA podria explicar la composició del verí observada en ambdós individus.
3. La predicció computacional de miRNA va donar lloc a 3 seqüències provinents de l'individu adult inhibidors de transcrits codificants de PLA2 i a 2 miRNA provinents de l'individu juvenil inhibidors de seqüències SVMP.
4. La divergència en els fenotips de les variants geogràfiques del Pacífic i del Carib de *Bothrops asper* podria tindre el seu origen en aquesta regulació post-transcripcional. Així, el nostre treball va localitzar transcrits codificants per a determinats pèptids exclusius d'una població en l'altra, en una clara indicació d'una regulació post-transcripcional.
5. S'ha caracteritzat la relació filogenètica entre una PLA2 no tòxica no descrita fins ara al verí de *Bothriechis lateralis* i les 26 seqüències de PLA2 que van mostrar una major homologia de seqüència.

ANNEX I

Estudi de la divergència molecular entre diferents poblacions de *Bothrops asper*

Introducció

Els resultats previs publicats a [147] ens van mostrar que els diferents mRNAs estan controlats d'una forma depenent d'espècie i que les diferències observades entre el transcriptoma i el proteoma podrien ser explicades per la diferent estabilitat i el diferent grau de traducció d'aquests missatgers.

Per altra banda, diferències geogràfiques i ontogenètiques de la composició final del verí han estat demostrades entre les poblacions costa-riquenyes del Pacífic i del Carib de *Bothrops asper* [180, 185]. Aquestes dues poblacions, separades per diferents cadenes muntanyoses exteses diagonalment al llarg de tot el país, representen comunitats reproductivament aïllades [186]. D'aquesta manera, l'existència de variabilitat intraespecífica, deixa entreveure el concepte que les espècies deurien ser considerades com un grup de metapoblacions.

A partir de la feina realitzada en el nostre laboratori en la caracterització dels transcriptomes de glàndules de verí de serps costa-riquenyes, es va intentar estudiar la història molecular de les principals toxines amb activitat enzimàtica de les poblacions del Carib i del Pacífic de *Bothrops asper*. L'anàlisi en detall de forma conjunta de les seqüències nucleotídiques deduïdes a partir del procés de seqüenciació explicat anteriorment i els pèptids obtinguts en la seva caracterització proteòmica, ens podria ajudar a explicar potencials mecanismes de regulació que governen la manera per la qual les dues poblacions mostren uns transcriptomes molt similars però difereixen en la composició relativa de les toxines en els seus proteomes. Aquests potencials mecanismes de regulació podrien correspondre a fenòmens de tipus post-transcripcional o traduccional, donant lloc a una expressió específica d'unes mateixes seqüències gèniques o, d'altra banda, l'aïllament d'ambdues poblacions per diferents cadenes muntanyoses fa 5-8 milions d'anys podria haver augmentat la complexitat genòmica d'aquestes seqüències, donant lloc a gens

específics de cada població, en un fenòmen normalment associat al reclutament de noves funcions o fenòmens d'especiació.

Mètodes

A partir del treball desenvolupat anteriorment i explicat al Capítol 1, es va deduir computacionalment la seqüència d'aminoàcids de les principals toxines amb activitat enzimàtica (segons la seva abundància) per a cadascuna de les dues poblacions de *Bothrops asper*.

En un intent d'estudiar la correlació entre el transcriptoma i el proteoma, aquestes seqüències d'aminoàcids, obtingudes a partir de la traducció de les seqüències nucleotídiques obtingudes mitjançant la seqüenciació del transcriptoma, i les seqüències d'aminoàcids obtingudes prèviament mitjançant tècniques proteòmiques [180], van ser usades en un alineament múltiple en el que anteriorment hem anomenat "proteotranscriptòmica" (veure Introducció, Transcriptòmica i proteòmica).

Donat que la majoria d'errors de la tecnologia de seqüenciació 454 han estat localitzats en regions d'homopolímers, els alineaments múltiples obtinguts van ser inspeccionats de forma visual, i allà on va ser possible, es van corregir canvis de pauta de lectura. Això va donar lloc a un conjunt de seqüències aminoacídiques i nucleotídiques lliures d'error adients per determinades anàlisis posteriors.

Resultats i discussió

Pocs treballs han usat la tecnologia 454 amb finalitats filogeogràfiques donat que les lectures resultants del procés de seqüenciació són més curtes que els productes resultants de la clàssica seqüenciació Sanger, el nombre d'errors és més elevat i la tecnologia, tal i com hem comentat, és sensible a regions d'homopolímers. A més, l'ús d'aquestes lectures curtes en estudis filogenètics pot donar lloc a regions sense cobertura en l'alineament múltiple previ a qualsevol consideració filogenètica. Tanmateix, segons John J. Wiens, més que la manca de dades, sembla que els caràcters existents coincidents entre

diferents tàxons són crítics per a resoldre determinades relacions filogenètiques [187].
Tot i els incovenients esmentats, des que la tecnologia NGS ha esdevingut una eina ordinària en molts laboratoris, s'han dut a terme diferents intents per tal d'usar les lectures resultants en la inferència de determinades relacions filogenètiques. Així, Roeding i col·laboradors van dur a terme uns estudis filogenètics usant un conjunt de seqüències de cDNA d'hemocianines de l'escorpí *Pandinus imperator* [162] obtingudes mitjançant 454. Més recentment, John E. McCormack va concloure que seqüències obtingudes mitjançant 454 podrien ser útils per a estudis filogenètics, genètica de poblacions o amb objectius filogeogràfics, especialment per a espècies properes [188]. Així, tot i ser incompletes, la majoria de seqüències obtingudes mitjançant 454 resulten d'utilitat per a identificar quina toxina ha estat identificada en la caracterització del transcriptoma, i més concretament en el nostre cas, del transcriptoma de les glàndules del verí de les serps.
D'altra banda, cal esmentar que la majoria d'estudis s'han limitat a seqüències gèniques sintètiques mitocondrials, i pocs estudis han usat gens nuclears. Les seqüències mitocondrials, amb una més elevada taxa de canvi que les seqüències gèniques nuclears, podrien amagar l'estimació filogenètica d'espècies més allunyades, i pel que fa les toxines trobades al verí de les serps, només dos estudis s'han dut a terme mitjançant seqüències gèniques de serines proteases [189] i de disintegrines [190].

LAAO

La seqüenciació del transcriptoma descrita al Capítol 1 i l'anàlisi posterior de les dades obtingudes, va donar lloc a dues seqüències nucleotídiques codificants per a una L-aminoàcid oxidasa per a cada variant de *Bothrops asper*, l'anàlisi de les quals no va mostrar cap expressió diferencial entre les poblacions ja que cadascuna d'elles va trobar una homologia igual o superior al 98% a una seqüència de l'altra població.

Per la seva banda, els estudis proteòmics previs havien descrit 9 fragments peptídics comuns a ambdues poblacions [180]. Aquests pèptids van poder ser localitzats inequívocament en la traducció dels transcrits. Tanmateix, les seqüències obtingudes del transcriptoma de la variant del Pacífic de *Bothrops asper* estaven localitzades a l'extrem N-terminal de la proteïna, la traducció de les quals va donar lloc a un fragment de 89 aminoàcids de longitud. La manca de dades provinents del procés de seqüenciació per a la població del Pacífic podria ser la causa per la que només 4 pèptids d'un total de 9 van ser trobats realment en ambdues poblacions tal i com podem veure a la Taula 7.

Fragment peptídic	Població Pèptid	Traducció Carib	Traducció Pacífic
NPLEECFR	Ambdues	NPLEECFR	NPLEECFR
DPGVLEYPVKPSEVGK	Ambdues	DPGVLEYPVKPSEVGK	
FWEDDGIHGGK	Ambdues	FWEDDGIHGGK	FWEDDGIHGGK
EGWYANLPGMR	Ambdues	EGWYANLGPMR	
SAGQLYEESLQK	Ambdues	SAGQLYEESLQK	
IYFAGEYTAQAHGWIDSTIK	Ambdues	IYFAGEYTAQAHGWIDSTIK	
ETDYEEFLEIAK	Ambdues	ETDYEEFLEIAR	ETDYEEFLEIAK
ETLSVTADYVIVCTTSR	Ambdues	ETLSVTADYVIVCTTSR	
ADDRNPLEECFRETD	Ambdues	ADDRNPLEECFRETD	ADDRNPLEECFRETD

Taula 7. Fragments peptídics corresponents a L-aminoàcid oxidases localitzats a dades transcriptòmiques. Fragments peptídics de LAAOs de les poblacions del Pacífic i del Carib de *Bothrops asper* obtinguts mitjançant l'aproximació *venomics* (veure Introducció) per Alape-Girón i col·laboradors, van ser trobats de forma majoritària a la traducció de les seqüències nucleotídiques obtingudes per seqüenciació del transcriptoma d'ambdues poblacions.

SVSPs

Pel que fa les serines proteases, l'estudi de les dades transcriptòmiques va donar lloc a 15 seqüències nucleotídiques per a la població del Carib de *Bothrops asper*, mentre que només una única seqüència es va obtindre de la variant del Pacífic. La distribució desigual de lectures en el procés de seqüenciació entre ambdues poblacions, ja esmentada en el cas de les LAAOs, va donar lloc a un

conjunt incomplet de seqüències nucleotídiques, especialment en la població del Pacífic. Aquesta podria ser la raó per la qual, tal i com veurem a continuació, cap fragment peptídic va trobar una homologia significativa en la traducció de les dades transcriptòmiques obtingudes de la variant del Pacífic.
Tanmateix, cal destacar que l'alineament múltiple d'aquestes seqüències nucleotídiques ens va permetre la visualització de la tríada catalítica característica del lloc actiu de les serines proteases [191] així com de les 12 cisteïnes conservades que participen en la formació de 6 ponts disulfur, ja descrites en la caracterització mitjançant tècniques crital·logràfiques de la TSV-PA [192].
L'anàlisi en profunditat de les seqüències del transcriptoma d'ambdues poblacions, ens va mostrar una homologia del 100% entre els 31 aminoàcids obtinguts de la traducció de la seqüència de la variant del Pacífic i una seqüència de la població del Carib.
Pel que fa la correlació entre el transcriptoma i el proteoma, tal i com hem esmentat, no es va trobar cap homologia significativa entre fragments peptídics obtinguts mitjançant tècniques proteòmiques i la traducció de la seqüència nucleotídica obtinguda per a la població del Pacífic. D'altra banda, els 6 pèptids descrits en ambdós proteomes, els 6 pèptids exclusius de la població del Carib i un parell de pèptids exclusius de la població del Pacífic van poder ser trobats en la traducció de dades transcriptòmiques de la població del Carib.
A la llum d'aquests resultats, les diferències en l'expressió de les serines proteases no semblaven poder ser explicades només per un fenòmen d'especiació, ja que la degradació / inhibició de mRNA podria estar jugant un paper important. A més, el fet que una proteïna restringida a la població del Pacífic mitjançant tècniques proteòmiques aparegui en el transcriptoma de la població del Carib podria apuntar a una regulació post-transcripcional que deuria ser comprovada en d'altres proteïnes.

SVMPs

L'estudi de les dades provinents dels transcriptomes de la població del Carib i del Pacífic de *Bothrops asper* va donar lloc a 29 i a 5

seqüències codificants de SVMPs respectivament. Malgrat la distribució desigual de lectures en ambdues poblacions, ja comentada en el cas de les LAAOs i de les SVSPs, la majoria de lectures en ambdós casos semblaven poder ser agrupades en una única seqüència. D'aquesta manera, a partir dels treballs desenvolupats al Capítol 1, 3,590 lectures d'un total de 6,746 podien ser agupades en una única seqüència d'una SVMPs per a la població del Carib, mentre que 218 lectures d'un total de 276 van poder ser agrupades també en una única seqüència codificant per a una SVMPs en la variant del Pacífic.
Per altra banda, es van obtindre 12 fragments peptídics únics per a la població del Carib, 7 fragments peptídics únics per a la població del Pacífic, i 20 pèptids semblaven ser compartits entre ambdues poblacions.
Tal i com ja hem descrit per a les SVSPs, determinades característiques estructurals de les seqüències de les SVMPs com els residus de cisteïna, el domini peptidasa, el domini disintegrina (amb el loop RGD característic) i el domini semblant a disintegrina (de l'anglès, *disintegrin-like*) amb els residus SECD, entre d'altres motius característics de les SVMPs detallats a la Introducció, van poder ser descrits amb la intenció de trobar diferències significatives en l'expressió de les SVMPs entre les poblacions del Carib i del Pacífic de *Bothrops asper*. La Taula 8 mostra la naturalessa de les SVMPs més representades en ambdós transcriptomes. Tal i com es pot veure, tant el transcriptoma de la variant del Pacífic com de la població del Carib semblaven mostrar una única seqüència de SVMPs tipus PI i diferents seqüències tipus PIII, recolzant la teoria descrita anteriorment al nostre laboratori sobre el model evolutiu que explicaria la diversitat de les disintegrines [193]. Cal destacar que en cap cas es van trobar seqüències codificants de SVMPs tipus PIIb, PIIc, PIId, PIIe ni PIIIa.
Pel que fa la correlació entre el transcriptoma i el proteoma segons la nostra aproximació, 5 pèptids d'un total de 20 trobats al proteoma d'ambdues poblacions no van poder ser localitzats a la traducció de les dades transcriptòmiques de la variant del Pacífic. D'altra banda, 12 pèptids semblaven ser exclusius de la població del Carib per aproximacions proteòmiques. Tanmateix, només 3 d'ells no van poder ser trobats en les dades resultants de la traducció del transcriptoma de

la població del Pacífic. Finalment, tots els pèptids obtinguts de forma exclusiva en la població del Pacífic van poder ser trobats en el transcriptoma de la població del Carib. De fet, 2 d'aquests pèptids no van poder ser realment trobats en el transcriptoma del Pacífic, una indicació de la manca de cobertura de la llibreria de cDNA seqüenciada, tal i com ja hem comentat anteriorment.
Amb tot, la regulació de la transcripció podria estar al darrere del fenotip del verí expressat per les poblacions de *Bothrops asper*. El fet que tots els fragments peptídics van poder ser localitzats mitjançant la nostra aproximació "proteotranscriptòmica" a dades provinents del transcriptoma de la vessant del Carib, juntament amb el fet que es trobessin seqüències de nucleòtids en el transcriptoma de la població del Pacífic per a 9 pèptids únicament trobats al verí de la variant del Carib, podria apuntar a una regulació post-transcripcional intraespecífica més que no pas a una diferència d'arrel gènica.

Població del Carib:

Tipus de SVMPs	**Dis / Dis-like**	**Residus cisteïna**
PI		6
PIII	SECD	6
PIIa	RGD	6
PIII	DDCD	6
PIIIb	SECD	7

Població del Pacífic:

Tipus de SVMPs	**Dis / Dis-like**	**Residus cisteïna**
PI		6
PIII	DDCD	5*
PIIa		4*
PIII		*
PIII	SECD	*

Taula 8. Estudi estructural de les SVMPs trobades al transcriptoma de *Bothrops asper* del Carib i del Pacífic. La nostra aproximació va permetre la classificació de les seqüències nucleotídiques codificants de SVMPs trobades en els transcriptomes de les poblacions del Carib i del Pacífic. Es mostren els residus localitzats a les regions disintegrina o semblants a disintegrina així com el nombre de residus de cisteïna que donen lloc a ponts disulfur característics.
*, la seqüència del transcriptoma no va cobrir la totalitat del domini peptidasa.

