Defensins and the convergent evolution of platypus and reptile venom genes

Camilla M. Whittington[1], Anthony T. Papenfuss[2], Paramjit Bansal[3], Allan M. Torres[4], Emily S. W. Wong[1], Janine E. Deakin[5], Tina Graves[6], Amber Alsop[5], Kyriena Schatzkamer[6], Colin Kremitzki[6], Chris P. Ponting[7], Peter Temple-Smith[8], Wesley C. Warren[6], Philip W. Kuchel[3] & Katherine Belov[1+]

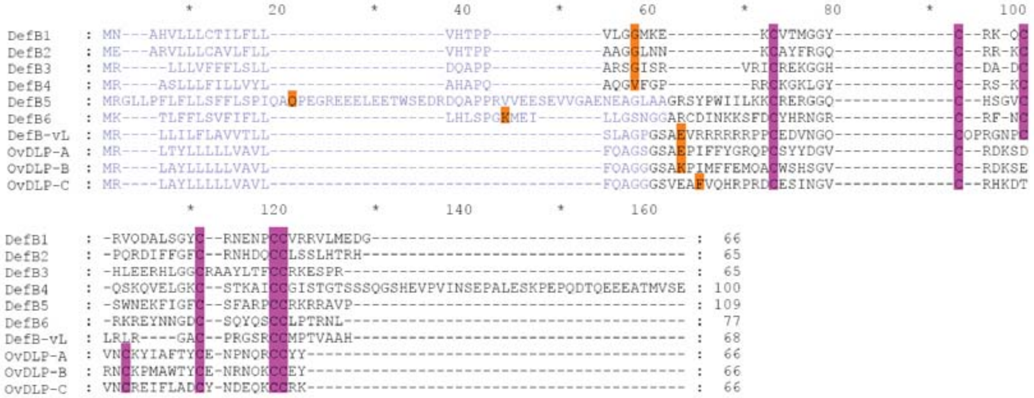
[+]Corresponding author: Katherine Belov, Faculty of Veterinary Science, University of Sydney NSW 2006 Australia, phone: +61 2 9351 3454, fax: +61 2 9351 3957, email: kbelov@vetsci.usyd.edu.au.

Supplemental Material

1. Platypus genome mining: defensins and OvDLPs

Based on the spacing of cysteines, *DefA4* may be an early alpha defensin (Figure 1). The first pair of cysteine residues in *DefA4* have an alpha-defensin-like spacing (CVC), but the fourth and fifth cysteine residues have more than the usual number of residues between them, and thus are slightly more beta-defensin-like in their spacing (CSRRRRICC). HMMer also suggested that these were more similar to beta-defensins, presumably due to the spacings between the fourth and fifth cysteines. It should be noted, however, that several mouse alpha-defensins (cryptidins) have longer intervals between their fourth and fifth cysteines (Patil *et al.* 2004), and the presence of an anionic propiece in this peptide places it within the alpha-defensin family (Liu and Ganz 1995). Based on this information, *DefA4* is an alpha-defensin with some beta-defensin-like characteristics, and since alpha-defensins are known to have evolved from beta-defensins (Liu *et al.* 1997; Patil *et al.* 2004), it may be an early alpha-defensin.

The structure of the OvDLPs and beta-defensins was elucidated based on the GenScan and GenomeScan predictions of these genes. All OvDLPs and beta-defensins contains two exons, and their introns are in phase 1. It can be seen in Supplemental Figure 1 that the cysteine spacing for *DefB-vL* is similar to the beta-defensins. However, its amino acid sequence is more similar to the OvDLPs (Supplemental Table 1). Sequences are available at http://bioinf.wehi.edu.au/platypus/.



```
                    *        20         *        40         *        60         *        80         *       100
DefB1    : MN---AHVLLLCTILFLL-------------------VHTPP------------VLGGMKE----------KCVTMGGY-------------C--RK-QC
DefB2    : ME---ARVLLLCAVLFLL-------------------VHTPP------------AAGGLNN----------KCAYFRGQ-------------C--RR-KC
DefB3    : MR-----LLLVFFFLSLL-------------------DQAPP------------ARSGISR--------VRICREKGGH-------------C--DA-DC
DefB4    : MR----ASLLLFILLVYL-------------------AHAPQ------------AQGVFGP---------RRCKGKLGY-------------C--RS-KC
DefB5    : MRGLLPFLFLLSFFLSPIQAQPEGREEELEETWSEDRDQAPPRVVEESEVVGAENEAGLAAGRSYPWIILKKCRERGGQ-------------C--HSGVC
DefB6    : MK----TLFFLSVFIFLL-------------------LHLSPGKMEI-------LLGSNGGARCDINKKSFDCYHRNGR-------------C--RF-NC
DefB-vL  : MR----LLILFLAVVTLL------------------------------------SLAGPGSAEVRRRRRRPPCEDVNGQ-------------CQPRGNPC
OvDLP-A  : MR----LTYLLLLLVAVL------------------------------------FQAGSGSAEPIFFYGRQPCSYYDGV-------------C--RDKSD
OvDLP-B  : MR----LAYLLLLLVAVL------------------------------------FQAGGGSAKPIMFFEMQACWSHSGV-------------C--RDKSE
OvDLP-C  : MR----LAYLLLLLVAVL------------------------------------FQAGGGSVEAFVQHRPRDCESINGV-------------C--RHKDT

                    *       120         *       140         *       160
DefB1    : -RVQDALSGYC--RNENPCCVRRVLMEDG---------------------------------- :  66
DefB2    : -PQRDIFFGFC--RNHDQCCLSSLHTRH----------------------------------- :  65
DefB3    : -HLEERHLGGCRAAYLTFCCRKESPR------------------------------------- :  65
DefB4    : -QSKQVELGKC--STKAICCGISTGTSSSQGSHEVPVINSEPALESKPEPQDTQEEEATMVSE : 100
DefB5    : -SWNEKFIGFC--SFARPCCRKRRAVP------------------------------------ : 109
DefB6    : -RKREYNNGDC--SQYQSCCLPTRNL------------------------------------- :  77
DefB-vL  : LRLR----GAC--PRGSRCCMPTVAAH------------------------------------ :  68
OvDLP-A  : VNCKYIAFTYCE-NPNQRCCYY----------------------------------------- :  66
OvDLP-B  : RNCKPMAWTYCE-NRNQKCCEY----------------------------------------- :  66
OvDLP-C  : VNCREIFLADCY-NDEQKCCRK----------------------------------------- :  66
```

