Current Biology

Exaptation as a Mechanism for Functional Reinforcement of an Animal Pheromone System

Highlights

- Persuasin was recruited to an already functional courtship pheromone system in newts

- Birth of persuasin involved a gene duplication and neofunctionalisation

- Molecular remodeling and a shift in expression assisted in its functional transition

- Recruitment of persuasin ultimately reinforced the existing pheromone blend

Authors

Margo Maex, Dag Treer,
Henri De Greve, Paul Proost,
Ines Van Bocxlaer, Franky Bossuyt

Correspondence

franky.bossuyt@vub.be

In Brief

Little is known about how additional molecules are recruited to functional pheromone systems. Maex et al. show how a new salamander courtship pheromone, persuasin, was co-opted alongside an ancient pheromone after a gene duplication and shift in expression. Subsequent evolution led the molecule along a pathway that reinforced the existing pheromone.

CellPress

Current Biology

Report

CellPress

Exaptation as a Mechanism for Functional Reinforcement of an Animal Pheromone System

Margo Maex,[1,5] Dag Treer,[1,5] Henri De Greve,[2,3] Paul Proost,[4] Ines Van Bocxlaer,[1] and Franky Bossuyt[1,6,*]

[1]Amphibian Evolution Lab, Biology Department, Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussels, Belgium
[2]Structural Biology Brussels, Department of Bioengineering Sciences, Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussels, Belgium
[3]VIB-VUB Center for Structural Biology, Vlaams Instituut voor Biotechnologie, Pleinlaan 2, 1050 Brussels, Belgium
[4]Laboratory of Molecular Immunology, Department of Microbiology and Immunology, Rega Institute, KU Leuven, Herestraat 49, 3000 Leuven, Belgium
[5]These authors contributed equally
[6]Lead Contact
*Correspondence: franky.bossuyt@vub.be
https://doi.org/10.1016/j.cub.2018.06.074

SUMMARY

Animal sex pheromone systems often exist as multicomponent signals [1–11] to which chemical cues have been added over evolutionary time. Little is known on why and how additional molecules become recruited and conserved in an already functional pheromone system. Here, we investigated the evolutionary trajectory of a series of 15 kDa proteins—termed persuasins—that were co-opted more recently alongside the ancient sodefrin precursor-like factor (SPF) courtship pheromone system in salamanders [9, 12]. Expression, genomic, and molecular phylogenetic analyses show that persuasins originated from a gene that is expressed as a multidomain protein in internal organs where it has no pheromone function but underwent gene duplication and neofunctionalisation. The subsequent evolution combined domain loss and the introduction of a proteolytic cleavage site in the duplicated gene to give rise to two-domain cysteine rich proteins with structural properties similar to SPF pheromones [12]. An expression shift to the pheromone-producing glands, where expression of persuasins was immediately spatiotemporally synchronized with the already available pheromone system, completed the birth of a new pheromone. Electrostatic forces between members of both protein families likely enhance colocalization and simultaneous activation of different female olfactory neurons, explaining why persuasins immediately had a selective advantage. In line with this, behavioral assays show that persuasins increase female receptivity on their own but also exert a cumulative or synergistic effect in combination with SPF, clearly reinforcing the pheromone system as a whole. Our study reveals molecular remodeling of an existing protein architecture as an evolutionary mechanism for functional reinforcement of animal pheromone systems.

Q1

RESULTS AND DISCUSSION

The oldest known vertebrate sex pheromone system is the "sodefrin precursor-like factor" (SPF) pheromone system, a series of proteins involved in the courtship process of many salamander families (Plethodontidae, Ambystomatidae and Salamandridae) [9, 12–17]. In aquatically reproducing newts (Salamandridae), males display an underwater courtship dance in which they direct a stream of these water-soluble protein pheromones from their cloaca toward the female [18, 19]. These courtship pheromones are produced in sexually dimorphic cloacal glands ("dorsal glands") that hypertrophy during, but almost completely regress outside, the mating season [20, 21]. As a consequence, secretory proteins that over the course of evolution become expressed in these seasonally enlarged glands are spatiotemporally co-expressed with SPF pheromones [12] and form ideal candidates to gain a pheromone function alongside the already functional pheromone system.

We combined transcriptome expression analyses with proteomic analyses (see STAR Methods for details) in two aquatically reproducing newt species—the palmate newt (*Lissotriton helveticus*) and the alpine newt (*Ichthyosaura alpestris*)—to characterize and purify a new family of protein pheromones that is abundantly secreted during courtship. We first investigated the full spectrum of courtship-specific proteins by comparing the protein content of water in which courting couples were held (hereafter designated as "courtship water") to water of non-courting couples. Reversed phase high-performance liquid chromatography (RP-HPLC) analyses demonstrated a specific peak in the courtship elution profile of both species (*L. helveticus*, Figure 1A, black; *I. alpestris*, Figure S1A), which is absent in non-courting male-female water (Figure 1A, gray). Using gel electrophoresis, mass spectrometry (MS), and N-terminal amino acid (aa) sequencing, we identified 15–16 kDa proteins in this dominant courtship-specific elution peak (Figure 1B and Table S1) that co-elute with previously identified 20 kDa beta-SPF proteins. Courtship water sampled at the end of the breeding season resulted in a peak with a much lower absorbance, illustrating that the expression and secretion of both SPF and the newly identified proteins are restricted to the breeding season (not shown). Whole-transcriptome sequencing (RNA-seq) on the pheromone-producing dorsal glands of
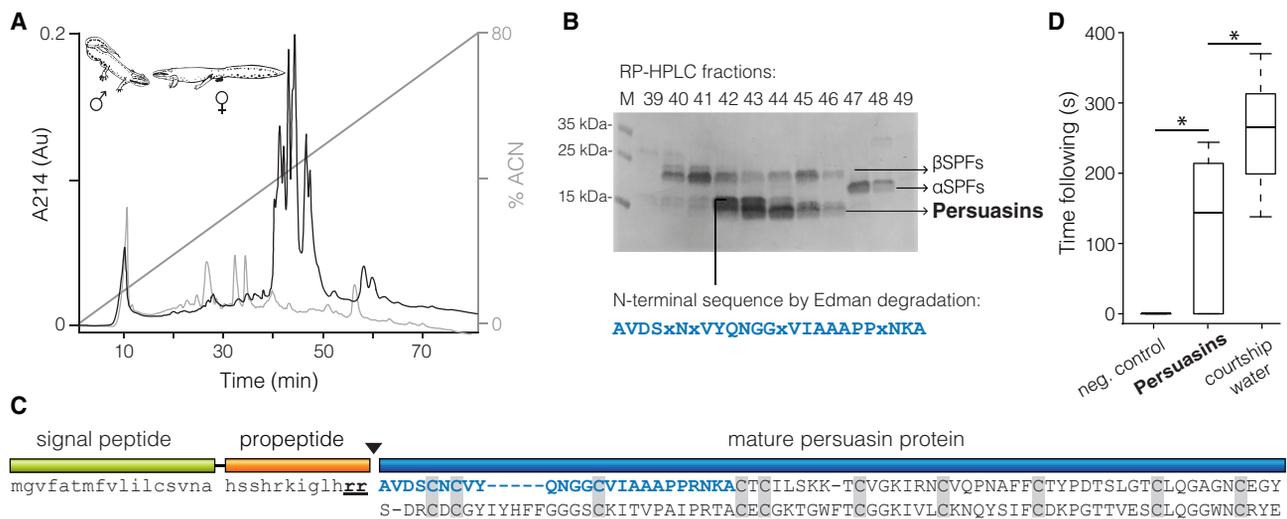
Figure 1. Identification and Characterization of Persuasins in the Palmate Newt, *Lissotriton helveticus*

(A) Reversed phase-HPLC elution profiles obtained from water in which couples were allowed to court (black) and water in which non-courting palmate newt males and females (*Lissotriton helveticus*) were held (gray). Courtship water yields a major elution peak (fr39–49) that is absent in male-female water, indicating the release of courtship-specific proteins.

