

Insights into Platypus Crural Gland Transcriptomics – Venom and Beyond: Supplementary Material

Table of Contents

Supplementary Tables	2
Table S1	2
Table S2	4
Table S3	5
Table S4	6
Table S5	7
Table S6	8
Table S7	9
Supplementary Figures	10
Figure S1	10
Figure S2	11
Figure S3	12
Figure S4	13
Figure S5	14

Supplementary Tables

Table S1. Adapter sequences used for trimming of raw sequencing read files.

Study	Adapters
Whittington, 2010 BioProject: PRJNA12885 Reference: (35)	>fastppredicted_illumina3prime AAGCAGTGGTAACAACGCATCCGAC
	>illumina3prime GATCGTCGGACTGTAGAACTCTGAAC
Brawand, 2011 & Lynch, 2015 BioProject: PRJNA143627 & PRJNA247824 Reference: (39) & (43)	>OligoPE1 GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
	>OligoPE2 ACACTCTTTCCCTACACGACGCTCTTCCGATCT
	>fastppredictedadapter AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
Julien, 2012 BioProject: PRJNA152927 Reference: (40)	>SmallRNAv1.5Adapter ATCTCGTATGCGTCTTCTGCTTG
	>SRAAdapter GUUCAGAGUUCUACAGUCCGACGAUC
Necsulea, 2014 BioProject: PRJNA186646 Reference: (41)	>TruSeq_Read1 AGATCGGAAGAGCACACGTCTGAACTCCAGTCA
	>TruSeq_Read2 AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Marin, 2017 BioProject: PRJNA381064 Reference: (42)	>IlluminaUniversalRead1adapter_alsofastppredictadapter AGATCGGAAGAGCACACGTCTGAACTCCAGTCA
	>IlluminaUniversalRead2adapter AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
	>IndexAdapter1 GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter2 GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter3 GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter4 GATCGGAAGAGCACACGTCTGAACTCCAGTCACTGACCAATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter5 GATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTGATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter6 GATCGGAAGAGCACACGTCTGAACTCCAGTCACGCCAATATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter7 GATCGGAAGAGCACACGTCTGAACTCCAGTCACCAGATCATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter8 GATCGGAAGAGCACACGTCTGAACTCCAGTCACACTTGAATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter9 GATCGGAAGAGCACACGTCTGAACTCCAGTCACGATCAGATCTCGTATGCCGTCTTCTGCTTG

	>IndexAdapter10 GATCGGAAGAGCACACGTCTGAACTCCAGTCACTAGCTTATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter11 GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGCTACATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter12 GATCGGAAGAGCACACGTCTGAACTCCAGTCACCTTGTAACTCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter13 GATCGGAAGAGCACACGTCTGAACTCCAGTCACAGTCAACAATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter14 GATCGGAAGAGCACACGTCTGAACTCCAGTCACAGTTCGGTATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter15 GATCGGAAGAGCACACGTCTGAACTCCAGTCACATGTCAGAATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter16 GATCGGAAGAGCACACGTCTGAACTCCAGTCACCCGTCCCGATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter18 GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTCCGCACATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter19 GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTGAAACGATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter20 GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTGGCCTTATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter21 GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTTTCGGAATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter22 GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGTACGTAATCTCGTATGCCGTCTTCTGCTTG
	>IndexAdapter23 GATCGGAAGAGCACACGTCTGAACTCCAGTCACGAGTGGATATCTCGTATGCCGTCTTCTGCTTG
	>Index Adapter 25 GATCGGAAGAGCACACGTCTGAACTCCAGTCACACTGATATATCTCGTATGCCGTCTTCTGCTTG
	>Index Adapter 27 GATCGGAAGAGCACACGTCTGAACTCCAGTCACATTCTTTATCTCGTATGCCGTCTTCTGCTTG
European Bioinformatics Institute, 2019 BioProject: PRJEB28810 Reference: (44)	>shortexpriboadapter AGATCGGAAGAGCACACGTCT
	>illuminauniversaladapter AGATCGGAAGAGCACACGTCTGAACTCCAGTCA
	>illuminauniversaladaptershort AGATCGGAAGAGCA

Number in brackets refers to manuscript reference numbers.

