



Integrative Organismal Biology

A Journal of the Society
for Integrative and
Comparative Biology

academic.oup.com/icb



OXFORD
UNIVERSITY PRESS



ARTICLE

Protein–lipid Association in Lizard Chemical Signals

M. Mangiacotti ^{*},¹ S. Baeckens [‡], M. Fumagalli,[§] J. Martín,[¶] S. Scali ^{||} and R. Sacchi ^{*}

^{*}Department of Earth and Environmental Sciences, University of Pavia, Via Taramelli 24, 27100 Pavia, Italy; [†]Functional Morphology Lab, Department of Biology, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium; [‡]Evolution and Optics of Nanostructures Group, Department of Biology, Ghent University, 9000 Gent, Belgium; [§]Department of Biology and Biotechnologies “L. Spallanzani”, University of Pavia, Via Ferrata 9, 27100 Pavia, Italy; [¶]Departamento de Ecología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, José Gutiérrez Abascal 2, E-28006 Madrid, Spain; ^{||}Sezione Erpetologia, Museo di Storia Naturale di Milano, Corso Venezia 55, IT-20121 Milano, Italy

¹Email: marco.mangiacotti@unipv.it

Synopsis Chemical communication in terrestrial vertebrates is often built on complex blends, where semiochemical and structural compounds may form an integrated functional unit. In lizards, many species have specialized epidermal glands whose secretions are waxy, homogeneous blends of lipids and proteins, both active in communication. The intimate co-occurrence of such compounds allows us to hypothesize that they should undergo a certain degree of covariation, considering both their semiochemical role and the support-to-lipid function hypothesized for the protein fraction. In order to assess the occurrence and level of protein–lipid covariation, we compared the composition and complexity of the two fractions in the femoral gland secretions of 36 lizard species, combining phylogenetically-informed analysis with tandem mass spectrometry. We found the composition and complexity of the two fractions to be strongly correlated. The composition of the protein fraction was mostly influenced by the relative proportion of cholestanol, provitamin D₃, stigmaterol, and tocopherol, while the complexity of the protein pattern increased with that of lipids. Additionally, two identified proteins (carbonic anhydrase and protein disulfide isomerase) increased their concentration as provitamin D₃ became more abundant. Although our approach does not allow us to decrypt the functional relations between the proteinaceous and lipid components, nor under the semiochemical or structural hypothesis, the finding that the proteins involved in this association were enzymes opens up to new perspectives about protein role: They may confer dynamic properties to the blend, making it able to compensate predictable variation of the environmental conditions. This may expand the view about proteins in the support-to-lipid hypothesis, from being a passive and inert component of the secretions to become an active and dynamic one, thus providing cues for future research.

(Italian) Riassunto La comunicazione chimica nei vertebrati terrestri è spesso costituita da miscele complesse, in cui composti semiochimici e strutturali possono formare un’unità funzionale integrata. Nelle lucertole, molte specie hanno ghiandole epidermiche specializzate le cui secrezioni sono miscele cerosi e omogenee di lipidi e proteine, entrambe attive nella comunicazione. L’intima coesistenza di tali composti lascia ipotizzare che essi debbano subire un certo grado di co-variazione, sia considerando il loro ruolo semiochimico, sia la funzione di supporto ai lipidi ipotizzata per la frazione proteica. Per valutare la presenza e il livello di covariatione proteine-lipidi, abbiamo confrontato la composizione e la complessità delle due frazioni nelle secrezioni della ghiandola femorale di 36 specie di lucertole, combinando l’analisi filogenetica comparativa con la spettrometria di massa. Abbiamo riscontrato una forte correlazione tra la composizione e la complessità delle due frazioni. La composizione della frazione proteica è stata influenzata soprattutto dalla proporzione relativa di colestanolo, provitamina D₃, stigmaterolo e tocoferolo, mentre la complessità del pattern proteico è aumentata con quella dei lipidi. Inoltre, due proteine identificate (anidrasi carbonica e disolfuro isomerasi) hanno aumentato la loro concentrazione al crescere dell’abbondanza della provitamina D₃. Sebbene il nostro approccio non consenta di decifrare le relazioni funzionali tra le componenti proteiche e lipidiche, né secondo l’ipotesi semiochimica né secondo quella strutturale, la scoperta che le proteine coinvolte in questa associazione sono enzimi apre a nuove prospettive sul ruolo delle proteine stesse: esse potrebbero conferire proprietà dinamiche alla miscela, rendendola capace di compensare le prevedibili variazioni delle condizioni ambientali. Questo può ampliare la visione

delle proteine nell'ipotesi che esse siano di supporto ai lipidi, da componente passiva e inerte delle secrezioni a componente attiva e dinamica, fornendo così spunti per ricerche future.

(Spanish) Resumen La comunicación química en los vertebrados terrestres se basa a menudo en mezclas complejas, donde los compuestos semioquímicos y estructurales pueden formar una unidad funcional integrada. En los lagartos y lagartijas, muchas especies tienen glándulas epidérmicas especializadas cuyas secreciones son mezclas cerosas y homogéneas de lípidos y proteínas, siendo ambos compuestos activos en la comunicación. La co-ocurrencia estrecha de tales compuestos permite plantear la hipótesis de que debería existir un cierto grado de covariación, considerando los posibles papeles de la fracción proteica tanto como semioquímicos y como apoyo a la función de los lípidos. Con el fin de evaluar la ocurrencia y el nivel de covariación entre proteínas y lípidos, comparamos la composición y complejidad de estas dos fracciones en las secreciones de la glándula femoral de 36 especies de lagartos y lagartijas, combinando análisis filogenéticos con espectrometría de masas en tándem. Encontramos que la composición y la complejidad de las dos fracciones estaban fuertemente correlacionadas. La composición de la fracción proteica estuvo influenciada principalmente por la proporción relativa de colestanol, provitamina D₃, estigmasterol y tocoferol, mientras que la complejidad del patrón proteico aumentó con el de los lípidos. Además, dos proteínas identificadas (anhidrasa carbónica y proteína disulfuro isomerasa) aumentaban su concentración a medida que la provitamina D₃ era más abundante. Aunque nuestro enfoque no permite descifrar las relaciones funcionales entre los componentes proteínicos y lipídicos, ni bajo la hipótesis semioquímica, ni bajo la estructural, el hallazgo de que las proteínas involucradas en esta asociación serían enzimas abre nuevas perspectivas sobre el papel de las proteínas. Estas podrían conferir propiedades dinámicas a la mezcla, haciéndola capaz de compensar la variación esperada debida a las condiciones ambientales. Estos resultados pueden ampliar la visión del papel de las proteínas en la hipótesis de apoyo a los lípidos, pasando de ser un componente pasivo e inerte de las secreciones a convertirse en uno activo y dinámico, proporcionando así ideas para investigaciones futuras.

Introduction

The chemical signals of animals often consist of a complex mixture of different classes of molecules with distinct properties (Alberts 1992; Lei and Vickers 2008; Wyatt 2010; Clifford and Riffell 2013; Junker et al. 2018). Such chemical blends typically combine volatile compounds, which spread easily and may travel far, with more heavy and stable non-volatiles, which may serve as the complementary or additional part of the information conveyed by the volatile fraction, or may favor signal stability, and prolong the durability of the volatile counterparts, or both (Alberts 1992; Nehring et al. 2013; Mucignat-Caretta and Caretta 2014; Wyatt 2014).

From an adaptive perspective, chemical mixtures, being built from different molecules, can be considered a multivariate trait (Lande 1979; Lande and Arnold 1983; Blows 2007; Murren 2012). If so, and blends represent a single, integrated functional unit, then selective forces are likely not acting separately on each single component, but rather on the covariation matrix of the mixture (Murren 2012; Armbruster et al. 2014; Felice et al. 2018). Indeed, several studies have shown that different components of a chemical mixture may influence each other (e.g., Briand et al. 2004; Mucignat-Caretta and Caretta 2014; Romero-Diaz et al. 2021), or that key messages (i.e., information) may be encoded in the relative proportions of compounds (e.g., Ozaki et al. 2005; di Mauro et al. 2015). These considerations highlight the need of an integrated approach in the evolution-

ary study of chemical blends used in inter- and intra-specific communication (Junker et al. 2018).

Over the last 30 years or so, lizards have become a model species to investigate the evolutionary forces shaping vertebrate chemical communication (e.g., Alberts 1991; Alberts et al. 1993; Martín and López 2000; Pruett et al. 2016; Baeckens et al. 2017; García-Roa et al. 2017a, 2017b; Campos et al. 2020; Ruiz-Monachesi et al. 2020; Baeckens and Whiting 2021; Mangiacotti et al. 2021). The reason for its popularity is partly due to the pivotal role of the chemical sensory modality in this taxon, which is not only characterized by readily observable chemoreception behavior of lizards (i.e., tongue-flicking; Schwenk 1995; Baeckens et al. 2017), but also by the occurrence, in most species, of epidermal glands in the cloacal region (García-Roa et al. 2017a; Cole 1966), used for inter- and intra-specific signaling (Martín and López 2014; Mayerl et al. 2015). Among them, femoral glands occur in the lizard lineages which do not suffer limb reduction (García-Roa et al. 2017a) and has been widely studied (Mayerl et al. 2015). Femoral glands consist of two symmetric series of structures along the inner part of the thighs (Cole 1966), which produce a waxy mixture of proteins (low-volatile) and lipids (high-volatile), in variable proportions (Alberts 1990; Escobar et al. 2001; Mangiacotti et al. 2019b). Both components are functional in communication (Alberts and Werner 1993a; Martín and López 2014, 2015; Mangiacotti et al. 2019a, 2020). The lipid fraction has been the focus of most studies

(e.g., Alberts et al. 1992; Khannoon et al. 2011; Heathcote et al. 2014; Martín et al. 2015a; Pruett et al. 2016; Zozaya et al. 2019; Donihue et al. 2020), which resulted in a relatively good understanding of the identity and partly of the signaling role of the involved lipids (Weldon et al. 2008; Martín and López 2011, 2014, 2015). Aside from specific molecules, many studies have also analyzed the variability of the whole lipid blend, and its association with environmental conditions (Martín et al. 2015b; Martín et al. 2017; Baeckens et al. 2018a; Ortega et al. 2019; Campos et al. 2020), phylogeny (Baeckens et al. 2018, 2018a), or individual traits (López and Martín 2005; Labra 2006; Ortega et al. 2019).

Conversely, proteins from femoral glands secretions have been far less studied (Martín and López 2014; Mayerl et al. 2015; Mangiacotti et al. 2017; Romero-Diaz et al. 2020), although they may represent most of the secretion mass (Alberts 1990; Mangiacotti et al. 2019b), they are homogeneously mixed with lipid blend (Alberts 1990; Mangiacotti et al. 2017), and can be detected by lizards (Alberts and Werner 1993a; Cooper et al. 2002;). On this basis, two main, not exclusive, functions have been theorized for secretion proteins: On the one hand, the support-to-lipid hypothesis, which states that proteins represent an inert structural matrix that holds and protects lipids in the secretion from fading (Cole 1966); and on the other hand, the semiochemical hypothesis, which considers the proteinaceous fraction to be directly involved in signaling (Alberts 1990; Alberts and Werner 1993a). Indeed, behavioral studies on the green iguana (*Iguana iguana*; Alberts and Werner 1993a), the common wall lizard (*Podarcis muralis*; Mangiacotti et al. 2019a), and Cyren's Rock Lizard (*Iberolacerta cyreni*; Mangiacotti et al. 2020) showed that proteins could play an active role in lizard chemical communication, at least conveying identity-related information. Recently, using mass-spectrometry coupled with database-searching, proteins from femoral glands have been preliminary characterized in the marine iguana, *Amblyrhynchus cristatus* (Tellkamp et al. 2020), and the sand lizard, *Lacerta agilis* (Ibáñez et al. 2022). Although both studies failed in finding proteins immediately relatable to semiochemical functions, they underscored their role in supporting the lipophilic fraction, with potential binding, regulatory, anti-oxidant, antibacterial, and immune functions (Tellkamp et al. 2020; Ibáñez et al. 2022).