(B) A Coomassie stained blot of the courtship water peak fractions reveals the presence of previously unknown 15 kDa proteins, termed persuasins, as well as known 20 kDa alpha and beta sodefrin precursors-like factor (SPF) protein pheromones. A persuasin N-terminal sequence obtained by Edman degradation is shown (Xs depict residues for which no signal was detected and most often represent cysteines, as cysteines were not alkylated). See also Figure S1A and Tables S1 and S2.

(C) Schematic representation of a persuasin precursor with a signal peptide (light green) and a propeptide (orange) that is cleaved at the dibasic Arg-Arg motif to yield a mature persuasin protein of about 145 amino acids (blue). The conserved cysteine residues in the mature protein are shaded, showing two tandem copies of the same cysteine pattern within a single persuasin.

(D) Non-parametric boxplots summarizing the median cumulative duration of female following behavior as a response on purified persuasin proteins compared to 10 mM tris pH 7 (negative control) and courtship water (positive control) (Kruskal-Wallis, p < 0.0001, X^2 = 21.03, n = 10; significance levels shown on the boxplots represent the corrected p values provided by the post hoc Dunn's test (* = p < 0.05).

See also Figures S1B and S1C.

sexually mature palmate and alpine newt males identified precursor transcripts, with an N-terminal signal peptide of 17 aas and a 12 aa propeptide from which the 15 kDa proteins are cleaved (GenBank: MG435350–MG435355; MH500757–MH500766). Two recurrent 10-cysteine motifs define the newly identified proteins in their mature cleaved form (Figure 1C). Furthermore, gene expression levels in both species confirmed that, next to SPF, these proteins belong to the top expressed transcripts (Table S2), confirming that the dorsal glands are their main source of synthesis and production. NCBI-BLAST searches of the newly identified sequences against the non-redundant nucleotide and protein databases did not return any significant hit in vertebrates, suggesting no substantial homology with any known vertebrate protein family. We therefore termed the proteins "persuasins," referring to their receptivity enhancing role in females (see below).

To assay the courtship pheromone function of persuasins, we first used immobilized lectins to selectively remove glycosylated SPF proteins from unglycosylated persuasins (Table S1) and obtained fractions with purified persuasins after a final, polishing RP-HPLC step (Figures S1B and S1C; see STAR Methods for details). We then adopted the same two-female test with palmate newts as previously employed for SPF to investigate the courtship pheromone function of persuasins [9]. When two receptive females (in absence of a male) are exposed to male courtship pheromones, they start showing courtship responses toward

each other (including closely following the movements of the other female) that are similar to how females follow a male during courtship before spermatopohore deposition [19]. Post hoc analyses of our two-female behavioral tests show that following behavior occurred significantly more in females that were exposed to persuasin than in the negative control solution (Kruskal-Wallis test followed by post hoc Dunn's multiple comparison test with holm correction; p = 0.0149, z = 2.17; Figure 1D). This renders persuasin a typical urodelan courtship pheromone that is privately delivered to increase receptivity in females during courtship.

To pinpoint the origin of persuasins as pheromones during salamander evolution, we profiled SPF and persuasin expression patterns relative to housekeeping gene *eukaryotic translation elongation factor 1 alpha 1* (*eef1a1*) in male pheromone-producing glands from four additional salamander species that diverged before the *Lissotriton-Ichthyosaura* split. Transcript expression analyses indicated that persuasins are not expressed in the pheromone-producing glands in any of these species, while SPFs showed the typical elevated expression pattern (Figure 2A). These findings indicate that persuasins enriched the already existing SPF courtship pheromone blend in the common ancestor of *Lissotriton* and *Ichthyosaura*, about 51 to 45 million years ago (mya) (Figure 2A; following the phylogeny of [22]). Next, we investigated the evolutionary origin of persuasins by further screening all publicly available urodelan transcriptomes and the available
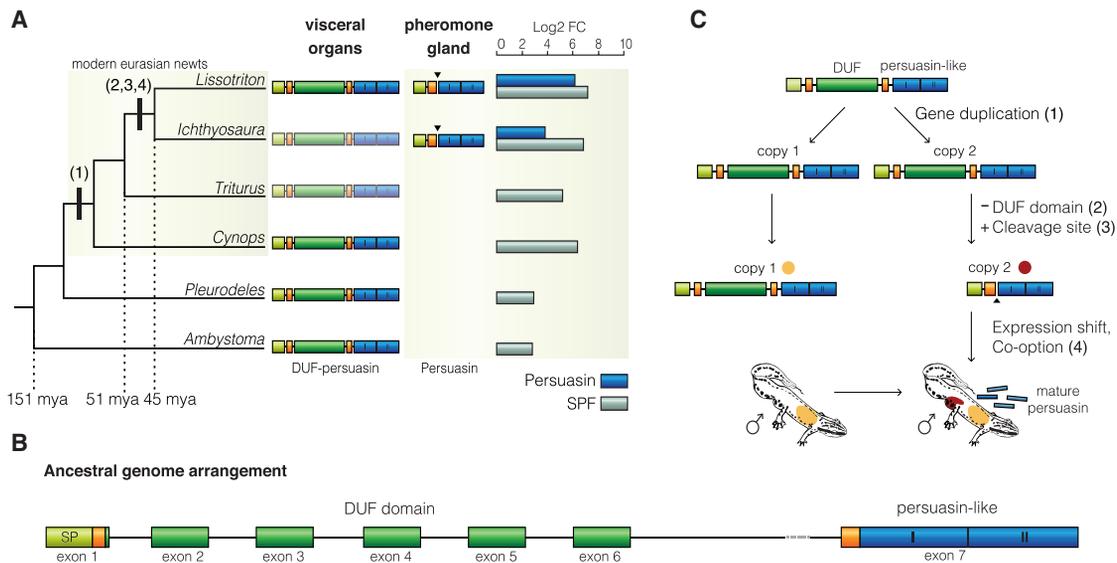
Figure 2. Evolution of Persuasin

(A) Comparative expression analyses demonstrate that elevated persuasin expression is restricted to the male pheromone-producing glands of *Lissotriton* and *Ichthyosaura*, whereas all examined salamanders show elevated sodefrin precursor-like factor (SPF) pheromone expression. The depicted salamander species and other salamanders also express a longer DUF-persuasin transcript in their visceral organs, which was involved in the origin of persuasin (see also Figure S2, S3 and Table S3; DUF = domain of unknown function). Numbers on the cladogram refer to evolutionary events represented on (C). Faded DUF-persuasin proteins are inferred.

(B) Ancestral genomic architecture of the DUF-persuasin gene from which persuasin originated. Colored exonic regions correspond to the signal peptide (SP, light green), parts homologous to the propeptide (orange, see Figure 1C and Figure S2), a domain of unknown function (DUF, dark green) and the two persuasin-like domains (blue). In the *Pleurodeles waltl* genome, all exons reside on the same scaffold, while in the axolotl, a gap in the sequence data is not conclusive on the genomic position of the persuasin domain.

(C) Persuasins evolved from a DUF-persuasin gene through gene duplication and neofunctionalisation. Losing the DUF domain and introducing a proteolytic cleavage site assisted in the functional transition of the duplicated gene. An expression shift to the pheromone-producing glands completed its co-option for pheromone activity.

salamander genomes of the axolotl (*Ambystoma mexicanum*) [23] and Spanish ribbed salamander (*Pleurodeles waltl*) [24]. We retrieved persuasin-containing transcripts from 14 additional species, which almost all differed from persuasin transcripts *sensu strictu* in having an additional stretch of about 230 aas (Figure 2A; Figure S2 and S3; see Table S3 for accession numbers). These correspond to parts of the persuasin propeptide and a protein domain that shows no similarity to any functionally characterized protein, henceforth referred to as the DUF domain ("domain of unknown function"). Genome screening shows that these DUF-persuasin transcripts are derived from a gene with seven exons, with the C-terminal persuasin part being fully contained within exon 7 (Figure 2B). Most of these transcripts were detected in visceral organs, such as the liver and lungs (Table S3).