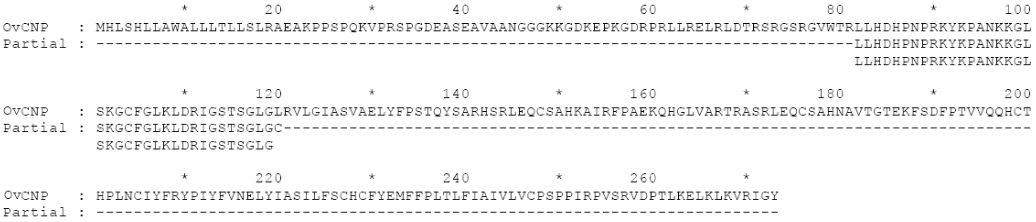
Supplemental Figure 1. Alignment of the platypus beta-defensins and OvDLPs. Alternating text colour denotes exons, and the first amino acid of the mature peptide is highlighted in orange. Conserved cysteine residues are highlighted in pink. It can be seen that the spacing of cysteines in *DefB-vL* appears to be similar to that of the other beta-defensins.

Supplemental Table 1. MEGA-generated sequence identity matrix of OvDLPs and platypus beta-defensin peptides aligned using MUSCLE.

	OvDLP-A	**OvDLP-B**	**OvDLP-C**	**DefB-vL**	**DefB1**	**DefB2**	**DefB3**	**DefB4**	**DefB5**	**DefB6**
OvDLP-A	ID	0.696	0.530	0.264	0.136	0.136	0.136	0.120	0.064	0.116
OvDLP-B	0.696	ID	0.545	0.235	0.136	0.136	0.106	0.110	0.036	0.129
OvDLP-C	0.530	0.545	ID	0.279	0.136	0.136	0.121	0.110	0.045	0.142
DefB-vL	**0.264**	**0.235**	**0.279**	**ID**	**0.147**	**0.147**	**0.191**	**0.090**	**0.110**	**0.155**
DefB1	0.136	0.136	0.136	0.147	ID	0.469	0.166	0.150	0.073	0.129
DefB2	0.136	0.136	0.136	0.147	0.469	ID	0.138	0.170	0.073	0.142
DefB3	0.136	0.106	0.121	0.191	0.166	0.138	ID	0.160	0.100	0.077
DefB4	0.120	0.110	0.110	0.090	0.150	0.170	0.160	ID	0.128	0.130
DefB5	0.064	0.036	0.045	0.110	0.073	0.073	0.100	0.128	ID	0.045
DefB6	0.116	0.129	0.142	0.155	0.129	0.142	0.077	0.130	0.045	ID

2. Platypus genome mining: OvCNPs and OvNGFs

We identified one OvCNP and one OvNGF in the platypus genome assembly (Supplemental Figures 2 and 3).

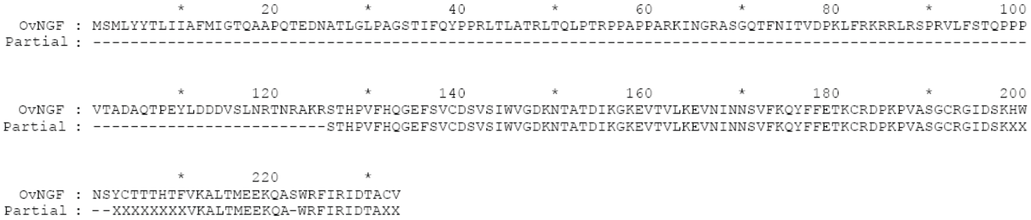


```
                   *        20         *        40         *        60         *        80         *       100
OvCNP   : MHLSHLLAWALLLTLLSLRAEAKPPSPQKVPRSPGDEASEAVAANGGGKKGDKEPKGDRPRLLRELRLDTRSRGSRGVWTRLLHDHPNPRKYKPANKKGL
Partial : ---------------------------------------------------------------------------------LLHDHPNPRKYKPANKKGL
                                                                                         LLHDHPNPRKYKPANKKGL

                   *       120         *       140         *       160         *       180         *       200
OvCNP   : SKGCFGLKLDRIGSTSGLGLRVLGIASVAELYFPSTQYSARHSRLEQCSAHKAIRFPAEKQHGLVARTRASRLEQCSAHNAVTGTEKFSDFPTVVQQHCT
Partial : SKGCFGLKLDRIGSTSGLGC--------------------------------------------------------------------------------
          SKGCFGLKLDRIGSTSGLG

                   *       220         *       240         *       260         *
OvCNP   : HPLNCIYFRYPIYFVNELYIASILFSCHCFYEMFFPLTLFIAIVLVCPSPPIRPVSRVDPTLKELKLKVRIGY
Partial : -------------------------------------------------------------------------
```

