Proteomics of Galápagos Marine Iguanas Links Function of Femoral Gland Proteins to the Immune System

Authors

Frederik Tellkamp, Franziska Lang, Alejandro Ibáñez, Lena Abraham, Galo Quezada, Stefan Günther, Mario Looso, Fabian Jannik Tann, Daniela Müller, Franz Cemic, Jürgen Hemberger, Sebastian Steinfartz, and Marcus Krüger

Correspondence

marcus.krueger@uni-koeln.de;
marcus.krueger@uni-koeln.de

In Brief

Femoral glands secrete a wax-like substance on the inner side of lizard hind legs, which is thought to function as a mode of chemical communication. Though the minor volatile fraction is well studied, the major protein fraction remains enigmatic. Here, we use proteomics to analyze proteins in femoral gland secretions of the Galápagos marine iguana. Although we found no evidence for proteins and peptides involved in chemical communication, we found several immune-regulatory proteins which also demonstrate anti-microbial functions. Accordingly, we show that femoral gland proteins and peptides function as a barrier against microbial infection and may prevent the rapid degradation of volatile substances.

Graphical Abstract
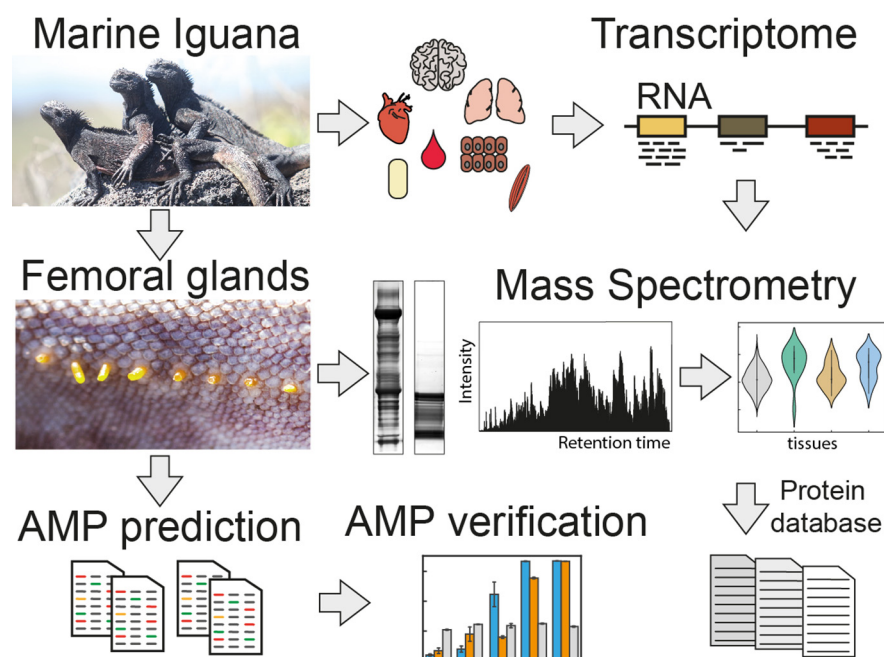
Highlights

- Comprehensive transcriptome and proteome of the Galápagos Marine Iguana.
- Analysis of femoral gland proteome.
- Identification of antimicrobial activity in femoral gland secretions.

Proteomics of Galápagos Marine Iguanas Links Function of Femoral Gland Proteins to the Immune System

Frederik Tellkamp[1,‡], Franziska Lang[2,‡], Alejandro Ibáñez[3,‡], Lena Abraham[1], Galo Quezada[4], Stefan Günther[5], Mario Looso[5], Fabian Jannik Tann[6], Daniela Müller[6], Franz Cemic[6], Jürgen Hemberger[6], Sebastian Steinfartz[7,*], and Marcus Krüger[1,8,*]

Communication between individuals via molecules, termed chemosignaling, is widespread among animal and plant species. However, we lack knowledge on the specific functions of the substances involved for most systems. The femoral gland is an organ that secretes a waxy substance involved in chemical communication in lizards. Although the lipids and volatile substances secreted by the femoral glands have been investigated in several biochemical studies, the protein composition and functions of secretions remain completely unknown. Applying a proteomic approach, we provide the first attempt to comprehensively characterize the protein composition of femoral gland secretions from the Galápagos marine iguana. Using samples from several organs, the marine iguana proteome was assembled by next-generation sequencing and MS, resulting in 7513 proteins. Of these, 4305 proteins were present in the femoral gland, including keratins, small serum proteins, and fatty acid-binding proteins. Surprisingly, no proteins with discernible roles in partner recognition or inter-species communication could be identified. However, we did find several proteins with direct associations to the innate immune system, including lysozyme C, antileukoproteinase (ALP), pulmonary surfactant protein (SFTPD), and galectin (LGALS1) suggesting that the femoral glands function as an important barrier to infection. Furthermore, we report several novel anti-microbial peptides from the femoral glands that show similar action against *Escherichia coli* and *Bacillus subtilis* such as oncocin, a peptide known for its effectiveness against Gram-negative pathogens. This proteomics data set is a valuable resource for future functional protein analysis and demonstrates that femoral gland secretions also perform functions of the innate immune system.

Femoral glands are small epidermal skin glands located on the lateral inner side of the hind legs of many lizards and amphibians and play a central role in the chemical signaling systems of nonophidian squamates (1, 2). Additionally femoral gland secretions are used as chemosignals for communication, territory marking, signaling of male quality, as well as self-, sex- and species recognition (3, 4). Femoral glands are holocrine glands and release a waxy plug through their pores, which is composed of secreted proteins and also terminally differentiated glandular cells. As the secretory plug emerges from the body surface, mechanical forces such as friction on rock surfaces disperse the secreted material and release cellular compounds and volatile substances. Gland secretions are a mixture of proteins ($\sim$80%), lipids ($\sim$20%), and volatile molecules. Although the chemical composition and chemosignaling functions of the lipid fraction have been analyzed in a range of biological contexts (5, 6), the identity and functions of the proteins in femoral gland secretions remain completely unknown (7).

Marine iguanas (*Amblyrhynchus cristatus*) are endemic to the Galápagos islands and are one of the few squamate species in which single males defend a resource-less territory to attract females during the reproductive season. This lek system can be found in several species including birds, fish, reptiles, and mammals (8) but is rare in lizards.

Marine iguanas exhibit territorial behavior and sexual dimorphic coloration. Males attract females via specific lipophilic substances released by their femoral glands. Certain lipids of femoral gland secretions of male marine iguanas are thought to be involved in territory marking and for communication of male quality to females (9).

Mass spectrometry-based proteomics is a useful approach to investigate protein expression in tissues and body fluids. However, proteomics requires established databases containing annotations from a reference genome, which were not previously available for the marine iguana. Until now, the phylogeny and population structure of marine iguanas have mainly been studied in an evolutionary context by analysis of mitochondrial genes and standard nuclear genes, or by detection of genome-wide polymorphisms based on restriction site-associated DNA (RAD) sequencing (10). The UniProtKB counts only 75 protein entries (73 of them unreviewed) for marine iguanas, illustrating our rudimentary knowledge of protein annotations for this unique iguanid species.