To examine how persuasin transcripts in pheromone glands relate to the longer DUF-persuasin transcripts in internal organs, we performed phylogenetic analyses on the aa and nucleotide sequences of all available persuasin and persuasin-like sequences. Maximum parsimony, maximum likelihood, and bayesian analyses consistently revealed a gene duplication in the ancestor of modern Eurasian newts as the first step toward the origin of the new pheromones (Figure S3; Figures 2A and 2C). DUF-persuasins and persuasins are transcribed from different gene copies, indicating that alternative splicing is not responsible for the formation of persuasin in extant newts

(Figure S3, green and blue, respectively). Instead, all modern Eurasian newts appear to have retained one copy of the ancestral DUF-persuasin-like gene, while the other copy underwent neo-functionalisation (Figures 2A and 2C). This involved loss of the DUF domain, giving rise to precursors that exclusively encompass the cysteine-rich persuasin domain, and the introduction of a proteolytic cleavage site leading to the release of mature persuasin proteins (Figures 2A and 2C; Figure S2). Since the latter domain originated from a protein that is expressed in visceral organs, it was initially not involved in chemical communication but likely served another yet unknown function. Therefore, the birth of persuasin represents an exaptation in which a pre-existing protein domain was co-opted to function as a courtship pheromone. Interestingly, the DUF domain loss and the introduction of a proteolytic cleavage site created a protein in which 14% of the aas are cysteines, enabling the formation of up to ten disulphide (S-S) bridges within a single 15 kDa protein (Figure 1C). The two-domain cysteine-rich structure is similar to that of SPF pheromone proteins, which carry up to 9 disulphide bonds spread over two similar domains. A high number of cysteines serve to stabilize and increase the longevity of the protein pheromones [25] in pond water, which, among the wealth of molecules produced in pheromone glands, may have been an important selective advantage in becoming a courtship pheromone.

But why would a second pheromone have been retained if SPF proteins by themselves already form a fully functional pheromone

CellPress

A

B
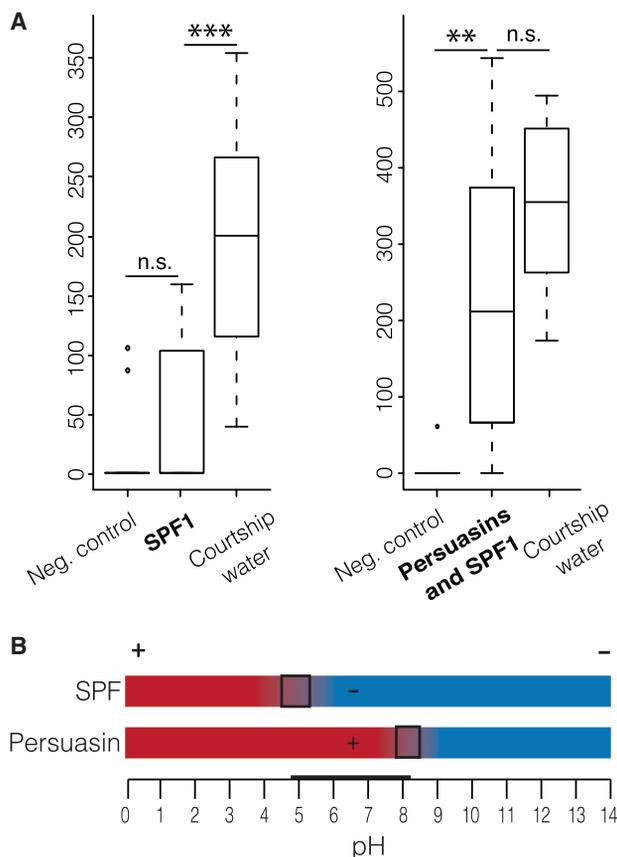
Figure 3. Female Behavioral Responses on Single and Combined Components of the Palmate Newt Male Courtship Pheromone Blend

(A) Non-parametric boxplots showing the median cumulative duration of female following behavior in the palmate newt (*Lissotriton helveticus*) in response to a single beta SPF isoform (left; Kruskal-Wallis, p < 0.001, X^2 = 17.4, n = 10), a combination of persuasins with the same SPF isoform (right; Kruskal-Wallis, p < 0.0001, X^2 = 18.9, n = 10), and their corresponding negative controls (0.1 mM bis-tris propane, 0.1 mM piperazine and 0.1 mM N-methyl piperazine, pH 7.5 for SPF1 and 10 mM of tris, pH 7 for the combination of persuasins and SPF1) and positive controls (courtship water). Significance levels shown on the boxplots represent the corrected p values provided by the post hoc Dunn's test (n.s. = p > 0.05, ** = p < 0.01, *** = p < 0.001). See also Figure S1D.

(B) Graphical representation of the theoretical isoelectric points (pI) and net charges of persuasins and SPFs over the entire pH range. Squares on the gradient depict the interval in which their pIs reside, while the bold line on the pH axis illustrates the pH range in which the overall net charge is negative for SPFs and positive for persuasins.

system? We hypothesized that a combination of SPF and persuasins would provide a selective advantage if both together induce stronger effects than each protein family alone. We therefore examined female behavioral responses on the most abundant SPF isoform alone, as well as on a combined SPF-persuasin pheromone blend. Two extant clades of SPF pheromones, termed alpha and beta, arose through an ancient gene duplication event followed by a series of additional duplications. Although earlier behavioral tests showed that a single alpha SPF is able to increase female receptivity, beta SPFs are clearly more abundant in palmate newts [9]. Here, we used anion exchange chromatography to isolate the highest expressed beta SPF protein (SPF1; Figure S1D) and subsequently investigated

its receptivity-enhancing capacity in a two-female behavioral assay. Interestingly, we observed no significant increase in female courtship responses when females were exposed to the SPF1 isoform alone (p = 0.2936, z = −0,54; Figure 3A, left panel). Next, SPF1 and persuasins (fraction 45, Figure 1B) were combined to assay the impact of the two protein families together on female receptivity. This blend induced an increase in female courtship behavior (p = 0.0023, z = 3.05; Figure 3A, right panel) that was not significantly different from the response observed in the full pheromone blend (courtship water) (p = 0.1247, z = 1.15; Figure 2A and 2D). Our combined behavioral assays (Figures 1D and 3A) thus suggest that beta SPFs and persuasins act in a cumulative or even synergistic way. Our estimates of the theoretical isoelectric points of SPFs (4.5–5.3) and persuasins (7.8–8.5) [26] and the adsorption of some SPF molecules to anion exchange columns at a pH of 7 [9] show that SPFs and persuasins carry opposite charges, with SPFs being negatively charged and persuasins positively charged at neutral or slightly acidic pH (Figure 3B). In freshwater, where there are hardly any ions to break ionic interactions, attraction between oppositely charged protein groups, in addition to other non-covalent forces, may accommodate the formation of transient protein complexes. Some regulatory proteins, known as scaffolds, co-localize protein components and increase local concentrations of otherwise freely diffusing molecules [27]. We hypothesize that persuasins and SPFs attract each other, delaying diffusion of the hydrophilic components in the aquatic environment. Enhanced co-localization of different components or evolutionary diverged isoforms would entail that the pheromone blend is delivered as a whole, leading to simultaneous activation of distinct olfactory neurons and a more coherent representation of the stimulus in the brain [1, 4, 28, 29]. Furthermore, the co-secretion of both protein families might already have been a selective advantage before persuasins were recruited as courtship pheromones by attracting SPF molecules into dimeric or multimeric complexes.