Table S2. Search queries for contamination and toxin database construction.

Database	Source	Search String	Date Accessed	Number of Sequences
Contamination Database	NCBI	mRna[All Fields] AND (bacteria[filter] AND biomol_mrna[PROP])	30/07/24	20642
	NCBI	mRna[All Fields] AND (viruses[filter] AND biomol_mrna[PROP])	30/07/24	47938
	NCBI	mRna[All Fields] AND (protists[filter] AND biomol_mrna[PROP])	02/08/24	1076140
	NCBI	mRna[All Fields] AND (fungi[filter] AND biomol_mrna[PROP])	02/08/24	6314295
	SILVA (https://www.arb-silva.de/SILVA_138.2_L_SURef_NR99_tax_silva_trunc.fasta.gz)	NA	11/07/24	95279
	SILVA (https://www.arb-silva.de/SILVA_138.2_S_SURef_NR99_tax_silva_trunc.fasta.gz)	NA	11/07/24	510495
Platypus Toxin Database	UniProt	(organism_id:33208) AND (cc_tissue_specificity:venom) OR (keyword:KW-0800) AND (reviewed:true) AND (organism_id:9258)	13/08/24	6
Tox-Prot Venom Toxin Database	Uniprot	(taxonomy_id:33208) AND (cc_tissue_specificity:venom) AND (reviewed:true)	09/09/24	7222
Platypus UniProt Database	UniProt	(organism_id:9258)	09/09/24	33077
Swiss-Prot Database	UniProt	(reviewed:true)	15/08/24	571864

Table S3. Redundancy analysis and variance partitioning analysis on tissue samples.

	Adjust R²	df	Variance	F	p-value
<i>Redundancy analysis</i>					
Gene count ~ study + tissue type	0.7382718	9	63968	11.656	0.001***
Tissue	-	6	59930	16.3807	0.001***
Study	-	3	4038	2.2072	0.001***
Residual	-	25	15244	-	-
<i>Variance partitioning analysis</i>					
[a] = Tissue Study	0.66685	6	51820	14.164	0.001***
[b] = Study Tissue	0.03385	3	4037.6	2.2072	0.008**
[c] = Tissue + Study	0.03757	0	-	-	-
[d] = Residuals	0.26173	-	-	-	-

Adjust R² indicates the percentage of variation attributed to each variable/s in each row. Significance indicated by asterisks.

Table S4. Genes with >1000-fold upregulation in the crural gland compared to other tissue types.