By assembling the first proteomic data set for marine iguanas, we sought to identify and determine the biological functions of the proteins present in femoral gland secretions of lizards. Next-generation sequencing of multiple tissues and *de novo* assembly of the transcriptome have been successfully applied in several other organisms with unknown genomes, including the newt (*Notophthalmus viridescens*) (11) and the African clawed frog (*Xenopus laevis*) (12). By sequencing the transcriptome of various tissues, we were able to establish a comprehensive protein database and identify over 7,500 proteins in the marine iguana. Furthermore, sequence homology searches enabled us to correlate proteins secreted by the femoral gland with functions relating to lipid transport, protease inhibition, and activation of the immune system.

EXPERIMENTAL PROCEDURES

See SI Methods for complete details.

Tissue Harvesting—Several tissues and secretions (brain, erythrocytes, femoral gland secretions, heart, lung, muscle, serum, sperm, and skin tissue) were harvested from an individual marine iguana from Genovesa island that had died of natural causes. The Galápagos National Park authority granted the research permission for this study.

Transcriptomics—Sequencing was performed on a NextSeq500 instrument (Illumina, San Diego, California) using v2 chemistry, resulting in an average of 80 to 240 M reads per library with the 2 × 150 bp paired end setup. The resulting raw reads were assessed for quality, adapter content, and duplication rates using FastQC (13). Trimmomatic version 0.33 was employed to trim reads if quality dropped below a mean of Q18 in a window of five nucleotides (14). Only reads above 30 nucleotides were selected for further analyses.

Assembly of ORFs—For *de novo* assembly of mRNA reads, Trinity (v2.0.6) was used in RF-stranded mode (–SS_lib_type RF) with minimum contig criteria (–min_contig_length 300 –min_kmer_cov 2) (15). ORF detection of assembled contigs was performed with the transcoder package (http://transdecoder.github.io) using the TransDecoder.LongOrfs command with option -S for strand specific assemblies, only analyzing the top strand. Duplicate ORFs were removed and interlaced ORFs were melted using the CD-HIT package (http://weizhongli-lab.org/cd-hit/) employing the parameters (-c, 1.00 and -n, 5).

Protein Isolation, Digestion and LC–MS/MS Analysis—After in-gel digestion with trypsin and peptide desalting, reverse phase liquid chromatography was used to separate peptides using an Easy nLC 1000 UHLPC. A hybrid quadrupole Orbitrap instrument (QExactive, Thermo Fisher Scientific, Waltham, Massachusetts) was used for mass spectrometric analysis (16). Each gel fraction was measured with a 150 min LC–MS/MS gradient.

MS Data Processing—Raw MS data files were analyzed using MaxQuant software (1.5.3.8) (17). We used the translated ORFs from the marine iguana as protein database, combined with known common contaminations. Default settings were used and peptides and proteins were identified using a false discovery rate (FDR) of 1% with intensity based absolute quantification (iBAQ) being enabled. Minimum peptide length was 7 amino acids. We allowed single peptide identifications and identifications by degenerated peptides, termed razor peptides, in this manuscript and defined this way proteins groups. Proteins or protein groups identified by a single peptide are only annotated as such in Additional file 1.

Protein Annotation—We performed protein BLAST searches against the UniProt and SwissProt databases using the Bio.Blast module of Biopython (18). We searched against 11 organisms, including the American alligator, Chinese alligator, green iguana, land iguana, marine iguana, Carolina anole, chicken, human, mouse, and rat, with an E-value cutoff of E-20. BLAST-hits with the lowest E-value were extracted for every organism and used for annotation. Protein sequences without similarities were searched against the UniProt and SwissProt *E. coli* databases to exclude bacterial contaminants.

Experimental Design and Statistical Rationale—One dead animal was used for the initial assessment of the samples and statistical analysis was not applicable. Data analysis and visualization were applied as follows:

Data analysis was performed using Perseus (19), the R environment (20), and Instant Clue (21) (http://www.instantclue.uni-koeln.de/). Human gene ontology terms for marine iguana proteins were based on the closest UniProt hits in BLAST searches. To identify proteins enriched in femoral gland secretions in an unbiased manner, we employed a strategy combining kmeans clustering and principal component analysis ([PCA]; see SI Appendix, SI Materials and Methods). We identified proteins with a valid intensity value in femoral gland secretions and at least two more values in any other tissue. Because we identified a small number of proteins in serum, we did not include this sample in our later analysis.

To further characterize the remaining proteins for which homologs were not detected, we performed database searches using several online tools. DeepLoc was used to predict subcellular localization using neuronal networks (22) and signal peptide cleavage sites were predicted using SignalP (23). To provide functional annotations for proteins without any sequence-related proteins in other organisms, we used HMMER to search against the Pfam protein profile/domain database (24). In addition a BLAST search was performed against UniProt and SwissProt protein databases without any E-value cutoff. Resulting hits were sorted according to their E-value and grouped into three confidence classes: (1) E-values <1E-20; (2) E-values 1E-20 – 1E-05; (3) E-values >1E-05.

In Silico AMP Prediction—Four support vector machines (SVM) with rbf (radial basis function, gaussian), linear, polynomial and sigmoid kernels were trained using Pythons scikit-learn (25) for antimicrobial peptide sequences (AMPs). Known AMP sequences were retrieved from the CAMP (Collection of Anti-Microbial Peptides) database (26) and APD (Antimicrobial Peptide Database) (27). Sequences had to fulfill the following requirements: manually curated and experimentally validated CAMP sequences, length of 5–100 amino acids and sequence identity <90% to each other.

FIG. 1. **Deciphering the Galápagos marine iguana proteome by integration of transcriptomics and proteomics.** Workflow from tissue isolation (*A*) to database generation (*B*) and to data analysis (C).
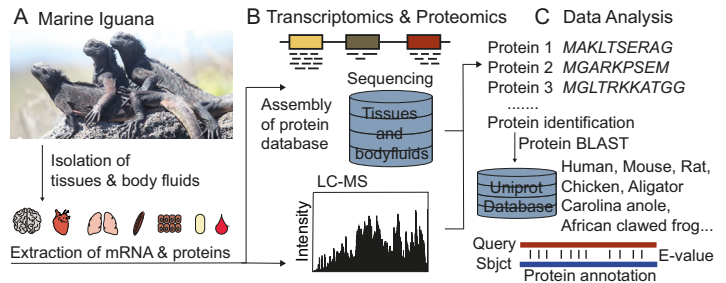
Physicochemical features, including Chou's pseudo amino acid composition (28, 29) were calculated with Biopython (18) and Propy (30). Feature selection was performed with a principle component analysis in R. The SVMs were trained with an exhaustive grid-search and 3-fold cross-validation (CV). The accuracy of the best kernels given by sklearn and the corresponding area under the curve (AUC) of the receiver operation characteristic are shown in (Additional file 2). The list of predicted AMPs was further inspected manually and only peptides with a predicted high water solubility and chemical stability were chosen for the synthesis.

Antimicrobial Activity—Test strains of Bacillus subtilis, Escherichia coli K12 or Escherichia coli BL21 were grown in Miller-Hinton broth (MHB) for 18–20 h at 37 °C with moderate shaking. An inoculum of $1–2 \times 10^7$ cfu/ml was prepared in fresh MHB and 100 µl of the suspension was pipetted into microtiter plate wells together with 20 µl test substance. Test substances were measured in duplicates. The plates were covered by a plastic seal to prevent evaporation and incubated for further 18–24 h at 37 °C with moderate shaking. The test substances were dissolved in water at a concentration of 1 mg/ml and a 2n dilution series down to 31 µg/ml was prepared in sterile PBS. After incubation, the bacterial growth was estimated with a Photometer at 600 nm and the relative growth inhibition was calculated compared with a control without test substance. As positive control the well-studied AMP oncocin was used.