Although chemical cues of all kinds can be co-opted over evolutionary time to give rise to multicomponent pheromones, our study shows that molecules with specific structural properties (such as a cysteine-rich protein domain) have an important selective advantage to become recruited as a pheromone. Moreover, immediate selective advantages, such as co-localization of pheromone components, likely play important roles in the early stages of co-option by instantly reinforcing the extant pheromone blend.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animal collection and husbandry
  - Ethics
- METHOD DETAILS
  - Collection of protein courtship pheromones
  - Purifications of stimuli assayed in experiments
  - Two-female behavioral experiments

- ○ Molecular and computational protein characterization
- ○ RNA extraction and RNA-Seq library construction
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ○ Statistical analyses of behavioral experiments
  - ○ Transcriptome and genome analyses
  - ○ Phylogenetic analyses
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and three tables and can be found with this article online at https://doi.org/10.1016/j.cub.2018.06.074.

AUTHOR CONTRIBUTIONS

Conceptualisation, M.M., D.T., I.V.B., and F.B.; Methodology, M.M., D.T., I.V.B., F.B., H.D.G., and P.P.; Software: M.M., I.V.B., and F.B.; Formal Analysis: M.M. and D.T.; Investigation, M.M., D.T., and P.P.; Resources, I.V.B., F.B., H.D.G., and P.P.; Writing – original draft, M.M. and F.B.; Writing – Review & Editing, M.M., F.B., D.T., I.V.B., H.D.G., and P.P.; Visualization, M.M., F.B., and I.V.B.; Supervision, F.B. and I.V.B.; Funding Acquisition, F.B., I.V.B, H.D.G., and P.P.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Wyatt, T.D. (2014). Introduction to Chemical Signaling in Vertebrates and Invertebrates. In Neurobiology of Chemical Communication, C. Mucignat-Caretta, ed. (Boca Raton, FL: CRC Press), pp. 1–22.

2. Wyatt, T.D. (2014). Pheromones and animal behavior: Chemical signals and signatures, Second Edition (Cambridge, UK: Cambridge University Press).

3. Wyatt, T.D. (2003). Pheromones and Animal Behaviour: Communication by Smell and Taste (Cambridge, UK: Cambridge University Press).

4. Wilburn, D.B., Doty, K.A., Chouinard, A.J., Eddy, S.L., Woodley, S.K., Houck, L.D., and Feldhoff, R.C. (2017). Olfactory effects of a hypervariable multicomponent pheromone in the red-legged salamander, *Plethodon shermani*. PLoS ONE *12*, e0174370.

5. Lim, H., and Sorensen, P.W. (2012). Common carp implanted with prostaglandin F2α release a sex pheromone complex that attracts conspecific males in both the laboratory and field. J. Chem. Ecol. *38*, 127–134.

6. Levesque, H.M., Scaffidi, D., Polkinghorne, C.N., and Sorensen, P.W. (2011). A multi-component species identifying pheromone in the goldfish. J. Chem. Ecol. *37*, 219–227.

7. Novotny, M.V., Ma, W., Žídek, L., and Daev, E. (1999). Recent biochemical insights into puberty acceleration, estrus induction and puberty-delay in the house mouse. In Advances in Chemical Signals in Vertebrates, R.E. Johnston, D. Müller-Schwarze, and P.W. Sorensen, eds. (New York: Kluwer Academic/Plenum Press), pp. 99–116.

8. Novotny, M.V. (2003). Pheromones, binding proteins and receptor responses in rodents. Biochem. Soc. Trans. *31*, 117–122.

9. Van Bocxlaer, I., Treer, D., Maex, M., Vandebergh, W., Janssenswillen, S., Stegen, G., Kok, P., Willaert, B., Matthijs, S., Martens, E., et al. (2015). Side-by-side secretion of Late Palaeozoic diverged courtship pheromones in an aquatic salamander. Proc. Biol. Sci. *282*, 20142960.

10. Lassance, J.M. (2010). Journey in the *Ostrinia* world: from pest to model in chemical ecology. J. Chem. Ecol. *36*, 1155–1169.

11. Allison, J.D., and Cardé, R.T., eds. (2016). Pheromone Communication in Moths: Evolution, Behavior and Application (Berkeley, CA: University of California Press).

12. Janssenswillen, S., Vandebergh, W., Treer, D., Willaert, B., Maex, M., Van Bocxlaer, I., and Bossuyt, F. (2015). Origin and diversification of a salamander sex pheromone system. Mol. Biol. Evol. *32*, 472–480.

13. Van Bocxlaer, I., Maex, M., Treer, D., Janssenswillen, S., Janssens, R., Vandebergh, W., Proost, P., and Bossuyt, F. (2016). Beyond sodefrin: evidence for a multi-component pheromone system in the model newt *Cynops pyrrhogaster* (Salamandridae). Sci. Rep. *6*, 21880.

14. Janssenswillen, S., Willaert, B., Treer, D., Vandebergh, W., Bossuyt, F., and Van Bocxlaer, I. (2015). High pheromone diversity in the male cheek gland of the red-spotted newt *Notophthalmus viridescens* (Salamandridae). BMC Evol. Biol. *15*, 54.

15. Maex, M., Van Bocxlaer, I., Mortier, A., Proost, P., and Bossuyt, F. (2016). Courtship Pheromone Use in a Model Urodele, the Mexican Axolotl (*Ambystoma mexicanum*). Sci. Rep. *6*, 20184.

16. Houck, L.D., Watts, R.A., Mead, L.M., Palmer, C.A., Arnold, S.J., Feldhoff, P.W., and Feldhoff, R.C. (2008). A candidate vertebrate pheromone, SPF, increases female receptivity in a salamander. In Chemical Signals in Vertebrates II, J.L. Hurst, R.J. Beynon, S.C. Roberts, and T.D. Wyatt, eds. (New York: Springer), pp. 213–221.

17. Treer, D., Maex, M., Van Bocxlaer, I., Proost, P., and Bossuyt, F. (2018). Divergence of species-specific protein sex pheromone blends in two related, nonhybridizing newts (Salamandridae). Mol. Ecol. *27*, 508–519.

18. Malacarne, G., and Vellano, C. (1987). Behavioral evidence of a courtship pheromone in the crested newt, *Triturus cristatus carnifex* Laurenti. Copeia *1987*, 245–247.

19. Treer, D., Van Bocxlaer, I., Matthijs, S., Du Four, D., Janssenswillen, S., Willaert, B., and Bossuyt, F. (2013). Love is blind: indiscriminate female mating responses to male courtship pheromones in newts (Salamandridae). PLoS ONE *8*, e56538.

20. Verrell, P.A., Halliday, T.R., and Griffiths, M.L. (1986). The annual reproductive cycle of the Smooth newt *Triturus vulgaris* in England. J. Zool. *210*, 101–119.

21. Sever, D.M., Verrell, P.A., Halliday, T.R., Griffiths, M., and Waights, V. (1990). The cloaca and cloacal glands of the male smooth newt, *Triturus vulgaris vulgaris* (Linnaeus), with especial emphasis on the dorsal gland. Herpetologica *46*, 160–168.

22. Zhang, P., Papenfuss, T.J., Wake, M.H., Qu, L., and Wake, D.B. (2008). Phylogeny and biogeography of the family Salamandridae (Amphibia: Caudata) inferred from complete mitochondrial genomes. Mol. Phylogenet. Evol. *49*, 586–597.

23. Nowoshilow, S., Schloissnig, S., Fei, J.F., Dahl, A., Pang, A.W.C., Pippel, M., Winkler, S., Hastie, A.R., Young, G., Roscito, J.G., et al. (2018). The axolotl genome and the evolution of key tissue formation regulators. Nature *554*, 50–55.