Overall, the above findings suggest the idea that lipids and proteins may show a certain degree of covariation. Unfortunately, no study has investigated whether and how the composition of the lipophilic fraction of lizard chemical signals is correlated to that of its proteinaceous fraction. By systematically focusing

on solely one aspect of the lipid–protein mix, our comprehension of lizard chemical communication has been potentially biased, preventing us from investigating the potentiality for covariation effects. In the present study we examined, for the first time, the pattern of covariation between lipids and proteins from femoral glands secretions, with the specific aim of assessing whether the composition of the proteinaceous fraction reflects the variability of the lipophilic counterpart. We focused on the Lacertidae family, an Old World lizard group distributed in Europe, Africa, and Asia, including about 350 species of small-to-medium-sized lizards (with few exceptions) (Sindaco and Jeremcenko 2008; Roll et al. 2017; Garcia-Porta et al. 2019). Chemical communication plays an important role in this group (Baeckens et al. 2015; Mangiacotti et al. 2021), and it has been extensively studied (Baeckens 2019). We assembled a comprehensive dataset of the protein and lipid composition of femoral gland secretions of 36 lacertid lizard species, and analyzed the protein–lipid relation under a phylogenetic comparative framework. A tentative identification of the proteins targeted by the comparative analysis was carried out using mass-spectrometry coupled with database-searching.

Material and methods

Dataset assembly

In order to compare the covariation in lipid and protein components of lizard femoral gland secretions, we set the observation scale to the species level, and combined already published data to new, *ad hoc* ones, which covered information gaps.

Overall, the dataset included complete information about composition and richness of both proteins and lipids from femoral gland secretions of 36 lacertid species (Table S1; note on sample collections are summarized in supplementary materials), together with a recent reconstruction of their phylogenetic relations (Garcia-Porta et al. 2019). Protein and lipid samples came from the same individuals (only males) and populations (Baeckens et al. 2018a; Mangiacotti et al. 2021). The final species list was mainly driven by data availability, mostly concerning the protein fraction, which was the actual limiting factor. Below the detailed description of the data collected.

Proteins

Information about femoral gland proteins were obtained from already published data (Mangiacotti et al. 2021), and consisted in species-level one-dimensional normalized average electrophoretic profiles (electrophoretogram, EPG) from 36 lacertid species. Each specific EPG was made of 300 values representing

the relative amount of protein clusters ordered by their molecular weight (Garfin 2009). Since protein identification is still challenging because it requires species-specific databases to enable the interpretation of tandem mass spectrometry data (Eng et al. 2011; Tellkamp et al. 2020), as a working hypothesis, we assumed EPG to represent a raw proxy of the protein composition (as in Alberts 1991; Alberts et al. 1993; Mangiacotti et al. 2019b, 2021). From EPG, we also obtained an estimation of the complexity of the protein mix as (1) the number of identifiable peaks (R_P ; Table S1), corresponding to the number of bands in the electrophoretic profile, and therefore correlated to protein richness, and (2) the Shannon's entropy (H_P ; Table S1), an information measure of EPG complexity. R_P was extracted from each EPG using the function `detect_localmaxima` of the R package `scorepeak` (Ochi 2019), setting the searching window to three ($w = 3$) and selecting as true peaks those being at least 10% as high as the highest peak. H_P was directly computed from EPG using the formula $H_P = -\sum_i p_i \cdot \log(p_i)$, where p_i represents the normalized protein abundance for the i th molecular weight bin (all EPGs were series of 300 values; Mangiacotti et al. 2021). All operations were implemented in R v3.5.2 (R Core Team 2018) adapting the functions available in Mangiacotti et al. (2019b).

Lipids

For each of the 36 lacertid species for which protein data were available, we retrieved three kinds of information about lipids. Firstly, we obtained average specific lipid composition at chemical class level (Weldon et al. 2008) from already published data (Baeckens et al. 2018b). For each species, we organized all compounds in 11 main classes, based on their chemical nature and the length of their carbon chains (i.e., volatility; Baeckens et al. 2018a): low weight alcohols ($C \leq 16$), high weight alcohols ($C > 16$), aldehydes, low weight fatty acids ($C \leq 16$), high weight fatty acids ($C > 16$), waxy esters, furanones, ketones, terpenoids, tocopherols, and steroids. This grouping allowed, including both identified and confirmed compounds, and compounds that, although their specific identities were unknown, could be assigned to a chemical class based on their mass spectra characteristics. Moreover, with this grouping, we characterized the lipid fraction at the same resolution already used in previous studies (Gabirot et al. 2008; Khannoon et al. 2011, 2013; Martín et al. 2013, 2015b; Baeckens et al. 2017; García-Roa et al. 2017b; Baeckens et al. 2018, 2018a). The reported average percentage in each class was used as the lipid profile of each species. Secondly, we focused on the relative abundance of eight main compounds which are known to be active in chemical communication or in hold-

ing signal efficacy (Weldon et al. 2008; Martín and López 2014, 2015; García-Roa et al. 2017b): cholesterol, campesterol, stigmasterol, ergosterol (provitamin D₂), 9,12-octadecadienoic acid (linoleic acid), α -tocopherol (vitamin E), cholestanol, and cholesta-5,7-dien-3-ol (provitamin D₃). Despite absent in some species, the chosen compounds are shared (at different concentration) among all individuals of the same species (Martín and López 2014; García-Roa et al. 2017b), not leading to artifact average profiles. For half of the species, data were directly available from García-Roa et al. (2017b). For the remnant species, for which data were not available from literature, the original GC-MS data from Baeckens et al. (2018a) were re-analyzed, and the relative abundance of the targeted compounds was calculated as the relative proportion of the total ion current, in exactly the same way as in García-Roa et al. (2017b). Thirdly, we included two overall measures of the complexity of the lipid mixture (Baeckens et al. 2018a): the total number of distinct compounds (identified and unidentified) found in the GC-MS analysis defined the lipid richness R_L (Baeckens et al. 2018a); the Shannon's entropy of the lipid profiles (H_L), computed applying the same formula used for H_P , defined lipid complexity.

Phylogeny

The phylogenetic tree, including all the studied species was obtained starting from the phylogeny of lacertid lizards available in García-Porta et al. (2019). The original tree, based on phylogenomic analysis of 246 species, was pruned to fit the 36 species of the present study using the `keep.tip` function of the `ape` R package (Paradis and Schliep 2019).

Statistical analysis

Before analysis, some data were transformed to account for their compositional nature (Aitchison 1982; van den Boogaart and Tolosana-Delgado 2013). Notably, we applied the centred-log-ratio transformation (Aitchison 1982) to the matrix of normalized EPG (proxy for protein composition) and that of lipid composition. Similarly, the matrix of the abundance of the eight main lipids was first closed (i.e., normalized to sum to one), in order to obtain the relative proportion of each compound compared to the others; then, the centred-log-ratio transformation was applied, obtaining the final matrix. In the above cases, a back-transformation of the data was applied when interpreting results (Aitchison 1982; van den Boogaart and Tolosana-Delgado 2013). R_P , H_P , R_L , and H_L were kept in their original scale.

As a first step, we assessed the degree of covariation between proteins and lipids using a phylogenetic partial least square analysis (Adams and Felice 2014), implemented by the function `phylo.integration` of the R pack-

age geomorph (Adams and Otárola-Castillo 2013). This analysis quantifies the degree of covariation of two phenotypic trait matrixes, after accounting for the phylogenetic dependence among the observations (species), which is used to transform the original trait values under a Brownian motion model of evolution (Adams and Felice 2014). Besides being conventionally used as a “null-model” for trait evolution (Felsenstein 1985), Brownian motion was selected for analogy with the evolutionary mode of lipids from femoral gland secretions, as assessed at single compound level across the lizard tree (García-Roa et al. 2017b). The partial least squares output is a correlation coefficient (r_{PLS}), whose statistical significance is assessed via permutations (Adams and Felice 2014). The approach is well suited for high-dimensional matrixes (i.e., more variables than observations; Adams and Felice 2014; Collyer et al. 2015), as it is the case for EPG (36 rows \times 300 columns). We set the number of permutations to 9999. A second analysis was run considering EPG and the relative abundance of the main eight lipid compounds as covarying matrixes.

Secondly, we used a phylogenetic regression via generalized least squares (Adams 2014) to investigate whether the abundance of the eight specific lipids affects the EPG, and which molecular weight regions are more influenced. We used the function `procD.pgls` from `geomorph` package, with normalized EPGs as the response matrix and the absolute abundances of the eight main compounds as predictors. Significance of predictors was assessed via residual permutations, using marginal (type III) sum of square and cross product computation (Adams and Collyer 2015). Model predictions were then used to identify the EPG regions with the highest response to increase abundance of significant lipids.

Thirdly, phylogenetic regression was also used to correlate protein- to lipid-complexity: in turn, protein richness (R_p) or complexity (H_p) entered the model as dependent variables, with corresponding lipid richness (R_L) and complexity (H_L) used as the only predictor. Although a parametric model is available for univariate response variables, we preferred using the same, multivariate, distance-based analysis, to keep a homogeneous approach to parameter estimation and significance assessment, considering that the two methods give numerically identical results for univariate data (Adams 2014).

Protein identification

As phylogenetic regression indicated a statistically significant relationship between EPG and the abundance of provitamin D₃ (see results), we investigated the proteinaceous fraction in more detail with the aim to identify the specific proteins involved. First, we localized the molecular weight regions of EPG predicted to

be related to provitamin D₃ increase; then, we searched specific EPGs bearing the highest expression in those three weight regions. We found the EPG of *Podarcis muralis* to have such characteristics. Thus, we retrieved the original protein samples used in the analysis from (Mangiacotti et al. 2021), which were still available as frozen PBS solution. Following the same methodology described in (Mangiacotti et al. 2019b), we performed a novel Sodium dodecyl sulphate-polyacrylamide gel electrophoresis of the samples. From the freshly obtained gels, we carefully excised the bands occupying the molecular weight regions of interest, which were put in distinct tubes (one for each region) and analyzed with mass spectrometry (see supplementary material for protocol details). We performed protein identification using peptide-spectrum matching (Eng et al. 2011; Nesvizhskii 2014) against UniProt *Podarcis muralis* reference proteome (UP000472272; Bateman et al. 2021), with MS-GF + v2022.01.17 (Kim et al. 2008; Kim and Pevzner 2014) (see supplementary material for adopted settings and procedures). Protein identification was achieved when at least two different peptides match the same database entry with false detection rate below 0.01 (Mangiacotti et al. 2019).

Results

Protein–lipid covariation

Phylogenetic partial least squares failed to identify a significant correlation between protein profiles (EPGs) and the class-level lipid composition ($r_{\text{PLS}} = 0.625$; $P = 0.670$; 9999 permutations), but revealed a significant association between the EPG pattern and the relative abundance of the eight lipids that are presumably active in chemical communication ($r_{\text{PLS}} = 0.747$; $P = 0.012$; 9999 permutations; Fig. 1A). Notably, protein variability seemed most related to the variation in the relative abundance of cholesterol, provitamin D₃ (cholesta-5,7-dien-3-ol), stigmasterol, and vitamin E (α -tocopherol) according to their scores along the first axes (Fig. 1B). The most responsive regions of the protein profile (Fig. 1C) were roughly four: at 60 kDa, with a single peak involved; between 30 and 40 kDa, with three main peaks; between 12 and 20 kDa, with three recognizable peaks; and below 10 kDa, with two peaks. However, pGLS confirmed a significant main effect on EPG only for provitamin D₃ (Table 1). As the abundance of this steroid lipid increased, the predicted electrophoretic pattern showed more intense and sharpen bands in three molecular weight regions (Fig. 2), respectively, at 47.7, 38.8, and between 14.5 and 16.2 kDa. The last two of these regions corresponded to two out of the four previously highlighted by partial least square analysis.

