

SCIENTIFIC REPORTS



OPEN

Beyond sodefrin: evidence for a multi-component pheromone system in the model newt *Cynops pyrrhogaster* (Salamandridae)

Received: 21 September 2015

Accepted: 18 January 2016

Published: 03 March 2016

Ines Van Bocxlaer¹, Margo Maex¹, Dag Treer¹, Sunita Janssenswillen¹, Rik Janssens², Wim Vandebergh¹, Paul Proost² & Franky Bossuyt¹

Sodefrin, a decapeptide isolated from the male dorsal gland of the Japanese fire belly newt *Cynops pyrrhogaster*, was the first peptide pheromone identified from a vertebrate. The fire belly salamander and sodefrin have become a model for sex pheromone investigation in aquatically courting salamanders ever since. Subsequent studies in other salamanders identified SPF protein courtship pheromones of around 20 kDa belonging to the same gene-family. Although transcripts of these proteins could be PCR-amplified in *Cynops*, it is currently unknown whether they effectively use full-length SPF pheromones next to sodefrin. Here we combined transcriptomics, proteomics and phylogenetics to investigate SPF pheromone use in *Cynops pyrrhogaster*. Our data show that not sodefrin transcripts, but multiple SPF transcripts make up the majority of the expression profile in the dorsal gland of this newt. Proteome analyses of water in which a male has been courting confirm that this protein blend is effectively secreted and tail-fanned to the female. By combining phylogenetics and expression data, we show that independent evolutionary lineages of these SPF's were already expressed in ancestral *Cynops* species before the origin of sodefrin. Extant *Cynops* species continue to use this multi-component pheromone system, consisting of various proteins in addition to a lineage-specific peptide.

Sodefrin, from the Japanese word “sodefuri” meaning soliciting, was the first peptide pheromone identified from a vertebrate¹. This small decapeptide (~1 kDa) was isolated from the sexually dimorphic dorsal glands of the Japanese fire belly newt *Cynops pyrrhogaster* (Salamandridae). Males of this species use an elaborate courtship display to tail-fan the pheromones produced in these glands via their cloaca to the female (see movie S1). Since the discovery of sodefrin, this decapeptide and its species *Cynops pyrrhogaster* have become a model system for pheromone investigation in aquatically courting salamanders^{1–11}. Variants of this pheromone were also found in a different population of this species (aonirin) and in its congeneric *C. ensicauda* (silefrin)^{7,12}. In behavioural tests, these decapeptides were shown to function as female attractants in a species-specific and even population-specific way^{3,6,7,12}.

Subsequent studies on chemical communication during courtship in urodelans identified three families of larger, uncleaved proteins as courtship pheromones in lungless salamanders (Plethodontidae). Interestingly, in addition to the family-specific Plethodontid Modulating Factor^{13–17} and Plethodontid Receptivity Factor^{18,19}, these studies revealed protein precursors that are largely similar to sodefrin precursors and were termed Sodefrin Precursor-like Factors²⁰ (SPF, Fig. 1). The sodefrin decapeptide and SPF protein pheromones are thus derived from the same gene family, here referred to as the SPF-family (see Fig. 1 for consistency in use of terminology in this paper). Behavioural tests with isolated SPF proteins in both lungless salamanders (Plethodontidae) and newts (Salamandridae) have demonstrated that they function as courtship pheromones that enhance female receptivity^{20,21}. Recent studies on the expression and evolution of SPF-family transcripts in cloacal glands of Salamandridae showed that, compared to SPF transcripts, sodefrin transcripts contain an extra 62 basepair insertion (Fig. 1, insert indicated in black)²². This causes a frameshift in protein translation and generates a sequence

¹Amphibian Evolution Lab, Biology Department, Vrije Universiteit Brussel (VUB), Pleinlaan 2, B-1050 Brussels, Belgium. ²Laboratory of Molecular Immunology, Department of Microbiology and Immunology, KU Leuven - University of Leuven, Minderbroedersstraat 10 - box 1030, B-3000 Leuven, Belgium. Correspondence and requests for materials should be addressed to I.V.B. (email: ivocxla@vub.ac.be)

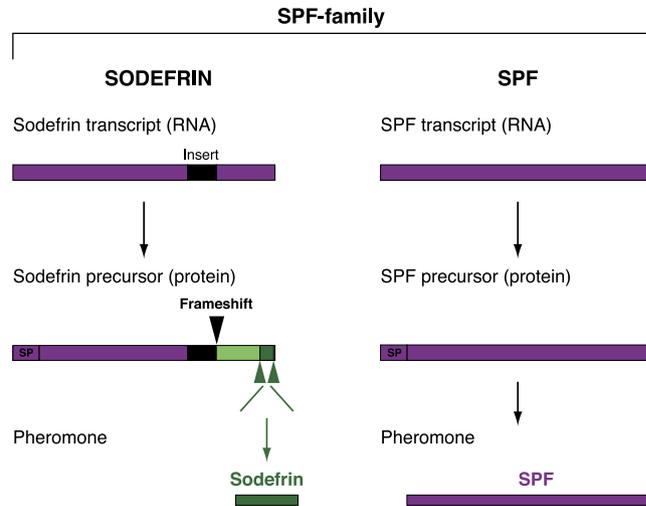


Figure 1. Schematic comparison of sodefrin and SPF. Sodefrin and SPF pheromones belong to the same gene family, here referred to as the SPF (pheromone) family, and are homologues (indicated in purple) at the nucleotide level. The major difference is that sodefrin RNA transcripts contain a 62 base pair insert (indicated in black) that causes a translational frame shift at the C-terminal end of the protein (indicated in light + dark green). This leads to the origin of cleavage sites (indicated with green arrows) that allow generating the decapeptide sodefrin (indicated in dark green). SPF transcripts do not have such an insert, and become expressed as full-length proteins. The pheromones that result from each precursor type therefore show no amino acid sequence similarity. SP = signal peptide.