Supplemental Figure 2. Alignment of partial amino acid sequence of *OvCNP*, previously obtained by Edman degradation, with the full amino acid sequence found in the platypus genome.



```
                   *        20         *        40         *        60         *        80         *       100
  OvNGF : MSMLYYTLIIAFMIGTQAAPQTEDNATLGLPAGSTIFQYPPRLTLATRLTQLPTRPPAPPARKINGRASGQTFNITVDPKLFRKRRLRSPRVLFSTQPPP
Partial : ----------------------------------------------------------------------------------------------------

                   *       120         *       140         *       160         *       180         *       200
  OvNGF : VTADAQTPEYLDDDVSLNRTNRAKRSTHPVFHQGEFSVCDSVSIWVGDKNTATDIKGKEVTVLKEVNINNSVFKQYFFETKCRDPKPVASGCRGIDSKHW
Partial : -------------------------STHPVFHQGEFSVCDSVSIWVGDKNTATDIKGKEVTVLKEVNINNSVFKQYFFETKCRDPKPVASGCRGIDSKXX

                   *       220         *
  OvNGF : NSYCTTTHTFVKALTMEEKQASWRFIRIDTACV
Partial : --XXXXXXXXVKALTMEEKQA-WRFIRIDTAXX
```

Supplemental Figure 3. Alignment of partial amino acid sequence of *OvNGF*, previously obtained by Edman degradation, with the full amino acid sequence found in the platypus genome.

3. Phylogenetic analysis

The accession numbers for the peptides used in the phylogenetic trees are shown in Supplemental Tables 2-4. The OvCNP, OvNGF and alpha-defensin neighbour-joining trees are shown in Supplemental Figures 4, 5 and 6.

It is important to note that the OvDLP-like peptide neighbour-joining tree was used for two different reasons. Firstly, this tree was used to show independent origin of platypus venom OvDLPs and snake venom crotamines. Secondly, the tree was used in combination with FISH results to infer likely clustering of beta-defensin genes in the platypus and thus speculate on the evolutionary processes that resulted in the expansion of beta defensins genes from a single cluster in chicken to four clusters in eutherians. To this end, two different Bayesian trees were also created for the OvDLP-like peptide library using MrBayes. The tree in Supplemental Figure 7 was created using the same alignment of beta-defensin and venom peptides, cropped to contain the six-cysteine motif, that was used in the neighbour-joining tree (Figure 2). This tree confirms the independent origin of the platypus and reptile venom components from different beta-defensin orthologs.

The second Bayesian tree (Supplemental Figure 8), created using the full-length sequences of mammalian beta defensins, was used to confirm the placement of platypus beta-defensins 3 and 5 into synteny group A (Figure 3). Unfortunately, orthologs of the other platypus beta-defensins could not be determined from this tree, but their placement into synteny groups was achieved using the neighbour-joining tree (Figure 2) and FISH mapping results.

Supplemental Table 2. GenBank accession numbers of peptides used in the OvCNP-like peptide phylogenetic analysis.

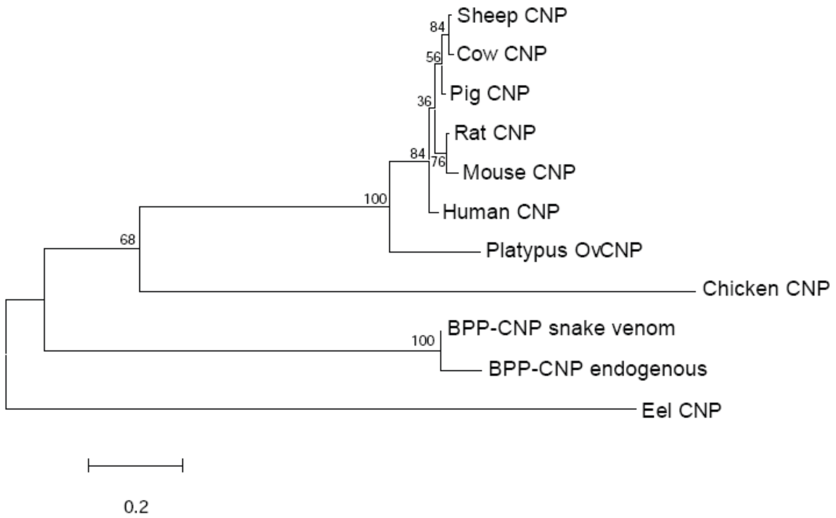
Name	Accession	Description	Species
Human NGF	AAH32517	NGF	Human
Gorilla NGF	BAA90439	NGF	Gorilla
Chimp NGF	BAA90438	NGF	Chimp
Monkey NGF	AAV74256	NGF	Bolivian squirrel monkey
Cow NGF	CAA70759	NGF	Cow
Mouse NGF	NP 038637	NGF	Mouse
Bird NGF	CAC87482	NGF	Streak-necked flycatcher
Taipan vNGF	ABA60117	venom NGF	*Oxyuranus microlepidotus*
Brown snake vNGF	ABA60119	venom NGF	*Psedonaja textilis*
Viper vNGF	AAV64846	venom NGF	*Macrovipera lebetina*
Rattlesnake vNGF	AAG30924	venom NGF	*Crotalus durissus terrificus*
Zebrafish vNGF	AA031815	NGF	Zebrafish