24. Elewa, A., Wang, H., Talavera-López, C., Joven, A., Brito, G., Kumar, A., Hameed, L.S., Penrad-Mobayed, M., Yao, Z., Zamani, N., et al. (2017). Reading and editing the *Pleurodeles waltl* genome reveals novel features of tetrapod regeneration. Nat. Commun. *8*, 2286.

25. Fass, D. (2012). Disulfide bonding in protein biophysics. Annu. Rev. Biophys. *41*, 63–79.

26. Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, J.M. Walker, ed. (Humana Press), pp. 571–607.

27. Buday, L., and Tompa, P. (2010). Functional classification of scaffold proteins and related molecules. FEBS J. *277*, 4348–4355.

28. Vickers, N.J., Christensen, T.A., Baker, T.C., and Hildebrand, J.G. (2001). Odour-plume dynamics influence the brain's olfactory code. Nature *410*, 466–470.

29. Bathellier, B., Gschwend, O., and Carleton, A. (2010). Temporal Coding in Olfaction. In The Neurobiology of Olfaction, A. Menini, ed. (Boca Raton, FL: CRC Press/Taylor & Francis), pp. 329–351.

30. Bryant, D.M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M.B., Payzin-Dogru, D., Lee, T.J., Leigh, N.D., Kuo, T.H., Davis, F.G., et al. (2017). A Tissue-Mapped Axolotl *De Novo* Transcriptome Enables Identification of Limb Regeneration Factors. Cell Rep. *18*, 762–776.

31. McElroy, K.E., Denton, R.D., Sharbrough, J., Bankers, L., Neiman, M., and Lisle Gibbs, H. (2017). Genome expression balance in a triploid trihybrid vertebrate. Genome Biol. Evol. *9*, 968–980.

32. Zhu, R., Chen, Z.Y., Wang, J., Yuan, J.D., Liao, X.Y., Gui, J.F., and Zhang, Q.Y. (2014). Thymus cDNA library survey uncovers novel features of immune molecules in Chinese giant salamander *Andrias davidianus*. Dev. Comp. Immunol. *46*, 413–422.

33. Irisarri, I., Baurain, D., Brinkmann, H., Delsuc, F., Sire, J.Y., Kupfer, A., Petersen, J., Jarek, M., Meyer, A., Vences, M., and Philippe, H. (2017). Phylotranscriptomic consolidation of the jawed vertebrate timetree. Nat Ecol Evol *1*, 1370–1378.

34. Che, R., Sun, Y., Wang, R., and Xu, T. (2014). Transcriptomic analysis of endangered Chinese salamander: identification of immune, sex and reproduction-related genes and genetic markers. PLoS ONE *9*, e87940.

35. Nourisson, C., Muñoz-Merida, A., Carneiro, M., and Sequeira, F. (2017). *De novo* transcriptome assembly and polymorphism detection in two highly divergent evolutionary units of Bosca's newt (*Lissotriton boscai*) endemic to the Iberian Peninsula. Mol. Ecol. Resour. *17*, 546–549.

36. Stuglik, M.T., and Babik, W. (2016). Genomic heterogeneity of historical gene flow between two species of newts inferred from transcriptome data. Ecol. Evol. *6*, 4513–4525.

37. Chen, M.Y., Liang, D., and Zhang, P. (2015). Selecting question-specific genes to reduce incongruence in phylogenomics: A case study of jawed vertebrate backbone phylogeny. Syst. Biol. *64*, 1104–1120.

38. Goedbloed, D.J., Czypionka, T., Altmüller, J., Rodriguez, A., Küpfer, E., Segev, O., Blaustein, L., Templeton, A.R., Nolte, A.W., and Steinfartz, S. (2017). Parallel habitat acclimatization is realized by the expression of different genes in two closely related salamander species (genus *Salamandra*). Heredity (Edinb) *119*, 429–437.

39. Farrer, R.A., Martel, A., Verbrugghe, E., Abouelleil, A., Ducatelle, R., Longcore, J.E., James, T.Y., Pasmans, F., Fisher, M.C., and Cuomo, C.A. (2017). Genomic innovations linked to infection strategies across emerging pathogenic chytrid fungi. Nat. Commun. *8*, 14742.

40. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

41. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659.

42. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

43. Zhao, Y., Tang, H., and Ye, Y. (2012). RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. Bioinformatics *28*, 125–126.

44. Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J., and Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. *36*, 3420–3435.

45. Gilbert, D. (2013). Gene-omes built from mRNA seq not genome DNA. In 7th Annual Arthropod Genomics Symposium 2013 (Notre Dame).

46. Abascal, F., Zardoya, R., and Telford, M.J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. *38*, W7-13.

47. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

48. Swofford, D.L. (1998). Phylogenetic analysis using parsimony (*and other methods). Version 4 (Sunderland, Massachusetts: Sinauer Associates).

49. Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. *61*, 539–542.

50. Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Gateway Computing Environments Workshop, pp. 1–8.

51. Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML Web servers. Syst. Biol. *57*, 758–771.

52. Drechsler, A., Bock, D., Ortmann, D., and Steinfartz, S. (2010). Ortmann's funnel trap - a highly efficient tool for monitoring amphibian species. Herpetol. Notes *3*, 13–21.

53. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., and Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. *31*, 3784–3788.

54. Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods *8*, 785–786.

55. Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240.

56. George, R.A., and Heringa, J. (2000). The REPRO server: finding protein internal sequence repeats through the Web. Trends Biochem. Sci. *25*, 515–517.

57. Dinno, A. (2017). dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums. 1–6.

Q2 Q3 STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Transcript from Transcriptome Shotgun Assembly – *Ambystoma mexicanum*	NCBI [30]	GFBM010561853
Transcript from Transcriptome Shotgun Assembly – *Ambystoma laterale*	NCBI [31]	GFLO01013857
Transcript from Transcriptome Shotgun Assembly – *Ambystoma texanum*	NCBI [31]	GFLJ01013195
Transcript from Transcriptome Shotgun Assembly – *Ambystoma tigrinum*	NCBI [31]	GFLI01009595
Transcript from Transcriptome Shotgun Assembly – Unisexual *Ambystoma*	NCBI [31]	GFLD01014822
Expressed sequence tag – *Andrias davidianus*	NCBI [32]	JZ574685 & JZ574685
Sequence Read Archive data – *Calotriton asper*	NCBI [33]	SRX2382496
Sequence Read Archive data – *Cynops pyrrhogaster*	NCBI	SRX682053
Transcript from Transcriptome Shotgun Assembly – *Hynobius chinensis*	NCBI [34]	GAQK01005567, GAQK01005569, GAQK01005571
Transcript from publicly available transcriptome – *Lissotriton boscai*	[35]	TR57513_c0_g1_i1 TR69586_c0_g1_i1
Sequence Read Archive data – *Lissotriton montandoni*	NCBI [36]	PRJNA316531
Sequence Read Archive data – *Paramesotriton hongkongensis*	NCBI [37]	SRX796492
Transcript from publicly available transcriptome – *Pleurodeles waltl*	Available upon request from [24]	TRINITY_DN2005931_c0_g1_i3
Sequence Read Archive data – *Salamandra lanzai*	NCBI [38]	SRX2775500
Transcript from Transcriptome Shotgun Assembly – *Tylototriton wenxianensis*	NCBI [39]	GESS01023677
Genome contigs – *A. mexicanum*	http://genome.axolotl-omics.org/ [23]	AMEXG_0030007534 & AMEXG_0030107421
Genome contigs – *P. waltl*	Available upon request from [24]	abyss_v4.2_66113332, abyss_v4.2_26981551 & abyss_v4.2_26981552
Sequence Read Archive data – *Lissotriton helveticus*	This paper	SRR7396737
Sequence Read Archive data – *Ichthyosaura alpestris*	This paper	SRR7396735 and SRR7396736
Sequence Read Archive data – *Triturus karelinii*	This paper	SRR7396734
Sequence Read Archive data – *Cynops pyrrhogaster*	This paper	SRR7396733
Sequence Read Archive data – *Pleurodeles waltl*	This paper	SRR7396732
Sequence Read Archive data – *A. mexicanum*	This paper	SRR7396731
Transcripts assembled from our own RNASeq data (*L. helveticus*, *I. alpestris* and *T. karelinii*)	This paper	MG435350-MG435355; MH500757-MH500767
Transcripts assembled from publicly available Sequence Read Archive datasets (*C. asper*, *C. pyrrhogaster*, *L. montandoni*, *P. hongkongensis*, *S. lanzai*)	This paper	https://doi.org/10.17632/9xyp88y6n5.1
Software and algorithms		
R3.4.1	Open Source	RRID: SCR_001905; http://www.r-project.org/
Trim Galore! v0.4.4	Babraham bioinformatics	RRID:SCR_011847; https://github.com/FelixKrueger/TrimGalore/releases
Trinity v2.5.1	[40]	RRID:SCR_013048; https://github.com/trinityrnaseq/trinityrnaseq/releases
CD-HIT	[41]	RRID:SCR_007105; https://github.com/weizhongli/cdhit/releases