(Fig. 1, indicated in dark and light green) from which the decapeptide pheromone can be cleaved (Fig. 1, indicated in dark green). The frameshift originated in an ancestral *Cynops* lineage, indicating that it is not present in other salamanders, and that sodefrin and its variants are likely restricted to the *Cynops* lineage²². Because the newly generated sodefrin peptide has no sequence homology with existing salamander SPF protein pheromones, this decapeptide represents an interesting case of the birth of a new pheromone²².

The urodelan SPF pheromone family is a complex one in which gene duplications have been very common²². The first duplication, already in the Late Palaeozoic, denotes the onset of diversification and gave rise to two SPF lineages that were termed alpha and beta^{21,22}. Today's descendant alpha and beta transcripts encode protein pheromones that are still secreted by courting male palmate newts²¹. Although SPF transcripts have been PCR-amplified in the genus *Cynops*^{22,23}, information on effective translation, precursor abundance, and release of these proteins into the water during courtship is inexistent. As a consequence, it is currently unknown whether *Cynops* species actively use full length SPF pheromones next to sodefrin, or whether the decapeptide has replaced the function of SPF in this genus.

Here we studied the use of SPF proteins during courtship in *Cynops pyrrhogaster*. We first performed RNA-Seq transcriptome analyses of the male pheromone-producing dorsal gland to study the significance of SPF with regard to sodefrin transcript expression in this species. To validate these results, we also performed proteome analyses of water in which a male Japanese fire belly newt had been tail fanning, and demonstrate that multiple SPF proteins are effectively secreted and tail-fanned to the female during courtship (see movie S1). Finally, we combined a phylogeny of SPF-family transcripts with our transcriptome and proteome data to show that both alpha and beta SPF's were already effectively secreted in water before the origin of the peptide sodefrin.

Results and Discussion

SPF transcripts are highly expressed in the male dorsal glands. To identify the SPF and sodefrin transcripts in the pheromone-producing dorsal gland of male *Cynops pyrrhogaster*, we performed RACE-PCR with a broad spectrum of primers designed on conserved regions of known SPF-family transcripts (Table S1). Obtained sequences were assembled in 99% contiguous sequences (contigs), which resulted in 25 cDNA sequences. (Table 1, GenBank numbers KU213615-KU213639). Two of these were identified as sodefrin transcripts by the presence of the typical 62 bp nucleotide insert (Table 1, CyPy_SPF_3 and 5), while all others are SPF transcripts. Translated sequences displayed up to 80% pairwise divergence, similar to the diversity found in the palmate newt²¹.

To further investigate the relative expression of sodefrin and SPF transcripts, we performed complete transcriptome sequencing (RNA-seq) of the dorsal gland. This yielded 101,402,858 reads, together constituting 5.2 Giga bases (Gb). Expression analyses resulted in mapping of 17,607,281 of these reads on the transcripts that were obtained by RACE PCR, indicating that at least 17.4% of the total RNA-Seq reads of the dorsal gland belong to the SPF-family. From this subset of reads, 15.2% of the expression could be attributed to the sodefrin transcripts. Investigation of the read mappings showed a relatively uniform distribution (including over the 62 bp insert range) with few aspecific reads, confirming that the expression level of these sodefrin transcripts is a realistic estimate. Uncleaved alpha and beta SPF transcripts accounted for 42.8% and 42% of the SPF-family expression

Name	Gene length	RPKM	% RPKM	Unique gene reads	Total gene reads
CyPy_SPF_1	753	458.927	34,41	5.913.474	6.084.698
CyPy_SPF_2	833	241.603	18,11	3.543.589	3.543.634
CyPy_SPF_3	738	142.199	10,66	786.539	1.847.790
CyPy_SPF_4	737	132.387	9,93	1.668.229	1.717.963
CyPy_SPF_5	734	60.250	4,52	344.312	778.674
CyPy_SPF_6	686	52.026	3,90	612.800	628.411
CyPy_SPF_7	662	41.813	3,13	232.970	487.380
CyPy_SPF_8	751	35.040	2,63	190.772	463.342
CyPy_SPF_9	743	32.691	2,45	244.395	427.685
CyPy_SPF_10	706	28.199	2,11	288.031	350.547
CyPy_SPF_11	520	16.174	1,21	127.386	148.092
CyPy_SPF_12	687	13.107	0,98	89.451	158.550
CyPy_SPF_13	718	12.924	0,97	91.112	163.385
CyPy_SPF_14	718	10.545	0,79	108.983	133.310
CyPy_SPF_15	720	9.150	0,69	103.095	115.999
CyPy_SPF_16	661	8.514	0,64	82.660	99.091
CyPy_SPF_17	677	8.237	0,62	83.855	98.193
CyPy_SPF_18	662	7.877	0,59	51.851	91.813
CyPy_SPF_19	673	6.641	0,50	66.292	78.696
CyPy_SPF_20	675	6.320	0,47	37.831	75.111
CyPy_SPF_21	684	3.033	0,23	18.493	36.528
CyPy_SPF_22	757	2.776	0,21	23.431	37.005
CyPy_SPF_23	668	1.467	0,11	7.402	17.255
CyPy_SPF_24	770	1.088	0,08	9.425	14.745
CyPy_SPF_25	685	778	0,06	4.664	9.384