Supplemental Table 3. GenBank accession numbers of peptides used in the OvNGF-like peptide phylogenetic analysis.

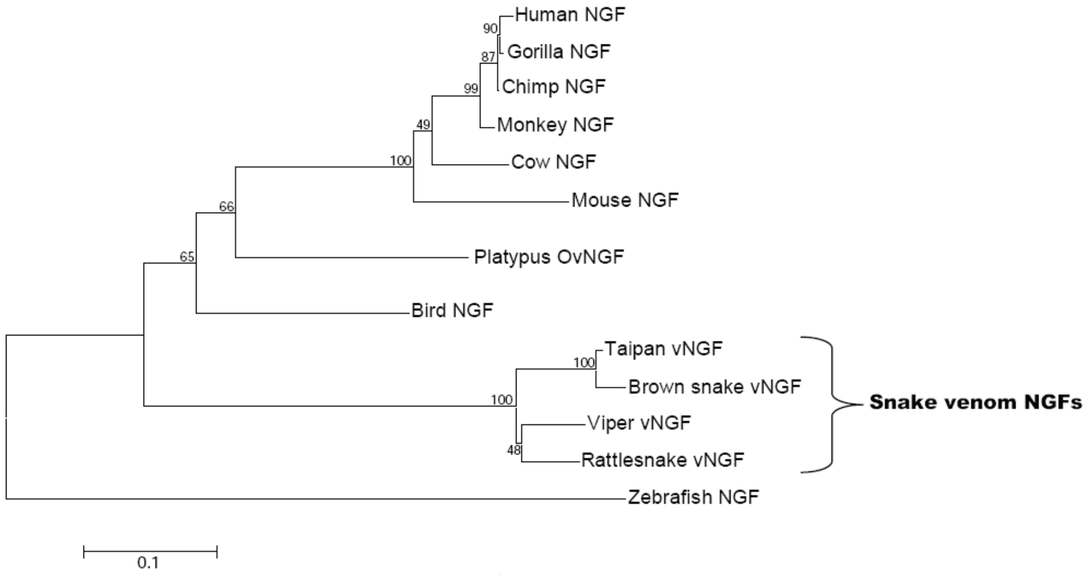
Name	Accession	Description	Species
Sheep CNP	AAB92261	CNP	Sheep
Cow CNP	Z48477.1	CNP	Cow
Pig CNP	M64758.1	CNP	Pig
Rat CNP	D90219.1	CNP	Rat
Mouse CNP	D28873.1	CNP	Mouse
Human CNP	M64710.1	CNP	Human
BPP-CNPsnake venom	Q9PW56	brain BPP-CNP	*Bothrops jararaca*
BPP-CNPsnake endogenous	AAD51326	venom BPP-CNP	*Bothrops jararaca*
Eel CNP	P18145	CNP	*Anguilla japonica*

Supplemental Table 4. GenBank accession numbers of reptile peptides used in the OvDLP-like peptide phylogenetic analysis.