Peptide Synthesis—Peptide synthesis was performed using an automatic, microwave-based solid phase peptide synthesizer using the FMOC method at 0.1 mMol scale (31). DIC/Oxyma served as activation solution, coupling was performed at 75 °C, deprotection was performed with 20% piperidine in DMF at 90 °C. After synthesis the beads were washed with DMF and lyophilized. Cleavage of the product from the resin was performed with 95% TFA, 2.5% H_2O, 2.5% TRIS for 4 h at room temperature. The resin was removed by filtration and the product precipitated with ether at −80 °C, centrifuged and washed twice with the same solvent. After removal of the ether, the peptide was dissolved in water. The peptides were purified by RP-HPLC on a C18/100 Å column using a water/acetonitrile gradient in the presence of 0.08% TFA. Identity of the peptides was verified by MALDI-TOF-MS. Peptides were stored frozen at −80 °C.

RESULTS

Database Generation and Proteomic Analysis of the Marine Iguana—We isolated several tissues, including brain, muscle, heart, and skin, from a freshly dead marine iguana found on Genovesa Island. Moreover, we collected femoral gland secretions, seminal fluid, and blood from living animals on the same island (Fig. 1). First, we performed transcriptomic analysis of five different tissues (brain, heart, lung, muscle, skin), as well as isolated erythrocytes to generate the refer-

ence protein database (11). Bioinformatic analysis of the data set revealed 499,046 contigs, mostly ranging in size from 300 – 1000 nucleotides, resulting in an N50 value of 1362 (supplemental Fig. S1*A*, S1*B*) (15). We *de novo* assembled the marine iguana transcriptome using all six potential reading frames (32), and the transcript sequences were translated into 105,810 protein open reading frames (ORFs), which served as a database for peptide identification (supplemental Fig. S1*A*).

Proteins were extracted from tissue samples (Fig.1*A*) and digested via in-gel digestion. Tryptic peptides were identified using 85 LC–MS/MS measurements, each with a gradient of 150 min. We identified 97,804 different peptides that matched to 7513 protein database entries with false discovery rates (FDRs) below 1% at the peptide and protein levels (Additional file 1). Half of the detected proteins were identified with a sequence coverage of > 23% (supplemental Fig. S1*C*). On average, we mapped ~4000 proteins per tissue with a molecular weight distribution resembling the human proteome (Fig. 2*A*) and ~900 proteins in the serum sample, indicating the high dynamic range for this body fluid (Fig. 2*B*) (33). Next, we overlapped the proteins identified in five solid organs and observed that 1551 proteins were expressed in all analyzed tissues (Fig. 2*C*), and 499 proteins were present in both the fluid samples (serum, seminal fluid, gland secretions) and isolated erythrocytes (supplemental Fig. S1*D*).

Because our transcriptome data set was based on a *de novo* assembly approach, ORF identification by itself does not provide any gene or protein information. Therefore, we sought to annotate the iguana proteome via BLAST protein sequence similarity searches against several related organisms, including the Carolina anolis (*Anolis caroliniensis*), alligator species (*Alligator mississippiensis*, *Alligator sinensis*), as well as chicken, mice, and human databases. Furthermore, we selected known marine iguana protein entries and those of two closely related iguana species (the green iguana *Iguana iguana* and the Galápagos land iguana *Conolophus subcristatus*) for identification of marine iguana proteins (supplemental Fig. S1*E*, S1*F*). Because of the evolutionary distance between the marine iguana and the selected species, we chose a stringent E-value < 10e-20 to ensure detection of highly similar proteins. In total, we transferred 7251 annotations from other
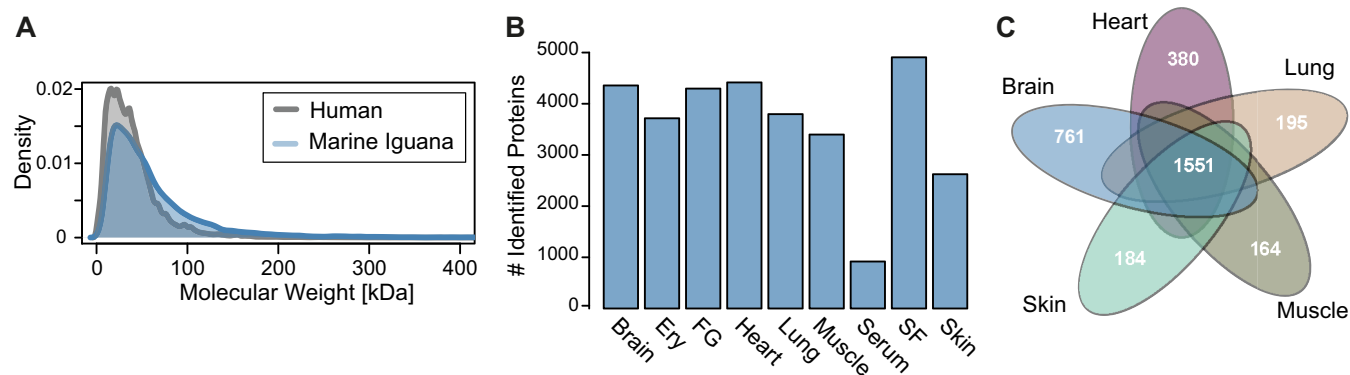
FIG. 2. **Metrics of marine iguana proteomics data.** *A*, Distribution of protein molecular weight of human proteome and marine iguana proteome. *B*, Number of proteins identified in tissues and secretions; Ery is erythrocytes, FG is femoral glands secretion and SF is seminal fluid. *C*, Overlap of identified proteins between tissues.

species, which represents an annotation efficiency of 96.5% (Additional file 1). To rule out annotation-bias toward highly abundant proteins, we clustered all proteins using their -log10 E-values, based on the sequence alignments against several organisms in the UniProt database. Protein intensities were plotted for each protein (supplemental Fig. S1*E*, S1*F*) on the resulting heat map. For example, comparison of muscle tissue from mouse and marine iguana revealed similar intensities for high abundance creatine kinase (CKM), whereas both samples showed low abundance of the E3 ubiquitin-protein ligase (UBR3) (supplemental Fig. S1*G*), supporting an unbiased protein annotation across protein intensities.

Notably, protein annotation using the protein BLAST search approach led to simultaneous annotation of several proteins to the same UniProt entry (Additional file 1). For example, seven marine iguana proteins were annotated to the human cytosolic phospholipase A2ε (PLA2G4E; supplemental Fig. S2*A*). Based on this observation, we screened the detected peptides and found unique peptides representing different phospholipase isoforms in the marine iguana data set. We therefore suggest this protein has greater diversity in the marine iguana than in mammalian species.