PLA2

A partir de les dades transcriptòmiques es van obtindre 9 seqüències d'aminoàcids per a la població del Carib i 4 per a la població del Pacífic. D'aquestes, només dues seqüències (una per a cada població) van mostrar homologia significativa. Per contra, 8 seqüències semblaven ser exclusives de la població del Carib mentre que 3 ho semblaven ser de la població del Pacífic.
Per altra banda, les tècniques proteòmiques van donar lloc a 12 fragments peptídics compartits per ambdues variants mentre que 8 i 6 pèptids van resultar exclusius per a les vessants del Carib i del Pacífic respectivament. Tal i com s'ha descrit per altres toxines, la totalitat dels pèptids identificats van poder ser localitzats en el transcriptoma de l'individu del Carib. Així, tot i que determinades seqüències peptídiques semblaven restringides a la població del Pacífic, el nostre treball va poder trobar els respectius transcrits codificants en el transcriptoma de la població del Carib. En el mateix sentit, només 2 dels 8 pèptids obtinguts en el proteoma de la variant del Carib semblaven ser exclusius d'aquesta població, ja que per a la resta es va poder localitzar un transcrit codificant en el transcriptoma de la població del Pacífic. De fet, tal i com ja s'ha explicat en les toxines anteriors, la manca de seqüències del transcriptoma de la població del Pacífic sembla ser la causa per la que pèptids descrits com a exclusius d'aquesta població no van ser trobats en el seu transcriptoma i sí, en canvi, en el transcriptoma de la població del Carib, posant de manifest, novament, la probable regulació post-transcripcional.

ANNEX II

Fontenay-aux-Roses, 07th of July, 2014

Object : Letter of activity

Dear colleagues,

I contact you as the main author of the newly accepted manuscript titled "**Transcriptomics and venomics: implications for medicinal chemistry**" by F. Ducancel, J. Durban and M. verdenaud, that has been accepted for publication as a Review in Future-Science. More precisely, it will be published in the special Issue "Venomics and drug Design" of Future Medicinal Chemistry. As corresponding author, I have been informed last week of its definitive acceptance for publication.

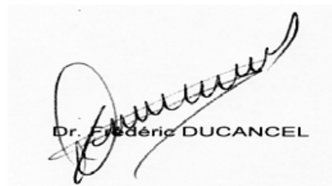
As specialist of molecular evolution of animal venoms compounds, Mr Jordi Durban has greatly contributed to this manuscript, as shown by the chapter dedicated to that topic that he totally written. As requested he wrote that chapter with the aim to illustrate the interest of using and exploiting transcriptomic data to further progress in our knowledge of venom compounds molecular evolution.

Mr Jordi Durban also accepted to carry out an internal review of the manuscript, identifying few points that have been corrected and/or completed.

To summarize, I can tell you, that Mr Jordi Durban actively contributed to that Review in a very professional manner, revealing an exhaustive knowledge of the domain.

I remain to your disposal if necessary

Regards,



Review

Transcriptomics and venomics: implications for medicinal chemistry

Frédéric Ducancel[*1], Jordi Durban[2] & Marion Verdenaud[3]

[1]Department of ImmunoVirology, iMETI/DSV/CEA, Bt 02, 18 rue du Panorama, BP 6, 92265, Fontenay-aux-Roses, France

[2]Instituto de Biomedicina de València, Jaume Roig, 11, 46010, València, Spain

[3]Department of Pharmacology & Immunoanalysis, iBiTec-Saclay/DSV/CEA, Bt 152, CEA de Saclay, 91191 Gif sur Yvette Cedex, France

[*]Author for correspondence:

Tel.: +33 01 46 54 83 18

frederic.ducancel@cea.fr

Abstract

Over the last three decades, transcriptomic studies of venom gland cells have continuously evolved, opening up new possibilities for exploring the molecular diversity of animal venoms, a prerequisite for the discovery of new drug candidates and molecular phylogenetics. The molecular complexity of animal venoms is much greater than initially thought. In this review, we describe the different technologies available for transcriptomic studies of venom, from the original individual cloning approaches to the more recent global Next Generation Sequencing strategies. Our understanding of animal venoms is evolving, with the discovery of complex and diverse bio-optimized cocktails of compounds, including mostly peptides and proteins, which are now beginning to be studied by academic and industrial researchers.

Defined key terms	
Biogenic amines	Biogenic amines consist of naturally biologically active enzymatic products that contain one or more primary amine groups, such as norepinephrine, histamine and serotonin. They act primarily as neurotransmitters.
Isoforms	An isoform is a natural mutant/variant of a peptide/protein of reference, which differs from that 'reference' sequence by a minimum of one amino acid. All these mutants/variants/isoforms belong to the same structural and functional family.
	Next Generation Sequencing technologies do not use the classical Sanger's strategy of DNA

NGS	sequencing. To date they correspond mainly to 454-pyrosequencing (Roche), to paired-end (Illumina-Solexa) or to semiconductor (Ion-Torrent™) sequencing strategies.
cDNA	Complementary DNA is generated by the reverse transcription of messenger RNAs (mRNAs) extracted from a tissue or pool of cells of interest. Messengers RNAs correspond to copies of genes that are translated via the ribosome machinery into polymers of amino acids: peptides or proteins. As such mRNAs are also named 'precursors'.
Synonymous vs nonsynonymous substitutions	A synonymous substitution corresponds to a silent mutation of a codon that results in an unchanged amino acid. On the other hand, a nonsynonymous substitution is an amino change within a given sequence that results from a nonsilent mutation of the corresponding codon.
Pseudogenes	Pseudogenes are genomic ubiquitous and abundant nonfunctional DNA sequences similar to normal and active genes. They are identified during genome annotation process and contain different type of modifications (mutations/insertions/deletions, etc) that result in their nonfunctionality. It is recognized that some of them play essential role in gene regulation of their parent and functional genes.

What is the natural venom resource?

Animal venoms are complex cocktails of several hundreds of components, most of which (~ 90%) are proteins or peptides [1]. Venom compounds are characterized by their capacity to recognize various targets, such as enzymes, ion channels and receptors. Their interaction with these targets results in direct or indirect effects on cell integrity, the central and/or peripheral nervous system, muscles and blood flow. This potential for biological activity is seen as a consequence of their continual refinement by natural evolution. Venoms are thus natural reservoirs containing many bioactive molecules that have been selected and recruited for their secretion, structural stability, functional plasticity and capacity to engage in precise molecular interactions with their targets (affinity, specificity and selectivity). The remaining 10% of venom components include various organic components, such as sugars, salts, amino acids, **biogenic amines**, nucleotides, nucleosides and nucleic acids (mRNAs, DNAs).

Recent technological developments in mass spectrometry for venom-based explorations (proteomic studies) and venom-gland gene expression profiling (transcriptomic studies) have revolutionized our perception of the nature of animal venoms. Indeed, the use of these approaches, either separately or together, has led to the discovery of hundreds of peptides and proteins in the venom of single venomous species, with cone snails and spiders appearing to possess the most diverse cocktails or 'arsenals' of compounds [2].

The global animal venom resource can therefore be seen as a collection of more than 40,000,000 compounds, mainly peptides and proteins, of which only ~ 3000 are known and have been studied to any extent. Additional levels of venom diversity have also been observed, due to intraspecific variations of the venom components [3-5]. Such variation may occur between specimens of a defined species as a result of changes in biotope, diet, age, development, sex, season or geographic location. Differences have also been observed between dissected and milked (injected) venoms. Finally, one intriguing observation from proteomic studies is the co-existence (within a single venom) of numerous 'related-sequences' of the same compound, of different lengths and concentrations. It remains unclear whether these longer and shorter venom molecule '**isoforms**' have any biological significance. It has become clear that venom composition varies considerably during the life cycle of venomous animals and that these changes may affect the pharmacology and toxicity of the venom.

These findings have greatly modified the toxinologist's view of venom gland systems, which now appear to be sophisticated, efficient and highly dynamic structures containing the venom itself, the venom-gland cells and associated

tissues (i.e., the salivary glands in cone snails). Together with increases in our capacity to synthesize peptides and recombinant proteins, these changes have opened up new perspectives for the study and use of this huge natural resource by academic researchers and industry.

Venoms: natural medicinal chemistry libraries & therapeutic potential

The amazing potential of venoms for drug discovery renders them of particular interest to the pharmaceutical industry, which is in dire need of innovation. In recent years, the successful release onto the market of therapeutic antibodies, proteins and peptides has increased interest in 'biologics' as potential novel drugs. There are currently 51 therapeutic peptides on the market and more than 100 natural (and a similar number of modified) therapeutic proteins have been approved for clinical use in the European Union and the USA [6]. In the domains of clinical and biotechnological applications, venom compounds in general, and peptides/toxins in particular, are naturally bio-optimized molecular tools that can be used for studies of their targets, potentially leading to the identification and development of novel candidate therapeutic molecules. Indeed, many toxin targets are involved in various human diseases, such as pain, cancer, neurodegenerative and cardiovascular diseases, diabetes, obesity and depression. Venoms contain stable bioactive molecules of high affinity, target subtype selectivity and large pharmacological spectra, yielding many promising candidates for innovative drug leads. Venom compounds constitute one of the most promising families of compounds for use in the diagnosis and treatment of human disease, due to their functional activities, small size, low immunogenicity and high stability, and the development of powerful strategies for their chemical synthesis or recombinant production. Furthermore, venom peptides and proteins bind to their targets through a large number of interactions, resulting in a decrease in the 'capacity' of the targeted system to escape and to 'resist' these ligands, through mutations affecting contact areas, for example (as observed in antibiotic resistance). Finally, the scaffold can be engineered for the design of compounds with modified biochemical, functional or biophysical properties, the adaptation of molecules for particular uses and for their labeling for *in vivo* imaging, for vectorization, or for the functionalization of nanoparticles [7-9].

Few peptide drugs of venom origin are currently available, but significant developments are occurring in the fields of pain, infection and cancer [10-12]. The analgesic ziconotide (Prialt®) is a peptide from marine snail (*Conus magus)* that was approved by the US FDA in 2004 and used as the last resort in the treatment of severe pain in patients refractory to morphine. Other drugs derived from snake-venom proteins are used in the control of hypertension and blood hemostasis. These drugs include the antihypertensive drug captopril (Capoten®), which mimics the action of the angiotensin-converting enzyme inhibitor peptide from a viper venom. On the other hand, eptifibatide (Integrilin®) and tirofiban (Aggrastat®) are used in the treatment and prevention of acute coronary syndrome when iodinated chlorotoxin from scorpion venom (TM-601®) has successfully undergone Phase II trials for the targeted treatment of glioma, a diffuse form of brain cancer. XEN2174 is an analgesic peptide derived from χ-conotoxin MrIA, which has successfully completed Phase IIa clinical trials for cancer-related pain and is currently undergoing testing in Phase IIb trials for postoperative pain. Many other venom peptides may one day lead to new drugs and more than 400 patents relating to venom peptides have been filed. Several peptides, mostly conotoxins, have reached preclinical or clinical trial stages. This field of research is very active, with rapid advances driven in particular by Australian groups in particular (University of Queensland and Xenome Ltd). A seminal development has recently occurred with the demonstration of oral activity in a rodent pain model for α-conotoxin Vc1.1, a nicotinic receptor

antagonist.

Venom proteins are also used for therapeutic applications. Several snake-venom proteins with pro-, anti-coagulant or fibrinolytic activities have found uses in medical applications. For instance, ancrod (Viprinex®), from the Malayan pit viper *Calloselasma rhodostoma,* is a defibrinogenating molecule currently under investigation as a possible treatment for stroke. Batroxobin (Pefakit® Reptilase®Time) from *Bothrops atrox* is used for investigations of the final phase of blood coagulation. Due to its heparin insensitivity, batroxabin can detect fibrinogen polymerization disorders even in the presence of heparin. Tirobifan is an antiplatelet drug derived from an *Echis carinatus* venom compound. Other proteins with anticoagulant properties from the venoms of the snakes *Bothrops atrox*, *Echis carinatus*, *Oxyuranus scutellatus* and *Daboia russelii* are used to diagnose coagulation disorders and to monitor anticoagulant treatment. Botrocetin is a strong platelet-aggregating protein found in *Bothrops atrox* venom. It is used for the diagnosis of several hemorrhagic vascular diseases of genetic origin, such as von Willebrand disease.

General strategies for venom profiling

Given the tremendous diversity and potential for application of venom compounds, major efforts are required to characterize the molecular and structural diversities of this natural repertoire of bioactive molecules [13]. Traditionally, such characterization has been based on the fractionation of venoms coupled with bioassay screening, followed by the purification and characterization of bioactive compounds. This process is time-consuming, limiting the extent of exploration possible and leading to a focus on the most abundant molecules. In addition, 'bioassay-guided' drug exploration is hampered by the availability of material, sample size (most venomous animals are small or very small) and the complexity and variability of venoms.

Sequence-driven approaches, based on mass spectrometry analyses with or without the cloning and sequencing of precursors [14-22], have recently been developed. These approaches were initially based on the transcriptomic exploration of venom glands with expressed sequence tag (EST) technology, but several groups have since successfully applied Next Generation Sequencing **(NGS)** strategies [23]. These highly sensitive and powerful strategies for precursor sequencing have facilitated more global and accurate transcriptomic explorations, paving the way for improvements in our understanding of the true molecular and structural diversity of the cocktails of molecules comprising animal venoms. NGS strategies, in addition to providing information about previously known families of venom compounds, are also, for the first time, providing access to the sequences of low abundance precursors. More importantly, exhaustive transcriptomics combined with annotation tools for bioinformatics is providing researchers with access to the unexplored 'Eldorado' of 'unknown' precursor sequences (encoding compounds with no equivalent in the available databases), which may account for as many as 20–40% of all venom compounds! These pools of precursors will undoubtedly include new sequences with original folding patterns associated with novel biological activities and target specificities.

Venom-gland cell transcriptomics: from individual cloning to global view

Individual cloning of animal toxin precursors

Toru Tamiya *et al.* was the first group to clone and sequence a precursor (mRNA) encoding an animal toxin, in 1985 [24]. This mRNA encoded a short-chain neurotoxin, 65 amino acids in length, stabilized by four disulfide bridges. This peptide,

erabutoxin a, is synthesized by a sea snake, *Laticauda semifasciata* (Figure 1). The next precursor to be studied was that of a phospholipase A_2 (PLA_2) from the venom of the sea snake *Laticauda laticaudata* [25]. Since these pioneering works, many animal toxin precursors have been cloned and sequenced, revealing both common and distinctive structural elements leading to different precursor organizations (Figure 2). Most animal toxin precursors are encoded by monocistronic sequences, with a single venom molecule encoded by a single mRNA. However, a few precursor sequences display a polycistronic organization, with long open reading frames encoding several toxins [26-30]. The proteins encoded may be isoforms of the same family, as for sarafotoxins [26], when precursors from *Bothrops jararaca* and *Lachesis muta muta*, revealed the presence of seven bradykinin-potentiating peptides, together with one sequence encoding a C-type natriuretic peptide [27,28]. Polycistronic precursors encoding antimicrobial/cytolytic peptides and neurotoxins have also been cloned from the spider *Lachesana tarabaevi* [29] and the sea anemone *Antheopsis maculata* [30], respectively.

```
 TCCGAAAAAGATCGCAAG ATG AAA ACT CTG CTG CTG ACC TTG GTG GTG GTG  51
                     M   K   T   L   L   L   T   L   V   V   V
ACA ATC GTG TGC CTG GAC TTA GGA TAC ACC AGG ATA TGT TTT AAC CAT  99
 T   I   V   C   L   D   L   G   Y   T   R   I   C   F   N   H
CAG TCA TCG CAA CCG CAA ACC ACT AAA ACT TGT TCA CCT GGG GAG AGC  147
 Q   S   S   Q   P   Q   T   T   K   T   C   S   P   G   E   S
TCT TGC TAT AAC AAG CAA TGG AGC GAT TTC CGT GGA ACT ATA ATT GAA  195
 S   C   Y   N   K   Q   W   S   D   F   R   G   T   I   I   E
AGG GGA TGT GGT TGC CCC ACA GTG AAG CCC GGT ATT AAA CTC AGT TGT  243
 R   G   C   G   C   P   T   V   K   P   G   I   K   L   S   C
TGC GAA TCA GAG GTC TGC AAC AAT TAG CTCTACGAGTGGCTAAATTCCTTGAGT  297
 C   E   S   E   V   C   N   N  stop
TTTACTCTCATTCATCAAGGACCATCCTTCAAAATGTATGCTTCTGGCCTTTACCACCACATG  360
GTCCATCATCCCCCTCTCCCCTGCTGTCTTTGACACCTCAACATCTTTCCCTTTTCTCTTGAT  423
CTGTAAGTTTCCTTCTGCTAGTTCTGTAGTTTGAGAATCAAATAAACCTCAGCATTCAAAAAA  486
AAAAAAAAAAAAAAAAAAAAA                                            507
```

Figure 1. Precursor encoding erabutatoxin a (Ea) from *Laticauda semifasciata*. The signal peptide is underlined. The polyadenylation site is underlined. Taken with permission from [24].