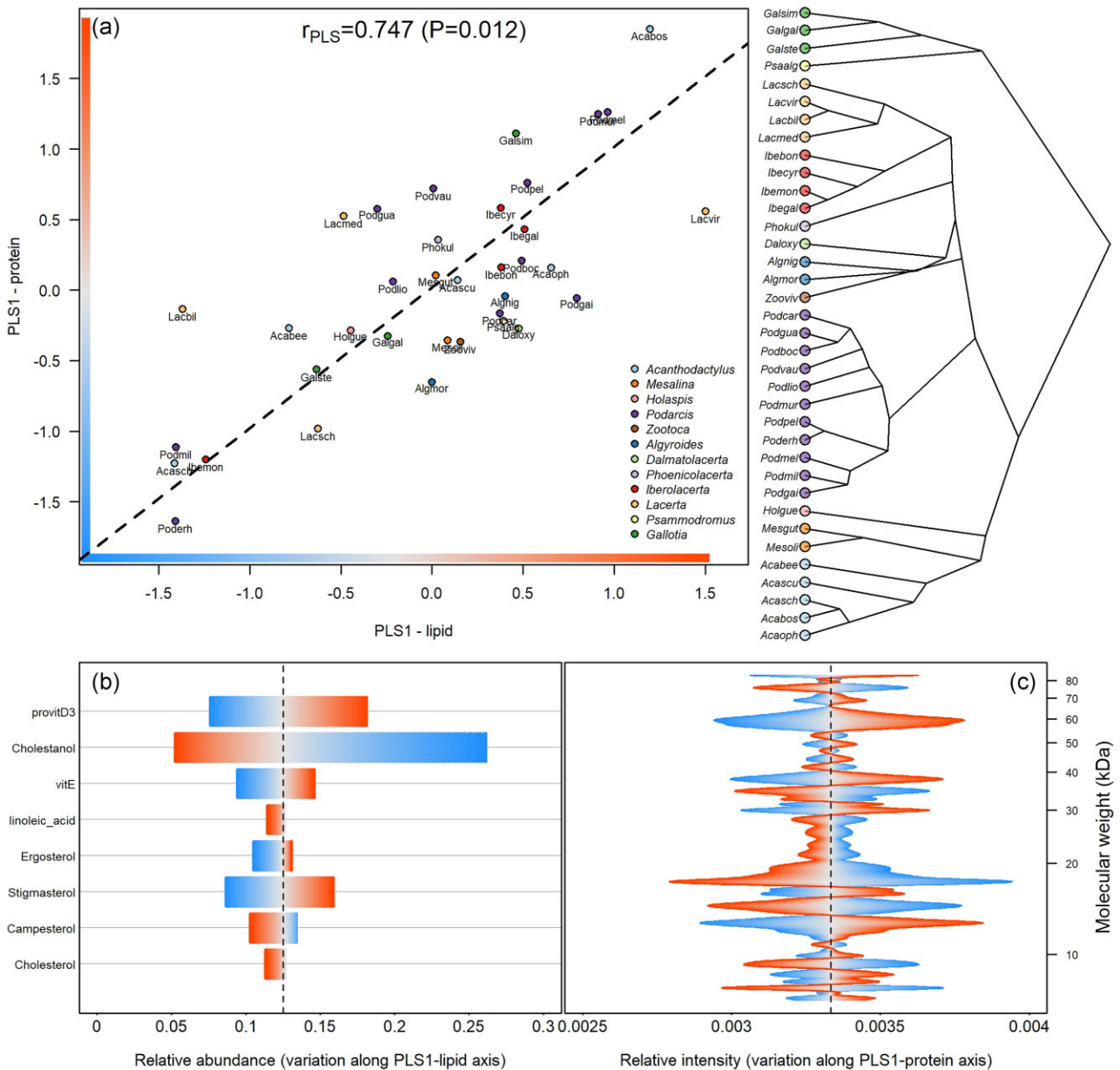


Fig. 1. Representation of the phylogenetic partial least square regression between the compositions of the protein and lipid fraction of the femoral gland secretions of 36 lacertid species. (a) Plot of the scores along the first axis (PLS1) of the lipid- and the protein-matrix, respectively; only the main eight lipids were considered; the blue-to-red bars represent increasing scores along both axes; regression line is dashed; points are colored by genus. Right panel: cladogram of the species according to the phylogeny by (García-Porta et al. 2019). (b) Relative abundance of the eight main lipids along the PLS1-lipid axis; color reflects score values as in panel A, in a way that light blue and orange correspond to low and high scores, respectively. (c) Relative intensity of the EPG along the molecular weight gradient associated with the PLS1-protein axis; light blue and orange coloration as in panels A and B.

Finally, the complexity of the protein pattern was found to be significantly correlated to lipid richness, whatever the chosen index were ($R_p \sim R_L$: pseudo- $F = 6.040$, $P = 0.017$; $H_p \sim H_L$: pseudo- $F = 7.552$, $P = 0.010$; 9999 permutations): as R_L increased, more peaks (i.e., protein clusters) were identified in the EPG

(R_p); the same occurred for the Shannon's indexes computed for protein- and lipid-profiles (Fig. 3).

Protein identification

The electrophoretic run on *Podarcis muralis* samples allowed isolating the three molecular weight regions

Table 1. ANOVA-like table from the multivariate phylogenetic regression model predicting protein composition (EPG) using the abundance of the eight main lipid compounds as independent variables. Type III sum of squares was used and significance of *F* values (pseudo-*F*) were tested using 9999 permutations (Adams and Collyer 2015). Bold denotes statistical significance ($P < 0.05$).

Variable	pseudo- <i>F</i>	<i>P</i>
Cholesterol	1.584	0.085
Campesterol	1.341	0.175
Stigmasterol	1.545	0.123
Ergosterol	1.177	0.288
Linoleic acid	1.450	0.122
Vitamin E	0.922	0.512
Cholestanol	1.161	0.280
Provitamin D₃	2.235	0.020

(bands at about 47.7, 38.8, and 14.5–16.2 kDa, respectively; Fig. S1) associated with the increased abundance in provitamin D₃. Tandem mass spectrometry provided analyzable spectra, which led to peptide identifications in two out of the three gel bands. Specifically, five

different peptides between 10 and 22 amino acids length were identified in the lighter band (A: 14.5–16.2 kDa; Table 2): three of them corresponded to the same database entry, namely “Protein disulfide-isomerase A6” (449 amino acids, 49.176 kDa), whose reported function is to catalyze the rearrangement of -S-S-bonds in proteins (Bateman et al. 2021). Despite the mismatch between predicted and observed molecular weight (about three times lower), the identification of multiple peptides from the same protein, each with low parental ion error and high score, makes the identification at protein level highly reliable. Two other peptides, not corresponding to the previous entry were also identified: a 12 amino acids sequence, showing low parental ion error and high score, and matching an uncharacterized protein (282 amino acids, 31.104 kDa; Table 2); a 22 amino acids length peptide, with high error and relatively low score, pointing to a large metallo-peptidase (1760 amino acids, 197.2 kDa; Table 2). The lack of multiple peptides matching the same protein, and the incompatible molecular weights (Table 2) suggested both identifications to be unlikely.

Twenty different peptides were identified in the second band (B: 38.8 kDa, Table 2), ranging from 7

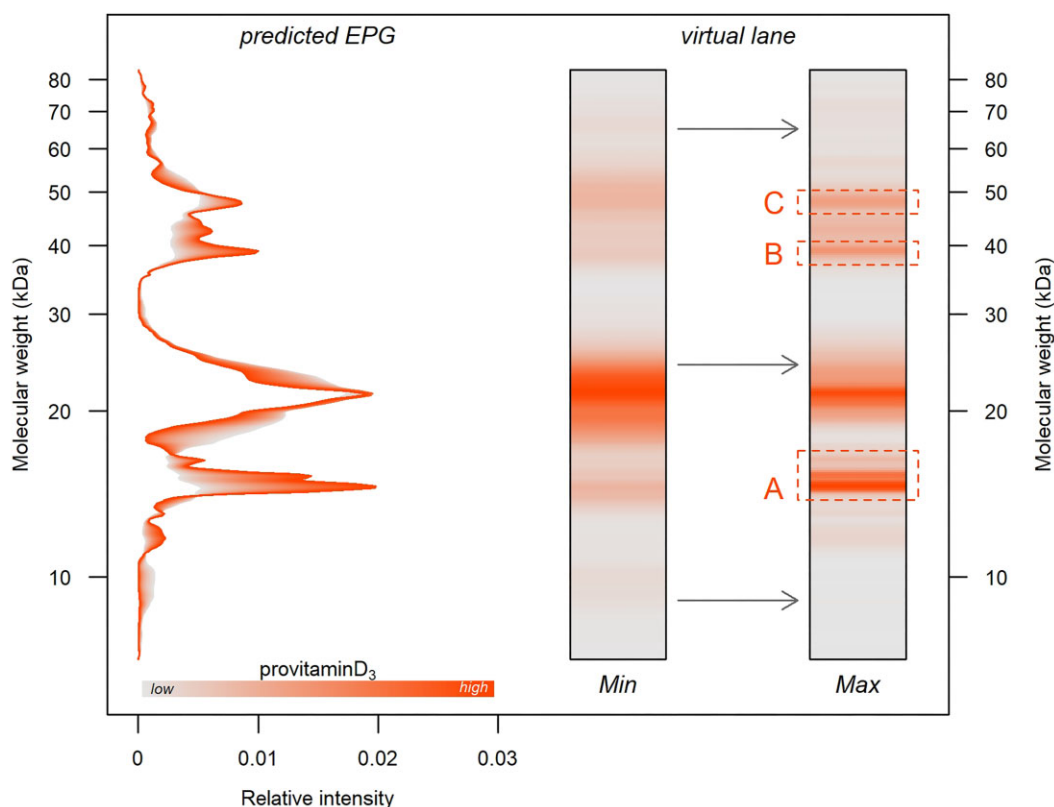


Fig. 2. Effect of increasing abundance of provitamin D₃ (grey-to-orange coloration) on the protein profile as predicted by the phylogenetic regression model. On the left, predicted electrophoretic profile (EPG); on the right, reconstruction of the electrophoretic running (virtual lane) corresponding to the minimum and maximum observed provitamin D₃ abundances. Orange boxes highlight the three molecular weight regions (A, B, and C) of maximum variation, which were used to target mass spectrometry analysis (Table 2).

Table 2. Results of peptide–spectrum matching. Only target peptides with false detection rate < 0.01 were shown. Identifications are grouped by the three molecular weight regions considered (Band; Fig. 2 and S1), and by descending scoring. MW region = molecular weight region of the gel band according to the phylogenetic regression model; Sequence = amino acids sequence of the identified peptide; Length = length (number of amino acids) of the identified peptide; Error = difference between the measured and calculated parental ion mass (Da); Score = MSGF + spectrum E-value ($-\log_{10}$ transformed); FDR = false detection rate at the peptide level; Accession = uniprotKB accession id; Description = uniprotKB protein description.