(*Continued on next page*)

CellPress

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Kallisto v0.43.1	[42]	https://pachterlab.github.io/kallisto/download
RAPSearch2	[43]	http://omics.informatics.indiana.edu/mg/RAPSearch2/
Blast2Go	[44]	RRID:SCR_005828; https://www.blast2go.com
EvidentialGene	[45]	http://arthropods.eugenes.org/EvidentialGene/trassembly.html
TranslatorX	[46]	RRID:SCR_014733; http://translatorx.co.uk/
MAFFT v7	[47]	RRID:SCR_011811; https://mafft.cbrc.jp/alignment/server/
PAUP* v4.0	[48]	RRID:SCR_014931; http://paup.phylosolutions.com/
MrBayes 3.2.232	[49]	RRID:SCR_012067; on the CIPRES Science Gateway v3.333 [50]
RAXML 7.0.4	[51]	RRID:SCR_006086; on the CIPRES Science Gateway v3.333 [50]
CIPRES Sceince Gateway V3.333	[46]	RRID:SCR_008439; http://www.phylo.org/sub_sections/portal/
Tracer 1.634	BEAST Developers	http://beast.community/tracer

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Franky Bossuyt (franky.bossuyt@vub.be).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animal collection and husbandry
Twenty adult males and 30 adult females of the alpine as well as palmate newt were caught in spring from ponds near Ternat, Belgium, using a modified Ortmann's funnel trap [52]. Animals were housed in single sex aquaria (lwh: 60 × 35 × 35 cm) with a maximum of 15 newts per aquarium. Every aquarium was filled with 30 l of water (50% demineralised, 50% tap water) and enriched with artificial plants and hiding places. Housing containers were kept in an air-conditioned room at 15-18°C under a light/dark regime of 13/11h. Every other day, newts were fed with live and defrosted *Daphnia*, bloodworms, *Artemia*, earthworms and maggots.

Ethics
A permit for the collection of palmate and alpine newts was issued by the 'Agentschap voor Natuur en Bos' (permits ANB/BL-FF/V12-00050 and ANB/BL-FF/V13-00134). All procedures are in compliance with EU (Annex IV, Directive 2010/63/EU of the European Parliament and of the Council of September 22nd 2010) and Belgian regulations (Annex VII, Belgian Law of May 29th 2013) concerning laboratory animal care and welfare. This research was approved by the Ethical Committee for Animal Experiments of the Vrije Universiteit Brussel (Project number EC14-220-35), and all experiments were performed in accordance with the approved guidelines and regulations.

METHOD DETAILS

Collection of protein courtship pheromones
Collection, extraction and crude separation of courtship pheromones were achieved as reported in [9]. Here, receptive males and females were allowed to court per couple in a plastic container filled with 600 mL water for approximately 1h. Non-courting male-female water was collected for each animal individually to avoid courtship between males and females. Molecules present in courtship and non-courting male-female water were extracted directly from the water using disposable reversed-phase solid phase extraction cartridges (RP-C8 and RP-C18, Sep-Pak plus cartridge, 400 mg sorbent, Waters, Milford, MA, USA). Using a vacuum pump, each cartridge was loaded with a maximum of 300 mL water to which 7.5 mL of 90% (v/v) acetonitrile containing 0.1% (v/v) trifluoroacetic acid was added. Molecules were manually eluted from the reversed-phase adsorbent resins with 7.5 mL 90% (v/v) acetonitrile with 0.1% (v/v) trifluoroacetic acid (TFA). After 1h of evaporation in a speedvac concentrator (SCV-100H, Savant instruments, Farmingdale, NY), samples of multiple couples were pooled and loaded onto a Source 5RPC column (4.6 × 150 mm, GE Healthcare Life Sciences) pre-equilibrated with 0.1% (v/v) TFA (solvent A) for reversed-phase high-performance liquid chromatography (RP-HPLC). After a wash step with solvent A, protein elution was attained using 80% acetonitrile in 0.1% TFA

(solvent B) in a linear gradient from 0% to 100% B in 80 min at 1 ml/min. Absorbance was measured at a wavelength of 214 nm and the eluate was collected in 1 mL fractions. The RP-HPLC profiles displayed in Figure 1A originates from six couples. However, more than 20 RP-HPLC separations were performed for 1 to 40 couples, consistently obtaining similar peak fractions. Fractions of interest were subjected to non-reducing SDS-PAGE using precast gels (Any kD Mini-PROTEAN TGX, Bio-Rad, Hercules, CA), transferred onto a polyvinylidene difluoride membrane by semi-dry blotting (Trans Blot Turbo System, Bio-Rad), and Coomassie stained (Coomassie brilliant blue R-250; Sigma, St. Louis, MO) to excise protein bands of interest for N-terminal amino acid sequencing by a 491 Procise cLC protein sequencer (Applied Biosystems, Foster city, CA). Protein masses in the RP-HPLC fractions were identified by electrospray ionisation mass spectrometry on an ESQUIRE LC/MS ion trap (Bruker, Brussels, Belgium). Averaged profile spectra of proteins were deconvoluted using Bruker's software DATA ANALYSIS v4.1.