Table 1. SPF-family transcript expression in the dorsal gland of *Cynops pyrrhogaster*, ranked according to reads per kilobase per million (RPKM) values. Transcript 3 and 5 are sodefrin transcripts, i.e. transcripts that contain an insertion leading to a frame shift from which a peptide pheromone can be cleaved, while all remaining are SPF transcripts.

in the dorsal gland, respectively. Our analyses therefore indicate that during the breeding period, SPF transcripts comprise the majority (84.8%) of expression of the SPF-family in the male dorsal gland in *Cynops pyrrhogaster*, indicating that sodefrin transcripts are expressed against this strong background.

Proteome analyses show high congruence with transcriptome expression data. To verify that the RNA-Seq expression data reflect the SPF proteins that are effectively secreted during courtship, we isolated and purified molecules directly from water in which male newts had been courting a female. To account for individual variation, we pooled water of 10 courting couples in which at least five minutes of male tail-fanning occurred. The RP-HPLC chromatogram revealed the presence of several proteins (Fig. 2a) that are absent in water of non-courting couples (see Supplementary Fig. S1). SDS-polyacrylamide gel electrophoresis of the fractions of interest showed the presence of multiple proteins around 20–25 kDa (Fig. 2b), and subsequent Edman degradation of the proteins in these fractions confirmed that they are indeed members of the SPF family (Table S2).

The N-terminal sequence already indicates that proteins eluting around minutes 69–81 belong to the beta SPF group, while proteins present in the two peaks around minutes 85–100 belong to the alpha SPF group (Fig. 2a). Mass spectrometry measurements indicate that these proteins are glycosylated (not shown), which is similar to SPF's in palmate newts²¹, and in line with our *Cynops* precursors showing several potential N-glycosylation sites. To link the proteins from courtship water to our transcripts from the dorsal gland, we aligned a data set combining the amino acid sequences obtained by Edman sequencing (10 to 30 amino acids, see Supplementary Table S2) with our translated transcripts, and estimated the pairwise similarities. The distance tree shows that the Edman sequences correspond to the transcripts that were identified as the most abundant in our RNA expression analyses (Table 1 and Fig. 3).

Since earlier studies (e.g.)⁶, have intensively focussed on recovering cleaved fragments of sodefrin precursors (i.e. sodefrin and peptide variants), our analyses particularly focussed on identifying full-length SPF proteins from courtship water. However, it is noteworthy mentioning that we retrieved a protein (<10 kDa) from courtship water (see Supplementary Table S2, fractions 71–73), which was cleaved from a sodefrin precursor at a lysine-arginine cleavage site. This indicates that, in addition to the decapeptide, other fragments of sodefrin precursors also end up in the water and have the potential of being sensed by a female. Altogether, our data show a high congruence between transcriptome and proteome, indicating that RNA-Seq expression data are a reliable representation of SPF proteins that are effectively tail-fanned from a male to a female during courtship.

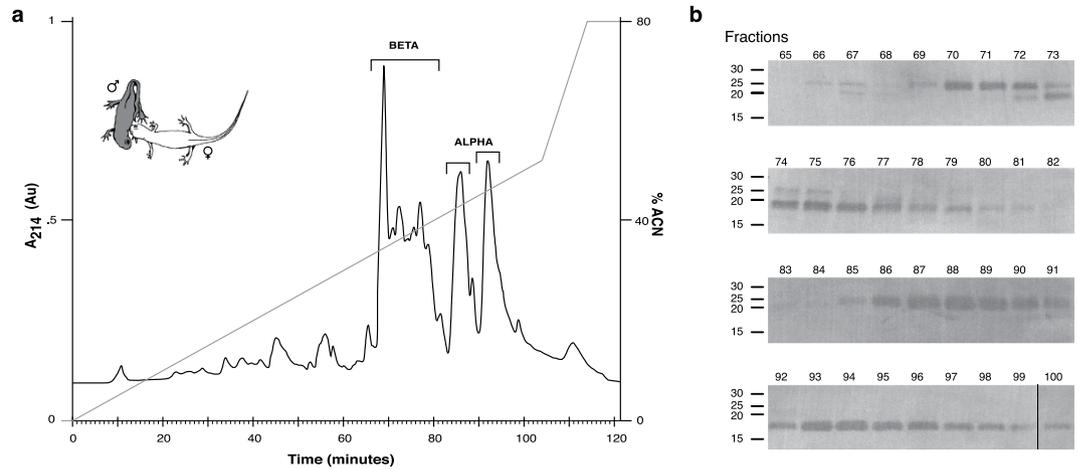


Figure 2. Isolation and identification of SPF proteins secreted during courtship. (a) RP-HPLC elution profile of water in which *C. pyrrhogaster* had been courting (drawing top left by Kim Roelants) and (b) Coomassie stained blot indicating the presence of multiple SPF proteins in courtship water. Molecular mass markers (in kDa) are indicated on the left and fraction numbers (1 fraction/minute) on top of the blots.