Band	MW region	Sequence	Length	Error	Score	FDR	Accession	Description
A	14.5–16.2	VGAVDADKHNLSLGGQYGV	19	0.013	24.861	0.000	A0A670J0 × 0	Protein disulfide-isomerase A6
		TSDAIVDAALSAIR	14	0.005	15.985	0.000		
		NRPEYQGG	10	0.005	9.465	0.000		
		LSIAVHAECTK	12	0.000	14.990	0.000	A0A670IPJ5	Uncharacterized protein
		QSSHRTQWRFGSWTPCSTTCGK	22	1.003	8.066	0.000	A0A670J7U0	ADAM metalloproteinase with thrombospondin type 1 motif 9
		YTQPDSCLGPLTSLGYEDR	19	0.016	22.735	0.000	A0A670JGA4	Carbonic anhydrase
		SFEAPFDGVVNFVFLASSLR	18	0.019	22.633	0.000		
		WCYDKPSCGGLTW	14	0.009	20.115	0.000		
		GNSVPLNGEFLSLLRPGGR	21	0.016	18.868	0.000		
		MQDNYRPVQPLNRR	15	0.003	18.426	0.000		
B	38.8	ELLNKGDYDVEMK	14	0.009	17.666	0.000		
		MQDNYRPVQPLNRR	14	0.007	17.569	0.000		
		SFEAPFDGVVNFVFLASSLR	19	0.021	17.380	0.000		
		EAAKDPQGFVVLGFFIQK	18	0.012	16.105	0.000		
		KFTNSLFTTK	11	0.006	14.234	0.000		
		FTNSLFTTK	10	0.005	13.374	0.000		
		KQSPINIVTR	10	0.006	12.254	0.000		
		QSPINIVTR	9	0.005	10.991	0.000		
		VHPYVVKK	8	0.002	10.500	0.000		
		VHPYVVK	7	0.001	10.190	0.000		
C	47.7	ITDKYYR	7	−0.001	9.810	0.000		
		HGGFGFKNSGRLLPGR	17	0.964	8.666	0.000	A0A670HZ81	Uncharacterized protein
		KSTIWQFFSRLLFSSSSPPPAK	22	1.021	8.237	0.000	A0A670JMD3	Mitogen-activated protein kinase 8 interacting protein 3
		KEGDGNWKK	9	0.017	7.937	0.000	A0A670IGN5	Sodium/potassium-transporting ATPase subunit beta
		WFAKLADFRLLFPPR	15	−0.013	7.926	0.000	A0A670JQ55	Acid sensing ion channel subunit 2
		No match						

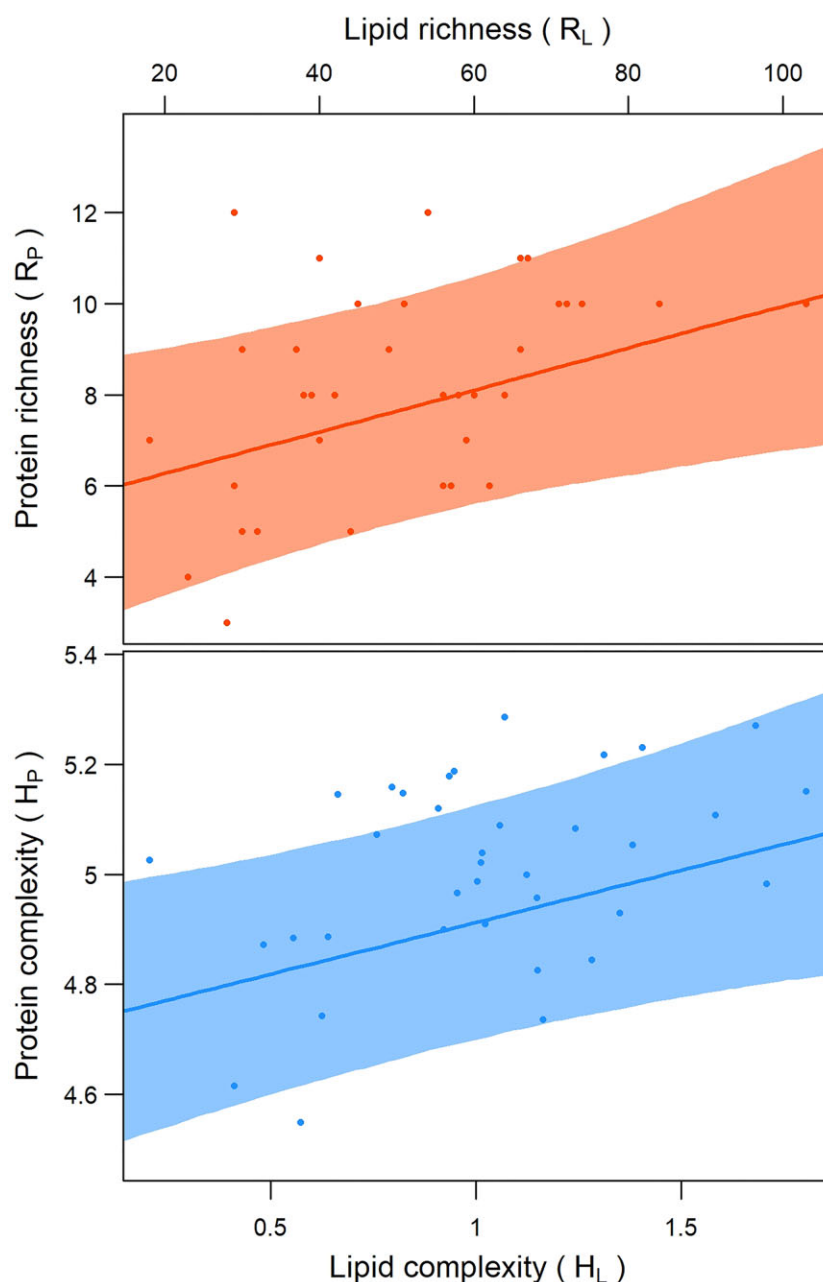


Fig. 3. Relation between protein and lipid complexity according to the phylogenetic regression models. Upper panel: protein richness (R_P), measured as the number of peaks along the EPG, on the y -axis, against lipid richness (R_L = no. of compounds), on the x -axis; solid line represents the average model prediction, grey area the 95% confidence interval of the line. Bottom panel: the same as the upper panel when using Shannon's entropy to measure protein (H_P , y -axis) and lipid complexity (H_L ; x -axis).

to 22 amino acids length. Eighteen corresponded to the same database entry, thus representing different fragments of the same protein, i.e., carbonic anhydrase (316 amino acids; 35.9 kDa; Table 2). The predicted molecular weight for this protein agreed with the position the band occupies in the SDS-PAGE according to the molecular weight ladder (Fig. S1), and it was not far from the weight region predicted by phylogenetic regression (Fig. 2). Together, these results made

the identification highly credible. The remnant four sequences (Table 2) matched different entries, some of them showing high parental ion error, and all scoring worse than those corresponding to carbonic anhydrase. Again, the lack of multiple sequences matches, made these identifications poorly reliable.

In the end, no relevant match against the *P. muralis* proteome was found for the peptides obtained from the heavier band (C: 47.7 kDa; Table 2).

Discussion

Protein–lipid covariation

Empirical and theoretical work suggest that the proteinaceous and lipid fraction of lizard gland secretions may be intertwined, and thereby contribute (even with potentially different roles) to the functionality of lizard chemical display (Alberts 1990; Alberts and Werner 1993a; Mangiacotti et al. 2017, 2019a, 2020; Tellkamp et al. 2020; Ibáñez et al. 2022). Considering the Lacertid family, the quantitative phylogenetically-informed analysis confirm such a co-variation between proteins and lipids to occur. We indeed found a significant correlation between the proteinic and lipid composition when focussing on a subset of eight specific lipid compounds with known structural or signalling properties. Notably, EPGs clearly responded to the change in the relative proportion of cholesterol, provitamin D₃, stigmaterol, and tocopherol, by increasing the concentration of proteins in certain molecular weight regions and decreasing it elsewhere. Additionally, an increase in provitamin D₃ absolute concentration corresponded to a sharper expression of three EPG bands, suggesting the abundance of specific lipids to be related to that of specific proteins. Interestingly, a similar relation, involving almost the same molecular weight regions, was previously observed in the seasonal variation of the EPG of *P. muralis*, where the change in protein profiles matched that of the secreted provitamin D₃ (Mangiacotti et al. 2019b). No relation was found when the composition of the lipid mixture is organized according to chemical classes (i.e., steroids, aldehydes, ketones, fatty acids, etc.). This may suggest that the protein–lipid association is not based on the occurrence in the lipid blend of specific chemical structure or functional groups, but rather of specific compounds, active in signal communication of pivotal to signal design (García-Roa et al. 2017b).

In the light of the support-to-lipid hypothesis, our finding of both general and specific correlations between the lipid and protein fraction in lizard secretions suggests additional, previously not theorized, roles for secretion proteins. Notably, proteins may contribute to enhance signal efficacy (*sensu* Endler 1993) by setting a “chemical environment” that inhibits more labile molecules (such as lipids) from oxidation, fast evaporation, or degradation (Alberts 1992, 1993; Wyatt 2010; Martín et al. 2016). In the iguanid *Amblyrhynchus cristatus* (Tellkamp et al. 2020) and the lacertid *Lacerta agilis* (Ibáñez et al. 2022), femoral gland secretions hold proteins with chemical properties that may serve in pH regulation (e.g., carbonic anhydrase), anti-oxidant function (e.g., thioredoxin), bacterial resistance (e.g., oncocin) or lipid-binding (e.g., fatty acid-binding pro-

tein 9), and not only in structural roles (e.g., keratin). This opens to the idea that proteins may be active in the mixture, and, by their quantitative and qualitative modulation, be involved in the very fine regulation of the efficacy, stability, and durability of the signal (Alberts 1992). Indeed, lizards typically deposit femoral secretions on bare substrates (Gabe and Saint Girons 1965; Cole 1966), and, consequently, secretions may experience hard and extreme conditions, ranging from dry to wet, over cold to hot, to high UV exposure and strong wind (Martín et al. 2016; Baeckens et al. 2018a; Campos et al. 2020). All above factors can heavily affect the ability of semiochemicals to withstand rapid degradation. Therefore, it would be worth a chemical blend enriched with compounds that counter the detrimental impact of climatic environment on the chemical properties of the blend itself (e.g., by regulating pH), or protecting signal molecules from them (e.g., from oxidation or biological degradation). A similar function has been already suggested for vitamin E (tocopherol), a lipophilic component of the secretions hypothesized to protect other semiochemicals against oxidation (López and Martín 2006; Gabirot et al. 2008). Contrary to the vitamin E case, where the function can be considered passive, i.e., protection comes from the reaction of tocopherol instead of other molecules, in the protein case, a higher-level interaction can be hypothesized, both because lipids may bond to proteins and because the latter may dynamically act on the chemical equilibrium of the mixture (Tellkamp et al. 2020; Ibáñez et al. 2022). These functional links may be expected to become gradually blurred by phylogenetic time (García-Porta et al. 2019), which may explain why only one out of the eight target lipids (i.e., provitamin D₃) was found to significantly predict EPG variation when explicitly tested in the regression model (Table 1; Fig. 2). Maybe, this relation may extend to more compounds if taxa sampling were at a lower phylogenetic level (e.g., within a single genus), where recent evolution and adaptation would be more dominant (García-Porta et al. 2019).

Alternatively, or additionally, proteins may directly contribute to the semiochemical function (semiochemical hypothesis), being themselves part of the signal (Wyatt 2014). It has been shown that the proteinaceous fraction can convey information about class- or individual-identity (Alberts and Werner 1993b; Mangiacotti et al. 2019a, 2020), and it is potentially able to allow species level discrimination (Alberts 1991; Mangiacotti et al. 2019, 2021). Further, lizards’ vomeronasal system is able to discriminate proteins (Schwenk 1995; Cooper et al. 2002). Although no protein easily relatable to a communication function has been identified so far (Tellkamp et al. 2020; Ibáñez et al. 2022), this might reflect a bias of the

database-driven protein identification process: Only the available annotated functions are associated with each identification, but multiple, unknown, roles cannot be excluded (Ibáñez et al. 2022). Even more so considering the peculiar chemo-physical properties of the secretions (i.e., dehydrated, lipophilic, and solid; Cole 1966; Mayerl et al. 2015), which are quite different from the physiological conditions where proteins usually exert their functions. If so, proteins might be part of a signature mixture (*sensu* Wyatt 2010) which, as a whole, may be used for individual recognition (Alberts 1992; Wyatt 2010; Mangiacotti et al. 2021), or may constitute single-component signals, as well (Hurst et al. 2001). The occurrence of a correlation between the two chemical components may support the former hypothesis, but targeted studies are needed to disentangle between these two alternatives.