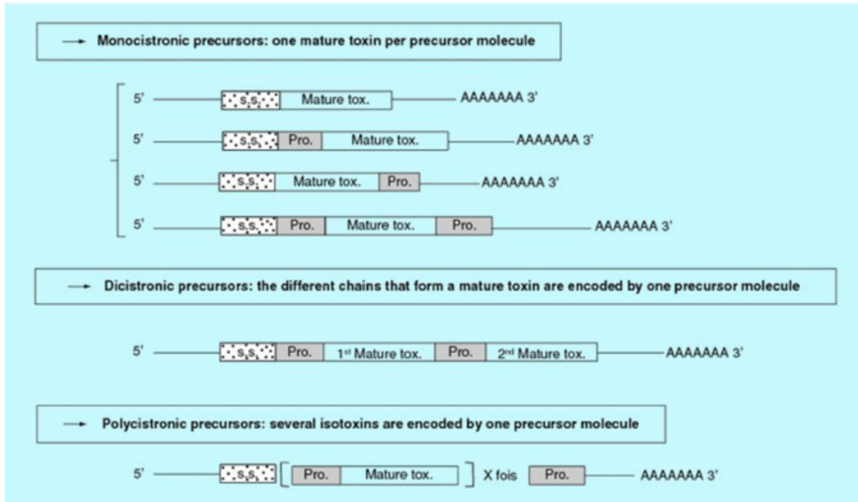


Figure 2. Animal toxin precursors. Overall structures and organizations of the different mRNA precursor molecules encoding animal toxins. The signal peptide sequence at the 5′ end of the open reading frame is indicated by 'S.S'. The propeptide sequences, if present, are indicated by 'pro' and are shown in light gray. The 5′ and 3′ untranslated sequences are represented as thin lines. The polyadenylated tail at the 3′ end of the precursors is indicated by 'AAAAAAA'.

Toward global transcriptomic studies of venom-gland cells: the EST strategy

EST analysis for transcriptomics first appeared toward the end of the 20th century. ESTs are short (10–400 base pairs) single DNA sequencing reads obtained from genomic DNA or complementary DNA (**cDNA**) libraries. They are obtained by the classical Sanger sequencing method and correspond to one-shot partial sequences of the 5′ and/or 3′ ends of cloned precursors. ESTs are useful for more global investigations of the gene-expression patterns and overall transcriptomic activity of tissues or cells (e.g., venom gland cells).

The first EST library for venom gland cells was described in 2001 [31]. In this pioneering work, the authors aimed to investigate the mechanisms by which cone snail had evolved. They analyzed and compared 170 different conopeptide-encoding ESTs from five different species of *Conus* snails. They demonstrated particularly rapid rates of nucleotide substitution within the sequences encoding the signal peptide, propeptide and mature peptide; a bias, with transversions favored over transitions in nucleotide substitutions (replacement of purines with pyrimidines or vice versa); the presence of cysteine codons in specific positions within the hypervariable regions; and a preponderance of **nonsynonymous over synonymous substitutions** in the mature peptide.

Many EST-based transcriptomic studies have since been carried out on various venomous animals: jellyfish [32], spiders [33-38], snakes [39-48], cone snails [49,50], fish [51], wasps [52], scorpions [53], centipedes [54], and more recently, a venomous ant [55]. In all cases, EST annotation led to the identification of several new isoforms of previously known families of peptide or protein toxins, revealing a predominance of some families of venom components over others. Interestingly, the EST annotation of nucleic acid and protein sequence databases also revealed the existence of two new

transverse (found in all venomous animal species) categories of venom-gland components. The first corresponds to sequences that have already been cloned, but the functions of which remain unknown. These sequences account for less than 10% of the annotated ESTs on average and, in many cases, they display similarity to sequences for components found in other tissues and phyla unrelated to venomous animals or to the venom apparatus. This observation has since been confirmed by NGS approaches (see below). These sequences probably correspond either to common cellular components or (in the case of secreted and disulfide-rich sequences) to particular peptide or protein scaffolds preferentially recruited by the venom gland to increase its molecular diversity [1,2].

Another particularly exciting discovery has been the detection, by EST-based strategies, of venom-apparatus-specific cellular and molecular actors displaying no sequence matches, for nucleotide or amino-acid sequences, with any sequence present in current databases. These molecules include compounds contributing to the originality and molecular specificity of the molecular machinery of the cells constituting the venom-gland tissues. Nevertheless, several ESTs are probably components of precursors encoding new peptides or proteins contributing to the richness of the venom arsenal developed by venomous animals. Future challenges in this field will include the development of tools and strategies for the exploration and study of these new sequences. Undoubtedly, these sequences will include some corresponding to new scaffolds and biological activities of potential relevance for new clinical applications.

However, although EST sequencing constitutes significant progress with respect to the single-precursor cloning approach, it is subject to several limitations. The exploration of gene expression remains partial, with information gleaned mostly about the more abundant families of compounds. Such studies remain time-consuming and expensive, due to the sequencing technology used (Sanger sequencing). Furthermore, for long precursors, such as those encoding venom enzymes, the coverage of cloned sequences by ESTs is far from total, and classical second-step cloning by PCR amplification is required to obtain the complete precursor sequence. However, in such cases, the PCR primers are mostly deduced from 5′ and 3′ EST sequences.

Toward global transcriptomic studies of venom-gland cells: NGS approaches

In 2005, an alternative DNA-sequencing technology (pyrosequencing) facilitating rapid, large-scale sequencing at low cost was described for the first time [56]. Various other approaches have since emerged and are grouped together under the umbrella term 'NGS' [57,58]. NGS uses various strategies to sequence DNA more efficiently than the traditional dideoxynucleotide method pioneered by Sanger [59]. NGS was initially developed for genomic sequencing projects, but was rapidly adopted for use in studies of the gene expression profiles (transcriptomics) of various tissues. For this purpose, mRNAs are extracted from venom-gland cells, converted into cDNAs, fragmented and sequenced. Whatever be the NGS technology used, the sequenced fragments or 'reads' (35 to 300–400 bp long) must be assembled into 'contigs', which are ultimately analyzed and annotated with bioinformatics labels. NGS techniques are high-throughput, with millions of sequencing reactions carried out in parallel.

Unlike previously described transcriptomic approaches, NGS techniques deliver a much broader view of the cocktail of compounds present in venom, as demonstrated by the increasing number of recent studies involving these approaches [60-82]. Over and above the compounds present in venom, these techniques can also be used to study the molecular machinery specifically used by the venom-gland cells in the production of the venom arsenal and to elucidate the post-transcriptional

mechanisms at work. In terms of candidate drug discovery, deep-transcriptomics studies based on NGS are likely to revolutionize this field in terms of the discovery of new isoforms, the identification of toxin-related genes suggesting convergent recruitment by the venom-gland tissues of compounds from diverse taxa [83], the discovery of totally new compounds/scaffolds and the identification of a molecular signature specific to products of the venom-gland cell.

NGS was first used to explore the transcriptomic activity of an animal venom apparatus in 2009. This study concerned the common emperor scorpion, *Pandinus imperator* [60]. Many other venom transcriptomes have since been determined, in snakes, cones snails, spiders, two venomous mammals, one venomous crustacean and one ant. Most of these studies involved pyrosequencing with Roche 454 or 454 GS FLX machines, the Illumina GA/HiSeq System, the Illumina system together with 454 technology and, very recently, Ion-Torrent™ technology (see Table 1). As expected, these studies have resulted in deeper gene expression (transcriptomic) profiling of venom-gland cells. Researchers now have access to a very large range of precursor representations, which was not the case with previous technologies. Thanks to these advances, it is now possible to detect and sequence transcripts present at very low abundance (a few copies only). It has also become possible to generate exhaustive inventories of the molecular diversity characterizing families of known toxin peptides. For example, for *Conus consors*, we were able to show the predominance in the venom of three families of conopeptides: superfamilies A, O and M, with 132, 40 and 28 different isoforms, respectively [63]. Within the same transcriptome, precursors encoding conolysins, conantokins, contulakins, conotoxins P, S or T, conodipins and conopressins were detected even when there were fewer than four copies of the precursor present.

Table 1. Venomous animal species for which venom-gland Next Generation Sequencing transcriptomic analyses have been reported.

Animal species	454	454 GS FLX	Illumina	Ion Torrent	Reference
Snakes					
Crotalus adamanteus		X			(61)
		X	X		(66)
Ovophis okinavensis *Protobothrops flavoviridis*			X		(77)
8 Costa Rican snakes		X			(62)
Atractaspis aterrima		X			(69)
Crotalus horidus			X		(70)
Crotalus simus simus		X			(72)
Micrurus fluvius			X		(78)
Cone snails					
Conus consors		X			(63)
Conus pulicarius	X				(64)
Conus geographus		X			(65)
Conus miles		X			(71)
Conus marmoreus		X			(73)
Conus victoriae	X				(82)
Scorpions					
Pandinus imperator	X				(60)
Centruroides noxius		X			(67)
5 Australian scorpions		X			(74)
Spiders					
Latrodectus tredecimguttatus			X		(75)
Selenotypus plumipes			X		(80)
Gasteracantha arcuata *Gasteracantha hasselti* *Nasoonaria sinensis*			X		(79)
Mammals					
Ornithorhynchus anatinus *Tachyglosus aculeatus*			X		(76)
Crustaceans					
Speleonectes tulumensis			X		(68)
Ants					
Dinaponera quadriceps				X	(81)
Ticks					
Ixodes ricinus		X	X		(86)

Clearly, NGS technologies provide a more complete picture of the composition of the cocktail of chemicals in venom. They have shown that venom-gland cells synthesize an average of 200–400 different precursors, encoding peptides and proteins that are secreted into the lumen of the venom gland. Levels of transcriptomic activity may vary between studies. Intraspecific and interspecific variations are observed, but gene expression profiles may also differ between specimens of the same defined species of venomous animal.

As pointed out above, venoms appear to have a much broader molecular content than initially suspected, raising questions about the origin of this diversity. Deep transcriptomic analyses, particularly for *Conus*, have suggested a highly dynamic process of sequence diversification. One interesting study on *Conus miles* [71] focused on conopeptides and led to the identification of more than 650 putative conopeptide precursors. Durban *et al.* [72] studied the process driving venom evolution in the snake *Crotalus simus simus*, revealing the role played by certain populations of miRNAs (microRNAs) as modulators of the ontogenetic composition of venoms.

NGS technologies are also used to shed some light on animals not considered truly venomous. Thus, several tick species produce highly paralytic and lethal cocktails of proteinaceous molecules in their salivary glands [84]. Detrimental effects of tick bites, such as paralysis, allergic reactions and pathogen transmission, have been reported in both animals and humans. Tick saliva has clearly evolved to contain a complex cocktail of components counteracting the effects of the host immune system, including anticoagulants, prostaglandins, immunosuppressants, antihistamines and prostacyclin and calreticulins [85]. Furthermore, fatal cases of human envenomation have been reported in Australia. A transcriptome analysis was recently carried out with a combination of 454 GS-FLX and Illumina NGS technologies [86], to improve our understanding of tick saliva. This study established that the saliva-gland transcriptome of the Australian paralysis tick, *Ixodes ricinus,* contained housekeeping proteins (~ 23%), secreted proteins (~ 13%) and unknown compounds (~ 60%). The secreted proteins included enzymes, protease inhibitors (basic tail and Kunitz domain families), lipocalins, ixostatins, antimicrobial peptides and 14 other different families of compounds. Together, these results confirm the huge diversity of molecules present in tick saliva, highlighting the importance of studies of this fluid.

Finally, a general feature of these recent NGS transcriptomes is the presence of a large proportion (ranging from 20 to 40%) of reads and contigs corresponding to previously unknown sequences. None of these sequences match any of the sequences present in currently available databases. This discovery, initially made in studies based on ESTs (see above), is clearly one of the most exciting results generated by NGS. The challenge now is for researchers to explore these new 'repertoires', which undoubtedly contain precursors encoding venom compounds. Given the huge numbers of new sequences discovered in this way, toxinologists and bioinformaticians need to join forces, combining their expertise, knowledge and capacities in the development of new, efficient exploration algorithms. One possible strategy for identifying the precursors

encoding venom compounds would involve making use of the molecular and structural features characteristic of venom-compound precursors (see above and Figure 2). This approach would involve searching for and identifying signal peptides, cysteine-rich sequences and propeptide-like sequences. For example, the concomitant presence of cysteine residues and a signal sequence within a precursor, with or without N- and/or C-terminal putative propeptide sequences, should be considered as strong indications that the precursor encodes a venom peptide or protein. Cross-referencing with proteomics data will be useful, to confirm this prediction and to identify the start sites of mature sequences more precisely. On the basis of such results, candidate sequences could then be synthesized chemically or produced by recombinant technology, for assessments of their toxicity, biological activity and structure. This is the overall philosophy followed in the ongoing European project 'VENOMICS' presented below.

VENOMICS: an ambitious European project (2011–2015)

VENOMICS is a European FP7-Health project dedicated to the exploration of biodiversity for public health (http://www.venomics.eu/). It aims to explore animal venom compounds, with a view to the identification and development of novel biotherapeutics.

Bioassay-guided approaches have classically been used for drug discovery in venoms. However, this low-throughput strategy requires large amounts of venom and focuses principally on compounds abundant in venom. VENOMICS makes use of an innovative 'Omics' workflow involving cutting-edge high-throughput transcriptomics, proteomics and peptide production technologies to decipher venom diversity (Figure 3).