Both alpha and beta SPF's were expressed at the origin of sodefrin. To investigate the phylogenetic position of SPF-family transcripts of *Cynops* in relation to other known SPF's, we combined our translated transcripts (full coding sequence) with data from Ambystomatidae^{24,25}, Plethodontidae²⁶ and other Salamandridae^{21,22}. Alignment resulted in a data matrix of 95 OTUs and 225 characters, for which ProtTest²⁷ assigned JTT + G as the best fitting model under the AIC criterion. Subsequent Maximum Likelihood and Bayesian analyses (see Supplementary Fig. S2) in combination with Notung analyses resulted in a gene tree that is highly congruent with previously published relationships among SPF precursor sequences^{21,22,28} (Fig. 4). The earliest split is formed by the divergence of two supported clades corresponding to alpha and beta SPF protein precursors, whose divergence has been estimated around the Late Paleozoic²¹.

Our tree and associated support values show that currently expressed *C. pyrrhogaster* SPF-family precursors belong to seven independent lineages (Fig. 4: A–G). Six of these lineages exclusively contain SPF precursors (Fig. 4a; A–C, E–G), while a single clade (Fig. 4a; D) combines SPF and sodefrin precursors. Since grouping of any of these lineages would require violating strongly supported nodes in our tree, our analyses indicate that all of these lineages must have originated through gene duplications before the divergence of Asian salamanders (including *Cynops*) from European salamanders (including *Lissotriton* and *Ichthyosaura*)²⁹. Because the origin of sodefrin happened in an ancestor of *Cynops* after the divergence of these clades²², this indicates that these seven lineages of SPF proteins were existent in *Cynops* at the time of sodefrin origin.

Four lineages currently make up about 95% of the expression of the SPF pheromone family (Fig. 4, A–D). First, the closely related precursors in clade A, together accounting for 44.3% of the expression, belong to a strongly supported alpha clade of SPF proteins (Fig. 4a). These proteins fall in a clade that contains a highly expressed *Lissotriton helveticus* SPF (LiHe 003; see Table S3 and Fig. 4, indicated with *) for which behavioural assays have shown a courtship pheromone function²¹. Furthermore, salamandrid SPF precursors in this clade are orthologous with the main SPF courtship pheromones from the mental gland of plethodontids (Fig. 4, indicated with #)²⁶. Altogether, this suggests that alpha SPF's have continued to play an important role in the courtship behaviour throughout the evolution of internally fertilizing salamanders. Second, the beta-SPF lineages B, C and D account for more than half (50.3%) of the cumulative expression in our RNA-Seq data (Fig. 4), with sodefrin precursors making up the majority of expression in group D (15.2%) (Fig. 4b). In contrast to the alpha precursors, *Lissotriton* orthologs are either absent (lineage B) or only show low expression (lineage C: LiHe 009, 023, 012, 018, 013 lineage D: LiHe 021 & 032) (Fig. 4a; Table S3), while we found no *C. pyrrhogaster* orthologs in the clade that shows the highest expression in *Lissotriton* (Fig. 4a, LiHe 001, 002 and 003, clade of Pleurodelinae indicated with an arrow). Additional BLAST searches for *L. helveticus* orthologs in *de novo* assemblies of our *Cynops* RNA-Seq data only resulted in hits with other SPF precursor clades, confirming that fire belly newts currently make no use of these SPF proteins in courtship. Since the use of multiple individuals for our protein analyses significantly reduces the possibility that the observed expression data are biased by individual variation, our joint analyses suggest that expression shifts between different beta SPF clades have occurred after the divergence of European and Asian newts. In addition to sequence evolution, expression shifts are an alternative way of obtaining species-specificity of pheromone blends in multi-component peptide and protein pheromone systems³⁰. Since no expression data are available for newt species that diverged before the *Cynops-Lissotriton* split, our data currently do not allow estimating individual transcript abundance in their common ancestor. However, they do support that beta protein pheromones were expressed in the ancestor of *Cynops*, and thus were also present at the timing of the origin of sodefrin. Our combined results thus indicate that *Cynops pyrrhogaster* already effectively secreted both alpha and beta proteins before the origin of sodefrin.