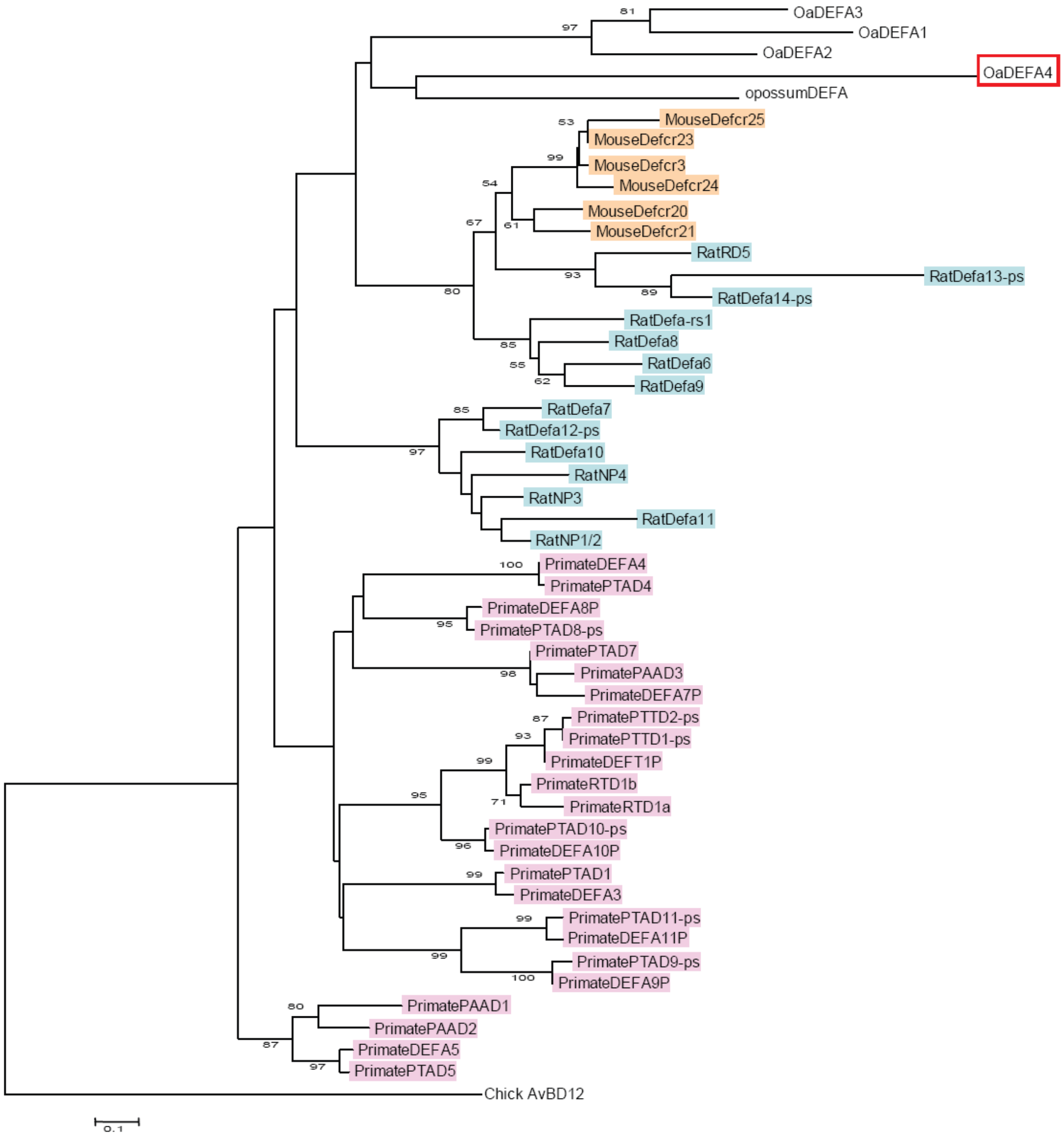
Name	Accession	Description	Species
Lizard vCLP3U	AAZ75613	vCLP3 maxillary	*Pogona barbata*
Lizard vCLP1U	AAZ75614	vCLP1 maxillary	*Pogona barbata*
Lizard vCLP2U	AAZ75615	vCLP2 mandibular	*Pogona barbata*
Lizard vCLP2L	AAZ75612	vCLP2 maxillary	*Pogona barbata*
Snake crotasin	AAT47437	crotasin	*Crotalus durissus terrificus*
Snake vCrotamine	P01475	crotamine	*Crotalus durissus terrificus*
Snake vCrotamine4	P24334	crotamine4	*Crotalus durissus terrificus*
Snake vCrotamine2	P24332	crotamine2	*Crotalus durissus terrificus*
Snake vCrotamine3	P24333	crotamine3	*Crotalus durissus terrificus*