Gene ID	NCBI predicted function	limma false discovery rate	limma log2 fold change	edgeR false discovery rate	edgeR log2 fold change	Previously confirmed platypus venom component
AWGGnovel_OvDLP-B	NA	3.40e ⁻¹⁶	12.4498	1.01e ⁻¹⁴	11.5060	Yes
HOXC12	Homeobox C12	8.30e ⁻²⁰	11.4517	9.45e ⁻²⁴	13.0800	No
KLK1	Kallikrein 1	4.14e ⁻¹⁷	13.3415	3.89e ⁻²³	11.7713	No
LOC100078598	Vomer nasal type-2 receptor 26-like	1.68e ⁻¹³	10.44644	1.59e ⁻¹⁶	9.962626	No
LOC100083360	Corticotropin releasing hormone binding protein	3.93e ⁻¹⁶	14.3950	5.42e ⁻¹¹	11.1027	Yes
LOC100681151	Prostaglandin-H2 D-isomerase-like	3.40e ⁻¹⁶	13.3490	1.96e ⁻⁰⁹	12.1788	No
LOC100681171	Epididymal-specific lipocalin-5-like	1.95e ⁻²⁵	19.2336	3.86e ⁻²⁸	18.3840	No
LOC107547358	Uteroglobin-like	2.84e ⁻¹⁴	11.6612	0.001172	10.4639	No
LOC114811285	NA	1.20e ⁻¹¹	11.0620	0.000343	10.1794	No
LOC114813694	Tissue factor pathway inhibitor	6.69e ⁻²⁴	18.5704	1.42e ⁻²⁰	17.1033	Yes
LOC120638161	Uncharacterised predicted lncRNA	3.21e ⁻¹⁶	10.7665	9.56e ⁻²⁹	10.87045	No
MSTRG.16525	NA	1.16e ⁻²⁷	19.4598	1.03e ⁻⁷⁶	18.6756	No
MSTRG.16527	NA	3.40e ⁻¹⁶	10.47097	2.13e ⁻¹²	11.21954	No
MSTRG.16533	NA	8.24e ⁻¹⁹	12.0617	3.65e ⁻¹⁸	12.4759	No
MSTRG.26387	NA	9.51e ⁻¹⁵	13.7885	1.54e ⁻⁰⁵	11.0532	No
MSTRG.26388	NA	4.24e ⁻¹²	11.0099	5.05e ⁻¹¹	10.2028	No
MSTRG.42377	NA	3.34e ⁻¹⁴	11.1294	6.47e ⁻⁰⁹	10.6294	No
MSTRG.46790	NA	1.39e ⁻¹³	10.95943	0.001502	10.41471	No
MSTRG.8257	NA	3.19e ⁻¹²	10.01912	2.42e ⁻¹³	9.890072	No
NPPC	C-type natriuretic peptide	4.22e ⁻¹⁶	12.9097	2.24e ⁻²⁸	11.1950	Yes
SPAM1	Hyaluronidase PH-20	8.58e ⁻²⁴	15.8120	1.69e ⁻¹⁷	14.9990	Yes
TYRP1	Tyrosinase related protein 1	1.71e ⁻¹⁴	10.03263	1.10e ⁻²⁵	10.16721	No

Genes included had >1000-fold upregulation in both limma and edgeR differential expression methods. Fold change is rounded to 4 decimal places.

Table S5. Putative toxins found by the DeTox pipeline from different transcriptomic subsets.

Transcriptome subset	# putative toxins
Crural Gland only	5942
Crural Gland DE only	151
Crural Gland Tau only	63
Crural Gland DE + Tau combined	37

Table S6. Physicochemical properties of putative platypus toxin sequences.

ID	InterPro Family/ Superfamily	Molecular weight (kDa)	# of AAs	Cationic residue %	Charge at pH 7	Theoretical pI	GRAVY score	Hydrophobic %
XR_486810.3_ORF.3	Beta defensin	4.94	43	14	-1	5.63	-0.458	30.23
XM_029046826.2_ORF.1	Calreticulin/calnexin	48.67	419	1.4	-27.9	4.82	-1.126	32.94
XM_001513234.4_ORF.1	FGF binding 1	23.12	211	17.1	10.1	9.13	-1	25.59
XM_029081728.2_ORF.1	Haem peroxidase, animal-type	78.96	693	12	7.7	8.41	-0.399	43.72
XM_029049771.2_ORF.1	Kazal domain	12.18	114	12.3	4	8.47	-0.166	31.58
XM_016227951.3_ORF.1	Leucine-rich repeat domain	37.81	331	10.6	3.2	8.14	0.276	45.32
XM_003431054.3_ORF.1	Lipocalin	18.69	162	13.6	-1.4	5.95	-0.621	34.57
XM_029064327.1_ORF.1	Peptidase S1A, chymotrypsin family	25.58	233	10.7	2.7	7.7	-0.441	36.05
XM_029064328.1_ORF.5	Peptidase S1A, chymotrypsin family	24.95	228	11	1	6.84	-0.457	35.53
XM_029064368.2_ORF.9	Peptidase S1A, chymotrypsin family	26.57	240	5.8	-6.7	5.26	-0.173	41.67
XM_016226360.3_ORF.1	Secretoglobin	7.11	64	12.5	-4.1	4.55	-0.452	40.63
MSTRG8465.1_ORF.3	Secretoglobin	7.15	64	12.5	-2.1	4.93	-0.27	46.88
MSTRG8471.1_ORF.3	Secretoglobin	10.09	87	13.8	-4.6	4.98	-0.589	39.08

Negative grand average of hydropathicity (GRAVY) score indicates hydrophilicity.

Table S7. Functional prediction of putative platypus toxin sequences using CSM-peptide machine learning method.