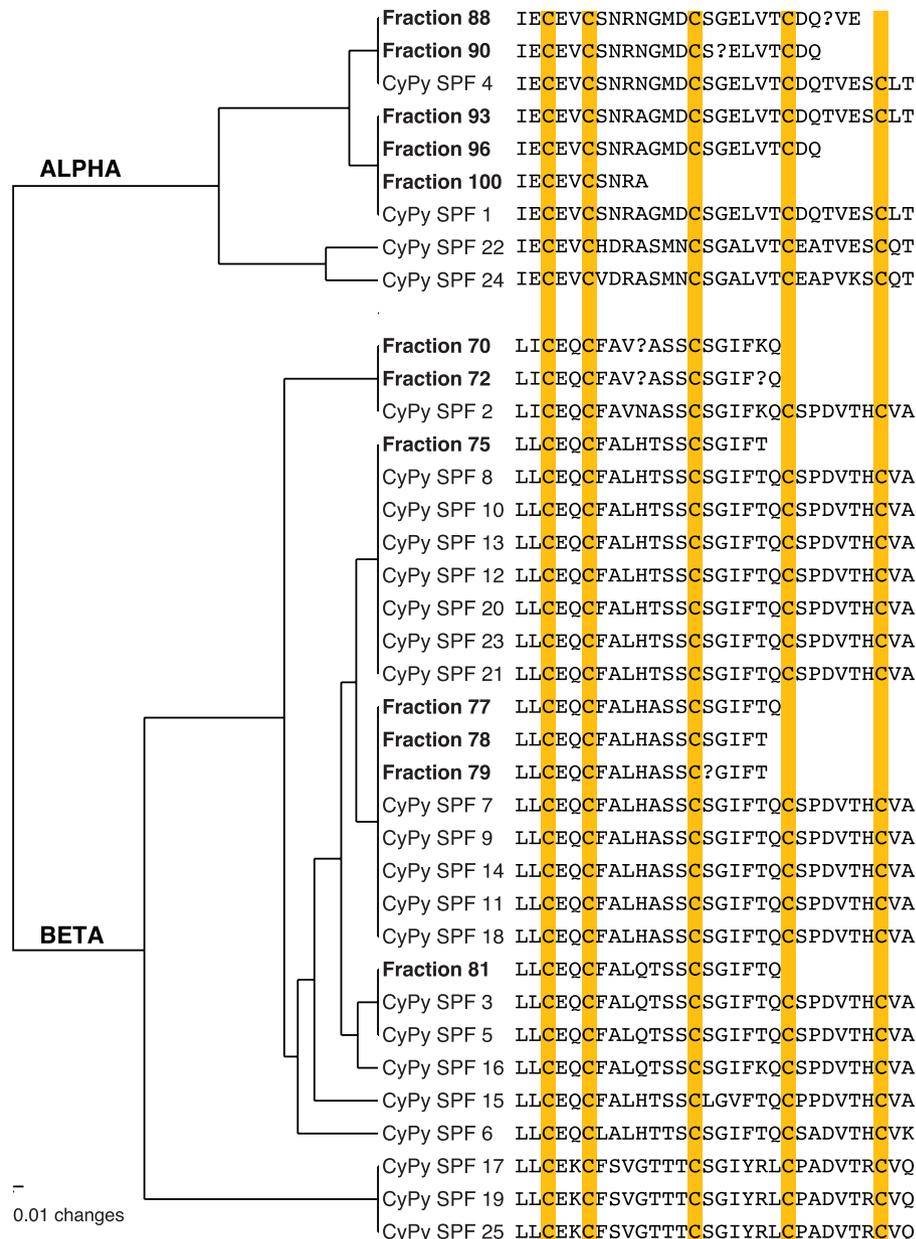


Figure 3. High congruence between transcriptome and proteome expression data. CyPy 1–25 are the SPF-family transcripts found in the dorsal gland of *C. pyrrhogaster*. Transcript names correspond to those in Table 1. Fraction numbers correspond to Fig. 2 and are the SPF proteins that were isolated and Edman sequenced from courtship water. The UPGMA tree with uncorrected distances illustrates a high congruence between expression of SPF transcripts in the dorsal gland and the effectively tail-fanned proteins i.e. the highest expressed SPF transcripts are also retrieved as proteins from courtship water.

Conclusion

Despite the fact that sodefrin in the Japanese fire belly newt has served as a model for newt pheromone research since its discovery, the alongside secretion of multiple uncleaved SPF proteins has remained undetected for a long time. Our combined transcriptomic, proteomic and phylogenetic approaches show that this species secretes multiple uncleaved SPF proteins that make up the majority (about 85%) of expression in the SPF pheromone family, and that ancestral forms of both alpha and beta SPF's were already expressed at the time of sodefrin origin. Because these SPF's likely continued to perform their pheromone function, the newly originated decapeptide sodefrin could be preserved, even without immediately acquiring a biological function^{22,31}. Both the place and timing of expression were identical, and the association with functional SPF pheromones likely predetermined sodefrin to develop into a sex pheromone²².