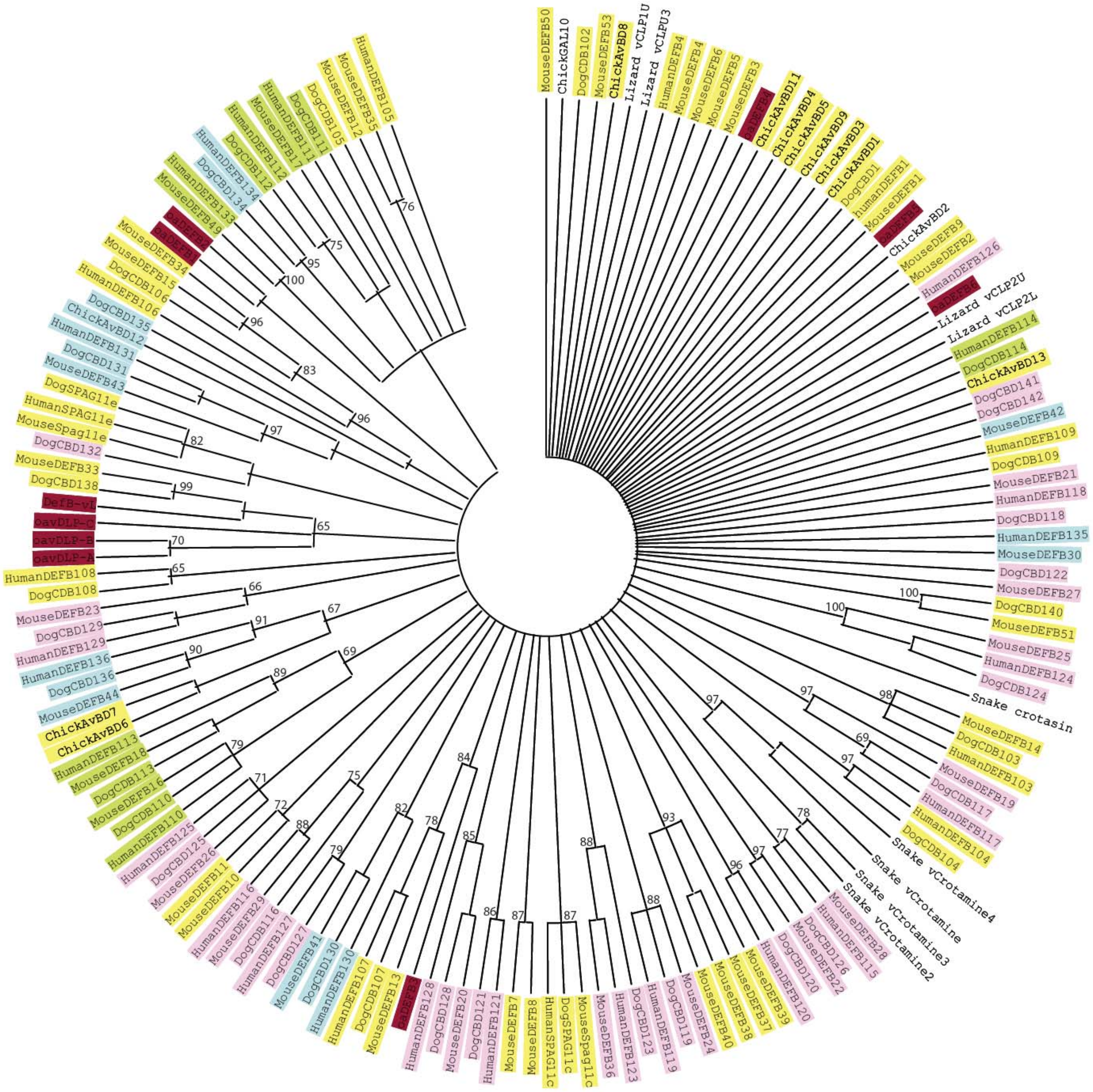
Supplemental Figure 4. Neighbour joining tree of eutherian CNPs, chicken CNP, platypus *OvCNP*, and two snake BPP-CNPs, one of which is from venom. The platypus *OvCNP* does not group with the snake vBPP-CNPs. The tree is rooted with Eel CNP.



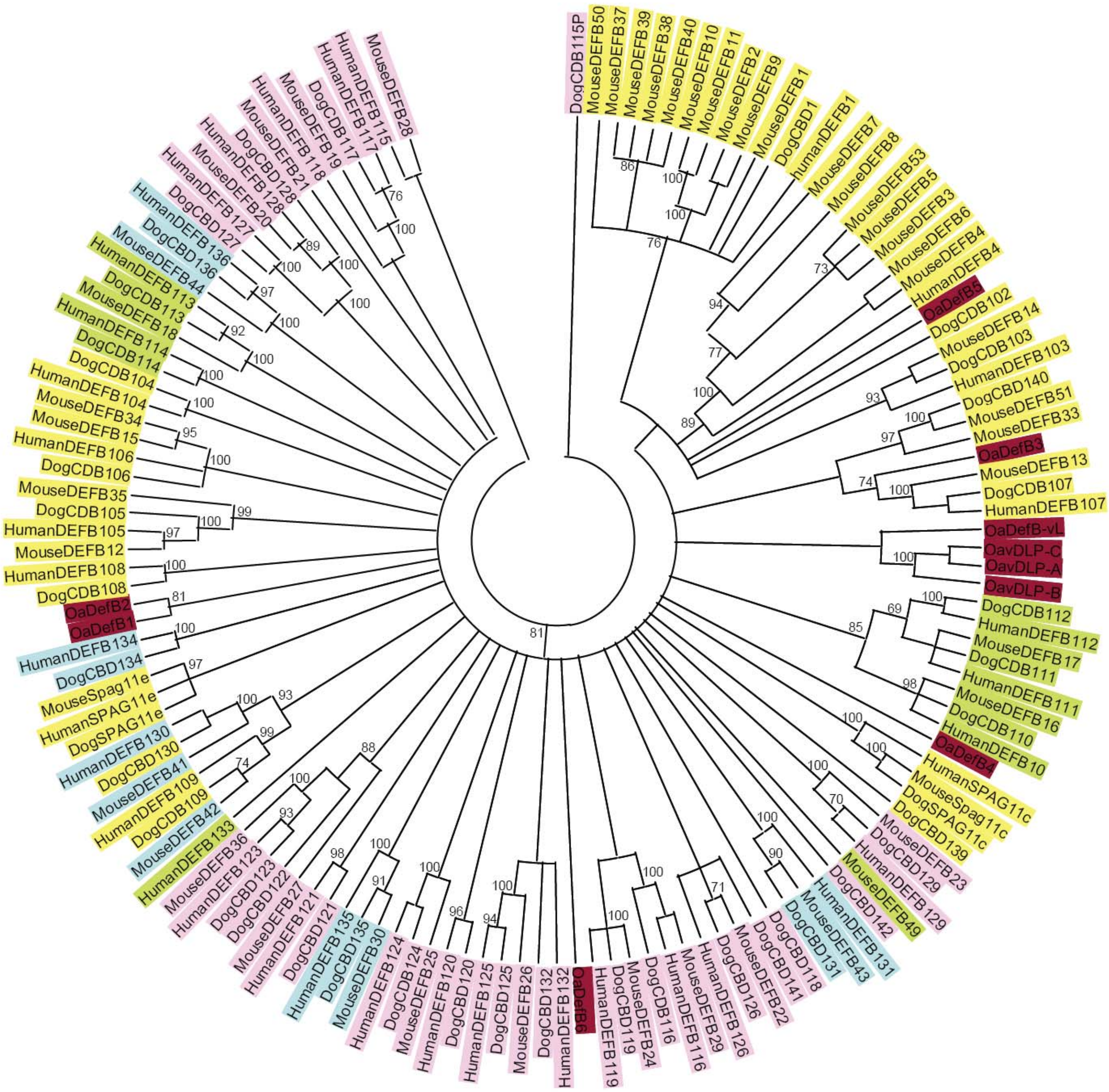
Supplemental Figure 5. Neighbour joining tree of eutherian NGFs, a bird NGF, snake vNGFs, and platypus *OvNGF*. The platypus *OvNGF* has arisen independently from the snake vNGFs. The tree is rooted with Zebrafish NGF.