Identification of Proteins with No Orthologs to Other Organisms—To characterize proteins that show no recognizable sequence similarity to other species (Additional file 1), we predicted their subcellular localization using DeepLoc software (22). Thirty-two proteins were predicted to localize in the nucleus, 21 to the mitochondria, and 15 to contain signal peptides that indicate extracellular localization. Among these 15 candidates, two were also enriched in femoral gland secretions (supplemental Fig. S5; Additional file 1). Last, a domain search (24) identified 38 distinct Pfam domains on 28 proteins, of which four are restricted to femoral gland secretions. For example, the protein (TR246911|c0_g3_i1|m.144323) contains a four-disulfide core domain, similar to the antileukoproteinase SLPI. This class of proteins, also described as whey acidic proteins (WAP), has also been detected in crustaceans (34), as well as in mammalian lungs (35). In addition, a single BLAST

search with TR246911|c0_g3_i1|m.144323 only revealed an overlap to the nawaprin protein (E-value of 2e-14; 60% identity) expressed in the black-necked spitting cobra (*Naja nigricollis*). This extracellular protein—with potential peptidase inhibitor activity—might damage bacterial cell membranes.

Using these methods, we found no homology for 121 proteins (supplemental Fig. S5). To further shed light on the identity of these proteins, we performed a BLAST search using an extended set of species, including the Coral snakes *Micrurus spixii* and *M. surinamensis*. To identify also distant proteins we used the BLAST search without any E-value restriction. The resulting list of potential homologs contained proteins with E-values ranging from 0.31 to 2.5E-157. We grouped these E-values to different confidence classes, with 25 high confidence homologs, 48 low confidence homologs and 49 proteins without homologs (E-value >1E-05) (supplemental Fig. S5). We speculate that the latter 49 proteins may represent proteins specific for the Iguanidae. However, a more detailed comparison with proteomic data from other related lizard species is required to verify whether these proteins are also present in other members of the Iguanidae or if they are specific to the marine iguana.

Proteome of the Marine Iguana Femoral Gland Secretions Contains Epidermal Marker and Immunregulatory Proteins—To analyze the protein compounds, femoral gland secretions were isolated by carefully squeezing the femoral glands of the marine iguana (Fig. 3*A*). The emerging substance was immediately homogenized in a denaturing buffer and separated by SDS-PAGE gel electrophoresis. A Coomassie staining of separated proteins revealed lower protein mass distribution in marine iguanas (ranging from ~15–40 kDa), when compared with protein extracts from the heart of the same animal (Fig. 3*C*).

We characterized the marine iguana femoral gland secretions by protein BLAST searches, protein abundance, and similarity to other tissues to identify proteins that are femoral gland specific or enriched in femoral gland secretions. MS analysis identified 4305 proteins in femoral gland secretions of which 4111 proteins could be identified with at
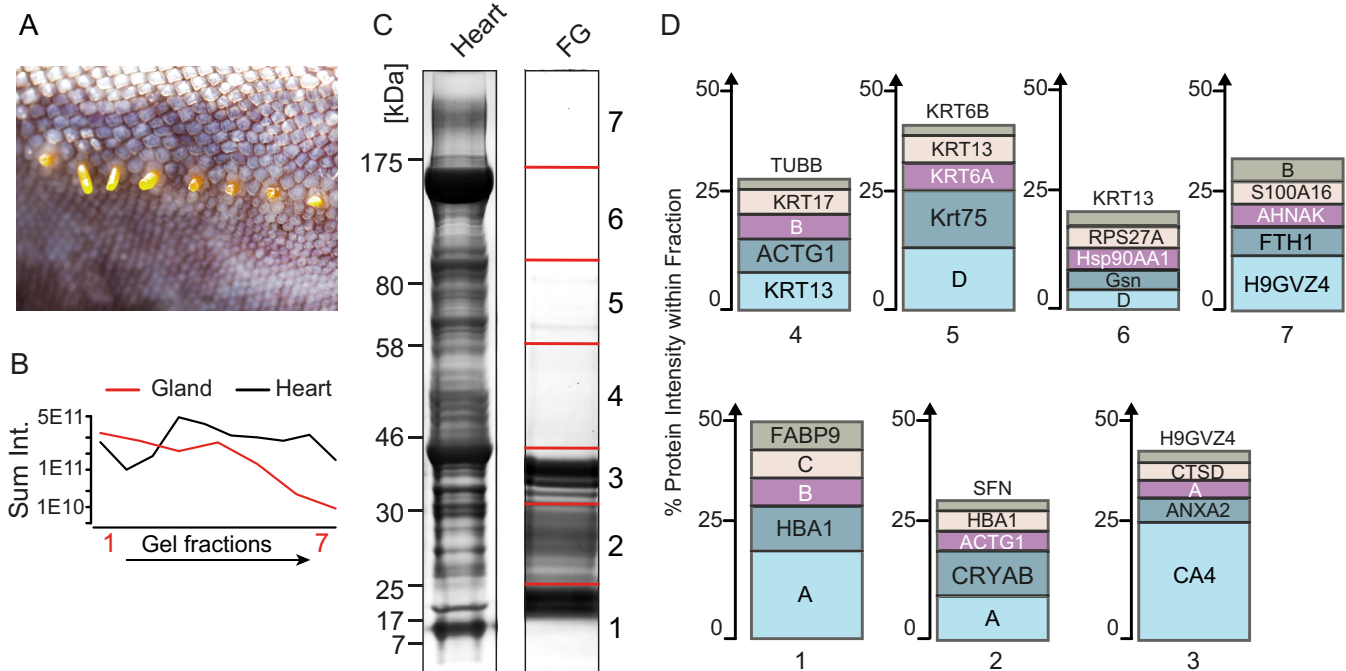
FIG. 3. **Identification of proteins enriched in the marine iguana femoral gland.** *A*, The femoral glands with secretions. *B*, Sum intensity of proteins per gel fraction in marine iguana femoral gland secretions and heart. *C*, Coomassie-stained protein gel of femoral gland secretions from heart and marine iguana. Molecular weight indicated on the left in kDa, Fraction indicated on the right. FG is femoral gland secretion. *D*, Most abundant proteins in each fraction of the femoral gland secretions. *A*: TR63421|c0_g1_i1|m.7934; *B*: WAP four-disulfide core domain protein 3-like (from the Chinese Alligator); *C*: Beta-microseminoprotein (from the American Alligator) *D*: TR238240|c0_g1_i1|m.134007.

least two peptides (see Additional file 1). Consistent with the Coomassie protein staining results, the total protein intensity and number of proteins identified per gel fraction substantially declined by one order of magnitude from band 1, containing the lowest molecular weight proteins to band 7 with the highest molecular weight (Fig. 3*B*).

We determined the five most abundant proteins in each marine iguana femoral gland secretion gel fraction. In the first fraction, the most abundant protein (A) was a 12 kDa protein, which represented ~20% of the total intensity (TR63421|c0_g1_i1|m.7934) (Fig. 3*D*). This protein was identified by 16 unique peptides and is detectable in all tissues analyzed, with the highest abundance in the femoral gland (Fig. 4*A*). Moreover, sequence alignments to other organisms at DNA and protein levels did not reveal any ortholog candidate or conserved protein domain. This protein may indicate the product of a fast evolving orphan gene, which lost sequence similarity within a short evolutionary timespan (36, 37). Furthermore, we found carbonic anhydrase 4 (CA4) as the most abundant protein with about 25% in fraction 3; this protein likely functions to regulate the pH of secretions by balancing the bicarbonate level. The detection of several keratins in the fractions 4-6 and 14-3-3 sigma (also known as epithelial cell marker protein 1) reflect the epidermal origin of femoral glands.