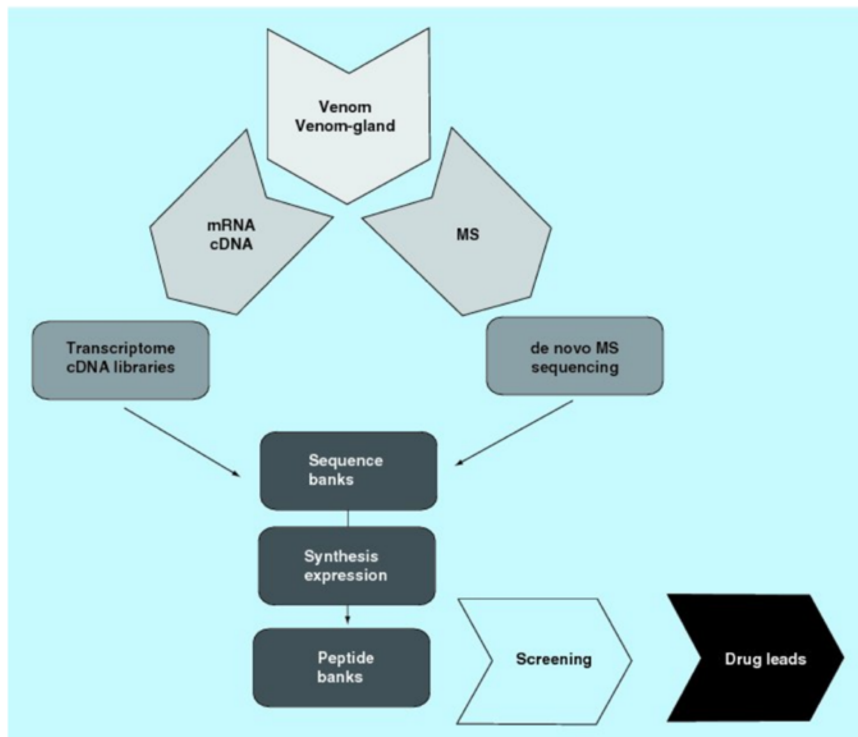


Figure 3. General overview of the 'Omics-based high-throughput lead generation' process developed in the European VENOMICS project.

The core objective of VENOMICS is to recreate *in vitro* collections of venom peptides that can be used as a resource for high-throughput screening, for the more efficient isolation of novel drug leads. This approach is akin to recreating 'synthetic venoms' in the laboratory. VENOMICS aims to reach this goal, by generating venoms and venom-gland biobanks, corresponding to 200 venomous animal species; sequencing venom peptides by proteomics and transcriptomics (Illumina technology) approaches, to create a database of 50,000 sequences for mature venom proteins and peptides; high-throughput *in vitro* chemical synthesis or the recombinant expression of selected peptides, to generate a bank of several thousand peptides; and pharmacological screening of the peptide bank against selected molecular targets and drug lead generation.

This strategy represents a new paradigm for venom-based drug discovery, differentiating the VENOMICS approach from the classical bioassay-guided process. VENOMICS focuses on disulfide bridge-rich venom peptides no longer than about 100 amino acids long. Candidate peptides of less than 40 amino acids in length are generated by chemical synthesis, whereas larger peptides are produced principally by bacteria, with recombinant technologies.

Venom-derived peptide libraries offer the advantage of containing only highly stable natural bioactive and bio-optimized molecules of high affinity and target subtype selectivity. They therefore have a much greater potential for drug discovery than the randomly generated peptide libraries obtained by phage display or combinatorial chemistry techniques. The

VENOMICS consortium consists of research laboratories and SMEs from Belgium, Denmark, France, Portugal and Spain.

Transcriptomics & molecular phylogenetic studies

Phylogenetics is used to study or trace the evolutionary history underlying biological diversity in groups of organisms. It has been said that "Nothing in biology makes sense except in the light of evolution" (Theodosius Grygorovych Dobzhansky) [87]. Phylogenetics therefore plays a key role in various areas of biology, including population genetics, ecology and animal behaviour, but also in the clinical and medical contexts, giving rise to what has been called 'evolutionary medicine' [88]. Phylogeny has been used to determine the origin and spread of a contagious disease from a molecular epidemiology standpoint [89,90] and to understand the adaptive evolution of viral pathogens, to facilitate vaccine design [91]. However, one of the most interesting fields in which phylogenetics is proving increasingly valuable is pharmaceutical research for the identification of new drugs or natural products. Wang *et al.* [92] used this approach to identify specific peptides from the tammar wallaby that effectively killed multidrug-resistant bacteria. Phylogenetic analysis could also be used to find genes with a common phylogenetic profile involved in a similar biological pathway or sharing a similar biochemical function. Komatsu *et al.* [93] used this approach to determine which species of *Panax* were most closely related to other medicinal species and might therefore have similar medicinal qualities.

In this context, the FP7-Health project VENOMICS aims to explore venoms from non-model organisms, to identify new therapeutic compounds. As described above, the profiling of those organisms, for which genome sequences are unavailable, has been based on transcriptomic data in particular, adding fuel to the controversy between morphologists and molecular biologists in the field of phylogenetic systematics [94,95], since classical phylogenetic approaches used morphological data to determine taxonomic relationships. Nevertheless, molecular data are becoming a valuable source of information and, although taxonomy is still based largely on Linnaean principles and morphological characters, information from DNA and proteins has been used to call into question previous taxonomic classifications based purely on morphological traits.

Studies of the changes in gene expression underlying phenotypic divergence have successfully increased our understanding of transcriptome evolution in several organisms, including mammals [96], fishes [97], mosquitoes [98], molluscs [99], turtles [100] and plants [101]. In systematics, information of this kind can be used to clarify unexpected evolutionary relationships, such as those for Colubridae, Viperidae and Elapidae snake families. However, in the context of the VENOMICS project, drug discovery could benefit from phylogenetic studies of the transcriptomes of venomous species [69], because venom toxins probably evolved from proteins with normal physiological function, and comparative phylogenetics might provide clues about disease-related proteins.

Analyses of transcriptome data for non-model organisms for phylogenetic purposes have been driven by the rapid development of sequencing technology. Several phylogenetic surveys [102-106] have been performed on ESTs. At this point, it should be pointed out that ESTs obtained from different taxa cannot contain overlapping genes, given the low probability of finding orthologous sequences in a reduced set of sequences. This makes it more difficult to attain comparative phylogenetics goals.

Since NGS technologies have become easily affordable due to the drop of per base sequencing costs, RNA-Seq [107], i.e. the characterization of the complete set of transcripts by massive parallel sequencing processes, is becoming a highly valuable tool for exploring the complexity of organisms at the genome-wide scale, giving rise to the field of phylogenomics

[108]. This field, as the intersection of evolution and genomics, involves the inference of the phylogenetic history of certain organisms from genome-wide data. Chan *et al.* [109] coined the term 'next-generation phylogenomics' for such large-scale phylogenetics sequencing projects, and Lin *et al.* [110] recently performed a phylogenomic analysis of subterranean mammals, using four *de novo* RNA-seq libraries.

Since the first analysis with 454 pyrosequencing methods for phylogenetic purposes performed by Roeding *et al.* [60], this technology has been applied for phylogeographic purposes by McCormack *et al.* [111] and by Rokyta *et al.* [112], in order to describe the positive selection (reflected by nonsynonymous substitutions) imposed on most venom proteins in *Crotalus adamanteus*. Similarly, Dutertre *et al.* [113] recently showed that *Conus geographus* has defence-evoked and predation-evoked venoms, and that the conotoxins found in these two different types of venoms have evolved rapidly under positive Darwinian selection. However, improvements in the read length, cost per Mb sequenced, total throughput and speed of Illumina technology have made this platform the technology of choice and it has recently been used in several phylogenetic studies [110,114,115].

Given NGS transcriptomic data from a non-model organism, it would be interesting to identify the specific features to be taken into account for phylogenetic surveys. Transcriptome sequences generated with high-throughput techniques provide a rich set of characters for phylogenetic studies in eukaryotes. However, the alignment of multiple partial sequences might result in an alignment with large numbers of gaps, potentially compromising the conclusions of any phylogenetic study [116]. It remains unclear to what extent the reliability of tree reconstruction is increased by maximizing either the number of taxa or the number of characters studied. McCormack *et al.* [111] recently concluded that the loci identified from 454 pyrosequencing data could be useful for phylogenetics, population genetics or phylogeographic purposes, particularly for closely related species. Despite being incomplete, most of the 454 sequences were useful for determining which toxin protein sequence had been identified in snake-venom-gland transcriptomes. According to Wiens [117], missing data are less critical than complete characters tied to other taxa in the tree for resolving phylogenetic relationships, and limited taxonomic sampling could be problematic in the downstream phylogenetic analysis. An increase in the number of taxa sampled could, therefore, increase confidence levels for the assignment of certain orthologues.

Several features should also be taken into account when carrying out phylogenetic analyses with transcriptome data:

-Downstream bioinformatics analyses of NGS data cannot currently identify potential **pseudogene** sequences, relics of the evolution that might lead to incorrect conclusions about phylogenetic relationships.

-Gene trees and species trees are not the same [118]. If there were duplications or polymorphic alleles, then phylogenies for genes will not match those for organisms. In such cases, the phylogenetic reconstruction of orthologous sequences could be useful, to generate the species tree.

-Conclusions for nuclear genes may conflict with those for mitochondrial genes. Moreover, it seems that the combined use of mitochondrial and nuclear sequences yields better results, without artifacts for nodes for which mitochondrial and nuclear gene datasets used separately generate conflicting topologies [119].

Conclusion & future perspective: venom-apparatus transcriptomics

Our understanding of animal venoms is clearly evolving, with the revelation that venoms are complex and diverse bio-

optimized cocktails of compounds, mostly peptides and proteins, that we are only just beginning to explore. These changes in our perception are closely linked to technological advances and one of the current difficulties facing researchers is establishing ways to analyze, interpret and valorize the huge numbers of sequences being generated.

Efficient data analysis will require the use of bioinformatics knowhow and concepts, together with the development from scratch of new strategies and scripts for the exploration of 'unknown' fractions in particular. One of the difficulties is identifying, as precisely as possible, the true N-terminus of the mature compounds encoded by the totally new precursor sequences emerging from NGS transcriptomic studies. With this goal in mind, the combination of this approach with proteomic analyses of the corresponding venoms may facilitate identification, but only for compounds present at sufficiently high concentrations in the venom studied.

Valorization will require high-throughput strategies for the synthesis or production of compounds of interest in their native forms. This implies a knowledge of the exact amino-acid sequences of these molecules, the number and location of the disulfide bridges, and the nature and position of post-translational modifications, when present. Again, the combination of these techniques with proteomics appears to be the most appropriate strategy. In the case of a totally new scaffold, one key issue is the native fold adopted by the molecule of interest. If present in sufficiently large amounts in the venom, the new compound could be extracted, purified and studied by X-ray diffraction or nuclear magnetic resonance. Compounds present at too low a concentration for this approach could be produced by recombinant technologies, by following a strategy allowing the formation of disulfide bridges *in vivo* before the initiation of 3D structure studies.

The development of more systematic and automated strategies for screening for biological activity is also an important issue, for the identification of new biotherapeutic hits. In particular, the development of straightforward screening tools compatible with high-throughput technologies is crucial, to increase the chances of successful exploration. Two principal approaches appear pertinent: concentrating the screening of compound diversity on one target of interest or presenting the libraries of molecules to be tested to libraries of targets. The hits identified in this way then enter a phase of exhaustive exploration of their biochemical, functional, structural and biological activities. These studies will, in many instances, require the mutation, engineering of labeling of compounds of interest.

In conclusion, although the techniques for exploring the resources provided by venomous animals are maturing, these analyses remain time-consuming and must be combined with complementary approaches. Nevertheless, the exploration of **large** and **naturally bio-optimized** libraries of compounds, such as those produced by the venom-gland systems, clearly constitutes a major advance in medicinal biochemistry.

Executive summary

- Given the existence of several thousand of venomous animal species, the venom resource is huge, opening tremendous perspectives in the fields of fundamental research and development of therapeutics.
- Technological developments such as high-throughput proteomics and transcriptomics result in a more complex molecular vision of venoms that appear to be highly diverse cocktails of peptides and proteins mainly whose composition is susceptible to vary upon different criteria or stimuli.
- Resulting from a long process of molecular evolution and selection, venom peptides and proteins are naturally bio-optimized scaffolds.

- Due their functional activities, reduced sizes, low immunogenicity and their high stability, venom peptides constitute one of the most promising families of compounds for use in the diagnosis and treatment of human diseases.
- Venom peptide drugs are associated to significant developments occurring in the fields of pain, infection and cancer, when venom proteins target mainly the human cardiovascular system.
- The study of the transcriptomic activity of venom-gland cells has started in 1985 with individual cloning of toxin precursors, to reach today global descriptions of the high activity of synthesis of venom-gland cells.
- Expressed sequence tags and then Next Generation Sequencing strategies are responsible for that evolution.
- Aside previously known families of venom peptides and proteins, these more global approaches have revealed new groups of compounds among which several display totally new sequences whose activity and structure are unknown to date! Exploration of these 'Unknown' fractions require the development of new bioinformatics tools to tentatively identify precursors encoding venom components.
- VENOMICS is an ambitious international/European project that aims at applying high-throughput technologies to explore the molecular diversity of venoms with the objective of reproducing artificially a part of it for drug-candidate screening.
- Next Generation Sequencing data require new strategies of sequence alignments leading to next-generation phylogenomics.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as: • of interest; ••• of considerable interest

1 Casewell NR, Wüster W, Freek J *et al.* Complex cocktails: the evolutionary novelty of venoms. *Trends Ecol. Evol.* **28**(4), 219–229 (2013).

•• A review that illustrates how, through the development of 'omic' technologies, our perception of the venom systems is evolving.

2 Fry BG, Scheib H, van der Weerd L *et al.* Evolution of an arsenal. *Mol. Cell. Proteomics* **7**(2), 215–246 (2008).

•• An illustration of how toxin scaffolds have been recruited and evolved from endogenous families of peptides or proteins.

3 Abdel-Rahman MA, Omran MA, Bdel-Nabi IM *et al.* Intraspecific variation in the Egyptian scorpion *Scorpio mauruspalmatus* venom collected from different biotopes. *Toxicon* **53**, 349–359 (2009).

4 Creer S, Malhotra A, Thorpe RS *et al.* Genetic and ecological correlates of intra-specific variation in pitviper venom composition detected using matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF-MS) and isoelectric focusing. *J. Mol. Evolution* **56**, 317–329 (2003).

5 Dutertre S, Biass D, Stöcklin R *et al.* Dramatic intraspecimen variations within the injected venom of *Conus consors*: an unsuspected contribution to venom diversity. *Toxicon* **55**, 1453–1462 (2010).

6 Kola I. The state of innovation in drug development. *Clin. Pharmacol. Ther.* **83**(2), 227–230 (2008).

7 Rennert R, Neundorf I, Beck-Sickinger AG. Synthesis and application of peptides as drug carriers. *Methods Mol. Biol.* **535**, 389–403

(2009).

8 Shaji J, Patole V. Protein and peptide drug delivery: oral approaches. *Indian J. Pharm. Sci.* **70**(3), 269–277 (2008).

9 Hayashi MAF, Ducancel F, Konno K. Natural peptides with potential applications in drug development, diagnosis, and/or biotechnology. *Int. J. Pept.* doi: 10.1155/2012/757838 (2012) (Epub ahead of print).

10 Lewis RJ, Garcia ML. Therapeutic potential of venom peptides. *Nat. Rev. Drug Discov.* **2**, 790–802 (2003).

•• A survey of the huge pharmacology diversity of venom peptides and their therapeutic potentials.

11 Cossins D. From toxins to therapeutics. *The Scientist* March 19, (2013).

12 King GF. Venoms as a platform for human drugs: translating toxins into therapeutics. *Expert Opin. Biol. Ther.* **11**(11), 1469–1484 (2011).

•• An illustration of the great interest of using selected toxin scaffolds to identify and develop therapeutics.

13 Calvete JJ. Snake venomics: from the inventory of toxins to biology. *Toxicon* **75**, 44–62 (2013).

14 Escoubas P, Sollod B, King GF. Venom landscapes: mining the complexicity of spider venoms via a combined cDNA and mass spectrometric approach. *Toxicon* **47**(6), 650–663 (2006).

15 Calvete JJ, Juarez P, Sanz L. Snake venomics. Strategy and applications. *J. Mass Spectrom.* **42**(11), 1405–1414 (2007).