ID	InterPro Domain	Anti-Angiogenic	Anti-Bacterial	Anti-Cancer	Anti-Inflammatory	Anti-Viral	Cell Penetrating	Quorum Sensing	Surface Binding
XR_486810.3_ORF.3	Beta defensin type	0.34	0.49	0.25	0.79	0.54	0	0	0.25
XM_029046826.2_ORF.1	Calreticulin/calnexin	0.42	0.11	0.17	0.97	0.64	0.03	0	0.22
XM_001513234.4_ORF.1	FGF binding 1	0.53	0.29	0.28	0.83	0.40	0.24	0.01	0.13
XM_029081728.2_ORF.1	Haem peroxidase, animal-type	0.47	0.15	0.14	0.89	0.62	0.02	0	0.28
XM_029049771.2_ORF.1	Kazal domain	0.46	0.59	0.1	0.82	0.54	0.02	0	0.35
XM_016227951.3_ORF.1	Leucine-rich repeat	0.35	0.09	0.17	0.84	0.6	0.02	0	0.27
XM_003431054.3_ORF.1	Lipocalin/cytosolic fatty-acid binding domain	0.38	0.13	0.06	0.82	0.71	0.01	0	0.31
XM_016226360.3_ORF.1	Secretoglobin	0.40	0.26	0.17	0.84	0.71	0.01	0	0.32
MSTRG8465.1_ORF.3	Secretoglobin	0.38	0.26	0.20	0.86	0.67	0.01	0	0.25
MSTRG8471.1_ORF.3	Secretoglobin	0.31	0.17	0.22	0.96	0.63	0	0	0.25
XM_029064327.1_ORF.1	Serine proteases, trypsin domain	0.39	0.22	0.09	0.85	0.59	0.12	0	0.31
XM_029064328.1_ORF.5	Serine proteases, trypsin domain	0.43	0.23	0.07	0.91	0.57	0.18	0	0.32
XM_029064368.2_ORF.9	Serine proteases, trypsin domain	0.32	0.12	0.13	0.85	0.64	0.02	0	0.34