As second support to the protein–lipid integration hypothesis comes from the positive correlation between the complexity of the two mixtures: As the number of lipid compounds raises, the structure of the protein electrophoretic patterns becomes more and more articulated, both considering the number of bands (i.e., identifiable protein clusters; R_p), and their complexity (i.e., Shannon's Entropy; H_p). The complexity of the lipid matrix responds to the environmental factors, increasing in mesic- compared to xeric-conditions (Baeckens et al. 2018a). It has been suggested that humidity may favor more complex signals due to the better performance of the chemoreceptive organs in humid conditions, which eventually promotes the recruitment of more compounds to increase redundancy or conveyable information (Baeckens et al. 2018a). In this scenario, proteins may do the same as lipids, both being a part of the signalling molecules, or providing the suitable chemical environment for lipids. In the latter case, moisture may require a wider range of proteins to contrast the faster degradation potential due to the wetter conditions (oxidation, pH modifications, and bio-degradation; Chamlin 1941; Bell 2020; Tapia et al. 2020). Although other causes may explain a so high variability of the lipophilic fraction (Baeckens et al. 2018a), the crucial point is the co-occurrence of the same pattern of variation in the lipophilic and proteinaceous counterparts, which strongly supports the hypothesis of a functional covariation between the two components, irrespective of the exact role of the proteins in the blend (semiochemical- or support-to-lipid hypothesis).

Protein identification

Out of the three electrophoretic bands whose expression was related to the provitamin D₃ abundance, two (A and B, Fig. 2 and S1) led to a reliable protein

identification using tandem mass spectrometry. In both cases, a single protein is associated with each band, suggesting the actual number of proteins can be approximated by the number of bands in the gels (Alberts 1991; Mangiacotti et al. 2017).

The band A corresponds to protein disulfide-isomerase A6 (PDI), a 449 amino acid length protein known to catalyze the rearrangement of both intrachain and interchain disulfide bonds in proteins, and usually operating in the endoplasmic reticulum lumen (Ferrari and Söling 1999; Ali Khan and Mutus 2014). PDI has been already identified in femoral gland secretions of the sand lizard (Ibáñez et al. 2022) and marine iguana (Tellkamp et al. 2020), making our identification biologically relevant. Actually, the expected molecular weight (49.176 kDa; Bateman et al. 2021) is about three times the one estimated by the band position in the gel (between 14.4 and 16.0), but all the three identified peptides (which cover about 10% of the overall amino acid sequence) fall within the first thioredoxin domain of the protein (position: 21–138; Bateman et al. 2021), covering more than 36% of it. This let us hypothesize that the protein might not be integrally preserved in the secretion, but rather modified and fragmented: the estimated weight considering only the first domain ranges between 12.83 kDa (including only the 118 strictly pertaining amino acids) and 15.99 kDa (extending the sequence from the beginning to the second domain), which well matches the observed weight in the gel.

Concerning PDI functions in the mixture, notably related to provitamin D₃, any possible hypotheses is subjected to the retaining its functional activity in a completely different environment compared to the usual ones (Ali Khan and Mutus 2014) and/or with only a part of its structure preserved. If so, different conjectures can be made: (1) it might stabilize other (undetected) proteins, linked to provitamin D₃, by acting on their folding through the rearrangement of disulfide bonds (chaperone activity; Ali Khan and Mutus 2014); (2) it might control provitamin D₃ oxidation through the thioredoxin-like domain (Gruber et al. 2006; Karala et al. 2009); and (3) it might be directly associated with provitamin D₃, as a binding protein, since PDI can bind steroids (notably, 17 β -estradiol; Fu et al. 2011).

The band B corresponds to further catalytic protein, i.e., carbonic anhydrase (CA), 316 amino acids length, 35.858 kDa weight (Bateman et al. 2021), whose function is to accelerate the carbon dioxide hydration and its reverse (i.e., bicarbonate dehydration; Tripp et al. 2001; Supuran 2016). CA has been detected in the femoral gland secretions of sand lizard (Ibáñez et al. 2022) and marine iguana, where it is among the most abundant proteins (Tellkamp et al. 2020). CA predicted weight

agrees with the band position in the gel (Fig. S1), and the 22 identified peptides cover homogeneously 49% of the overall amino acid sequence, making its identification highly reliable. Assuming CA to be enzymatically active in the mixture, it is hard to specifically link its function to the provitamin D₃ abundance. Besides a generic effect on pH regulation (Tellkamp et al. 2020; Ibáñez et al. 2022), indeed, CA might play an indirect role, by acting on the stability/activity of other compounds (proteins or lipids) which, in turn, may affect provitamin D₃. For example, CA, by controlling pH, can regulate PDI catalytic activity (Wang and Narayan 2008).

Finally, for both CA and PDI, a semiochemical role cannot be excluded (see previous section of the discussion), in which case it should be expected to be related to that of provitamin D₃.

Conclusions

Our analysis revealed a relevant association between the lipid and a part of the protein fraction of the femoral gland secretions of lacertid lizards, which support the idea of an intertwined, more complex, system than previously thought. The exact functional relations between the two components are still far from being decrypted, but the occurrence of proteins with enzymatic behavior let us hypothesize that proteins may play an active role in the mixture, possibly conferring dynamic properties to the signal, i.e., making it able to somehow “react” to the predictable variation of the environmental conditions experienced by such cues. Of course, this hypothesis needs experimental supports from targeted studies: firstly, about the actual catalytic activity of the identified proteins in the secretions; secondly on the effect of such enzymatic functions on the blend properties. At the same time, we cannot exclude that proteins contribute to the semiochemical function of the blend, maybe forming with lipids a complex, individual-specific chemical profile (Wyatt 2010). Possibly, a correlative approach at a lower phylogenetic level may still be worthwhile, as it could help identifying other proteins and driving functional hypothesis formulation. As often the case, the availability of novel information leads to more questions, but our work provides new perspectives on the study of chemical communication mechanisms.

Acknowledgements

We thank three anonymous reviewers for their valuable comments, which contributed improving an early draft of the manuscript.

Funding

M.M. was funded by Italian Ministry of University and Research, “Fondo Sociale Europeo REACT EU—Programma Operativo Nazionale Ricerca e Innovazione 2014–2020”. SB was funded by the Fonds Wetenschappelijk Onderzoek (FWO 12I8822N).

Supplementary Data

Supplementary Data available at [IOB](#) online.

Data Availability

The data underlying this article are available in Zenodo repository, doi: 10.5281/zenodo.7036727.

Authors' Contributions

M.M., R.S., and S.S. conceived the ideas. M.M., M.F., J.M., and S.B. designed methodology; S.B., J.M., M.M., R.S., and S.S. collected the data; M.M., J.M., and M.F. analyzed the data; M.M. and S.B. led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Declaration of Competing Interest

The authors declare no competing interests.

Conflict of Interest

The authors declared no conflicts of interest.

References

- Adams DC. 2014. A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. *Evolution* 68:2675–88.
- Adams DC, Collyer ML. 2015. Permutation tests for phylogenetic comparative analyses of high-dimensional shape data: what you shuffle matters. *Evolution* 69:823–9.
- Adams DC, Felice RN. 2014. Assessing trait covariation and morphological integration on phylogenies using evolutionary covariance matrices. *PLoS One* 9:e94335.
- Adams DC, Otárola-Castillo E. 2013. Geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Met Ecol Evol* 4:393–9.
- Aitchison J. 1982. The statistical analysis of compositional data. *J R Stat Soc Series B Stat Methodol* 44:139–77.
- Alberts AC. 1990. Chemical properties of femoral gland secretions in the desert iguana, *Dipsosaurus dorsalis*. *J Chem Ecol* 16:13–25.
- Alberts AC. 1991. Phylogenetic and adaptive variation in lizard femoral gland secretions. *Copeia* 1991:69–79.
- Alberts AC. 1992. Constraints on the design of chemical communication systems in terrestrial vertebrates. *Am Nat* 139:S62–89.
- Alberts AC. 1993. Chemical and behavioral studies of femoral gland secretions in iguanid lizards. *Brain Behav Evol* 41:255–60.
- Alberts AC, Phillips JA, Werner DI. 1993. Sources of intraspecific variability in the protein composition of lizard femoral gland secretions. *Copeia* 1993:775–81.

- Alberts AC, Sharp TR, Werner DI, Weldon PJ. 1992. Seasonal variation of lipids in femoral gland secretions of male green iguanas (*Iguana iguana*). *J Chem Ecol* 18:703–12.
- Alberts AC, Werner DI. 1993a. Chemical recognition of unfamiliar conspecifics by green iguanas: functional significance of different signal components. *Anim Behav* 46:197–9.
- Alberts AC, Werner DI. 1993b. Chemical recognition of unfamiliar conspecifics by green iguanas: functional significance of different signal components. *Anim Behav* 46:197–9.
- Armbruster WS, Pélabon C, Bolstad GH, Hansen TF. 2014. Integrated phenotypes: understanding trait covariation in plants and animals. *Philos Trans R Soc Lond, B, Biol Sci* 369:20130245.
- Baeckens S. 2019. Evolution of animal chemical communication: insights from non-model species and phylogenetic comparative methods. *Belg J Zool* 149:63–93.
- Baeckens S, Edwards S, Huyghe K, Van Damme R. 2015. Chemical signalling in lizards: an interspecific comparison of femoral pore numbers in Lacertidae. *Biol J Linn Soc* 114:44–57.
- Baeckens S, García-Roa R, Martín J, Van Damme R. 2017. The role of diet in shaping the chemical signal design of lacertid lizards. *J Chem Ecol* 43:902–10.
- Baeckens S, Herrel A, Broeckhoven C, Vasilopoulou-Kampitsi M, Huyghe K, Goyens J, Van Damme R. 2017. Evolutionary morphology of the lizard chemosensory system. *Sci Rep* 7: 10141.
- Baeckens S, Martín J, García-Roa R, Pafilis P, Huyghe K, Van Damme R. 2018a. Environmental conditions shape the chemical signal design of lizards. *Funct Ecol* 32:566–80.
- Baeckens S, Martín J, García-Roa R, Pafilis P, Huyghe K, Van Damme R. 2018b. Data from: how environmental conditions shape the chemical signal design of lizards. Dryad, Dataset, doi: 10.5061/dryad.340g8.
- Baeckens S, Martín J, García-Roa R, Van Damme R. 2018. Sexual selection and the chemical signal design of lacertid lizards. *Zool J Linn Soc* 183:445–57.
- Baeckens S, Whiting MJ. 2021. Investment in chemical signalling glands facilitates the evolution of sociality in lizards. *Proc R Soc B Biol Sci* 288:20202438.
- Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bursteinas B et al. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 49:D480–89.
- Bell LN. 2020. Moisture effects on food's chemical stability. In: *Water Activity in Foods*. Chicago: John Wiley & Sons, Ltd. p. 227–53.
- Blows MW. 2007. A tale of two matrices: multivariate approaches in evolutionary biology. *J Evol Biol* 20:1–8.
- Briand L, Blon F, Trotier D, Pernollet JC. 2004. Natural ligands on hamster aphrosdidin. *Chem Senses* 29:425–30.
- Campos SM, Pruett JA, Soini HA, Jaime Zúñiga-Vega J, Goldberg JK, Vital-García C, Hews DK, Novotny MV, Martins EP. 2020. Volatile fatty acid and aldehyde abundances evolve with behavior and habitat temperature in *Sceloporus* lizards. *Behav Ecol* 31:978–91.
- Chamlin G. 1941. The effect of moisture on chemical reactions. *J Chem Educ* 18:386.
- Clifford MR, Riffell JA. 2013. Mixture and odorant processing in the olfactory systems of insects: a comparative perspective. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 199:911–28.
- Cole CJ. 1966. Femoral glands in lizards: a review. *Herpetologica* 22:199–206.
- Collyer ML, Sekora DJ, Adams DC. 2015. A method for analysis of phenotypic change for phenotypes described by high-dimensional data. *Heredity* 115:357–65.
- Cooper WE, Pérez-Mellado V, Vitt LJ. 2002. Responses to major categories of food chemicals by the lizard *Podarcis lilfordi*. *J Chem Ecol* 28:709–20.
- di Mauro G, Perez M, Lorenzi MC, Guerrieri FJ, Millar JG, d'Ettorre P. 2015. Ants discriminate between different hydrocarbon concentrations. *Front Ecol Evol* 3:2296–701X.
- Donihue CM, Herrel A, Martín J, Fougopoulos J, Pafilis P, Baeckens S. 2020. Rapid and repeated divergence of animal chemical signals in an island introduction experiment. *J Anim Ecol* 89:1458–67.
- Endler JA. 1993. Some general comments on the evolution and design of animal communication systems. *Philos Trans R Soc Lond B Biol Sci* 340:215–25.
- Eng JK, Searle BC, Clauser KR, Tabb DL. 2011. A face in the crowd: recognizing peptides through database search. *Mol Cell Proteomics* 10:R111.009522.
- Escobar CA, Labra A, Niemeyer HM. 2001. Chemical composition of preloacal secretions of *Liolaemus* lizards. *J Chem Ecol* 27:1677–90.
- Felice RN, Randau M, Goswami A. 2018. A fly in a tube: macroevolutionary expectations for integrated phenotypes. *Evolution* 72:2580–94.
- Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat* 125:1–15.
- Ferrari DM, Söling HD. 1999. The protein disulphide-isomerase family: unravelling a string of folds. *Biochem J* 339:1–10.
- Fu X-M, Wang P, Zhu BT. 2011. Characterization of the estradiol-binding site structure of human protein disulfide isomerase (PDI). *PLoS One* 6:e27185.
- Gabe M, Saint Girons H. 1965. Contribution à la morphologie comparée du cloaque et des glandes épidermoïdes de la région cloacale chez les lépidosauriens. *Mém Mus Natl hist Nat, Sér. A—Zoologie* 33:149–292.
- Gabirot M, López P, Martín J, de Fraipont M, Heulin B, Sinerro B, Clobert J. 2008. Chemical composition of femoral secretions of oviparous and viviparous types of male common lizards *Lacerta vivipara*. *Biochem Syst Ecol* 36:539–44.
- García-Porta J, Irisarri I, Kirchner M, Rodríguez A, Kirchhof S, Brown JL, MacLeod A, Turner AP, Ahmadzadeh F, Albaladejo G et al. 2019. Environmental temperatures shape thermal physiology as well as diversification and genome-wide substitution rates in lizards. *Nat Commun* 10:4077.
- García-Roa R, Jara M, Baeckens S, López P, Van Damme R, Martín J, Pincheira-Donoso D. 2017a. Macroevolutionary diversification of glands for chemical communication in squamate reptiles. *Sci Rep* 7:9288.
- García-Roa R, Jara M, López P, Martín J, Pincheira-Donoso D. 2017b. Heterogeneous tempo and mode of evolutionary diversification of compounds in lizard chemical signals. *Ecol Evol* 7:1286–96.
- Garfin DE. 2009. One-dimensional gel electrophoresis. In: RR Burgess, MPBT-M in E Deutscher, editors. *Methods in enzymology*. 2nd ed. Amsterdam: Academic Press. p. 497–513.
- Gruber CW, Čemažar M, Heras B, Martin JL, Craik DJ. 2006. Protein disulfide isomerase: the structure of oxidative folding. *Trends Biochem Sci* 31:455–64.