Purifications of stimuli assayed in experiments

Persuasin purification

In order to purify naturally secreted persuasins, courtship water from thirty couples was collected using the same methods as described above, however, only using RP-C8 cartridges to extract the secreted proteins from the water. Next, immobilised lectins (HiTrap con A 4B column, GE Healthcare Life Sciences) were employed to selectively capture glycosylated SPF proteins from the pheromone blend. To prepare the sample for lectin affinity, the RP-C8 cartridge eluate was made less acidic by adding 1 M tris pH 7.4 (up to approximately a pH of 7, checked by pH paper). Next, complete buffer exchange was attained using PD-10 desalting columns (GE Healthcare Life Sciences) equilibrated with 20 mM Tris-HCl, 0.5 M NaCl, 1 mM $MnCl_2$, 1 mM $CaCl_2$, pH 7.4, enabling affinity chromatography under high salt buffer conditions. The sample was loaded onto a 1 mL HiTrap con A 4B column (GE Healthcare Life Sciences) a constant flow rate of 0.1 mL/min at 4°C, washed with the equilibration buffer and step-eluted with 0.5 M methyl-α-D-glucopyranoside (98+%, Alfa Aesar), 20 mM Tris-HCl, 0.5 M NaCl, pH 7.4 at room temperature. After acidification of both the eluate (adsorbed fraction) and the flow-through (non-adsorbed fraction) with TFA to 0.1% (v/v), RP-HPLC was conducted as a polishing step, eradicating lectins that leaked from the column. Both samples were loaded separately onto a 15RPC SOURCE column (4.6 × 100 mm, GE Healthcare Life Sciences) pre-equilibrated with 0.1% TFA at a constant flow rate of 0.4 mL per minute, washed, and eluted with 80% acetonitrile in 0.1% TFA by applying the following gradient: (1) 0%–30% B in 12 min, (2) 30%–65% B in 48 min (increase of 0.5% acetonitrile per minute), (3) 65%–100% B in 14 min. Absorbance was measured at a wavelength of 214 nm and the eluates were collected in 0.5 mL fractions. Pooled fractions were analyzed by non-reducing SDS-PAGE using Any kD Mini-PROTEAN TGX precast gels (Bio-Rad) and proteins were visualized by silver staining (Silverquest Silver Staining kit, Invitrogen, Carlsbad, CA). RP-HPLC fractions 28 to 42 of the lectin flow-through were used in the behavioral assay to characterize persuasins functionally (see supplementary material Figure S1C). Acetonitril was removed from these fractions under vacuum by rotational evaporation and 1 M tris pH 7.4 was added to the samples until a pH of 7 was reached (verified on a paper pH strip).

Purification of a single beta SPF isoform

Courtship water was collected as described above, and SPF1 (GenBank: KJ402326) was purified similar to how SPFs were purified in [9]. In short, courtship water of forty couples was retained for application on RP-C8 and RP-C18 cartridges. Next, pooled and concentrated samples were loaded onto a Source 5RPC column pre-equilibrated with 0.1% (v/v) TFA, washed with the same buffer, and eluted with 80% acetonitrile in 0.1% TFA (solvent B), going from 30%–65% of solvent B in 56 min, which effects a sequential elution and separation of the different SPF isoforms [9]. Using a combination of SDS-PAGE, silver staining, Edman degradation and mass spectrometry only RP-HPLC fractions that contained SPF1, and no other SPF isoforms, were selected for ion exchange chromatography to eliminate persuasins. Acetonitrile was evaporated from all fractions of interest and the concentrated sample was brought to a pH of 7.5 by diluting it in a buffer containing 20 mM bis-tris propane, 20 mM piperazine and 20 mM, N-methyl piperazine (Sigma). The sample was loaded onto a 1 mL Hitrap DEAE Fast Flow (GE Healthcare Life Sciences) column pre-equilibrated with binding buffer containing 15 mM bis-tris propane, 15 mM piperazine and 15 mM N-methyl piperazine (buffer A, pH 7.5) and washed for at least 10 min with the same buffer until all material in the effluent disappeared. Persuasins - with theoretical isoelectric points ranging between 7.8 and 8.5 - were not retained on the column while SPF1 was (isoelectric point, pI = 4.64) [9]. Proteins were eluted using 15 mM bis-tris propane, 15 mM piperazine and 15 mM N-methyl piperazine (buffer B, pH 3) by applying a linear gradient (from 0% to 100% B in 20 min). Detection of eluting proteins was performed at a wavelength of 280 nm and the eluate was collected in fractions of 1 mL. Fractions were analyzed by non-reducing SDS-PAGE using Any kD Mini-PROTEAN TGX precast gels and proteins were visualized by silver staining (Figure S1D). Eluted peak fractions used for the behavioral assay were brought to a pH of 7 using 1 M tris pH 7.4 (verified on a paper pH strip).

Protein purification of SPF1 combined with persuasins

An RP-HPLC run with courtship water of six courting couples was performed as described in section 'Collection of courtship pheromones' to attain a combination of a single beta SPF isoform and persuasins. The RP-HPLC fraction with beta SPF1 and several persuasins was selected by means of N-terminal sequencing and mass spectrometry (fraction 45, Figure 1A). Prior to the two-female behavioral assay, all acetonitrile was evaporated in a speedvac concentrator and the sample was brought to a pH of 7 by adding 1 M tris (pH 7.4).

Two-female behavioral experiments

Pheromone candidates were assayed in a two-female set-up [19]. Here, two females are put together in a testing container filled with water to which a chemical stimulus is added. When two females are exposed to courtship water (i.e., water in which a palmate newt

male courted a female and released courtship pheromones), females show natural female mating behavior toward each other, including following behavior [19]. In the present study, female courtship behavior was recorded and quantified as the cumulative amount of time in which at least one of the two females was following the other female within the 7 min time span of the test. Experiments were carried out with 1) purified persuasins, 2) a single beta SPF isoform, and 3) a combination of both. For each behavioral trial, two females were put together in a plastic container (lwh: 25 × 16 × 14 cm) filled with 800 mL of aged tap water (50% demineralised, 50% tap water) in which purified pheromone candidates were dissolved or to which a control solution was added (up to additional 10 mL). All experiments were performed under the same light (artificial daylight) and temperature (15-18°C) conditions, and only water was used that had been kept in the room where the animals were housed, ensuring similar temperature conditions. To evaluate the suitability of female palmate newts for a behavioral test, all experiments were preceded by a receptivity test as previously described [19]. Ten female couples were used to test each stimulus, and its corresponding negative and positive control. Freshly collected courtship water was used as a positive control. Negative controls were water with 10 mM of tris, pH 7 (for persuasins, and beta SPF1 combined with persuasins) or water with 0.1 mM bis-tris propane, 0.1 mM piperazine and 0.1 mM N-methyl piperazine, pH 7.5 (i.e., anion exchange buffer, used as a negative control for beta SPF1). These negative controls represent the same buffer concentrations as when the buffered test stimulus is added to 800 mL of water. A single set of behavioral experiments (stimulus and corresponding controls) was always performed in a short time frame (2-4 days) within a single breeding season.

Molecular and computational protein characterization

To characterize persuasins on a molecular level, *de novo* persuasin transcripts (see below) were translated into protein sequences using the ExPASy Translate tool [53]. Signal peptide motifs were predicted using SignalP (v4.1) [54], InterProScan 5 was used to identify protein domains [55] and REPRO to identify 10-cysteine motif repeats [56]. To infer orthology to known vertebrate proteins, BLASTp, tBLASTn and BLASTn searches were performed against the NCBI vertebrate nucleotide collection and non-redundant protein sequence database (https://www.ncbi.nlm.nih.gov/). Theoretical relative molecular masses of persuasins without signal peptide and propeptide were calculated using the pI/Mr tool on ExPASy. The formation of all disulfide bridges was taken into account, subtracting twice the relative molecular mass of H^+ (1.008 Da) per possible disulfide bridge.