Supplementary Figures

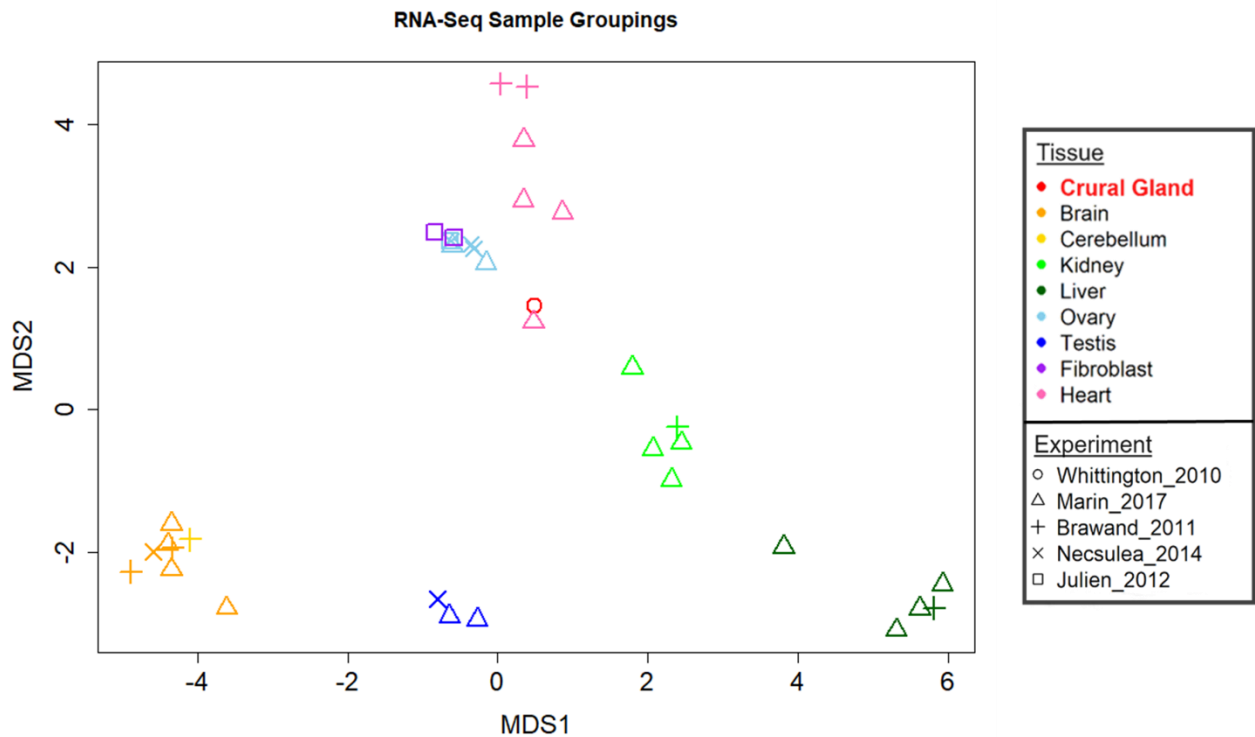


Figure S1. Multi-dimensional scaling plot of the 45 platypus RNA samples.

Colours indicate tissue type (crural gland label in red and bold), and symbols indicate the study from which the data originated. Each symbol represents a biological replicate, with any technical replicates condensed into their respective biological replicates.

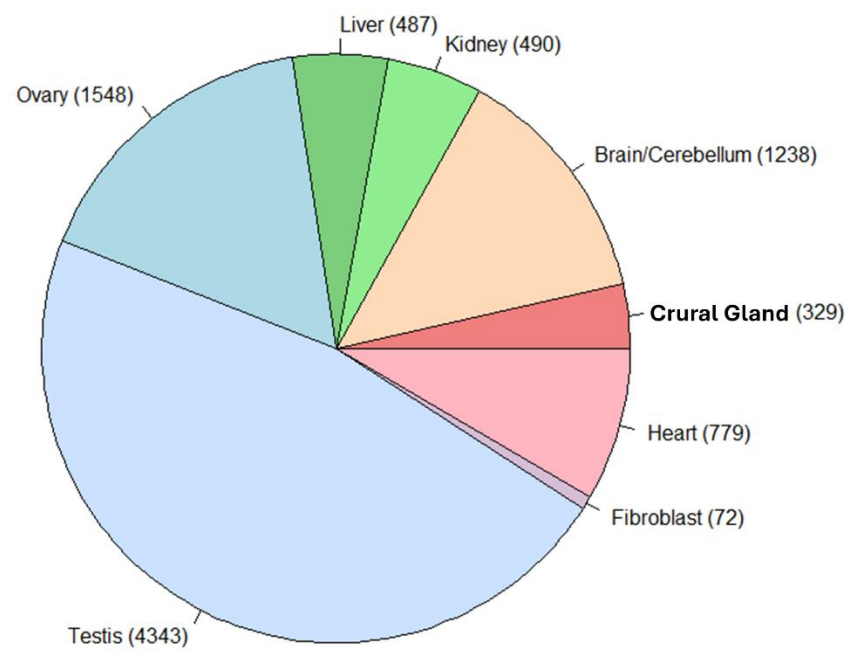


Figure S2. Tissue specific genes in the platypus.

All genes included ($n = 9286$) have high/absolute ($\text{Tau} \geq 0.85$) specificity to one tissue type. Crural gland label in bold.