- Heathcote RJP, Bell E, d’Ettorre P, While GM, Uller T. 2014. The scent of sun worship: basking experience alters scent mark composition in male lizards. *Behav Ecol Sociobiol* 68: 861–70.
- Hurst JL, Payne CE, Nevison CM, Marie AD, Humphries RE, Robertson DHL, Cavaggoni A, Beynon RJ. 2001. Individual recognition in mice mediated by major urinary proteins. *Nature* 414:631–4.
- Ibáñez A, Skupien-rabian B, Jankowska U, Kędracka-krok S, Zając B. 2022. Functional protein composition in femoral glands of sand lizards (*Lacerta agilis*). *Molecules* 27:1–13.
- Junker RR, Kuppler J, Amo L, Blande JD, Borges RM, van Dam NM, Dicke M, Dötterl S, Ehlers BK, Etl F et al. 2018. Covariation and phenotypic integration in chemical communication displays: biosynthetic constraints and eco-evolutionary implications. *New Phytol* 220:739–49.
- Karala A, Lappi A, Saaranen MJ, Ruddock LW. 2009. Efficient peroxide-mediated oxidative refolding of a protein at physiological pH and implications for oxidative folding in the endoplasmic reticulum. *Antioxid Redox Signal* 11:963–70.
- Khannoon ER, El-Gendy A, Hardege JD. 2011. Scent marking pheromones in lizards: cholesterol and long chain alcohols elicit avoidance and aggression in male *Acanthodactylus boskianus* (Squamata: lacertidae). *Chemoecology* 21: 143–9.
- Khannoon ER, Lunt DH, Schulz S, Hardege JD. 2013. Divergence of scent pheromones in allopatric populations of *Acanthodactylus boskianus* (Squamata: lacertidae). *Zoolog Sci* 30: 380–5.
- Kim S, Gupta N, Pevzner PA. 2008. Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. *J Proteome Res* 7:3354–63.
- Kim S, Pevzner PA. 2014. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun* 5:5277.
- Labra A. 2006. Chemoreception and the assessment of fighting abilities in the lizard *Liolaemus monticola*. *Ethol* 112: 993–9.
- Lande R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* 33:402.
- Lande R, Arnold SJ. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210.
- Lei H, Vickers N. 2008. Central processing of natural odor mixtures in insects. *J Chem Ecol* 34:915–27.
- López P, Martín J. 2005. Female Iberian wall lizards prefer male scents that signal a better cell-mediated immune response. *Biol Lett* 1:404–6.
- López P, Martín J. 2006. Lipids in the femoral gland secretions of male Schreiber’s green lizards, *Lacerta schreiberi*. *Z Naturforsch, C, J Biosci* 61:763–8.
- Mangiacotti M, Baeckens S, Scali S, Martín J, Van Damme R, Sacchi R. 2021. Evolutionary and biogeographical support for species-specific proteins in lizard chemical signals. *Biol J Linn Soc* 134:912–28.
- Mangiacotti M, Fumagalli M, Cagnone M, Viglio S, Bardoni AM, Scali S, Sacchi R. 2019. Morph-specific protein patterns in the femoral gland secretions of a colour polymorphic lizard. *Sci Rep* 9:8412.
- Mangiacotti M, Fumagalli M, Scali S, Zuffi MAL, Cagnone M, Salvini R, Sacchi R. 2017. Inter- and intra-population variability of the protein content of femoral gland secretions from a lacertid lizard. *Curr Zool* 63:657–65.
- Mangiacotti M, Gaggiani S, Coladonato AJ, Scali S, Zuffi MAL, Sacchi R. 2019a. First experimental evidence that proteins from femoral glands convey identity-related information in a lizard. *Acta Ethol* 22:57–65.
- Mangiacotti M, Martín J, López P, Reyes-Olivares CV, Rodríguez-Ruiz G, Coladonato AJ, Scali S, Zuffi MAL, Sacchi R. 2020. Proteins from femoral gland secretions of male rock lizards *Iberolacerta cyreni* allow self—but not individual—recognition of unfamiliar males. *Behav Ecol Sociobiol* 74:68.
- Mangiacotti M, Pezzi S, Fumagalli M, Coladonato AJ, d’Ettorre P, Leroy C, Bonnet X, Zuffi MAL, Scali S, Sacchi R. 2019b. Seasonal variations in femoral gland secretions reveals some unexpected correlations between protein and lipid components in a lacertid lizard. *J Chem Ecol* 45:673–83.
- Martín J, Castilla AM, López P, Al-Jaidah M, Al-Mohannadi SF, Al-Hemaidi AAM. 2016. Chemical signals in desert lizards: are femoral gland secretions of male and female spiny-tailed lizards, *Uromastix aegyptia microlepis* adapted to arid conditions? *J Arid Environ* 127:192–8.
- Martín J, López P. 2000. Chemoreception, symmetry and mate choice in lizards. *Proc R Soc B Biol Sci* 267:1265–9.
- Martín J, López P. 2011. Pheromones and reproduction in reptiles. In: KH Lopez, DO Norris, editors. *Hormones and reproduction of vertebrates*. London: Academic Press. p. 141–67.
- Martín J, López P. 2014. Pheromones and chemical communication in lizards. In: JL Rheubert, DS Siegel, SE Trauth, editors. *Reproductive biology and phylogeny of lizards and tuatara*. New York, NY: Taylor and Francis Group USA. p. 54–88.
- Martín J, López P. 2015. Condition-dependent chemosignals in reproductive behavior of lizards. *Horm Behav* 68:14–24.
- Martín J, Martínez-Silvestre A, López P, Ibáñez A, Rodríguez-Domínguez MÁ, Verdaguer I. 2015a. Lipophilic compounds in femoral secretions of males and females of the El Hierro giant lizard *Gallotia simonyi* (Lacertidae). *Biochem Syst Ecol* 61:286–92.
- Martín J, Ortega J, López P. 2013. Lipophilic compounds in femoral secretions of male collared lizards, *Crotaphytus bicinctores* (Iguania, Crotaphytidae). *Biochem Syst Ecol* 47:5–10.
- Martín J, Ortega J, López P. 2015b. Interpopulational variations in sexual chemical signals of Iberian wall lizards may allow maximizing signal efficiency under different climatic conditions. *PLoS One* 10:e0131492.
- Martín J, Zamora-Camacho FJ, Reguera S, López P, Moreno-Rueda G. 2017. Variations in chemical sexual signals of *Psammotromus algerus* lizards along an elevation gradient may reflect altitudinal variation in microclimatic conditions. *Sci Nat* 104:16.
- Mayerl C, Baeckens S, Van Damme R. 2015. Evolution and role of the follicular epidermal gland system in non-ophidian squamates. *Amphibia-Reptilia* 36:185–206.
- Mucignat-Caretta C, Caretta A. 2014. Message in a bottle: major urinary proteins and their multiple roles in mouse intraspecific chemical communication. *Anim Behav* 97:255–63.
- Murren CJ. 2012. The integrated phenotype. *Integr Comp Biol* 52:64–76.
- Ali Khan H, Mutus B. 2014. Protein disulfide isomerase a multifunctional protein with multiple physiological roles. *Front Chem* 2:70.
- Nehring V, Wyatt TD, D’Ettorre P. 2013. Noise in chemical communication. In: H Brumm, editor. *Animal communication and noise*. Berlin: Springer. p. 373–405.