Sodefrin and its variants have been shown to be effective as attractants, but since these molecules are specifically released at close distance during tail-fanning, they are likely to function as courtship pheromones in nature³². Furthermore, several other fragments ending in the water after cleavage from sodefrin precursors,

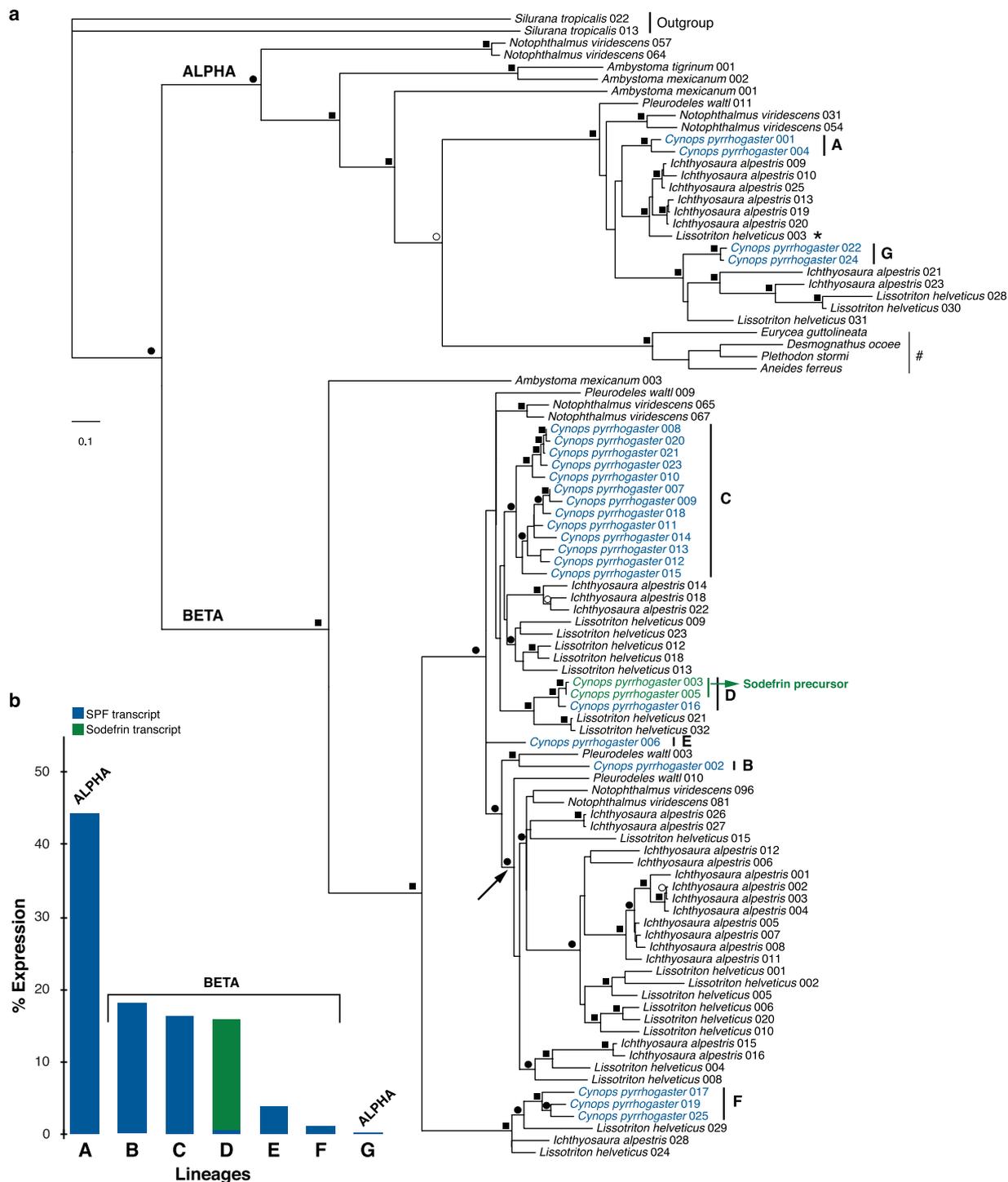


Figure 4. Evolution of precursors in the SPF pheromone family. (a) Maximum likelihood tree of translated SPF and sodefrin transcripts after Notung analyses. Squares on the branches indicate Bayesian posterior probabilities (BPP) equal to or higher than 0.95 in combination with ML bootstrap support equal to or higher than 70. Black and white circles indicate support for BPP and ML alone, respectively. (b) Comparison of expression (percentage) in the dorsal gland of *C. pyrrhogaster* for seven transcript lineages in the SPF pheromone family.

some of which still showing homology with SPF's, have not been tested in behavioural assays that would allow demonstrating a courtship pheromone role. Such tests will be indispensable for a complete understanding of how this multi-component courtship pheromone system originated and evolved.

Methods

Ethics statement. *C. pyrrhogaster* individuals (N = 21) were purchased from a local pet shop (Squama, Herent, Belgium). One male was first anesthetized by immersion in 0.5 g/L buffered MS-222 (Sigma-Aldrich) and then euthanized by decapitation and pithing of the brain and spinal canal. This procedure does not violate any European convention (European Convention for the protection of Vertebrate animals used for experimental and other scientific purposes; CETS #123), Belgian law (Art. 2.6 of the Belgian Law of May 4th 1995), or institutional regulation. This research was approved by the Ethical Committee for Animal Experiments of the Vrije Universiteit Brussel (Project number 14-220-35) and carried out in accordance with these approved guidelines.