Femoral glands are holocrine glands, thus their secretion is composed of whole cells and includes cytosolic, cytoplasmic and secreted proteins. Metabolic and house keeping proteins in femoral gland secretions are likely shared with other cell types and tissues, which prompted us to decipher femoral glands specific proteins. To define a gland-specific proteome, we overlapped all detected femoral gland proteins with solid tissue and body fluid proteins, and found 337 proteins that are unique to femoral gland secretions (Additional file 1). To complement the set of unique femoral gland (FG) proteins with enriched FG proteins, we performed k-means clustering and identified a cluster of proteins enriched in FG, and to a lesser extent, in the seminal fluid (SF) (supplemental Fig. S3*A*–S3*B*). To further specify femoral gland enriched proteins, we applied a principal component analysis (PCA) of z-score transformed iBAQ protein intensities for SF and FG (supplemental Fig. S3*C*). We used PCA loadings and k-means cluster information to define a subset of 428 proteins (supplemental Fig. S3*D*, green circles) that are enriched in femoral gland secretions (Additional file 1).

Ranking these femoral gland-specific proteins (337 + 428 = 765) by their protein intensity enables the identification of several epidermis specific proteins, including epiplakin (EPPK1) and cornifelin (Fig. 4*B*, Additional file 1). In addition, we identified a large number of proteins with lipid-binding and lipid-modifying properties, which is consistent with the hydrophobic nature of these secretions.

Interestingly, the FG specific proteome contained proteins with human homologs that have immune responsive or host defensive properties, such as Lysozyme C, Galectin-1, and
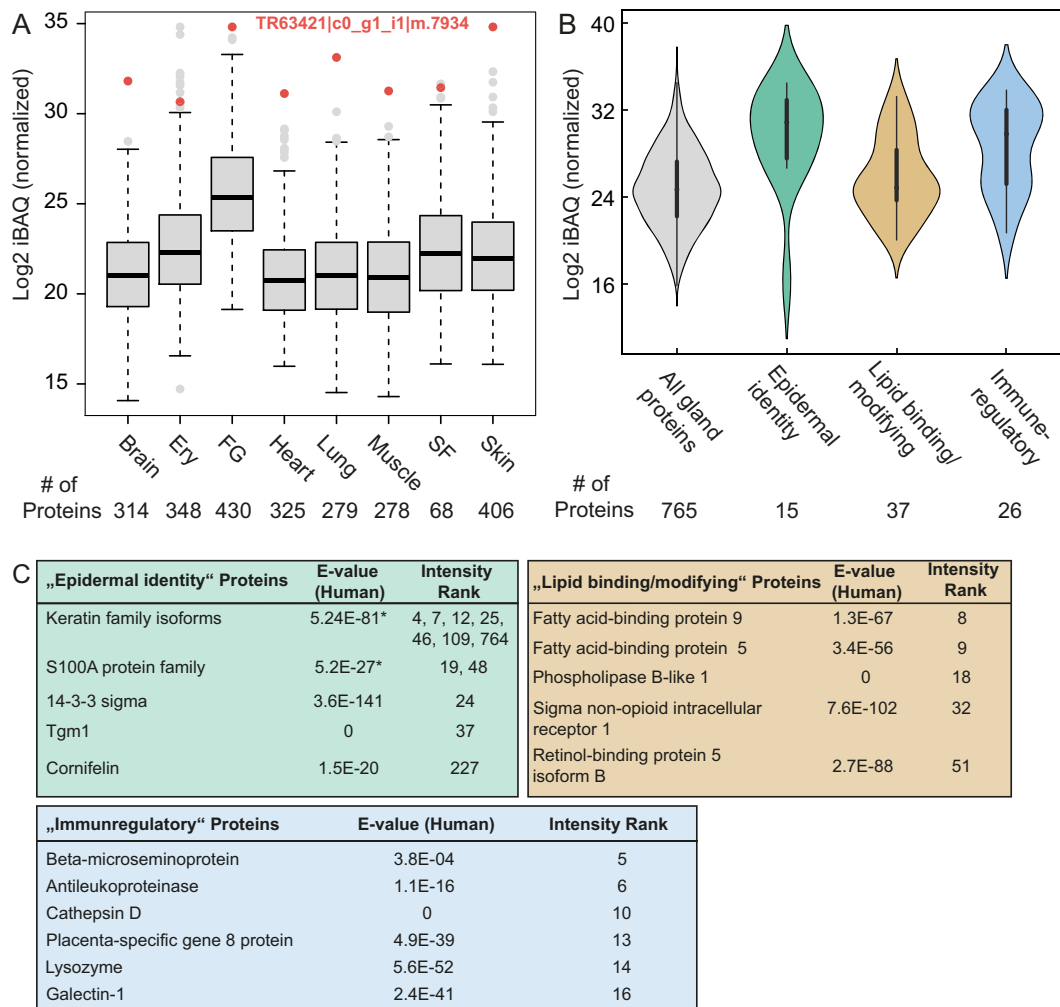
A

# of Proteins:
Brain 314, Ery 348, FG 430, Heart 325, Lung 279, Muscle 278, SF 68, Skin 406

TR63421|c0_g1_i1|m.7934

B

# of Proteins:
All gland proteins 765, Epidermal identity 15, Lipid binding/modifying 37, Immune-regulatory 26

C

„Epidermal identity" Proteins	E-value (Human)	Intensity Rank
Keratin family isoforms	5.24E-81*	4, 7, 12, 25, 46, 109, 764
S100A protein family	5.2E-27*	19, 48
14-3-3 sigma	3.6E-141	24
Tgm1	0	37
Cornifelin	1.5E-20	227

„Lipid binding/modifying" Proteins	E-value (Human)	Intensity Rank
Fatty acid-binding protein 9	1.3E-67	8
Fatty acid-binding protein 5	3.4E-56	9
Phospholipase B-like 1	0	18
Sigma non-opioid intracellular receptor 1	7.6E-102	32
Retinol-binding protein 5 isoform B	2.7E-88	51

„Immunregulatory" Proteins	E-value (Human)	Intensity Rank
Beta-microseminoprotein	3.8E-04	5
Antileukoproteinase	1.1E-16	6
Cathepsin D	0	10
Placenta-specific gene 8 protein	4.9E-39	13
Lysozyme	5.6E-52	14
Galectin-1	2.4E-41	16

FIG. 4. **Expression of immunoregulatory proteins in femoral gland secretions.** *A*, iBAQ intensity of the proteins enriched in femoral gland secretions in solid tissues and body fluid samples. FG femoral gland secretion and SF is seminal fluid. *B*, Intensity distribution of proteins enriched in femoral gland secretions. *C*, Selected secretion-specific proteins. *Depicted E-value corresponds to the protein with the lowest rank number.

Antileukoproteinase (Fig. 4*B*–4*C*). To investigate this further, we annotated human homologs with Gene Ontology (GO) terms and found 68 proteins annotated with "activation of innate immune response," and 32 proteins with "adaptive immune response" (GO: 0002250). We manually inspected the proteins with highest expression values in femoral glands and found that some proteins harbor antimicrobial properties, including Protein S100-A7, beta-microseminoprotein (MSMB, PSP94) (38), and the complement system components C5-9. In addition, three proteins were annotated as potential orthologs of mouse C-type lectin pulmonary surfactant-associated proteins, which are important components of the innate immune system as a direct defense against pathogens, and are present in the fluid lining the surface of lung alveoli. Pulmonary surfactant-associated protein D (SFTPD) belongs to the collectin protein family, and contains an N-terminal collagen-like domain and C-terminal lectin domain (39).