- Nesvizhskii AI. 2014. Proteogenomics: concepts, applications and computational strategies. *Nat Methods* 11: 1114–25.
- Ochi S. 2019. scorepeak: peak functions for peak detection in univariate time series. R package version 0.1.2. <https://CRAN.R-project.org/package=scorepeak>.
- Ortega J, Martín J, Crochet P-A, López P, Clobert J. 2019. Seasonal and interpopulational phenotypic variation in morphology and sexual signals of *Podarcis liolepis* lizards. *PLoS One* 14:e0211686.
- Ozaki M, Wada-Katsumata A, Fujikawa K, Iwasaki M, Yokohari F, Satoji Y, Nisimura T, Yamaoka R. 2005. Behavior: ant nestmate and non-nestmate discrimination by a chemosensory sensillum. *Science* 309:311–4.
- Paradis E, Schliep K. 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–8.
- Pruett JA, Zúñiga-Vega JJ, Campos SM, Soini HA, Novotny MV, Vital-García C, Martins EP, Hews DK. 2016. Evolutionary interactions between visual and chemical signals: chemosignals compensate for the loss of a visual signal in male *Sceloporus* lizards. *J Chem Ecol* 42:1164–74.
- R Core Team. 2018. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Roll U, Feldman A, Novosolov M, Allison A, Bauer AM, Bernard R, Böhm M, Castro-Herrera F, Chirio L, Collen B et al. 2017. The global distribution of tetrapods reveals a need for targeted reptile conservation. *Nat Ecol Evol* 1:1677–82.
- Romero-Diaz C, Campos SM, Herrmann MA, Lewis KN, Williams DR, Soini HA, Novotny MV, Hews DK, Martins EP. 2020. Structural identification, synthesis and biological activity of two volatile cyclic dipeptides in a terrestrial vertebrate. *Sci Rep* 10:4303.
- Romero-Diaz C, Campos SM, Herrmann MA, Soini HA, Novotny MV, Hews DK, Martins EP. 2021. Composition and compound proportions affect the response to complex chemical signals in a spiny lizard. *Behav Ecol Sociobiol* 75: 42.
- Ruiz-Monachesi MR, Cruz FB, Valdecantos S, Labra A. 2020. Unravelling associations among chemosensory system components in *Liolaemus* lizards. *J Zool* 312:148–57.
- Schwenk K. 1995. Of tongues and noses: chemoreception in lizards and snakes. *Trends Ecol Evol* 10:7–12.
- Sindaco R, Jeremcenko VK. 2008. The reptiles of the western palearctic: annotated checklist and distributional atlas of the turtles, crocodiles, amphisbaenians and lizards of Europe, North Africa, Middle East and Central Asia. Latina (Italy): Edizioni Belvedere.
- Supuran CT. 2016. Structure and function of carbonic anhydrases. *Biochem J* 473:2023–32.
- Tapia MS, Alzamora SM, Chirife J. 2020. Effects of water activity (a_w) on microbial stability as a hurdle in food preservation. In: *Water activity in foods*. Chicago: John Wiley & Sons, Ltd. p. 323–55.
- Tellkamp F, Lang F, Ibáñez A, Abraham L, Quezada G, Günther S, Looso M, Tann FJ, Müller D, Cemic F et al. 2020. Proteomics of Galápagos marine iguanas links function of femoral gland Proteins to the immune system. *Mol Cell Proteomics* 19:1523–32.
- Tripp BC, Smith K, Ferry JG. 2001. Carbonic anhydrase: new insights for an ancient enzyme. *J Biol Chem* 276:48615–8.
- van den Boogaart KG, Tolosana-Delgado R. 2013. Analyzing compositional data with R Berlin: Springer.
- Wang Y-H, Narayan M. 2008. pH Dependence of the isomerase activity of protein disulfide isomerase: insights into its functional relevance. *Protein J* 27:181–5.
- Weldon PJ, Flachsbarth B, Schulz S. 2008. Natural products from the integument of nonavian reptiles. *Nat Prod Rep* 25: 738–56.
- Wyatt TD. 2010. Pheromones and signature mixtures: defining species-wide signals and variable cues for identity in both invertebrates and vertebrates. *J. Comp. Physiol. A* 196:685–700.
- Wyatt TD. 2014. Proteins and peptides as pheromone signals and chemical signatures. *Anim Behav* 97:273–80.
- Zozaya SM, Higgie M, Moritz C, Hoskin CJ. 2019. Are pheromones key to unlocking cryptic lizard diversity? *Am Nat* 194:168–82.

M1. Note on samples collection

Although the specifics on the sampling collection have been reported in earlier studies (García-Roa et al. 2017; Baeckens et al. 2018a), we briefly summarise here the basic field protocol. Sampling was carried out between 2002 and 2016 (Table S1), during seasonal periods where glandular activity reaches its maximum (i.e., the breeding season; Cole 1966; Alberts et al. 1992; Mangiacotti, Pezzi, et al. 2019). Lizards were captured by hand or noose, and femoral glands secretions were obtained by gently pressing along the thigh with the help of a steel spatula, a non-invasive technique which causes minimal stress to the animals. No lizards were killed or injured during the study. Once transferred into individual vials, the secretions were kept at -20 °C until analysis. All lizards were released, healthy, at their capture point. The experimental design and procedure complied the ARRIVE guidelines (<https://arriveguidelines.org/>), and was in accordance with University of Antwerp (Belgium) animal welfare standards and protocols (ECD 2014-32). Permits or licences for each state are listed from (Baeckens et al. 2018a): Croatia (UP/I-612-07/14-48/111 & UP/I-612-07/14-48/33), The Netherlands (FF/74A/2015/009), Israel (2014/40323), SA Free State Province (S54C-515022511060), SA Eastern Cape Province (CRO 45/15CR & 46/15CR), SA Western Cape Province (0056-AAA041-00093), SA Northern Cape Province (FAUNA 229/2015 & 230/2015); SA Limpopo Province (0092-MKT001-00004); Spain (Captures of lizards and sampling procedures were performed under different licenses for the Environmental Agencies of the different Regional Governments of Spain where lizards were studied); Greece (all Greek species were collected in accordance with the Hellenic National Legislation, Presidential Decree 67/81).

M2. Mass Spectrometry and protein identification protocol

Sample preparation. From the Sodium dodecyl sulphate-polyacrylamide gel (Fig. S1), we carefully excised the bands occupying the molecular weight regions of interest, and put them in distinct tubes (one for each region). We prepared the samples for mass spectrometry analysis as follows: (1) we completely de-stained the gel pieces by washing twice with 200 µL of 100 mM ammonium bicarbonate buffer pH 7.8, 50 % acetonitrile (ACN) and keeping under stirring overnight; (2) we dehydrated the samples with 100 µL ACN; (3) we performed reduction of the samples by adding 100 µL of 10 mM Dithiothreitol (DTT) solution (30 min at 37°C), replaced with 100 µL of 55 mM iodoacetamide for 45 min at 60°C; (4) we washed the gel pieces twice with 200 µL of 100 mM ammonium bicarbonate for 10 min; (5) we dehydrated by addition of 100 µL of ACN; (6) we removed ACN and dried samples under vacuum; (7) we performed protein digestion overnight at 37°C by adding 100 µL of 100 mM ammonium bicarbonate buffer pH 7.8, containing 20 ng/µL sequencing grade trypsin (Promega, Madison, WI, USA); (8) we extracted the obtained peptides by adding twice 100 µL of 50 % ACN in water, 5 % formic acid (FA), and pooling the supernatant from each step after stirring and centrifuging; (9) we dried each pool and stored at -20°C until mass spectrometry analysis.

Mass spectrometry. Before analysis, peptide mixtures were first solubilized in 100µL of 0.1% formic acid (FA). All analyses were carried out with a LC unit (ExionLCTM AD) equipped with a column oven thermostated at 40°C, an autosampler cooled at 10°C and a binary gradient pump system. MS instrument consists of a high resolution QTOF mass spectrometer (AB Sciex X500B) equipped with a Turbo V Ion source and a Twin Sprayer ESI (electrospray ionization) probe, controlled by SCIEX OS 2.1 software. Peptides were separated by reverse phase (RP) HPLC on a Hypersil Gold (Thermo Fisher Scientific, USA) C18 column (150 × 2.1 mm, 3 µm particle size, 175 Å pore size) using a linear gradient (2-50 % solvent B in 15 min) in which solvent A consisted of 0.1 % aqueous FA and solvent B of acetonitrile (ACN) containing 0.1 % FA. Flow rate was 0.2 mL/min. Mass spectra were generated in positive polarity under constant instrumental conditions: ion spray voltage 4500 V, declustering potential 100 V, curtain gas 30 psi, ion source gas 1 40 psi, ion source gas 2 45 psi, temperature 350 °C, collision energy 10 V. Spectra were acquired with SCIEX OS 2.1 software.

Database searching. We used MS-GF+ v2022.01.17 (Kim et al. 2008; Kim and Pevzner 2014) for protein identification by peptide-spectrum matching (Eng et al. 2011; Nesvizhskii 2014). We set the algorithm as follows (Tellkamp et al. 2020): tolerance, 30 ppm; charge range, 1 – 6+; range of peptide length, 7 – 70; isotope error 0 – 1 Da; cleavage, tryptic; post translational modification, fix carbamidomethylation of cysteine (Creasy and Cottrell 2004; Eng et al. 2011; Schittmayer et al. 2016). Searching was performed against UniProt *Podarcis muralis* reference proteome (UP000472272; Bateman et al. 2021), consisting in 36,445 protein sequences obtained from the *P. muralis* genome (Andrade et al. 2019), to which pig trypsin and human keratin sequences (retrieved from uniprot.org) were added to account for contamination. To maximize identification power, a two-stages, target-decoy approach was run (Craig and Beavis 2003; Jeong et al. 2012; Mangiacotti, Fumagalli, et al. 2019). An unbiased false detection rate (FDR) was applied at peptide level (Everett et al. 2010; Bern and Kil 2011), using spectrum E-value as the reference score (Jeong et al. 2012), and peptides with FDR > 0.01 were filtered out from the final list. Protein identification was achieved when at least two different peptides match the same database entry (Mangiacotti, Fumagalli, et al. 2019). Call to MSGF+ and all post-processing analyses were done in R v3.5.2 (R Core Team 2018), using the packages `mzID` (Pedersen et al. 2016), `Biostrings` (Pagès et al. 2017), `stringr` (Wickham 2018).

Species	Latitude	Longitude	N	Month	Year	R _P	H _P	R _L	H _L
<i>Acanthodactylus beershebensis</i>	31.07	34.84	4	April	2014	8	4.91	56	1.02
<i>Acanthodactylus boskianus</i>	30.95	34.59	7	April	2014	8	5.19	38	0.95
<i>Acanthodactylus ophiodurus</i>	29.94	35.08	4	June	2014	12	5.18	29	0.93
<i>Acanthodactylus schreiberi</i>	32.00	34.79	5	April	2014	9	5.22	37	1.31
<i>Acanthodactylus scutellatus</i>	32.00	34.79	5	April	2014	8	4.96	39	1.15
<i>Algyroides moreoticus</i>	37.25	21.01	4	May	2014	9	5.15	30	0.66
<i>Algyroides nigropunctatus</i>	45.10	14.34	5	May	2013	11	4.99	67	1.00
<i>Dalmatolacerta oxycephala</i>	44.69	14.39	11	May	2013	6	5.03	29	0.20
<i>Gallotia galloti</i>	28.08	-16.62	9	April	2013	10	5.05	103	1.38
<i>Gallotia simonyi</i>	27.76	-17.99	22	June	2014	6	4.93	57	1.35
<i>Gallotia stehlini</i>	27.74	-15.60	5	April	2013	5	4.74	32	0.62
<i>Holaspis guentheri</i>	-3.81	37.63	17	August	2014	10	5.08	45	1.24
<i>Iberolacerta bonnali</i>	42.61	0.18	7	June	2013	8	5.07	42	0.76
<i>Iberolacerta cyreni</i>	40.79	-4.01	15	May	2003	5	4.55	44	0.57
<i>Iberolacerta galani</i>	42.41	-6.41	10	May	2010	9	4.97	66	0.95
<i>Iberolacerta monticola</i>	40.32	-7.60	19	May	2006	9	4.87	49	0.48
<i>Lacerta bilineata</i>	42.73	-3.41	7	May	2013	8	5.02	42	1.01
<i>Lacerta media</i>	33.21	35.77	6	April	2014	10	5.12	71	0.91
<i>Lacerta schreiberi</i>	40.78	-4.07	15	April	2005	10	5.09	51	1.06
<i>Lacerta viridis</i>	47.35	19.76	22	April	2008	7	5.00	40	1.12
<i>Mesalina guttulata</i>	30.83	34.75	5	April	2014	4	4.85	23	1.28
<i>Mesalina olivieri</i>	31.04	35.07	6	April	2014	3	4.83	28	1.15
<i>Phoenicolacerta kulzeri</i>	33.30	35.79	5	April	2014	5	4.88	30	0.55
<i>Podarcis bocagei</i>	42.74	-9.08	9	May	2007	6	5.15	56	0.82
<i>Podarcis carbonelli</i>	40.50	-6.09	6	May	2007	8	5.16	60	0.79
<i>Podarcis erhardii</i>	37.83	24.85	7	May	2014	11	5.04	40	1.02
<i>Podarcis gaigeae</i>	38.95	24.52	6	May	2014	10	5.15	72	1.80
<i>Podarcis guadarramae</i>	40.73	-4.04	21	April	2002	6	4.89	62	0.64
<i>Podarcis liolepis</i>	42.75	1.84	12	May	2012	7	4.90	59	0.92
<i>Podarcis melisellensis</i>	45.10	14.34	21	May	2013	12	5.29	54	1.07
<i>Podarcis milensis</i>	36.69	24.44	7	May	2014	10	5.11	74	1.58
<i>Podarcis muralis</i>	45.24	9.23	23	May	2011	8	4.98	64	1.71
<i>Podarcis peloponnesiacus</i>	37.24	22.46	5	May	2014	11	5.27	66	1.68
<i>Podarcis vaucheri</i>	35.18	-2.43	8	April	2012	10	5.23	84	1.40
<i>Psammodromus algirus</i>	40.73	-4.37	22	May	2004	8	4.74	58	1.16
<i>Zootoca vivipara</i>	44.41	3.74	35	April	2005	7	4.62	18	0.41

Table S1. List of the species included in the study ordered by their scientific name with information about the sampling locality, size (N), and period (month and year), and the chemical complexity of proteinaceous (R_P and H_P) and lipid (R_L and H_L) fractions of the femoral gland secretions: R_P = number of bands in the electrophoretic profile; H_P = Shannon entropy of the electrophoretic profile; R_L = total number of distinct lipid compounds; H_L = Shannon entropy of the lipid profile. Sampling data and R_L were taken from Baeckens et al. (2018b), R_P, H_P, and H_L were computed starting from available data. See methods for details.