Supplemental Figure 6. Neighbour joining tree of eutherian, opossum and platypus alpha-defensins (*OaDefA1-4*), rooted with a chicken beta-defensin (Chick AvBD12), showing species-specific alpha-defensin expansions. Mouse alpha-defensins (cryptidins) are in orange, rat apha-defensins in blue, and primate alpha-defensins in pink. Bootstrap values less than 50 are not shown. Platypus alpha-defensins 1-3 have clearly evolved recently via gene duplication. Platypus alpha-defensin 4 (boxed) is not positioned in the tree with any level of confidence, probably due to the fact that it has beta-defensin-like characteristics and may be an early alpha-defensin (Figure 1).



Supplemental Figure 7. Bayesian tree of the six-cysteine region beta-defensin and venom sequences. drawn in MrBayes, showing independent evolution of reptile and platypus venom. Posterior probabilities 65 and over are shown. Platypus sequences are highlighted in red, and chicken and eutherian sequences are highlighted according to their grouping into syteny groups (Figure 3). Synteny group A is in yellow, B is in blue, C is in green and D is in pink.



Supplemental Figure 8. Bayesian tree of the full sequence of beta-defensins, drawn in MrBayes. Posterior probabilities 65 and over are shown. Platypus sequences are highlighted in red, and chicken and eutherian sequences are highlighted according to their grouping into syteny groups (Figure 3). Synteny group A is in yellow, B is in blue, C is in green and D is in pink.

4. Rates of evolution and dating gene duplication of the OvDLPs

We date the duplication of the *DefB-vL* gene at a median of 192 million years ago (95% highest posterior density: 120.8, 277.4), and the duplication of the first of the vDLPs from the *DefB-vL* at a median of 98 million years ago (98% highest posterior density: 57.8, 143.5). A subsequent duplication of vDLPs occurred at a median of 47 million years ago (95% highest posterior density: 24.15, 75.53).

5. Fluorescence in-situ hybridization (FISH)

The results of FISH experiments are summarised in Supplemental Table 5. It can be seen that two of the alpha-defensins and all of the beta-defensins, bar DefB6, lie on chromosome X_2. Although the order of these genes on the chromosome was not elucidated, it can be inferred based on synteny information (Figures 2 and 3). The localization of both antimicrobial and venom genes to the platypus sex chromosome chain supports the idea that the chain of ten platypus sex chromosomes appears to have evolved from ordinary autosomes (Graves 2006; Warren *et al.* Submitted).

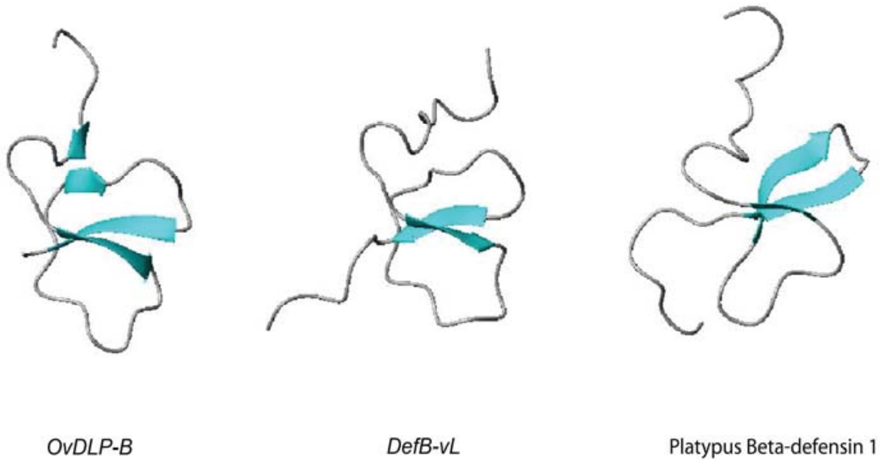
Supplemental Table 5. Results of FISH.