16 Escoubas P, King GF. Venomics as a drug discovery platform. *Expert Rev. Proteomics* **6**(3), 221–224 (2009).

17 Calvete JJ. Venomics, what else? *Toxicon* **60**(4), 427–433 (2012).

18 Prashanth JR, Lewis RJ, Dutertre S. Towards an integrated venomics approach for accelerated conopeptide discovery. *Toxicon* **60**(4), 470–477 (2012).

19 Favreau P, Menin L, Michalet S *et al.* Mass spectrometry strategies for venom mapping and peptide sequencing from crude venoms: case applications with single arthropod specimen. *Toxicon* **47**, 676–687 (2006).

20 Rodrigues RS, Boldrini-França J, Fonseca FPP *et al.* Combined snake venomics and venom gland transcriptomic analysis of *Bothropoides pauloensis*. *J. Proteomics* **75**, 2707–2730 (2012).

21 Margres MJ, McGivern JJ, Wray KP *et al.* Linking the trascriptome and proteome to characterize the venom of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *J. Proteomics* **96**, 145–158 (2014).

22 Calvete JJ. Snake venomics: from the inventory of toxins to biology. *Toxicon* **75**, 44–62 (2013).

23 Schuster SC. Next-generation sequencing transforms today's biology. *Nat. Methods* **5**, 16–18 (2008).

• A statement of the different technological evolutions that make transcriptomics to enter into a new age.

24 Tamiya T, Lamouroux A, Julien JF *et al.* Cloning and sequence analysis of the cDNA encoding a snake neurotoxin precursor. *Biochimie* **67**(2), 185–189 (1985).

25 Guignery-Frelat G, Ducancel F, Ménez A *et al.* Sequence of a cDNA encoding a snake venom phospholipase A2. *Nucleic Acids Res.* **15**(14), 5892 (1987).

26 Ducancel F. Endothelin-like peptides. *Cell Mol. Life Sci.* **62**(23), 2828–2839 (2005).

27 Murayama N, Hayashi MA, Ohi H *et al.* Cloning and sequence analysis of a *Bothrops jararaca* cDNA encoding a precursor of seven bradykinin-potentiating peptides and a C-type natriuretic peptide. *Proc. Natl. Acad. Sci. USA* **94**(4), 1189–1193 (1997).

28 Soares MR, Oliveira-Carvalho AL, Wermelinger LS *et al.* Identification of novel bradykinin-potentiating peptides and C-type natriuretic peptides from *Lachesis muta muta* venom. *Toxicon* **46**(1), 31–38 (2005).

29 Kozlov SA, Vassilevski AA, Foefanov AV *et al.* Latarcins, antimicrobial and cytolytic peptides from the venom of the spider *Lachesana tarabaevi* (Zodariidae) that exemplify biomolecular diversity. *J. Biol. Chem.* **281**(30), 20983–20992 (2006).

30 Honma T, Hasegawa Y, Ishida M *et al.* Isolation and molecular cloning of novel peptide toxins from the sea-anemone *Antheopsis maculata*. *Toxicon* **45**(1), 33–41 (2005).

31 Conticello SG, Gilad Y, Avidan N *et al.* Mechanisms for evolving hypervariability: the case of conopeptides. *Mol. Biol. Evol.* **18**(2), 120–131 (2001).

32 Yang Y, Cun S, Xie X *et al.* EST analysis of gene expression in the tentacle of *Cyanea capillata*. *FEBS Lett.* **538**(12), 183–191 (2003).

33 Kozlov S, Malyavka A, McCutchen B *et al.* A novel strategy for the identification of toxin-like structures in spider venom. *Proteins* **59**(1), 131–140 (2005).

34 Jiang L, Peng L, Chen J *et al.* Molecular diversification based on analysis of expressed sequence tags from the venom glands of the Chinese bird spider *Ornithoctonus huwena*. *Toxicon* **51**(8), 1479–1489 (2008).

35 Jiang L, Liu C, Duan Z *et al.* Transcriptome analysis of venom glands from a single fishing spider *Dolomedes mizhoanus*. *Toxicon* **73**, 23–32 (2013).

36 Chen J, Zhao L, Jiang L *et al.* Transcriptome analysis revealed novel possible venom components and cellular processes of the tarantula *Chilobrachys jingzhao* venom gland. *Toxicon* **52**(7), 794–806 (2008).

37 Tang X, Zhang Y, Hu W *et al.* Molecular diversification of peptide toxins from the tarantula *Haplopelma hainanum* (*Ornithoctonus hainana*) venom based on transcriptomic, peptidomic and genomic analyses. *J. Proteome Res.* **9**(5), 2550–2564 (2010).

38 McCowan C, Garb JE. Recruitment and diversification of an ecdysozoan family of neuropeptide hormones for the black widow spider venom expression. *Gene* **536**(2), 366–375 (2014).

39 Wagstaff SC, Harrison RA. Venom gland EST analysis of the saw-scaled viper *Echis ocellatus*, reveals novel alpha9beta1 integrin-binding motifs in venom metalloproteinases and a new group of putative toxins, renin-like aspartic proteases. *Gene* **377**, 21–32 (2006).

40 Junqueira-de-Azevedo IL, Ching AT, Carvalho E *et al. Lachesis muta* (Viperidae) cDNAs reveal diverging pit viper molecules and scaffolds typical of cobra (Elapidae) venoms: implications for snake toxin repertoire evolution. *Genetics* **173**(2), 877–889 (2006).

41 Cidade DA, Simão TA, Dávilla AM *et al. Bothrops jararaca* venom gland transcriptome: analysis of the gene expression pattern. *Toxicon* **48**(4), 437–461 (2006).

42 Jia Y, Cantu BA, Sánchez EE *et al.* Complementary DNA sequencing and identification of mRNAs from the venomous gland of *Agkistrodon piscivorus leucostoma*. *Toxicon* **51**(8), 1457–1466 (2008).

43 Jiang Y, Li Y, Lee W *et al.* Venom gland transcriptomes of two elapid snakes (*Bungarus multicinctus* and *Naja atra*) and evolution of toxin genes. *BMC Genomics* **12**, 1. doi: 10.1186/1471–2164–2112–2111 (2011) (Epub ahead of print).

44 Rodrigues RS, Boldrini-França J, Fonseca FPP *et al.* Combined snake venomics and venom gland transcriptomic analysis of Bothropoidespauloensis. *J. Proteomics* **75**(9), 2707–2720 (2012).

45 Chatrath ST, Chapeurouge A, Lin Q *et al.* Identification of novel proteins from the venom of a cryptic snake *Drysdalia coronoides* by a combined transcriptomics and proteomics approaches. *J. Proteome Res.* **10**(2), 739–750 (2011).

46 Zelanis A, Andrade-Silva D, Rocha MM *et al.* A transcriptomic view of the proteome variability of newborn and adult *Bothrops jararaca* snake venoms. *PLoS Negl. Trop. Dis.* **6**(3), e1554 (2012).

47 Ching ATC, Rocha MMT, PaesLeme AF *et al.* Some aspects of the venom proteome of the Colubridae snake *Philodryas olfersii* revealed from a Duvernoy's (venom) gland transcriptome. *FEBS Lett.* **580**(18), 4417–4422 (2006).

48 Ching ATC, PaesLeme AF, Zelanis A *et al.* Venomics profiling of *Thamnodynastes strigatus* unveils matrix metalloproteinases and other novel proteins recruited to toxin arsenal of rear-fanged snakes. *J. Proteome Res.* **11**(2), 1152–1162 (2012).

49 Pi C, Liu J, Peng C *et al.* Diversity and evolution of conotoxins based on gene expression profiling of *Conus litteratus*. *Genomics* **88**(6), 809–819 (2006).

50 Pi C, Liu J, Peng C *et al.* Analysis of expressed sequence tags from the venom ducts of *Conus striatus*: focusing on the expression profile of conotoxins. *Biochimie* **88**(2), 131–140 (2006).

51 Magalhães GS, Junqueira-de-Azevedo IL, Lopes-Ferreira M *et al.* Transcriptome analysis of expressed sequence tags from the venom glands of the fish *Thalassophryne nattereri*. *Biochimie* **88**(6), 693–699 (2006).

52 Baek JH, Lee SH. Identification and characterization of venom proteins of two solitary wasps, *Eumenes pomiformis* and *Orancistrocerus drewseni*. *Toxicon* **56**(4), 554–562 (2010).

53 Abdel-Rahman MA, Quintero-Hernandez V, Possani LD. Venom proteomic and venomous glands transcriptomic analysis of the Egyptian scorpion *Scorpion maurus palmatus* (Arachnida: Scorpionidae). *Toxicon* **74**, 193–207 (2013).

54 Liu ZC, Zhang R, Zhao F *et al.* Venomic and transcriptomic analysis of centipede *Scolopendra subspinipes dehaani*. *J. Proteome Res.* **11**(12), 6197–6212 (2012).

55 Bouzid W, Klopp C, Verdenaud M *et al.* Profiling the venom gland transcriptome of *Tetramorium bicarinatum* (Hymenoptera: Formicidae): the first transcriptome analysis of an ant species. *Toxicon* **70**, 70–81 (2013).

56 Margulies M, Egholm M, Altman WE *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**(15), 376–380.

57 Liu L, Li Y, Li S *et al.* Comparison of next-generation sequencing systems. *J. Biomed. Biotech.* 251364 doi: 10.1155/2012/251364. Epub 5 (2012) (Epub ahead of print).

58 Mardis ER. Next-Generation Sequencing Platforms. *Annu. Rev. Anal. Chem.* **6**, 287–303 (2013).

59 Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467

(1977).

60 Roeding F, Borner J, Kube M *et al.* A 454 sequencing approach for large-scale phylogenomic analysis of the common emperor scorpion (*Pandinus imperator*). *Mol. Phylogen. Evol.* **53**, 826–834 (2009).

61 Rokyta DR, Wray KP, Lemmon AR *et al.* A high-troughput venom-gland transcriptome for the Eastern diamondback rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes. *Toxicon* **57**, 657–671 (2011).

62 Durban J, Juarez P, Angulo Y *et al.* Profiling the venom gland transcriptomes of Costa Rican snakes by 454 pyrosequencing. *BMC Genomics* **12**, 259–275 (2011).

63 Terrat Y, Biass D, Dutertre S *et al.* High-resolution picture of a venom gland transcriptome: case study with the marine snail *Conus consors*. *Toxicon* , **59**, 34–46 (2012).

64 Lluisma AO, Milash BA, Moore B *et al.* Novel venom peptides from the cone snail *Conus pulicarius* discovered through next-generation sequencing of its venom duct transcriptome. *Mar. Genomics* **5**, 43–51 (2012).

65 Hu H, Bandyopadhyay PD, Olivera BM *et al.* Elucidation of the molecular envenomation strategy of the cone snail *Conus geographus* through transcriptome sequencing of its venom duct. *BMC Genomics* **13**, 284–296 (2012).

66 Rokyta DR, Lemmon AR, Margres MJ *et al.* The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics* **13**, 312 (2012).

67 Rendón-Anaya M, Delaye L, Possani LD *et al.* Global transcriptome analysis of the scorpion *Centruroides noxius*: new toxin families and evolutionary insights from an ancestral scorpion species. *PLoS One* **7**(8), e43331 (2012).

68 vonReumont BJ, Blanke A, Richter S *et al.* The first venomous crustacean revealed by transcriptomics and functional morphology: remipede venom glands express a unique toxin cocktail dominated by enzymes and a neurotoxin. *Mol. Biol. Evol.* doi: 10. 1093 Nov 7 (2013) (Epub ahead of print).

69 Terrat Y, Sunagar K, Fry BG *et al.* *Atractaspis aterrima* toxins: the first insight into the molecular evolution of venom in side-stabbers. *Toxins (Basel)* **5**(11), 1948–1964. doi: 10.3390/toxins5111948 (2013) (Epub ahead of print).

70 Rokyta DR, Wray KP, Margres MJ. The genesis of an exceptionally lethal venom in the timber rattlesnake (*Crotalus horridus*) revealed through comparative venom-gland transcriptomics. *BMC Genomics* **14**, 394 (2013).

71 Jin AH, Dutertre S, Kaas Q *et al.* Transcriptomic messiness in the venom duct of *Conus miles* contributes to conotoxin diversity. *Mol. Cell. Proteomics* **12**(12), 3824–3833 (2013).

72 Durban J, Pérez A, Sanz L *et al.* Integrated 'omics' profiling indicates that miRNAs are modulators of the ontogenetic venom composition shift in the Central American rattlesnake, *Crotalus simus simus*. *BMC Genomics* **14**, 234 (2013).

73 Dutertre S, Jin A, Kaas Q *et al.* Deep venomics reveals the mechanism for expanded peptide diversity in cone snail venom. *Mol. Cell. Proteomics* **12**(2), 312–329. doi: 10.1074/mcp.M112.021469. Epub 14. (2013) (Epub ahead of print).

74 Sunagar K, Undheim EAD, Chan AHC *et al.* Evolution stings: the origin and diversification of scorpion toxin peptides scaffolds. *Toxins (Basel)* **5**(12), 2456–2487. doi: 10.3390/toxins5122456 (2013) (Epub ahead of print).

75 He Q, Duan Z, Yu Y *et al.* The venom gland transcriptome of *Latrodectus tredecimguttatus* revealed by deep sequencing and cDNA library analysis. *PLoS One* **8**(11), e81357 (2013).

76 Wong ESW, Nicol S, Warren WC *et al.* Echidna venom gland trasncriptome provides insights into the evolution of monotreme venom. *PLoS One* **8**(11), e79092 (2013).

77 Aird SD, Watanabe Y, Villar-Briones A *et al.* Quantitative high-throughput profiling of snake venom gland transcriptomes and proteomes (*Ovophis okinavensis* and *Protobothrops flavoviridis*). *BMC Genomics* **14**, 790 (2013).

78 Margres MJ, Aronow K, Loyacano J *et al.* The venom-gland transcriptome of the eastern coral snake (*Micrurus fulvius*) reveals high venom complexity in the intragenomic evolution of venoms. *BMC Genomics* **14**, 531 (2013).

79 Zhao YJ, Zeng Y, Chen L *et al.* Analysis of transcriptomes of three orb-web spider species reveals gene profiles involved in silk and toxin. *Insect Sci.* doi: 10.1111/1747917.12068 (2013) (Epub ahead of print).

80 Wong ESW, Hardy MC, Wood D *et al.* SVM-based prediction of propeptide cleavage sites in the spider toxins identities toxin innovation in an Australian tarantula. *PLoS One* **8**(7), e66279 (2013).

81 Torres AF, Huang C, Chong CM *et al.* Transcriptome analysis in venom gland of the predatory giant ant *Dinoponera quadriceps*: insights into the polypeptide toxin arsenal of Hymenopterans. *PLoS One* **9**(1), e87556 (2014).

82 Robinson SD, Safavi-Hemami H, McIntosh LD *et al.* Diversity of conotoxin gene superfamilies in the venomous snail, *Conus victoria*. *PLoS One* **9**(2), e87648 (2014).

83 Fry BG, Roelants K, Champagne DE *et al.* The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annu. Rev. Genomics Hum. Genet.* **10**, 483–511 (2009).

84 Mans BJ, Louw AI, Neitz AW. Biochemical perspectives on paralysis and other forms of toxicoses causes by ticks. *Parasitology*

129(Suppl), S95–S111 (2004).

85 Steen NA, Barker SC, Alewood PF. Proteins in the saliva of the Ixodida (ticks): pharmacological features and biological significance. *Toxicon* **47**, 1–20 (2006).

86 Schwarz A, von Reumont BM, Erhart J *et al. De novo Ixodesricinus* salivary gland transcriptome analysis using two next-generation sequencing methodologies. *FASEB J.* **27**, 4745–4756 (2013).