Transcriptome analyses. *RNA sequencing and transcript expression.* For RNA sequencing (RNA-Seq), the part of the dorsal gland extending into the pleuroperitoneal coelom cavity (also known as abdominal gland) was sampled from a *Cynops pyrrhogaster* male in breeding mood. Total RNA was extracted using TRI Reagent (Sigma-Aldrich) and the RNAeasy mini kit (Qiagen). A pair-end cDNA sequencing library (PE50) was created by Baseclear (Leiden, The Netherlands) with Illumina TruSeq RNA Library Preparation Kit and fragments were sequenced on an Illumina HiSeq 2500 instrument. FastQ reads were generated after analyses with Illumina Casava pipeline (version 1.8.3), a post-filtering script (Baseclear) and FASTQC quality control tool (version 0.10.0) to remove low quality, PhiX-control and adapter reads. *De novo* transcriptome assembly was performed with Trinity³³ and CLC Genomics Workbench 6.0.4. Transcript expression levels were estimated by mapping reads to the transcripts obtained by RACE PCR, on the RNA-Seq module of the CLC Genomics Workbench 6.0.4, using a similarity fraction of 0.99 and length fraction of 1.00. To compare expression of *Cynops pyrrhogaster* with that of *Lissotriton helveticus*, we re-analysed the full-length transcripts from the dorsal gland of the latter species (data from)²¹ by mapping the RNAseq reads on the transcripts under the same conditions.

Rapid amplification of cDNA ends. We used the SMARTer-RACE cDNA amplification kit (Clontech) to create RACE cDNA by reverse transcription of 1 µg total RNA extracted from the dorsal gland tissue. To amplify the full coding sequence of the most abundant SPF isoforms, primers were designed on the 3' untranslated region of the highly expressed SPF isoforms obtained from the RNA sequencing procedure (see above). In addition, eleven degenerate primers from a previous study²¹ were used to obtain a large diversity of SPF protein sequences (Table S1). PCR products were amplified with FastStart High Fidelity Taq DNA polymerase (Roche) using a wide range of annealing temperatures. The following PCR conditions were used: one initial denaturation for 240 s at 94 °C, followed by 36 cycles with denaturation for 40 s at 94 °C, annealing for 60 s at different temperatures, and elongation for 60 s at 72 °C. To clone these amplification products, we used a pGEM-T Easy cloning vector (Promega). Vectors were transformed into TOP10 Competent Cells (Invitrogen). Colonies were picked randomly and inserts were amplified with Faststart Taq DNA polymerase using the same above-mentioned conditions. Amplification products (96 in total) were purified and sequenced, and sequence editing was performed with CodonCode Aligner version 3.7.1.1. Contigs were constructed with 99% identical bases and translated into amino acid sequences.

Proteome analyses. *Collection of proteins from courtship water.* Pheromone collection was done by placing a male and female for 1 h in a plastic container (25 × 16 × 14 cm) filled with 600 ml of ultrapure water (Milli Q, Millipore, Billerica, MA). During sampling, couples were monitored for courtship behaviour and the amount of time a male fanned his tail was measured. We retained water of 10 courting couples in which at least five minutes of male tail-fanning occurred and in which the female followed the courting male, indicating the presence of pheromones (henceforth termed courtship water). Pheromones were extracted by applying courtship water of a single couple onto two separate solid phase extraction cartridges (RP-C8 and RP-C18 Sep-Pak plus cartridge, 400 mg sorbent, Waters, Milford, MA, USA) using a vacuum pump. As a negative control, this procedure was repeated for 10 non-courting couples that had spent the same amount of time in water as the 10 courting couples. Proteins were eluted from the cartridges with 90% (v/v) acetonitrile containing 0.1% (v/v) trifluoroacetic acid (TFA). All acetonitrile was evaporated using a SpeedVac concentrator (SCV-100 H, Savant instruments, Farmingdale, NY). After concentration, samples were pooled and subjected to reversed-phase high-performance liquid chromatography (RP-HPLC) for additional purification.

Purification of proteins from courtship water using RP-HPLC. Peptides and proteins were purified using RP-HPLC. After collecting the molecules, pooled and concentrated samples (see above) were loaded onto a Source 5RPC column (4.6 × 150 mm, GE Healthcare Life Sciences, Uppsala, Sweden) pre-equilibrated with 0.1% (v/v) TFA (A). After loading, the column was washed for 10 minutes at a constant flow rate of 1 ml/min using the same solvent. Proteins were eluted with 80% acetonitrile in 0.1% TFA (B) by applying following gradient: from 0–65% B in 104 minutes and from 65–100% B in 10 minutes, at 1 ml/min. Detection of eluting proteins was performed at a wavelength of 214 nm and the eluate was collected in fractions of 1 ml. Fractions of interest were subjected to non-reducing SDS-PAGE using precast gels (Any kD Mini-PROTEAN TGX, Biorad, Hercules, CA, USA). Proteins were visualized by silver staining (Silverquest Silver Staining kit, Invitrogen, Carlsbad, CA, USA, data not shown).

Amino acid sequencing. After SDS-PAGE, proteins were transferred from the gel onto a PVDF membrane by semi-dry blotting (Trans Blot Turbo System, Bio-Rad) and stained with 0.1% Coomassie Brilliant Blue R-250 (Sigma, St. Louis, MO, USA; membrane). Protein bands of interest were excised from the blot for N-terminal sequencing on a 491 Procise cLC protein sequencer (Applied Biosystems, Foster City, CA, USA).