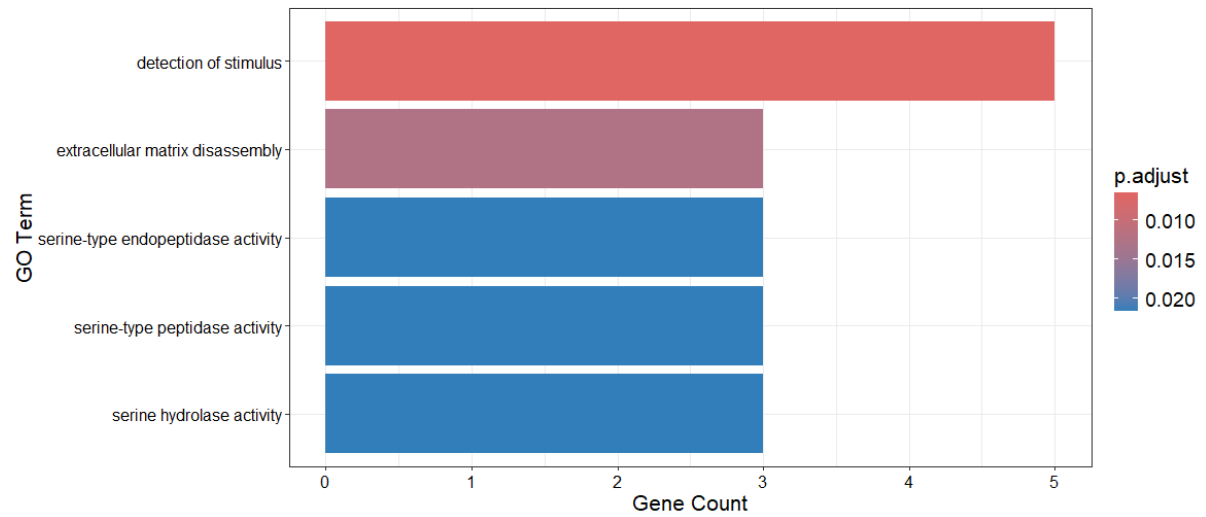
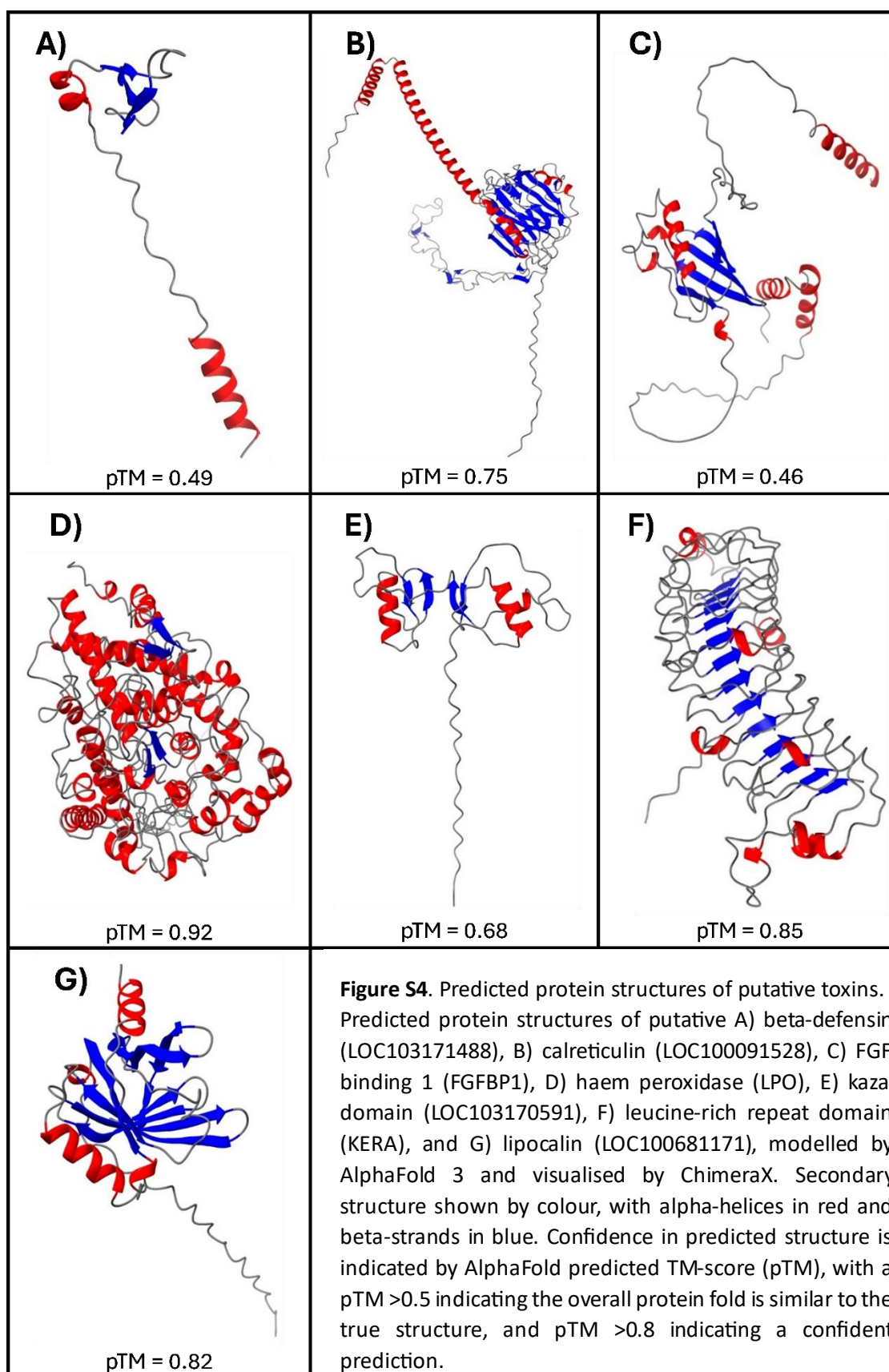