Peptides from Femoral Glands with Antimicrobial Activity—We were intrigued by these antimicrobial signatures and sought to systematically analyze the proteome of femoral glands. To this end, we utilized the marine iguana protein database of all detected proteins, and generated a library of potential antimicrobial peptides (AMP) by *in-silico* sliding window digestion. Here, we generated sequence windows of 8 to 15 amino acids and resulting peptides were predicted for antimicrobial activity by a trained support vector machine (as described in the methods) (Fig. 5). This step yielded in 8675 potential candidates and we selected a subset of 17 AMPs from 13 proteins with high water solubility and stability to avoid problems in solid phase peptide synthesis (Additional file 2). Of the selected peptides with predicted antimicrobial activity, nine were derived from proteins that showed an accumulation in the femoral glands and four were from proteins evenly expressed in all tissues analyzed. To experimentally test the

Fig. 5. **Screen for anti-microbial peptides in marine iguana proteome.** *A*, Working scheme for prediction of anti-microbial peptides (AMPs). *B*, Minimal inhibitory concentration of marine iguana peptides "Iguana-04" and "Iguana-02" in *Escherichia coli* K12 cultures. *C*, Minimal inhibitory concentration of marine iguana peptides "Iguana-04" and "Iguana-02" in *Bacillus subtilis* cultures. The known AMP oncocin served as a positive control for bacterial growth reduction in figures (*B*) and (*C*).
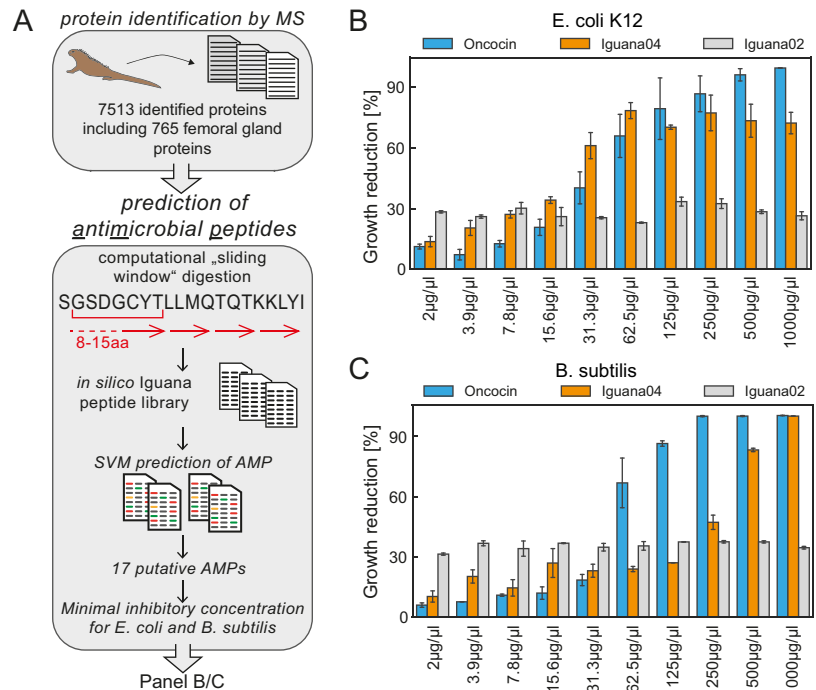
antimicrobial properties of these peptides, we determined growth rate reductions of bacterial cultures in the presence of increasing candidate peptide concentration. As positive control, we used the well-established AMP oncocin (40). As negative control, we have picked four peptides predicted to have no AMP activity. The analysis showed a moderate inhibitory activity for 14 peptides, partly without dose dependence (Additional file 2). However, peptide #4: KKLYISKGCMSKSLC demonstrated a strong antimicrobial effect against *E.coli* and *B. subtilis* cultures with a dose dependent maximal growth reduction of approx. 75% in *E. coli* cultures and 95% in *B. subtilis* cultures. Although *B. subtilis* cultures require high concentrations of "Iguana #4" for maximal growth reduction, this peptide exceeded even the oncocin activity in *E. coli* cultures at low concentrations between 3 and 60 μg/μl. The selected negative controls did not show any antimicrobial activity (Additional file 2).

DISCUSSION

The ancestral lineage of Galápagos iguanas diverged around 5–6 million years ago into the land and marine iguana lineages. Unique among lizard species worldwide, marine iguanas have evolved several distinctive adaptations to cope with the conditions of the marine environment. To date, there has been no systematic attempt to unravel the functions underlying these adaptations through transcript or protein analysis. Here, we report the first comprehensive *de novo* transcriptome assembly, together with a protein atlas for several tissues and body fluids of this sea-going iguana. In total,

we identified over 100,000 transcripts and 7513 experimentally verified protein sequences.

Among the different organs, the femoral gland plays a central role in chemical communication and territorial demarcation in lizards and amphisbaenians. Although earlier studies identified a plethora of secreted lipophilic and volatile substances, work to identify the proteins that constitute the major fraction of femoral gland secretions is entirely lacking. Proteins have been assumed to play also a major role in chemical communication (7, 9), similar to the function of secreted volatiles, which serve as olfactory pheromones. Because substrate licking is present in several members of the Iguanidae family, marine iguanas can possibly detect proteins via their vomeronasal organ to recognize occupied territories and to find potential mating partners. Perception of nonvolatile substances, such as proteins and peptides via olfactory receptor neurons or the vomeronasal organ for attraction and recognition of individuals are well described for mammals. The major urinary proteins (MUPs), for example, provide information to other individuals when recognized via olfactory sensing. Secreted MUPs function as pheromone transporters, stabilize volatile substances, provoke aggression, and play a role in attraction of mating partners (41). At a first glance, the analysis of more than 4000 femoral gland-secreted proteins indicated no functional correlates with MUPs or other proteins involved with communication, therefore, communication may occur through currently unknown secreted proteins.

Mass spectrometry showed a prevalence of femoral gland-secreted proteins within a low molecular mass range (below

~40 kDa). Degradation can be excluded, because the protein migration patterns correlated with their predicted masses (supplemental Fig. S2D). However, conclusive evidence explaining why smaller proteins are overrepresented in femoral gland secretions is lacking and this topic should be addressed in future studies. Also, the comparison of femoral gland proteins to other squamate species may help to unravel whether such low molecular weights are restricted to marine iguana secretions or is a general pattern across squamates.

The high number of cellular proteins reflects the presence of keratinocytes, granular cells, and epithelial cells in the glandular parenchyma as holocrine organs (42). Not surprisingly, keratin isoforms comprised ~10% of the secreted proteome and together with cholesterol are the major components of femoral gland matrix. Further, we found two fatty acid binding proteins (FABP5, FABP9) with high signal intensities. FABP5 is expressed in the epidermis and transports fatty acids and other lipophilic substances, such as eicosanoids. FABPs also function as lipid chaperones and are involved in the conversion of fatty acids to eicosanoid intermediates, which are present at high levels in marine iguana femoral gland secretions signaling male quality (9).