87 Dobzhansky T. Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teach.* **35**, 125–129 (1973).

88 Abu-Asab M, Chaouchi M, Amri H. Evolutionary medicine: A meaningful connection between omics, disease, and treatment. *Proteomics Clin. Appl.* **2**, 122–134 (2008).

89 Segovia M, Carrasco HJ, Martinez CE *et al.* Molecular epidemiologic source tracking of orally transmitted Chagas disease, Venezuela. *Emerg. Infect. Dis.* **19**, 1098–1101 (2013).

90 Holmes EC. Molecular epidemiology and evolution of emerging infectious diseases. *Br. Med. Bull.* **54**, 533–543 (1998).

91 Ojosnegros S, Beerenwinkel N. Models of RNA virus evolution and their roles in vaccine design. *Immunome Res.* **6**, S5 (2010).

92 Wang J, Wong ES, Whitley JC *et al.* Ancient antimicrobial peptides kill antibiotic-resistant pathogens: Australian mammals provide new options. *PLoS ONE* **6**, e24030 (2011).

93 Komatsu K, Zhu S, Fushimi H *et al.* Phylogenetic analysis based on 18S rRNA gene and matK gene sequences of Panaxvietnamensis and five related species. *Planta Med.* **67**, 461–465 (2001).

94 Patterson C. In *Molecules and Morphology in Evolution: Conflict or Compromise?* Colin P (Ed.) Cambridge University Press, NY , USA (1987).

95 Crowe TM. Molecules vs morphology in phylogenetics: anon-controversy. *Trans. R. Soc. South Afr.* **46**, 317–334 (1988).

96 Brawand D, Soumillon M, Necsulea A *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).

97 Goetz F, Rosauer D, Sitar S *et al.* A genetic basis for the phenotypic differentiation between siscowet and lean lake trout (*Salvelinus namaycush*). *Mol. Ecol.* **19**, 176–196 (2010).

98 Hittinger CT, Johnston M, Tossberg JT *et al.* Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. *Proc. Natl. Acad. Sci. USA* 200910449 (2010) doi:10.1073/pnas.0910449107 (Epub ahead of print).

99 Smith SA, Wilson NG, Goetz FE *et al.* Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature* **480**, 364–367 (2011).

100 Chiari Y, Cahais V, Galtier N *et al.* Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biol.* **10**, 65 (2012).

101 Wen J, Xiong Z, Nie ZL *et al.* Transcriptome sequences resolve deep relationships of the grape family. *PLoS ONE* **8**, e74394 (2013).

102 Bapteste E, Brinkmann H, Lee JA *et al.* The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium, Entamoeba*, and Mastigamoeba. *Proc. Natl. Acad. Sci. USA* **99**, 1414–1419 (2002).

103 Hughes J, Longhorn SJ, Papadopoulou A *et al.* Dense taxonomic EST sampling and its applications for molecular systematics of the Coleoptera (Beetles). *Mol. Biol. Evol.* **23**, 268–278 (2006).

104 Bourlat SJ, Juliusdottir T, Lowe CJ *et al.* Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444**, 85–88 (2006).

105 Roeding F, Hagner-Holler S, Ruhberg H *et al.* EST sequencing of Onychophora and phylogenomic analysis of Metazoa. *Mol. Phylogenet. Evol.* **45**, 942–951 (2007).

106 Sukkapan P, Jia Y, Nuchprayoon I *et al.* Phylogenetic analysis of serine proteases from Russell's viper (*Daboia russelli siamensis*) and *Agkistrodon piscivorus leucostoma* venom. *Toxicon* **58**, 168–178 (2011).

107 Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).

108 Delsuc F, Brinkmann H, Philippe H. Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* **6**, 361–375 (2005).

109 Chan CX, Ragan MA. Next-generation phylogenomics. *Biol. Direct* **8**(3), 1745–6150 (2013).

•• An argumentation of how next-generation sequencing data requires new strategies of sequence alignments leading to next-generation phylogenomics.

110 Lin GH, Wang K, Deng XG *et al.* Transcriptome sequencing and phylogenomic resolution within Spalacidae (Rodentia). *BMC Genomics* **15**, 32 doi: 10.1186/1471–2164 (2014) (Epub ahead of print).

111 McCormack JE, Maley JM, Hird SM *et al.* Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. *Mol. Phylogenet. Evol.* **62**, 397–406 (2012).

112 Rokyta DR, Wray KP, Lemmon AR, Lemmon EM, Caudle SB. A high-throughput venom-gland transcriptome for the eastern diamondback rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes. *Toxicon* **57**, 657–671 (2011).

113 Dutertre S, Jin AH, Vetter I *et al.* Evolution of separate predation- and defence-evoked venoms in carnivorous cone snails. *Nat. Commun.* **5**, 3521, doi: 10.1038/ncomms4521 (2014) (Epub ahead of print).

• A molecular demonstration of how cone snails are able to rapidly switch between distinct venoms cocktails in response to predatory or defensive stimuli.

114 Schwarz A, Cabezas-Cruz A, Kopecky J *et al.* Understanding the evolutionary structural variability and target specificity of tick salivary Kunitz peptides using next generation transcriptome data. *BMC Evol. Biol.* **14**, 4, doi: 10.1186/1471-2148 (2014) (Epub ahead of print).

115 Li P, Deng W, Li T *et al.* Illumina-xbased *de novo* transcriptome sequencing and analysis of *Amanita exitialis* basidiocarps. *Gene* **532**, 63–71 (2013).

116 Hartmann S, Vision TJ. Using ESTs for phylogenomics: Can one accurately infer a phylogenetic tree from a gappy alignment? *BMC Evol. Biol.* **8**, 95 (2008).

117 Wiens JJ. Missing data and the design of phylogenetic analyses. *J. Biomed. Inform.* **39**, 34–42 (2006).

118 Nichols R. Gene trees and species trees are not the same. *Trends Ecol. Evol.* **16**, 358–364 (2001).

119 Caravas J. Phylogenetic utility of mitochondrial and nuclear genes: a case study in the Diptera (true flies). *Wayne State Univ. Diss.* (2012). http://digitalcommons.wayne.edu/oa_dissertations/4

REFERÈNCIES

1. Kelly CMR, Barker NP, Villet MH: **Phylogenetics of advanced snakes (Caenophidia) based on four mitochondrial genes**. *Syst Biol* 2003, **52**:439–459.

2. Pyron RA, Burbrink FT, Colli GR, de Oca ANM, Vitt LJ, Kuczynski CA, Wiens JJ: **The phylogeny of advanced snakes (Colubroidea), with discovery of a new subfamily and comparison of support methods for likelihood trees**. *Mol Phylogenet Evol* 2011, **58**:329–342.

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

4. Lawson R, Slowinski JB, Crother BI, Burbrink FT: **Phylogeny of the Colubroidea (Serpentes): new evidence from mitochondrial and nuclear genes**. *Mol Phylogenet Evol* 2005, **37**:581–601.

5. Pyron RA, Burbrink FT, Wiens JJ: **A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes**. *BMC Evol Biol* 2013, **13**:93.

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

7. Uetz, P, Etzold, T: **The EMBL/EBI Reptile Database**. *Herpetol Rev 27* 1996:174–175.

8. Lomonte B, Fernández J, Sanz L, Angulo Y, Sasa M, Gutiérrez JM, Calvete JJ: **Venomous snakes of Costa Rica: Biological and medical implications of their venom proteomic profiles analyzed through the strategy of snake venomics**. *J Proteomics* 2014.

9. Vonk FJ, Casewell NR, Henkel CV, Heimberg AM, Jansen HJ, McCleary RJR, Kerkkamp HME, Vos RA, Guerreiro I, Calvete JJ, Wüster W, Woods AE, Logan JM, Harrison RA, Castoe TA, de

Koning APJ, Pollock DD, Yandell M, Calderon D, Renjifo C, Currier RB, Salgado D, Pla D, Sanz L, Hyder AS, Ribeiro JMC, Arntzen JW, van den Thillart GEEJM, Boetzer M, Pirovano W, et al.: **The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system**. *Proc Natl Acad Sci U S A* 2013, **110**:20651–20656.

10. Quijada-Mascareñas, A., Wüster, W.: **Recent Advances in Venomous Snake Systematics**. In *Handb Venoms Toxins Reptil*. CRC Press; 2010:25 – 64.

11. Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDA, King GF, Nevalainen TJ, Norman JA, Lewis RJ, Norton RS, Renjifo C, de la Vega RCR: **The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms**. *Annu Rev Genomics Hum Genet* 2009, **10**:483–511.

12. Fry BG, Casewell NR, Wüster W, Vidal N, Young B, Jackson TNW: **The structural and functional diversification of the Toxicofera reptile venom system**. *Toxicon Off J Int Soc Toxinology* 2012, **60**:434–448.

13. Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG: **Complex cocktails: the evolutionary novelty of venoms**. *Trends Ecol Evol* 2013, **28**:219–229.

14. 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, Kochva E: **Early evolution of the venom system in lizards and snakes**. *Nature* 2006, **439**:584–588.

15. Vidal N, Hedges SB: **The phylogeny of squamate reptiles (lizards, snakes, and amphisbaenians) inferred from nine nuclear protein-coding genes**. *C R Biol* 2005, **328**:1000–1008.

16. Calvete JJ, Escolano J, Sanz L: **Snake venomics of Bitis species reveals large intragenus venom toxin composition variation:**

application to taxonomy of congeneric taxa. *J Proteome Res* 2007, **6**:2732–2745.

17. Markland FS: **Snake venoms and the hemostatic system**. *Toxicon Off J Int Soc Toxinology* 1998, **36**:1749–1800.

18. Fox JW, Serrano SMT: **Snake toxins and hemostasis**. *Toxicon* 2005, **45**:949. [*Snake Toxins and Hemostasis*]

19. Chippaux JP, Williams V, White J: **Snake venom variability: methods of study, results and interpretation**. *Toxicon Off J Int Soc Toxinology* 1991, **29**:1279–1303.

20. Barlow A, Pook CE, Harrison RA, Wüster W: **Coevolution of diet and prey-specific venom activity supports the role of selection in snake venom evolution**. *Proc Biol Sci* 2009, **276**:2443–2449.

21. Daltry JC, Wüster W, Thorpe RS: **Diet and snake venom evolution**. *Nature* 1996, **379**:537–540.

22. Gibbs HL, Sanz L, Chiucchi JE, Farrell TM, Calvete JJ: **Proteomic analysis of ontogenetic and diet-related changes in venom composition of juvenile and adult Dusky Pigmy rattlesnakes (Sistrurus miliarius barbouri)**. *J Proteomics* 2011, **74**:2169–2179.

23. Gibbs HL, Sanz L, Sovic MG, Calvete JJ: **Phylogeny-based comparative analysis of venom proteome variation in a clade of rattlesnakes (Sistrurus sp.)**. *PloS One* 2013, **8**:e67220.

24. Casewell NR, Wagstaff SC, Wüster W, Cook DAN, Bolton FMS, King SI, Pla D, Sanz L, Calvete JJ, Harrison RA: **Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms**. *Proc Natl Acad Sci U S A* 2014, **111**:9205–9210.

25. Casewell NR, Huttley GA, Wüster W: **Dynamic evolution of venom proteins in squamate reptiles**. *Nat Commun* 2012, **3**:1066.

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

27. Fry BG, Vidal N, van der Weerd L, Kochva E, Renjifo C: **Evolution and diversification of the Toxicofera reptile venom system**. *J Proteomics* 2009, **72**:127–136.

28. Ohno M, Ménez R, Ogawa T, Danse JM, Shimohigashi Y, Fromen C, Ducancel F, Zinn-Justin S, Le Du MH, Boulain JC, Tamiya T, Ménez A: **Molecular evolution of snake toxins: is the functional diversity of snake toxins associated with a mechanism of accelerated evolution?**. *Prog Nucleic Acid Res Mol Biol* 1998, **59**:307–364.

29. Ohno,M., Ogawa,T., Oda-Ueda,N., Chijiwa, T., Hattori, S.: **Accelerated and regional evolution of snake venom-gland isozyme**. In *Perspect Mol Toxinology*. New York, USA: Wiley; 2002:387–400.

30. Kordis D, Gubensek F: **Adaptive evolution of animal toxin multigene families**. *Gene* 2000, **261**:43–52.

31. Lynch VJ: **Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes**. *BMC Evol Biol* 2007, **7**:2.

32. Ogawa T, Chijiwa T, Oda-Ueda N, Ohno M: **Molecular diversity and accelerated evolution of C-type lectin-like proteins from snake venom**. *Toxicon Off J Int Soc Toxinology* 2005, **45**:1–14.

33. Fox JW, Serrano SMT: **Structural considerations of the snake venom metalloproteinases, key members of the M12 reprolysin family of metalloproteinases**. *Toxicon Off J Int Soc Toxinology* 2005, **45**:969–985.

34. Casewell NR: **On the ancestral recruitment of metalloproteinases into the venom of snakes**. *Toxicon Off J Int Soc Toxinology* 2012, **60**:449–454.

35. SATAKE M, MURATA Y, SUZUKI T: **Studies on snake venom. XIII. Chromatographic separation and properties of three proteinases from Agkistrodon halys blomhoffii venom**. *J Biochem (Tokyo)* 1963, **53**:438–447.

36. 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.

37. Kini RM: *Venom Phospholipase A2 Enzymes: Structure, Function, and Mechanism*. John Wiley; 1997.

38. Six DA, Dennis EA: **The expanding superfamily of phospholipase A(2) enzymes: classification and characterization**. *Biochim Biophys Acta* 2000, **1488**:1–19.

39. Guo C, Liu S, Yao Y, Zhang Q, Sun M-Z: **Past decade study of snake venom L-amino acid oxidase**. *Toxicon Off J Int Soc Toxinology* 2012, **60**:302–311.

40. Don J. Phillips, Stephen D. Swenson, Francis S. Markland, Jr: **Thrombin-like Snake Venom Serine Proteinases**. In *Handb Venoms Toxins Reptil*. CRC Press/Taylor & Francis Group. Boca Raton; 2009:139–154.

41. Gulland JM, Jackson EM: **5-Nucleotidase**. *Biochem J* 1938, **32**:597–601.

42. Aird SD: **Ophidian envenomation strategies and the role of purines**. *Toxicon Off J Int Soc Toxinology* 2002, **40**:335–393.

43. Caccin P, Pellegatti P, Fernandez J, Vono M, Cintra-Francischinelli M, Lomonte B, Gutiérrez JM, Di Virgilio F, Montecucco C: **Why myotoxin-containing snake venoms possess powerful nucleotidases?**. *Biochem Biophys Res Commun* 2013, **430**:1289–1293.

44. Pawlak J, Manjunatha Kini R: **Snake venom glutaminyl cyclase**. *Toxicon* 2006, **48**:278–286.

45. Georgieva D, Ohler M, Seifert J, von Bergen M, Arni RK, Genov N, Betzel C: **Snake venomic of Crotalus durissus terrificus--correlation with pharmacological activities**. *J Proteome Res* 2010, **9**:2302–2316.

46. Ferreira SH, Greene LJ, Alabaster VA, Bakhle YS, Vane JR: **Activity of Various Fractions of Bradykinin Potentiating Factor against Angiotensin I Converting Enzyme**. *Nature* 1970, **225**:379–380.

47. Ferreira SH: **A bradykinin-potentiating factor (bpf) present in the venom of bothrops jararaca**. *Br J Pharmacol Chemother* 1965, **24**:163–169.

48. Camargo ACM, Ianzer D, Guerreiro JR, Serrano SMT: **Bradykinin-potentiating peptides: beyond captopril**. *Toxicon Off J Int Soc Toxinology* 2012, **59**:516–523.