Figure S3. Significant GO terms associated with upregulated and crural gland-specific genes. Bar chart colour indicates adjusted p-value. Note only 25 genes were included in this analysis, since 86.19% of genes failed to map to gene symbols/Entrez IDs.



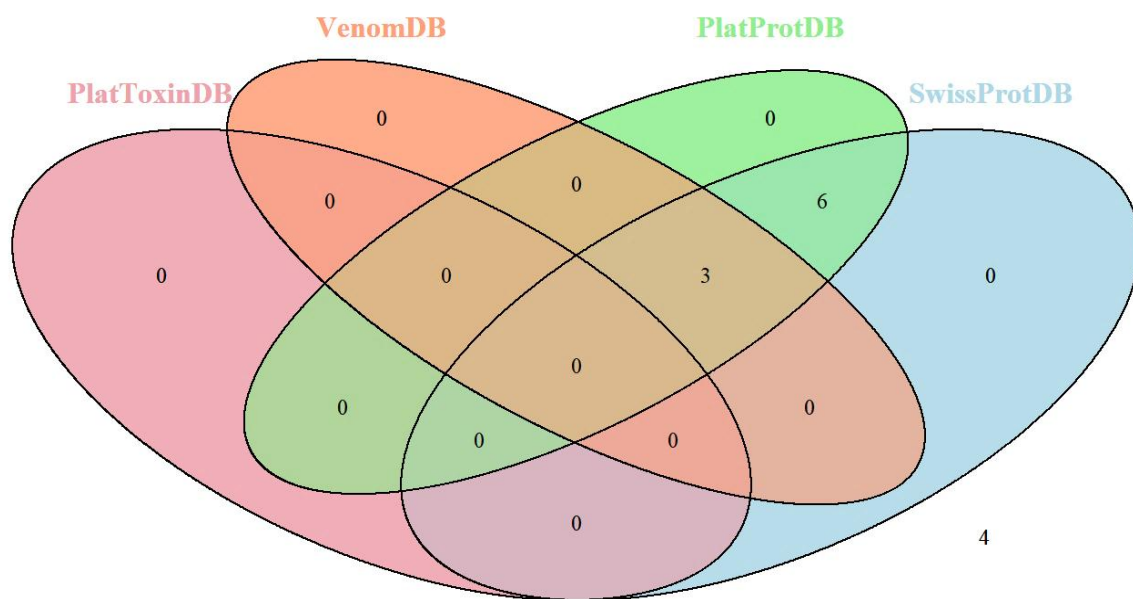


Figure S5. Database matches for DeTox identified platypus putative toxins. BLAST hits (e-value threshold $1e^{-10}$) of DeTox identified toxins against databases of known platypus toxins (pink, PlatToxinDB), Tox-Prot venom proteins (orange, VenomDB), all reviewed and unreviewed platypus proteins (green, PlatProtDB), and reviewed Swiss-Prot proteins (blue, SwissProtDB). Four proteins did not have hits to any of the investigated databases, but all had hits with a higher e-value threshold of $1e^{-3}$.