Femoral Gland Secretions as a Barrier to Microbial Infection and a Preservative Matrix—We identified in the femoral gland secretions a highly abundant protein with a high level of homology to human lysozyme C. This immune-related protein with a bacteriolytic function provides the first hint that femoral gland secretions may also function as a barrier to microbial infection. Earlier studies showed that secreted galectins (*e.g.* LGALS1) modulate cell-matrix adhesion (43). LGALS1 also functions as an immunosuppressor for T cells, and has recently been shown to be expressed at high levels in secretory granules released from cytotoxic T lymphocytes (44). Detection of placenta-specific gene 8 (PLAC8) and three orthologs of SFTPD (supplemental Fig. S4) further indicates the presence of components of the innate immune system. Thus, we speculate that marine iguana femoral glands are equipped with a battery of immune cells that protect marine iguanas against infections through the openings of glands.

As a side effect, anti-microbial proteins may also prevent the rapid degradation of femoral gland secretions by bacteria. The fatty secretions produced by the femoral glands, which are located on the ventral hind limbs of males, are actively or passively smeared onto the ground or rocks to mark a males' territory. Such volatile substances could be rapidly degraded by bacteria under the hot and humid conditions of the marine iguana's terrestrial environment, therefore an anti-microbial mixture of proteins and peptides could conserve the secretions and thus sustain the signaling function for a longer duration.

Although our sequence alignment efforts identified direct orthologs in other species for more than 98% of the proteins, several proteins did not show any correlation with existing databases, indicating the existence of marine iguana-specific proteins. For example, TR63421|c0_g1_i1|m.7934 has no homology with other species, no conserved sequence motifs, and is expressed at the highest level in the femoral gland. We would like to emphasize that other members of the Iguanidae family could express similar proteins. The generation of transcriptome databases for the Galápagos land iguana and lizards of the genus Ctenosaura—the phylogenetic sister group of Galápagos iguanas from mainland South America—could clarify whether these proteins have evolved in the Galápagos lineage or only in marine iguanas. Further studies are also needed to explore whether these proteins play a specific role in chemical communication or serve other biological function.

In contrast to the overall expectation that proteins secreted by femoral glands of lizards should function in chemical communication, we found instead clear evidence for a complex and rich molecular barrier to microbes and other potential pathogens. This work adds to previous studies that have identified protein fragments with antimicrobial activity (45–47). Based on a comprehensive catalog of femoral gland-secreted proteins, this study opens up a new avenue of research addressing the functions and potential applications of such proteins. The detection of several novel protease inhibitors and C-type lectin proteins may provide a basis for identification of new therapeutic proteins and inhibitors for protease-induced inflammatory responses in human lung- and skin diseases.

In addition, our proteomics study showed that the femoral glands might be a source for novel biologically active AMPs. A more focused analysis of processed peptides from the femoral secretions by the use of specific protocols, including removal of proteins, deactivation of enzymes, and specific peptide separation techniques such as capillary electrophoresis and top-down fragmentation (48) should be utilized to identify more AMPs. Because many AMPs are only expressed and processed after a bacterial infection (49, 50), a screening of the gland secretions of animals with varying health statuses could be another option to identify candidates and thus lead toward the development of potential new antibiotics.

CONCLUSIONS

By the use of proteomics we analyzed proteins in femoral gland secretions of the Galápagos marine iguana. In this species, femoral gland secretions contribute to male quality signaling and mate recognition. Although we found no evidence for proteins and peptides involved in chemical communication, we identified several immune-regulatory proteins, which demonstrate anti-microbial functions. Accordingly, we show that femoral gland proteins and peptides function as a barrier against microbial infection and may prevent the rapid degradation of volatile substances.

DATA AVAILABILITY

The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (51) with the data set identifier PXD018909.

REFERENCES

1. Cole, C. J. (1966) Femoral glands of the lizard, Crotaphytus collaris. *J. Morphol.* **118,** 119–135
2. Mayerl, C., Baeckens, S., and Van Damme, R. (2015) Ecolution and role of the follicular epidermal gland system in non-ophidian squamates. *Amphib. Reptilia* **36,** 185–206
3. Imparato, B. A., Antoniazzi, M. M., Rodrigues, M. T., and Jared, C. (2007) Morphology of the femoral glands in the lizard Ameiva ameiva (teiidae) and their possible role in semiochemical dispersion. *J. Morphol.* **268,** 636–648
4. Weldon, P. J., Flachsbarth, B., and Schulz, S. (2008) Natural products from the integument of nonavian reptiles. *Nat. Prod. Rep.* **25,** 738–756
5. Barbosa, D., Font, E., Desfilis, E., and Carretero, M. A. (2006) Chemically mediated species recognition in closely related Podarcis wall lizards. *J. Chem. Ecol.* **32,** 1587–1598
6. Khannoon, E. R., Dollahon, N. R., and Bauer, A. M. (2013) Comparative study of the pheromone-manufacturing femoral glands in two sympatric species of lacertid lizards (Acanthodactylus). *Zool. Sci.* **30,** 110–117
7. Alberts, A. C. (1990) Chemical properties of femoral gland secretions in the desert iguana, Dipsosaurus dorsalis. *J. Chem. Ecol.* **16,** 13–25
8. Vitousek, M. N., Rubenstein, D. R., Nelson, K. N., and Wikelski, M. (2008) Are hotshots always hot? A longitudinal study of hormones, behavior, and reproductive success in male marine iguanas. *Gen. Comp. Endocrinol.* **157,** 227–232
9. Ibanez, A., Menke, M., Quezada, G., Jimenez-Uzcategui, G., Schulz, S., and Steinfartz, S. (2017) Diversity of compounds in femoral secretions of Galapagos iguanas (genera: Amblyrhynchus and Conolophus), and their potential role in sexual communication in lek-mating marine iguanas (Amblyrhynchus cristatus). *PeerJ* **5,** e3689
10. MacLeod, A., Irisarri, I., Vences, M., and Steinfartz, S. (2016) The complete mitochondrial genomes of the Galapagos iguanas, Amblyrhynchus cristatus and Conolophus subcristatus. *Mitochondrial DNA A DNA Mapp. Seq. Anal.* **27,** 3699–3700
11. Looso, M., Preussner, J., Sousounis, K., Bruckskotten, M., Michel, C. S., Lignelli, E., Reinhardt, R., Hoffner, S., Kruger, M., Tsonis, P. A., Borchardt, T., and Braun, T. (2013) A de novo assembly of the newt transcriptome combined with proteomic validation identifies new protein families expressed during tissue regeneration. *Genome Biol.* **14,** R16
12. Wühr, M., Freeman, R. M., Presler, M., Horb, M. E., Peshkin, L., Gygi, S., and Kirschner, M. W. (2014) Deep Proteomics of the Xenopus laevis Egg using an mRNA-derived Reference Database. *Curr. Biol.* **24,** 1467–1475
13. Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data
14. Bolger, A. M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120
15. Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., and Regev, A. (2013) De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat. Protoc.* **8,** 1494–1512
16. Wiederstein, J. L., Nolte, H., Gunther, S., Piller, T., Baraldo, M., Kostin, S., Bloch, W., Schindler, N., Sandri, M., Blaauw, B., Braun, T., Holper, S., and Kruger, M. (2018) Skeletal Muscle-Specific Methyltransferase METTL21C Trimethylates p97 and Regulates Autophagy-Associated Protein Breakdown. *Cell Rep.* **23,** 1342–1356
17. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10,** 1794–1805
18. Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and de Hoon, M. J. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25,** 1422–1423
19. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., and Cox, J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13,** 731–740
20. Team, R. C. (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing
21. Nolte, H., MacVicar, T. D., Tellkamp, F., and Kruger, M. (2018) Instant Clue: A software suite for interactive data visualization and analysis. *Sci. Rep* **8,** 12648
22. Almagro Armenteros, J. J., Sonderby, C. K., Sonderby, S. K., Nielsen, H., and Winther, O. (2017) DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* **33,** 3387–3395
23. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10,** 1–6
24. Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., Bateman, A., and Eddy, S. R. (2015) HMMER web server: 2015 update. *Nucleic Acids Res.* **43,** W30–W38
25. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011) Scikit-learn. Machine learning in Python. *J. Machine Learning Res.* **12,** 2825–2830
26. Waghu, F. H., Barai, R. S., Gurung, P., and Idicula-Thomas, S. (2016) CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Res.* **44,** D1094–D1097
27. Wang, G., Li, X., and Wang, Z. (2016) APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* **44,** D1087–D1093