49. Juárez P, Comas I, González-Candelas F, Calvete JJ: **Evolution of snake venom disintegrins by positive Darwinian selection**. *Mol Biol Evol* 2008, **25**:2391–2407.

50. 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 (Mosc)* 2002, **41**:14248–14254.

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

52. Clemetson KJ: **Snaclecs (snake C-type lectins) that inhibit or activate platelets by binding to receptors**. *Toxicon Off J Int Soc Toxinology* 2010, **56**:1236–1246.

53. Arlinghaus FT, Eble JA: **C-type lectin-like proteins from snake venoms**. *Toxicon Off J Int Soc Toxinology* 2012, **60**:512–519.

54. Cohen S, Levi-Montalcini R: **A NERVE GROWTH-STIMULATING FACTOR ISOLATED FROM SNAKE VENOM**. *Proc Natl Acad Sci U S A* 1956, **42**:571–574.

55. Hogue-Angeletti RA, Frazier WA, Jacobs JW, Niall HD, Bradshaw RA: **Purification, characterization, and partial amino acid sequence of nerve growth factor from cobra venom**. *Biochemistry (Mosc)* 1976, **15**:26–34.

56. McDonald NQ, Lapatto R, Rust JM, Gunning J, Wlodawer A, Blundell TL: **New protein fold revealed by a 2.3-Å resolution crystal structure of nerve growth factor**. *Nature* 1991, **354**:411–414.

57. Sunagar K, Fry BG, Jackson TNW, Casewell NR, Undheim EAB, Vidal N, Ali SA, King GF, Vasudevan K, Vasconcelos V, Antunes A: **Molecular evolution of vertebrate neurotrophins: co-option of the highly conserved nerve growth factor gene into the advanced snake venom arsenalf**. *PloS One* 2013, **8**:e81827.

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

59. Kierszenbaum AL, Lea O, Petrusz P, French FS, Tres LL: **Isolation, culture, and immunocytochemical characterization of epididymal epithelial cells from pubertal and adult rats.** *Proc Natl Acad Sci U S A* 1981, **78**:1675–1679.

60. Harvey AL, Bradley KN, Cochran SA, Rowan EG, Pratt JA, Quillfeldt JA, Jerusalinsky DA: **What can toxins tell us for drug discovery?**. *Toxicon Off J Int Soc Toxinology* 1998, **36**:1635–1640.

61. Koh DCI, Armugam A, Jeyaseelan K: **Snake venom components and their applications in biomedicine**. *Cell Mol Life Sci CMLS* 2006, **63**:3030–3041.

62. Adrienne Mayor: *The Poison King: The Life and Legend of Mithradates, Rome's Deadliest Enemy*. New Jersey (USA): Princeton University Press; 2010.

63. Francesco Redi: *Osservazioni intorno alle vipere*. All'Insegna della Stella; 1668.

64. Hawgood BJ: **Abbé Felice Fontana (1730-1805): founder of modern toxinology**. *Toxicon Off J Int Soc Toxinology* 1995, **33**:591–601.

65. Cushman DW, Ondetti MA: **Design of angiotensin converting enzyme inhibitors**. *Nat Med* 1999, **5**:1110–1112.

66. Lewis RJ, Garcia ML: **Therapeutic potential of venom peptides**. *Nat Rev Drug Discov* 2003, **2**:790–802.

67. Cossins, D.: **From toxins to therapeutics**. *The Scientist* 2013.

68. King GF: **Venoms as a platform for human drugs: translating toxins into therapeutics**. *Expert Opin Biol Ther* 2011, **11**:1469–1484.

69. Diochot S, Baron A, Salinas M, Douguet D, Scarzello S, Dabert-Gay A-S, Debayle D, Friend V, Alloui A, Lazdunski M, Lingueglia E: **Black mamba venom peptides target acid-sensing ion channels to abolish pain**. *Nature* 2012, **490**:552–555.

70. Marsh NA: **Diagnostic uses of snake venom**. *Haemostasis* 2001, **31**:211–217.

71. Albert Calmette: **Contribution à l'étude du venin des serpents. Immunisation des animaux et traitement de l'envenimation**. *Ann Inst Pasteur* , **VIII**:275–291.

72. Albert Calmette: **Propriétés du sérum des animaux immunisés contre le venin des serpents et thérapeutique de l'envenimation**. *Comptes Rendus Académie Sci* , **118**:720–722.

73. Warrel, D.A.: **Geographical and intraspecies variation in the clinical manifestations of envenoming by snakes**. In *Venom Snakes Ecol Evol Snakebite*. Oxford, UK: Clarendon Press; 1997.

74. **WHO | WHO Guidelines for the Production, Control and Regulation of Snake Antivenom Immunoglobulins** [http://www.who.int/bloodproducts/snake_antivenoms/snakeantivenomguide/en/]

75. Gutiérrez JM, Solano G, Pla D, Herrera M, Segura Á, Villalta M, Vargas M, Sanz L, Lomonte B, Calvete JJ, León G: **Assessing the preclinical efficacy of antivenoms: from the lethality neutralization assay to antivenomics**. *Toxicon Off J Int Soc Toxinology* 2013, **69**:168–179.

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

77. **WHO | WHO highlights critical need for life-saving antivenoms** [http://www.who.int/mediacentre/news/notes/2010/antivenoms_20100504/en/]

78. Chippaux, JP: **Bulletin of the World Health Organization**. 1998, **76**:515–524.

79. Chippaux J-P: **Estimate of the burden of snakebites in sub-Saharan Africa: a meta-analytic approach**. *Toxicon Off J Int Soc Toxinology* 2011, **57**:586–599.

80. Bochner R, Struchiner CJ: **[Exploratory analysis of environmental and socioeconomic factors related to snakebite incidence in Rio de Janeiro from 1990 to 1996]**. *Cad Saúde Pública* 2004, **20**:976–985.

81. Harrison RA, Hargreaves A, Wagstaff SC, Faragher B, Lalloo DG: **Snake Envenoming: A Disease of Poverty**. *PLoS Negl Trop Dis* 2009, **3**:e569.

82. **WHO | Snakebite** [http://www.who.int/neglected_diseases/diseases/snakebites/en/]

83. **WHO | The 17 neglected tropical diseases** [http://www.who.int/neglected_diseases/diseases/en/]

84. **United Nations Millennium Development Goals** [http://www.un.org/millenniumgoals/]

85. **WHO | Global Plan to Combat Neglected Tropical Diseases** [http://whqlibdoc.who.int/hq/2007/WHO_CDS_NTD_2007.3_eng.pdf]

86. Gutiérrez JM, Warrell DA, Williams DJ, Jensen S, Brown N, Calvete JJ, Harrison RA, for the Global Snakebite Initiative: **The Need for Full Integration of Snakebite Envenoming within a Global Strategy to Combat the Neglected Tropical Diseases: The Way Forward**. *PLoS Negl Trop Dis* 2013, **7**:e2162.

87. Williams D, Gutiérrez JM, Harrison R, Warrell DA, White J, Winkel KD, Gopalakrishnakone P, Global Snake Bite Initiative Working Group, International Society on Toxinology: **The Global Snake Bite Initiative: an antidote for snake bite**. *Lancet* 2010, **375**:89–91.

88. **WHO | A multicomponent strategy to improve the availability of antivenom for treating snakebite envenoming** [http://www.who.int/bulletin/volumes/92/7/13-132431-ab/en/]

89. E.M. Fonseca Aizpurúa, E. García Piney, F.J. Nuño Mateo, A. Braña Rodríguez: **Mordeduras de víbora**. *Anu Med Interna* 2007, **24**:407–408.

90. Saz-Parkinson Z, del Pino Luengo M, López-Cuadrado T, Andújar D, Carmona-Alférez R, Flores RM, Amate JM: **Approach to the epidemiology of venomous bites in Spain**. *Toxicon Off J Int Soc Toxinology* 2012, **60**:706–711.

91. Chippaux J-P: **Epidemiology of snakebites in Europe: a systematic review of the literature**. *Toxicon Off J Int Soc Toxinology* 2012, **59**:86–99.

92. Gutiérrez JM, Williams D, Fan HW, Warrell DA: **Snakebite envenoming from a global perspective: Towards an integrated approach**. *Toxicon Off J Int Soc Toxinology* 2010, **56**:1223–1235.

93. Gutierrez JM, Lomonte B: **Efectos Locales en El Envenenamiento Ofídico en America Latina**. In *Animais Pessonhentos No Bras Biol*. *Volume 32*. Sao Paulo, Brazil: Sarvier; 2003:310–323.

94. Calvete JJ: **Snake venomics: from the inventory of toxins to biology**. *Toxicon Off J Int Soc Toxinology* 2013, **75**:44–62.

95. Calvete JJ: **Next-generation snake venomics: protein-locus resolution through venom proteome decomplexation**. *Expert Rev Proteomics* 2014, **11**:315–329.

96. Renuse S, Chaerkady R, Pandey A: **Proteogenomics**. *Proteomics* 2011, **11**:620–630.

97. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray**. *Science* 1995, **270**:467–470.

98. Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO, Davis RW: **Yeast microarrays for genome wide parallel genetic and gene expression analysis**. *Proc Natl Acad Sci U S A* 1997, **94**:13057–13062.

99. Costa V, Angelini C, De Feis I, Ciccodicola A: **Uncovering the complexity of transcriptomes with RNA-Seq**. *J Biomed Biotechnol* 2010, **2010**:853916.

100. Ho PL, Soares MB, Yamane T, Raw I: **Reverse Biology Applied to Micrurus corallinus, a South American Coral Snake**. *Toxin Rev* 1995, **14**:327–337.

101. Junqueira de Azevedo, Ho, PL: **Venom gland transcriptomic analysis**. In *Anim Toxins State Art Perspect Helath Biotechnol*. UFMG.; 2009:693 – 713.

102. Francischetti IMB, My-Pham 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, **337**:55–69.

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

104. Kashima S, Roberto PG, Soares AM, Astolfi-Filho S, Pereira JO, Giuliati S, Faria M Jr, Xavier MAS, Fontes MRM, Giglio JR, França SC: **Analysis of Bothrops jararacussu venomous gland transcriptome focusing on structural and functional aspects: I--gene expression profile of highly expressed phospholipases A2**. *Biochimie* 2004, **86**:211–219.

105. Cidade DAP, Simão TA, Dávila AMR, Wagner G, Junqueira-de-Azevedo I de LM, Ho PL, Bon C, Zingali RB, Albano RM: **Bothrops jararaca venom gland transcriptome: analysis of the gene expression pattern**. *Toxicon Off J Int Soc Toxinology* 2006, **48**:437–461.

106. Qinghua L, Xiaowei Z, Wei Y, Chenji L, Yijun H, Pengxin Q, Xingwen S, Songnian H, Guangmei Y: **A catalog for transcripts in**

the venom gland of the Agkistrodon acutus: identification of the toxins potentially involved in coagulopathy. *Biochem Biophys Res Commun* 2006, **341**:522–531.

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

108. Wagstaff SC, Harrison RA: **Venom gland EST analysis of the saw-scaled viper, Echis ocellatus, reveals novel alpha9beta1 integrin-binding motifs in venom metalloproteinases and a new group of putative toxins, renin-like aspartic proteases**. *Gene* 2006, **377**:21–32.

109. Junqueira-de-Azevedo ILM, Ching ATC, Carvalho E, Faria F, Nishiyama MY Jr, 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.

110. 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.

111. Jia Y, Cantu BA, Sánchez EE, Pérez JC: **Complementary DNA sequencing and identification of mRNAs from the venomous gland of Agkistrodon piscivorus leucostoma**. *Toxicon Off J Int Soc Toxinology* 2008, **51**:1457–1466.

112. Metzker ML: **Sequencing technologies - the next generation**. *Nat Rev Genet* 2010, **11**:31–46.

113. Freeman J. Dyson: *The Sun, the Genome, the Internet*. USA: Oxford University Press; 1999.

114. Brenner S: **Life sentences: Hunters and gatherers**. *Genome Biol* 2002, **3**:comment1003.1–comment1003.2.

115. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics**. *Nat Rev Genet* 2009, **10**:57–63.

116. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq**. *Nat Methods* 2008, **5**:621–628.

117. Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, Emanuelsson O, Zhang ZD, Weissman S, Snyder M: **What is a gene, post-ENCODE? History and updated definition**. *Genome Res* 2007, **17**:669–681.

118. ENCODE Project Consortium: **The ENCODE (ENCyclopedia Of DNA Elements) Project**. *Science* 2004, **306**:636–640.

119. ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M: **An integrated encyclopedia of DNA elements in the human genome**. *Nature* 2012, **489**:57–74.

120. Graur D, Zheng Y, Price N, Azevedo RBR, Zufall RA, Elhaik E: **On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE**. *Genome Biol Evol* 2013, **5**:578–590.

121. Lee RC, Feinbaum RL, Ambros V: **The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14**. *Cell* 1993, **75**:843–854.

122. Bartel DP: **MicroRNAs: target recognition and regulatory functions**. *Cell* 2009, **136**:215–233.

123. Zelanis A, Tashima AK, Pinto AFM, Paes Leme AF, Stuginski DR, Furtado MF, Sherman NE, Ho PL, Fox JW, Serrano SMT: **Bothrops jararaca venom proteome rearrangement upon neonate to adult transition**. *Proteomics* 2011, **11**:4218–4228.

124. Madrigal M, Sanz L, Flores-Díaz M, Sasa M, Núñez V, Alape-Girón A, Calvete JJ: **Snake venomics across genus Lachesis. Ontogenetic changes in the venom composition of Lachesis stenophrys and comparative proteomics of the venoms of adult Lachesis melanocephala and Lachesis acrochorda**. *J Proteomics* 2012, **77**:280–297.

125. Calvete JJ: **Proteomic tools against the neglected pathology of snake bite envenoming**. *Expert Rev Proteomics* 2011, **8**:739–758.

126. Georgieva D, Arni RK, Betzel C: **Proteome analysis of snake venom toxins: pharmacological insights**. *Expert Rev Proteomics* 2008, **5**:787–797.

127. Calvete JJ, Juárez P, Sanz L: **Snake venomics. Strategy and applications**. *J Mass Spectrom JMS* 2007, **42**:1405–1414.

128. Calvete JJ, Sanz L, Angulo Y, Lomonte B, Gutiérrez JM: **Venoms, venomics, antivenomics**. *FEBS Lett* 2009, **583**:1736–1743.

129. Pla D, Gutiérrez JM, Calvete JJ: **Second generation snake antivenomics: comparing immunoaffinity and immunodepletion protocols**. *Toxicon Off J Int Soc Toxinology* 2012, **60**:688–699.

130. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors**. *Proc Natl Acad Sci U S A* 1977, **74**:5463–5467.

131. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG,

Nadeau J, McKusick VA, Zinder N, et al.: **The sequence of the human genome**. *Science* 2001, **291**:1304–1351.

132. Mardis ER: **Next-generation sequencing platforms**. *Annu Rev Anal Chem Palo Alto Calif* 2013, **6**:287–303.

133. Mardis ER: **Next-generation DNA sequencing methods**. *Annu Rev Genomics Hum Genet* 2008, **9**:387–402.

134. Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M: **Comparison of next-generation sequencing systems**. *J Biomed Biotechnol* 2012, **2012**:251364.

135. Bainbridge MN, Warren RL, Hirst M, Romanuik T, Zeng T, Go A, Delaney A, Griffith M, Hickenbotham M, Magrini V, Mardis ER, Sadar MD, Siddiqui AS, Marra MA, Jones SJM: **Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach**. *BMC Genomics* 2006, **7**:246.