RNA extraction and RNA-Seq library construction

Animals were euthanised by spraying 0.2-0.5 mL of a 10% lidocaine solution on the head or by immersing them in a 0.5 g/L buffered MS-222 solution followed by decapitation and pithing of the brain. Pheromone-producing glands were excised from the body cavity and stored in RNALater solution (Sigma-Aldrich). Total RNA (1 μg) was extracted from the pheromone-producing glands using the RNeasy Plus Universal Mini Kit (QIAGEN). Whole transcriptome shotgun sequencing (RNA-seq, paired-end) was performed on the pheromone-producing dorsal glands of one palmate newt male (*Lissotriton helveticus*, wild-caught; 52 Mio 50 bp read pairs at Baseclear, Leiden, the Netherlands; reads obtained earlier in [9]), two alpine newt males (*Ichthyosaura alpestris*, wild-caught; 48 Mio 50 bp read pairs at Baseclear and 57 Mio 100 bp read pairs at DNAVision, Gosselies, Belgium; reads obtained earlier in [17]), one southern crested newt male (*Triturus karelinii*, captive bred; 159 Mio 100 bp read pairs at DNAVision) and one Japanese fire belly newt male (*Cynops pyrrhogaster*, captive bred; 51 Mio 50 bp read pairs at Baseclear; reads obtained earlier in [13]). Dorsal glands form a dorsal part of the cloaca that secretes into the posterior end of the cloacal cavity. In species that lack dorsal glands, whole transcriptome sequencing was conducted on the cloaca. This was done for a Spanish ribbed newt male (*Pleurodeles waltl*, captive bred; 58 Mio 100 bp read pairs at DNAVision) and an axolotl male (*Ambystoma mexicanum*, captive bred; 64 Mio 100 bp reads pair at DNAVision; reads obtained earlier in [15]). All libraries sequenced by Baseclear were constructed using an Illumina TruSeq RNA sample preparation kit for sequencing on the Illumina HiSeq 2500 platform (Illumina, San Diego, California); procedures for base calling, filtering, quality control and trimming are explained in detail elsewhere [9]. Libraries sequenced by DNAVision were created using the Illumina TruSeq RNA sample preparation kit and sequenced on an Illumina NextSeq500 instrument. Here, adaptor sequences and low quality bases were trimmed using Trim Galore! v0.4.4. Bases with an average quality score below 5 (Phred33) were filtered out and read pairs for which forward or reverse reads were trimmed to < 25 nucleotides were discarded.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses of behavioral experiments

Statistical comparisons were only made within a set of behavioral experiments (i.e., stimulus, positive control and negative control, depending on buffers in which the test stimulus resided). Data were submitted to Kruskal-Wallis tests followed by post hoc Dunn's tests for pairwise comparisons using the holm correction for multiple testing (dunn.test, version 1.3.4) [57]. Statistical analyses were performed using R statistical software (R3.4.1, R Foundation for Statistical Computing, Vienna, Austria). Statistical details can be found in the figures, figures legends, and main text. Throughout the figures, we indicate p values as follows: 'ns' if $p > 0.05$, * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$. 'n' represents the number of female couples tested. Results of behavioral experiments are summarized in box-and-whisker plots, displaying the minimum, first quartile, median, third quartile and maximum duration of female following.

Transcriptome and genome analyses

De novo transcriptomes of male pheromone-producing glands were assembled in Trinity v2.5.1 using default parameters [40] and clustered at 95% sequence identity level using CD-HIT [41]. Trancipt expression levels were estimated using kallisto (version 0.43.1) [42] and expression values were quantified as transcripts per million (TPM). *De novo* persuasin, DUF-persuasin and SPF sequences were identified using RAPSearch2 [43], using a custom-made local database. SPF sequences identified by RAPSearch2 were re-evaluated using tBLASTn and tBLASTx in Blast2Go [44]. Persuasin and SPF expression values were each calculated by adding up the TPM values of all identified persuasin or SPF transcripts, and dividing them by the TPM value of housekeeping gene *eukaryotic translation elongation factor 1 alpha 1* (*eef1a1*), obtaining the fold increase in expression relative to the housekeeping gene. Fold changes are represented in logarithm base two (log2FC). RAPSearch2 was used to identify the highest expressed transcript of housekeeping gene *eef1a1* in each dataset. One DUF-persuasin transcript was identified in the dorsal gland of *Triturus karelinii*, but since its expression was lower than 2 TPM level (0.62 TPM, log2FC = −12.43), this transcript was discarded from the expression analyses (Figure 2A).

To extract less truncated, full-length transcripts for mass prediction, molecular characterization, and phylogenetic analyses (see below), we also examined transcriptomes built by EvidentialGene [45] in combination with Trinity v2.5.1 [40] (Table S2 with expression values for palmate and alpine newt). We reconstructed three assemblies indicating (1) no strand-specificity, (2) RF strand-specificity and (3) FR strand-specificity and combined these in an over-assembly. The over-assembly was processed to a more accurate and concise assembly in EvidentialGene [45]. Although our libraries are not strand-specific, the De Bruyn graphs in Trinity are less complicated when indicating strand-specificity (http://groups.google.com/forum/#!msg/trinityrnaseq-users/2SY_L7tKx88/Qj7hqBVTAQAJ), and transcripts with longer open reading frames are built. However, sequencing depth is reduced by approximately 50% and transcripts are assembled in both orientations. Therefore, EvidentialGene was used to filter out the artifacts created by opting for strand-specificity (i.e., transcripts being made in both directions), while preserving the intentioned sequencing depth by including a trinity assembly with default parameters. Some full-length transcripts used for mass prediction and further characterization were retrieved from EviGene's 'dropset' or 'okay-alts'.

The axolotl genome was screened for the presence of persuasin or DUF-persuasin genes using the axolotl genome browser (https://genome.axolotl-omics.org) [23]. The pleurodeles genome [24] was screened using the BLASTn and tBLASTn command line applications (NCBI).

Phylogenetic analyses

To expand our phylogenetic analyses, we screened all publicly available urodelan transcriptomes as well as all urodelan RNA-seq raw read datasets for persuasin or DUF-persuasin transcripts. In sum, we extracted transcripts from our own assembled transcriptomes (see Table S3 for accession numbers), from readily available, online transcriptomes (9 species; Transcriptome Shotgun Assembly, NCBI or other resources; consult Key Resources Table and Table S3 for accession numbers), but also reconstructed *de novo* persuasin or DUF-persuasin transcripts from publicly available raw reads (5 species; Short Read Archive on NCBI; accession numbers are listed in Key Resources Table and Table S3). Urodelan RNA-seq short read datasets were screened with tBLASTn for the presence of reads that map on persuasin or DUF-persuasin transcripts, retrieved using the SRA toolkit (https://www.ncbi.nlm.nih.gov/sra) and assembled into a transcriptome using EvidentialGene [45] in combination with Trinity v2.5.1 [40] as described above.

A protein-coding-guided multiple nucleotide sequence alignment was made in TranslatorX [46] using MAFFT version 7 (G-INS-1 method allowing gappy regions) for the protein alignment [47]. Phylogenetic analyses were performed on amino acid sequences of the signal peptide and persuasin(-like) part, as well as on nucleotide sequences of the signal peptide, propeptide, and persuasin(-like) part (hence in both cases omitting the DUF domain). Maximum parsimony (MP) and maximum-likelihood (ML) analyses were run in PAUP* v4.0 [48], using the model assigned by MrBayes in the latter. Bayesian analyses were run in MrBayes 3.2.232 [49] at the CIPRES Science Gateway v3.333 [50], with a mixed prior for the substitution models and gamma correction for among-site rate heterogeneity. Two parallel runs of four Markov chain Monte Carlo were executed for 10,000,000 generations, with trees sampled every 1,000th generation and the first 5,000 generations discarded as burn-in. Convergence of parallel runs was confirmed by split frequency standard deviations (less than 0.01) and potential scale reduction factors (approximating 1.0) for all model parameters. Adequate posterior sampling for each run was verified using Tracer 1.634, by examining the effective sampling sizes of all model parameters. Clade support under ML was assessed by 1,000 replicates of rapid bootstrapping using RAXML 7.0.4 [51] on the CIPRES Science Gateway v3.333.

DATA AND SOFTWARE AVAILABILITY

Mass spectrometry and gene expression results are presented in the supplementary material and main article. Nucleotide and protein sequences have been deposited in GenBank under accession numbers GenBank: MG435350–MG435355 and GenBank: MH500757–MH500767. In the absence of an NCBI database for sequences assembled by third party analyses (assembly of publicly available reads), an additional set of sequences was deposited on the Mendeley Data repository (https://doi.org/10.17632/9xyp88y6n5.1). Raw reads have been deposited in the NCBI Sequence Reads Archive (SRA) database under BioProject NCBI: PRJNA476935, BioSamples SAMN09461956-SAMN09461962 and Runs SRR7396731- SRR7396737.