28. Chou, K. C. (2001) Prediction of protein cellular attributes using pseudo-amino acid composition. *Proteins* **43,** 246–255

29. Chou, K. C. (2009) Pseudo Amino Acid Composition and its Applications in Bioinformatics. *CP*. **6,** 262–274

30. Cao, D. S., Xu, Q. S., and Liang, Y. Z. (2013) propy: a tool to generate various modes of Chou's PseAAC. *Bioinformatics* **29,** 960–962

31. Collins, J. M., Porter, K. A., Singh, S. K., and Vanier, G. S. (2014) High-efficiency solid phase peptide synthesis (HE-SPPS). *Org. Lett*. **16,** 940–943

32. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., and Pachter, L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc*. **7,** 562–578

33. Anderson, N. L., and Anderson, N. G. (2002) The Human Plasma Proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* **1,** 845–867

34. Du, Z. Q., Yuan, J. J., and Ren, D. M. (2015) A novel single WAP domain-containing protein isoform with antibacterial relevance in Litopenaeus vannamei. *Fish Shellfish Immunol*. **44,** 478–484

35. Vogelmeier, C., Hubbard, R. C., Fells, G. A., Schnebli, H. P., Thompson, R. C., Fritz, H., and Crystal, R. G. (1991) Anti-neutrophil elastase defense of the normal human respiratory epithelial surface provided by the secretory leukoprotease inhibitor. *J. Clin. Invest*. **87,** 482–488

36. Domazet-Loso, T., and Tautz, D. (2003) An evolutionary analysis of orphan genes in Drosophila. *Genome Res*. **13,** 2213–2219

37. Schmid, K. J., and Aquadro, C. F. (2001) The evolutionary analysis of "orphans" from the Drosophila genome identifies rapidly diverging and incorrectly annotated genes. *Genetics* **159,** 589–598

38. Weiber, H., Andersson, C., Murne, A., Rannevik, G., Lindstrom, C., Lilja, H., and Fernlund, P. (1990) Beta microseminoprotein is not a prostate-specific protein. Its identification in mucous glands and secretions. *Am J Pathol* **137,** 593–603

39. Weis, W. I., Taylor, M. E., and Drickamer, K. (1998) The C-type lectin superfamily in the immune system. *Immunol. Rev*. **163,** 19–34

40. Knappe, D., Piantavigna, S., Hansen, A., Mechler, A., Binas, A., Nolte, O., Martin, L. L., and Hoffmann, R. (2010) Oncocin (VDKPPYLPRPRPPR-RIYNR-NH2): a novel antibacterial peptide optimized against gram-negative human pathogens. *J. Med. Chem*. **53,** 5240–5247

41. Hurst, J. L., Payne, C. E., Nevison, C. M., Marie, A. D., Humphries, R. E., Robertson, D. H., Cavaggioni, A., and Beynon, R. J. (2001) Individual recognition in mice mediated by major urinary proteins. *Nature* **414,** 631–634

42. Chamut, S., Valdez, V. G., and Manes, M. E. (2009) Functional morphology of femoral glands in the Tegu lizard, Tupinambis merianae. *Zool. Sci*. **26,** 289–293

43. Viguier, M., Advedissian, T., Delacour, D., Poirier, F., and Deshayes, F. (2014) Galectins in epithelial functions. *Tissue Barriers*. **2,** e29103

44. Clemente, T., Vieira, N. J., Cerliani, J. P., Adrain, C., Luthi, A., Dominguez, M. R., Yon, M., Barrence, F. C., Riul, T. B., Cummings, R. D., Zorn, T. M., Amigorena, S., Dias-Baruffi, M., Rodrigues, M. M., Martin, S. J., Rabinovich, G. A., and Amarante-Mendes, G. P. (2017) Proteomic and functional analysis identifies galectin-1 as a novel regulatory component of the cytotoxic granule machinery. *Cell Death Dis*. **8,** e3176

45. Dziuba, B., and Dziuba, M. (2014) New milk protein-derived peptides with potential antimicrobial activity: an approach based on bioinformatic studies. *Int. J. Mol. Sci*. **15,** 14531–14545

46. Ramada, M. H. S., Brand, G. D., Abrao, F. Y., Oliveira, M., Filho, J. L. C., Galbieri, R., Gramacho, K. P., Prates, M. V., and Bloch, C. Jr (2017) Encrypted antimicrobial peptides from plant proteins. *Sci. Rep* **7,** 13263

47. Lu, Y., Zhang, T. F., Shi, Y., Zhou, H. W., Chen, Q., Wei, B. Y., Wang, X., Yang, T. X., Chinn, Y. E., Kang, J., and Fu, C. Y. (2016) PFR peptide, one of the antimicrobial peptides identified from the derivatives of lactoferrin, induces necrosis in leukemia cells. *Sci. Rep* **6,** 20823

48. Mahlapuu, M., Hakansson, J., Ringstad, L., and Bjorn, C. (2016) Antimicrobial peptides: an emerging category of therapeutic agents. *Front. Cell. Infect. Microbiol*. **6,** 194

49. Erler, S., Popp, M., and Lattorff, H. M. (2011) Dynamics of immune system gene expression upon bacterial challenge and wounding in a social insect (Bombus terrestris). *PLoS ONE*. **6,** e18126

50. Shike, H., Lauth, X., Westerman, M. E., Ostland, V. E., Carlberg, J. M., Van Olst, J. C., Shimizu, C., Bulet, P., and Burns, J. C. (2002) Bass hepcidin is a novel antimicrobial peptide induced by bacterial challenge. *Eur. J. Biochem*. **269,** 2232–2237

51. Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Rios, D., Dianes, J. A., Sun, Z., Farrah, T., Bandeira, N., Binz, P. A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R. J., Kraus, H. J., Albar, J. P., Martinez-Bartolome, S., Apweiler, R., Omenn, G. S., Martens, L., Jones, A. R., and Hermjakob, H. (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature Biotechnol*. **32,** 223–226 Data set: PXD014706