Gene	Contig	Coordinates	FISH localisation
DefB1	Contig17484	7229-8481	X2-E3.q2
DefB2	Contig17484	10440-10858	X2-E3.q2
DefB3	Contig17484	13735-20044	X2-E3.q2
DefB4	Contig19787	6187-8273	X2-E3.q2
DefB5	Ultra279	274911-277518	X2-E3.q2
DefB6	Chromosome X1	2759953-2770606	X1
DefA1	Contig23157	3234-4026	X2-E3.q2
DefA2	Contig24085	7706-8555	X3-E5.q1 and X2-E3.q1.2
DefA3	Contig4716	643-2266	X2, Y1
DefA4	Contig23157	5544-7668	X2
DefB-vL	Contig4716	20599-22958	X2, Y1
Ov-DLPA	Contig4716	27522-28947	X2, Y1
Ov-DLPB	Contig4716	31586-32717	X2, Y1
Ov-DLPC	Contig4716	35166-36201	X2, Y1
NGF	Chromosome 7	34930447-34931184	7q1
CNP	Contig2134	1874-4966	1q5.3

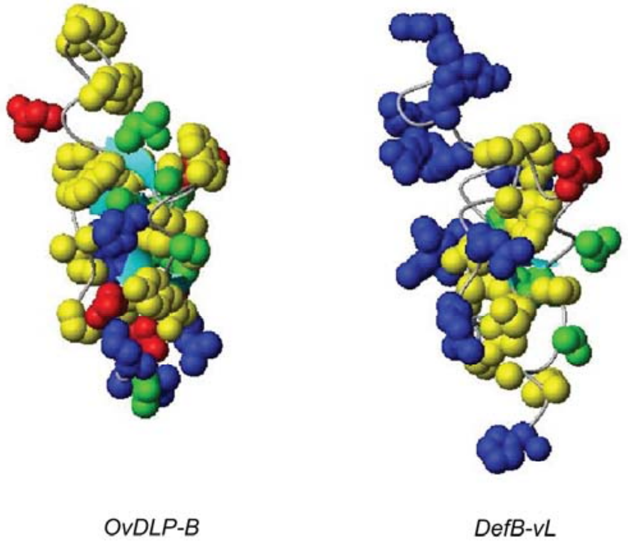
6. Modelling and synthesis of DefB-vL

The platypus beta-defensin 2 has 43% sequence identity to cynomolgus monkey beta defensin 118 (CAL68943), which does not have a reported tertiary structure at present, and 24% sequence identity with human beta-defensin 1. Results of the modelling in Supplemental Figures 9 and 10 show that the peptide folding structure of platypus beta-defensin 1, *OvDLP-B* and *DefB-vL* is very similar. DefB-vL and OvDLP-B were synthesised for NMR analysis and future antimicrobial assays, and the 1-D NMR spectrum of *DefB-vL* also showed similarity with those of OvDLPs, confirming these results.

Structural modelling of *DefB-vL* suggests that this protein, like the beta-defensins, is antimicrobial. The presence of 6 arginine residues on the N-terminus of the *DefB-vL* protein, with hydrophilic residues situated at one end of the molecule and hydrophobic residues at the other, indicate that this peptide is probably membrane active. It can be seen in Supplemental Figure 10 that the *DefB-vL* peptide, unlike the *OvDLP-B* peptide, has a clear separation of hydrophobic (yellow) and hydrophilic (blue) residues, possibly indicating antimicrobial activity (Zimmermann *et al.* 1995).The structural modelling of *DefB-vL* suggests significant similarities in their structural folds due to their size and in the cross-liking combinations of the six cysteine residues; this further confirms the relatedness of these molecules and supports the idea that *DefB-vL* was the ancestral beta-defensin that gave rise to the OvDLPs.



Supplemental Figure 9. Ribbon diagrams of *OvDLP-B, DefB-vL* and platypus beta-defensin 1, showing that their structures are very similar. Arrows indicate beta-sheets.



Supplemental Figure 10. Space-filling models of *OvDLP-B* and *DefB-vL*. Hydrophobic residues are shown in yellow and hydrophilic residues are shown in blue.

8. Supplemental References

Graves, JAM. 2006. Sex chromosome specialization and degeneration in mammals. *Cell* 124: 901-914.

Liu, L and Ganz, T. 1995. The Pro Region of Human Neutrophil Defensin Contains a Motif That Is Essential for Normal Subcellular Sorting. *Blood* 85: 1095-1103.

Liu, LD, Zhao, CQ, Heng, HHQ, and Ganz, T. 1997. The human beta-defensin-1 and alpha-defensins are encoded by adjacent genes: Two peptide families with differing disulfide topology share a common ancestry. *Genomics* 43: 316-320.

Patil, AA, Hughes, AL, and Zhang, G. 2004. Rapid evolution and diversification of mammalian alpha-defensins as revealed by comparative analysis of rodent and primate genomes. *Physiol Genomics* 20: 11-11.

Torres, AM and Kuchel, PW. 2004. The beta-defensin-fold family of polypeptides. *Toxicon* 44: 581-588.

Warren, W, Hillier, L, Marshall Graves, J, Birney, E, Ponting, C, Grutzner, F, Belov, K, Miller, W, Clarke, L, Chinwalla, A *et al.* Submitted. Genome analysis of the platypus reveals unique signatures of evolution.

Zimmermann, GR, Legault, P, Selsted, ME, and Pardi, A. 1995. Solution Structure of Bovine Neutrophil Beta-Defensin-12 - the Peptide Fold of the Beta-Defensins Is Identical to That of the Classical Defensins. *Biochemistry* 34: 13663-13671.