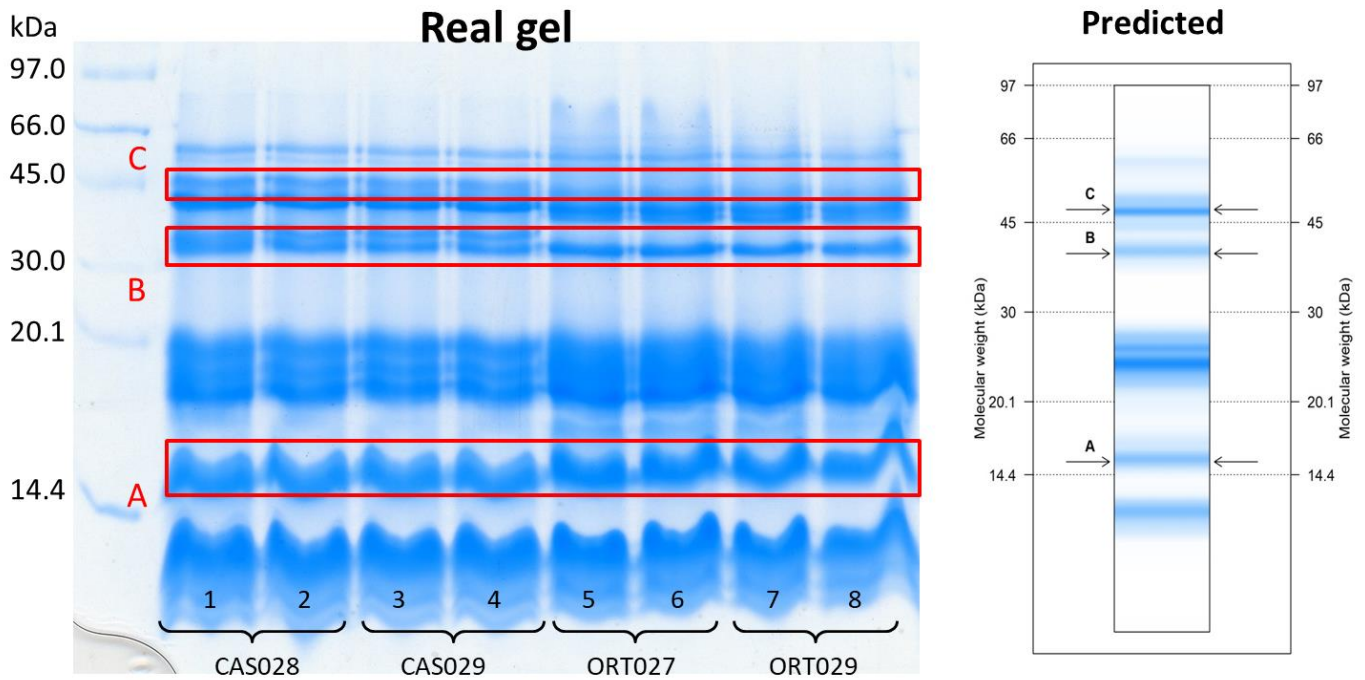


Figure S1. *On the left.* SDS-PAGE gel of the *Podarcis muralis* samples used to retrieve the bands correlated with pro-vitamin D₃ abundance as predicted by pGLS models. The gel pieces from the same molecular weight region were pooled before mass spectrometry analysis. Ladder lanes are marked with the corresponding molecular weights. We duplicated each of the four samples (ID code at the bottom of right panel) in two adjacent lanes to increase protein concentration in each region. *On the right.* Virtual lane reconstruction of the *Podarcis muralis* gel with arrows and letters indicating the predicted position of the bands responding to increase pro-vitamin D₃; it was used to drive bands identification in the real gel.

REFERENCES

- Alberts AC, Pratt NC, Phillips JA. 1992. Seasonal productivity of lizard femoral glands: Relationship to social dominance and androgen levels. *Physiology & Behavior* 51:729–33.
- Andrade P, Pinho C, Pérez i de Lanuza G, Afonso S, Brejcha J, Rubin C-J, Wallerman O, Pereira P, Sabatino SJ, Bellati A, Pellitteri-Rosa D, Bosakova Z, Bunikis I, Carretero MA, Feiner N, Marsik P, Paupério F, Salvi D, Soler L, While GM, Uller T, Font E, Andersson L, Carneiro M. 2019. Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard. *Proceedings of the National Academy of Sciences* 116:5633–42.
- Baeckens S, Martín J, García-Roa R, Pafilis P, Huyghe K, Van Damme R. 2018a. Environmental conditions shape the chemical signal design of lizards. *Functional Ecology* 32:566–80.
- Baeckens S, Martín J, García-Roa R, Pafilis P, Huyghe K, Van Damme R. 2018b. Environmental conditions shape the chemical signal design of lizards. *Functional Ecology* 32:566–80.
- Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bursteinas B, Bye-A-Jee H, Coetzee R, Cukura A, Silva A Da, Denny P, Dogan T, Ebenezer TG, Fan J, Castro LG, Garmiri P, Georghiou G, Gonzales L, Hatton-Ellis E, Hussein A, Ignatchenko A, Insana G, Ishtiaq R, Jokinen P, Joshi V, Jyothi D, Lock A, Lopez R, Luciani A, Luo J, Lussi Y, MacDougall A, Madeira F, Mahmoudy M, Menchi M, Mishra A, Moulang K, Nightingale A, Oliveira CS, Pundir S, Qi G, Raj S, Rice D, Lopez MR, Saidi R, Sampson J, Sawford T, Speretta E, Turner E, Tyagi N, Vasudev P, Volynkin V, Warner K, Watkins X, Zaru R, Zellner H, Bridge A, Poux S, Redaschi N, Aimo L, Argoud-Puy G, Auchincloss A, Axelsen K, Bansal P, Baratin D, Blatter MC, Bolleman J, Boutet E, Breuza L, Casals-Casas C, de Castro E, Echioukh KC, Coudert E, Cucho B, Doche M, Dornevil D, Estreicher A, Famiglietti ML, Feuermann M, Gasteiger E, Gehant S, Gerritsen V, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, Hyka-Nouspikel N, Jungo F, Keller G, Kerhornou A, Lara V, Le Mercier P, Lieberherr D, Lombardot T, Martin X, Masson P, Morgat A, Neto TB, Paesano S, Pedruzzi I, Pilbout S, Pourcel L, Pozzato M, Pruess M, Rivoire C, Sigrist C, Sonesson K, Stutz A, Sundaram S, Tognolli M, Verbregue L, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Garavelli JS, Huang H, Laiho K, McGarvey P, Natale DA, Ross K, Vinayaka CR, Wang Q, Wang Y, Yeh LS, Zhang J. 2021. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research* 49:D480–89.

- Bern M, Kil YJ. 2011. Comment on “unbiased statistical analysis for multi-stage proteomic search strategies.” *Journal of Proteome Research* 10:2123–27.
- Cole CJ. 1966. Femoral glands in lizards: a review. *Herpetologica* 22:199–206.
- Craig R, Beavis RC. 2003. A method for reducing the time required to match protein sequences with tandem mass spectra. *Rapid Communications in Mass Spectrometry* 17:2310–16.
- Creasy DM, Cottrell JS. 2004. Unimod: Protein modifications for mass spectrometry. *Proteomics* 4:1534–36.
- Eng JK, Searle BC, Clauser KR, Tabb DL. 2011. A Face in the Crowd: Recognizing Peptides Through Database Search. *Molecular & Cellular Proteomics* 10:R111.009522.
- Everett LJ, Bierl C, Master SR. 2010. Unbiased Statistical Analysis for Multi-Stage Proteomic Search Strategies. *Journal of Proteome Research* 9:700–707.
- García-Roa R, Jara M, López P, Martín J, Pincheira-Donoso D. 2017. Heterogeneous tempo and mode of evolutionary diversification of compounds in lizard chemical signals. *Ecology and Evolution* 7:1286–96.
- Jeong K, Kim S, Bandeira N. 2012. False discovery rates in spectral identification. *BMC bioinformatics* 13 Suppl 1:S2–S2.
- Kim S, Gupta N, Pevzner PA. 2008. Spectral probabilities and generating functions of tandem mass spectra: A strike against decoy databases. *Journal of Proteome Research* 7:3354–63.
- Kim S, Pevzner PA. 2014. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nature Communications* 5:5277.
- Mangiacotti M, Fumagalli M, Cagnone M, Viglio S, Bardoni AM, Scali S, Sacchi R. 2019. Morph-specific protein patterns in the femoral gland secretions of a colour polymorphic lizard. *Scientific Reports* 9:8412.
- Mangiacotti M, Pezzi S, Fumagalli M, Coladonato AJ, d’Ettorre P, Leroy C, Bonnet X, Zuffi MAL, Scali S, Sacchi R. 2019. Seasonal Variations in Femoral Gland Secretions Reveals some Unexpected Correlations Between Protein and Lipid Components in a Lacertid Lizard. *Journal of Chemical Ecology* 45:673–83.
- Nesvizhskii AI. 2014. Proteogenomics: concepts, applications and computational strategies. *Nat Methods* 11:1114–25.
- Pagès H, Aboyoun P, Gentleman R, DebRoy S. 2017. Biostrings: Efficient manipulation of biological strings. R package version 2460.
- Pedersen TL, with contributions from Laurent Gatto VAP, Gibb. S. 2016. mzID: An mzIdentML parser for R. .
- R Core Team. 2018. R: a language and environment for statistical computing. .
- Schittmayer M, Fritz K, Liesinger L, Griss J, Birner-Gruenberger R. 2016. Cleaning out the Litterbox of Proteomic Scientists Favorite Pet: Optimized Data Analysis Avoiding Trypsin Artifacts. *Journal of Proteome Research* 15:1222–29.
- Tellkamp F, Lang F, Ibáñez A, Abraham L, Quezada G, Günther S, Looso M, Tann FJ, Müller D, Cemic F, Hemberger J, Steinfartz S, Krüger M. 2020. Proteomics of Galápagos Marine Iguanas Links Function of Femoral Gland Proteins to the Immune System. *Molecular and Cellular Proteomics* 19:1523–32.
- Wickham H. 2018. stringr: Simple, Consistent Wrappers for Common String Operations. .