136. Check Hayden E: **Is the $1,000 genome for real?**. *Nature* 2014.

137. Wetterstrans KA: **DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)**. .

138. Wadman M: **James Watson's genome sequenced at high speed**. *Nature* 2008, **452**:788.

139. Ronaghi M, Uhlén M, Nyrén P: **A sequencing method based on real-time pyrophosphate**. *Science* 1998, **281**:363, 365.

140. Jarvie T, Harkins T: **Transcriptome sequencing with the Genome Sequencer FLX system**. *Nat Methods* 2008, **5**.

141. Pascal Mayer,: **DNA colony massively parrallel sequencing ams98 presentation**. 16:13:41 UTC.

142. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons

JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, et al.: **An integrated semiconductor device enabling non-optical genome sequencing**. *Nature* 2011, **475**:348–352.

143. Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska M, Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM, Karch H: **Prospective genomic characterization of the German enterohemorrhagic Escherichia coli O104:H4 outbreak by rapid next generation sequencing technology**. *PloS One* 2011, **6**:e22751.

144. Gilbert JA, Hughes M: **Gene expression profiling: metatranscriptomics**. *Methods Mol Biol Clifton NJ* 2011, **733**:195–205.

145. Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH: **Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing**. *Mol Ecol* 2008, **17**:1636–1647.

146. Schwartz TS, Tae H, Yang Y, Mockaitis K, Van Hemert JL, Proulx SR, Choi J-H, Bronikowski AM: **A garter snake transcriptome: pyrosequencing, de novo assembly, and sex-specific differences**. *BMC Genomics* 2010, **11**:694.

147. Durban J, Juárez P, Angulo Y, Lomonte B, Flores-Diaz M, Alape-Girón A, Sasa M, Sanz L, Gutiérrez JM, Dopazo J, Conesa A, Calvete JJ: **Profiling the venom gland transcriptomes of Costa Rican snakes by 454 pyrosequencing**. *BMC Genomics* 2011, **12**:259.

148. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**:403–410.

149. Aird SD, Watanabe Y, Villar-Briones A, Roy MC, Terada K, Mikheyev AS: **Quantitative high-throughput profiling of snake venom gland transcriptomes and proteomes (Ovophis okinavensis and Protobothrops flavoviridis)**. *BMC Genomics* 2013, **14**:790.

150. 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**:609–623.

151. Calvete, JJ, Sanz, L, Cid, P, De La Torre, P, Flores-Díaz, M, Dos Santos, MC, Borges, A, Bremo, A, Angulo, Y, Lomonte B, Alape-Girón, A: **Snake venomics of the Central American rattlesnake Crotalus simus and the South American Crotalus durissus complex points to neurotoxicity as an adaptative paedomorphic trend along Crotalus dispersal in South America**. *J Proteome Res* 2010, **9**:528–544.

152. Durban, J, Pérez, A, Sanz, L, Gómez, A, Bonilla, F, Rodríguez, S, Chacón, D, Sasa, M, Angulo, Y, Gutiérrez, JM, Calvete, JJ: **Integrated "omics" profiling indicates that miRNAs are modulators of the ontogenetic venom composition shift in the Central American rattlesnake, Crotalus simus simus**. *BMC Genomics* 2013, **14**:234.

153. Anton J Enright, Bino John, Ulrike Gaul, Thomas Tuschl, Chris Sander, Debora S Marks: **MicroRNA targets in Drosophila**. *Genome Biol* 2003, **5**(RI).

154. Guerra-Assunção, J., Enright, A. J.: **MapMi: automated mapping of microRNA loci**. *BMC Bioinformatics* 2010, **11**.

155. M. Vander Laat, J. Fernández, J. Durban, E. Villalobos, E. Camacho, J.J. Calvete, B. Lomonte: **Amino acid sequence and biological characterization of BlatPLA2, a non-toxic acidic phospholipase A2 from the venom of the arboreal snake Bothriechis lateralis from Costa Rica**. *Toxicon* 2013, **73**:71–80.

156. Rokyta DR, Lemmon AR, Margres MJ, Aronow K: **The venom-gland transcriptome of the eastern diamondback rattlesnake (Crotalus adamanteus)**. *BMC Genomics* 2012, **13**:312.

157. Rokyta DR, Wray KP, Lemmon AR, Lemmon EM, Caudle SB: **A high-throughput venom-gland transcriptome for the Eastern Diamondback Rattlesnake (Crotalus adamanteus) and evidence for pervasive positive selection across toxin classes**. *Toxicon Off J Int Soc Toxinology* 2011, **57**:657–671.

158. Terrat Y, Sunagar K, Fry BG, Jackson TNW, Scheib H, Fourmy R, Verdenaud M, Blanchet G, Antunes A, Ducancel F: **Atractaspis aterrima toxins: the first insight into the molecular evolution of venom in side-stabbers**. *Toxins* 2013, **5**:1948–1964.

159. Aird SD, Watanabe Y, Villar-Briones A, Roy MC, Terada K, Mikheyev AS: **Quantitative high-throughput profiling of snake venom gland transcriptomes and proteomes (Ovophis okinavensis and Protobothrops flavoviridis)**. *BMC Genomics* 2013, **14**:790.

160. Margres MJ, Aronow K, Loyacano J, Rokyta DR: **The venom-gland transcriptome of the eastern coral snake (Micrurus fulvius) reveals high venom complexity in the intragenomic evolution of venoms**. *BMC Genomics* 2013, **14**:531.

161. Rokyta, DR., Wray, KP., Margres, MJ.: **The genesis of an exceptionally lethal venom in the timber rattlesnake (Crotalus horridus) revealed through comparative venom-gland transcriptomics**. *BMC Genomics* 2013, **14**.

162. Roeding F, Borner J, Kube M, Klages S, Reinhardt R, Burmester T: **A 454 sequencing approach for large scale phylogenomic analysis of the common emperor scorpion (Pandinus imperator)**. *Mol Phylogenet Evol* 2009, **53**:826–834.

163. Weber APM, Weber KL, Carr K, Wilkerson C, Ohlrogge JB: **Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing**. *Plant Physiol* 2007, **144**:32–42.

164. Shawn T. O'Neil, Scott J. Emrich: **Assessing De Novo transcriptome assembly metrics for consistency and utility**. *BMC Genomics* 2013, **14**.

165. Wang, W., Wang, Y., Zhang, Q., Qi, Y., Guo, D.: **Global characterization of Artemisia annua glandular trichome transcriptome using 454 pyrosequencing**. *BMC Genomics* 2009, **10**.

166. Fox JW: **A brief review of the scientific history of several lesser-known snake venom proteins: l-amino acid oxidases, hyaluronidases and phosphodiesterases**. *Toxicon Off J Int Soc Toxinology* 2013, **62**:75–82.

167. Fernández J, Lomonte B, Sanz L, Angulo Y, Gutiérrez JM, Calvete JJ: **Snake venomics of Bothriechis nigroviridis reveals extreme variability among palm pitviper venoms: different evolutionary solutions for the same trophic purpose**. *J Proteome Res* 2010, **9**:4234–4241.

168. Castoe TA, Daza JM, Smith EN, Sasa MM, Kuch U, Campbell JA, Chippindale PT, Parkinson CL: **Comparative phylogeography of pitvipers suggests a consensus of ancient Middle American highland biogeography**. *J Biogeogr* 2009, **36**:88–103.

169. Freedman RB, Hirst TR, Tuite MF: **Protein disulphide isomerase: building bridges in protein folding**. *Trends Biochem Sci* 1994, **19**:331–336.

170. Kini RM, Doley R: **Structure, function and evolution of three-finger toxins: mini proteins with multiple targets**. *Toxicon Off J Int Soc Toxinology* 2010, **56**:855–867.

171. Jern P, Coffin JM: **Effects of retroviruses on host genome function**. *Annu Rev Genet* 2008, **42**:709–732.

172. Pung, YF, Wong, PTH, Kumar, PP, Hodgson, WC, Kini, RM: **Ohanin, a novel protein from king cobra venom, induces hypolocomotion and hyperalgesia in mice**. *J Biol Chem* 2005:13137 – 13147.

173. Regina Brockmann, Andreas Beyer, Jürgen J. Heinisch, Thomas Wilhelm: **Posttranscriptional Expression Regulation: What determintes translation Rates?**. *Plos Comput Biol* 2007, **3**:531–539.

174. Zhang B, Pan X, Cannon CH, Cobb GP, Anderson TA: **Conservation and divergence of plant microRNA genes**. *Plant J Cell Mol Biol* 2006, **46**:243–259.

175. Zhang BH, Pan XP, Wang QL, Cobb GP, Anderson TA: **Identification and characterization of new plant microRNAs using EST analysis**. *Cell Res* 2005, **15**:336–360.

176. Baohong Zhang, Xiaoping Pan, Qinglian Wang, George P. Cobb, Todd A. Anderson: **Computational identification of microRNAs and their targets**. *Comput Biol Biochem* 2006, **30**:395–407.

177. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ: **miRBase: tools for microRNA genomics**. *Nucleic Acids Res* 2008, **36**(Database issue):D154–158.

178. Wang X-J, Reyes JL, Chua N-H, Gaasterland T: **Prediction and identification of Arabidopsis thaliana microRNAs and their mRNA targets**. *Genome Biol* 2004, **5**:R65.

179. Min H, Yoon S: **Got target? Computational methods for microRNA target prediction and their extension**. *Exp Mol Med* 2010, **42**:233–244.

180. Alberto Alape-Girón, Libia Sanz, José Escolano, Marietta Flores-Díaz, Marvin Madrigal, Mahmood Sasa, JJ Calvete: **Snake venomics of the lancehead piviper Bothrops asper: geographic, individual, and ontogenetic variations.** *J Proteome Res* 2008, **7**:3556 – 3571.

181. Gutiérrez JM, Lomonte B: **Phospholipases A2: unveiling the secrets of a functionally versatile group of snake venom toxins**. *Toxicon Off J Int Soc Toxinology* 2013, **62**:27–39.

182. Silveira LB, Marchi-Salvador DP, Santos-Filho NA, Silva FP, Marcussi S, Fuly AL, Nomizo A, da Silva SL, Stábeli RG, Arantes EC, Soares AM: **Isolation and expression of a hypotensive and anti-platelet acidic phospholipase A2 from Bothrops moojeni snake venom**. *J Pharm Biomed Anal* 2013, **73**:35–43.

183. Santos-Filho NA, Silveira LB, Oliveira CZ, Bernardes CP, Menaldo DL, Fuly AL, Arantes EC, Sampaio SV, Mamede CCN, Beletti ME, de Oliveira F, Soares AM: **A new acidic myotoxic, anti-platelet and prostaglandin I2 inductor phospholipase A2 isolated from Bothrops moojeni snake venom**. *Toxicon Off J Int Soc Toxinology* 2008, **52**:908–917.

184. Bustillo S, Gay CC, García Denegri ME, Ponce-Soto LA, Bal de Kier Joffé E, Acosta O, Leiva LC: **Synergism between baltergin metalloproteinase and Ba SPII RP4 PLA2 from Bothrops alternatus venom on skeletal muscle (C2C12) cells**. *Toxicon Off J Int Soc Toxinology* 2012, **59**:338–343.

185. Jimenez Porras JM: **INTRASPECIFIC VARIATIONS IN COMPOSITION OF VENOM OF THE JUMPING VIPER, BOTHROPS NUMMIFERA**. *Toxicon Off J Int Soc Toxinology* 1964, **69**:187–195.

186. Sasa M, Wasko DK, Lamar WW: **Natural history of the terciopelo Bothrops asper (Serpentes: Viperidae) in Costa Rica**. *Toxicon Off J Int Soc Toxinology* 2009, **54**:904–922.

187. Wiens JJ: **Missing data and the design of phylogenetic analyses**. *J Biomed Inform* 2006, **39**:34–42.

188. McCormack JE, Maley JM, Hird SM, Derryberry EP, Graves GR, Brumfield RT: **Next-generation sequencing reveals**

phylogeographic structure and a species tree for recent bird divergences. *Mol Phylogenet Evol* 2012, **62**:397–406.

189. Sukkapan P, Jia Y, Nuchprayoon I, Pérez JC: **Phylogenetic analysis of serine proteases from Russell's viper (Daboia russelli siamensis) and Agkistrodon piscivorus leucostoma venom**. *Toxicon Off J Int Soc Toxinology* 2011, **58**:168–178.

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

191. Hedstrom L: **An overview of serine proteases**. *Curr Protoc Protein Sci Editor Board John E Coligan Al* 2002, **Chapter 21**:Unit 21.10.

192. Parry MA, Jacob U, Huber R, Wisner A, Bon C, Bode W: **The crystal structure of the novel snake venom plasminogen activator TSV-PA: a prototype structure for snake venom serine proteinases**. *Struct Lond Engl 1993* 1998, **6**:1195–1206.

193. Calvete JJ, Marcinkiewicz C, Monleón D, Esteve V, Celda B, Juárez P, Sanz L: **Snake venom disintegrins: evolution of structure and function**. *Toxicon Off J Int Soc Toxinology* 2005, **45**:1063–1074.

ABREVIATURES I ACRÒNIMS

Kbases	Milers de bases
Mbases	Milions de bases
Gbases	Milers de milions de bases
Tbases	Bilions de bases
ORF	Marc obert de lectura (de l’anglès *Open Reading Frame*)
UTR	Regió no traduïda (de l’anglès *Untranslated Region*)
miRNA	MicroRNA
NHGRI	National Human Genome Research Institute
PCR	Reacció en cadena de la polimerasa (de l’anglès *Polimerase Chain Reaction*)
NGS	Seqüenciació de Nova Generació (de l’anglès *Next Generation Sequencing*)
NCBI	National Center for Biotechnology Information
SRA	Sequence Read Archive
kDa	Kilodaltons
OMS	Organització Mundial de la Salut
FDA	US Food and Drug Administration
DNA	Àcid desoxiribonucleic (de l’anglès *Deoxyribonucleic Acid*)
RNA	Àcid ribonucleic (de l’anglès *Ribonucleic acid*)
EST	Expressed Sequence Tag
ATP	Adenosín trifosfat
emPCR	PCR en emulsió
CRT	Terminació cíclica reversible (de l’anglès *Cyclic Reversible Termination*)
SBL	Seqüenciació per ligació (de l’anglès, *Sequencing by ligation*)
SBS	Seqüenciació per síntesi (de l’anglès *Sequencing by synthesis*)
SMRT	Seqüenciació en temps real d’una molècula única (de l’anglès *Single Molecule Real Time Sequencing*)
cDNA	DNA complementari
mRNA	RNA missatger
CDS	Seqüencies codificants (de l’anglès *Coding sequences*)
NTD	Malaltia tropical descuidada (de l’anglès *Neglected Tropical Diseases*)

SVMP Metaloproteases del verí de les serps (de l'anglès, *Snake Venom Metalloproteases*)
SVSP Serines Proteases del verí de les serps (de l'anglès, *Snake Venom Serine Proteases*)
LAAO L-aminoàcid oxidases

seqüenciació verí serps NGS DNA espècies tipus RNA tecnologia família procés dades mossegada determinats proteïnes estat serp organismes caracterització expressió rendiment verins transcriptoma plataformes tècniques funció lectura recerca genoma total components Viperidae tractament seqüència generació plaquetària Ion majoria regulació temps descrits principals accidents Joana nucleòtids molecular estratègia pertanyents existència estudi plataforma base Life molècula pèptids seqüències funcions estratègies processos nombre informació efectes lectures mida evolució SVMPs coneixement mètode estudis anàlisi mètodes variabilitat coagulació terme famílies seqüenciador gènica llibreria longitud activitat composició milions sistema toxines elevada podem cost Laia treball creixement primer enzims capacitat única funcional Illumina diversitat bases antiverins enzim molècules amplificació incidència catalitzen emPCR PLA2 nucleotidases 9N

ALEXANDER PP VI V

CSIC