Phylogenetic reconstructions. We combined translated amino acid sequences of the 25 SPF-family transcripts from *C. pyrrhogaster* with translated SPF transcripts from four plethodontid, four ambystomatid and 60 other salamandrid sequences from Genbank (Table S4). Two frog Phospholipase A2 Inhibitor (PLI) sequences were chosen as outgroup (Table S4)²². Amino acid sequences were aligned using standard automatic parameters implemented in Mafft v7³⁴. ProtTest 2.4 was used to select the best fitting model of amino acid (AA) replacement for this data set according to an Akaike Information Criterion²⁷. Phylogenetic relationships were estimated under maximum likelihood (ML) with PAUP* under a JTT + G model (as assigned by ProtTest), and in a Bayesian framework using MrBayes 3.2.2^{35,36}. Bayesian analyses were conducted using a mixed prior for the AA substitution model and gamma correction for among-site rate heterogeneity. Two runs of four Markov chain Monte Carlo (MCMC) chains each were executed in parallel for 10,000,000 generations, with a sampling interval of 1,000 generations. Convergence of the parallel runs was confirmed by split frequency standard deviations (<0.01) and potential scale reduction factors (approximating 1.0) for all model parameters, as reported by MrBayes. Adequate posterior sampling was verified using Tracer 1.6³⁷, by checking if the runs had reached effective sampling sizes > 200 for all model parameters. A Bayesian consensus phylogram and Bayesian posterior probabilities (BPP) were inferred from the last 5,000 sampled trees of both runs. Clade support under ML was assessed by 1,000 replicates of rapid bootstrapping using RAXML 7.0.4³⁸ on the CIPRES Science Gateway v3.3³⁹. Finally, while constraining clades that were supported by BPP (≥ 0.95) or high bootstrap values (≥ 70), speciation-duplication analyses in Notung v2.6⁴⁰ were used to find the phylogenetic tree that minimizes the number of gene duplications and losses, and branch lengths were re-estimated from that tree.

References

- Kikuyama, S. *et al.* Sodefrin: a female-attracting peptide pheromone in newt cloacal glands. *Science*. **267**, 1643–1645 (1995).
- Yamamoto, K., Toyoda, F., Tanaka, S., Hayashi, H. & Kikuyama, S. Radioimmunoassay of a newt sex pheromone, sodefrin, and the influence of hormones on its level in the abdominal gland. *Gen. Comp. Endocr.* **104**, 356–363 (1996).
- Iwata, T. *et al.* Molecular cloning of newt sex pheromone precursor cDNAs: evidence for the existence of species-specific forms of pheromones. *FEBS Letter*. **457**, 400–404 (1999).
- Kikuyama, S. & Toyoda, F. Sodefrin: a novel sex pheromone in a newt. *Rev. Reprod.* **4**, 1–4 (1999).
- Toyoda, F. & Kikuyama, S. Hormonal influence on the olfactory response to a female-attracting pheromone, sodefrin, in the newt, *Cynops pyrrhogaster*. *Comp. Biochem. Physiol. B*. **126**, 239–245 (2000).
- Iwata, T. *et al.* Processing of multiple forms of preprosodefrin in the abdominal gland of the red-bellied newt *Cynops pyrrhogaster*: regional and individual differences in preprosodefrin gene expression. *Peptides*. **25**, 1537–1543 (2004).
- Nakada, T. *et al.* Isolation, characterization and bioactivity of a region-specific pheromone, [Val⁸]sodefrin from the newt *Cynops pyrrhogaster*. *Peptides*. **28**, 774–780 (2007).
- Nakada, T. *et al.* Evidence for processing enzymes in the abdominal gland of the newt, *Cynops pyrrhogaster*, that generate sodefrin from its biosynthetic precursor. *Zoo. Sci.* **24**, 521–524 (2007).
- Iwata, T. *et al.* Responsiveness of vomeronasal cells to a newt peptide pheromone, sodefrin as monitored by changes of intracellular calcium concentrations. *Peptides*. **45**, 15–21 (2013).
- Nakada, T. *et al.* Expression of G proteins in the olfactory receptor neurons of the newt *Cynops pyrrhogaster*: their unique projection into the olfactory bulbs. *J. Comp. Neurol.* **522**, 3501–3519 (2014).
- Wilburn, D. B. & Swanson, W. J. From molecules to mating: Rapid evolution and biochemical studies of reproductive proteins. *J. Proteomics*. doi: 10.1016/j.jprot.2015.06.007, (2015).
- Yamamoto, K. *et al.* Silefrin, a sodefrin-like pheromone in the abdominal gland of the sword-tailed newt, *Cynops ensicauda*. *FEBS Lett.* **472**, 267–270 (2000).
- Houck, L. *et al.* A new vertebrate courtship pheromone, PMF, affects female receptivity in a terrestrial salamander. *Anim. Behav.* **73**, 315–320 (2007).
- Palmer, C. A. *et al.* Plethodontid modulating factor, a hypervariable salamander courtship pheromone in the three-finger protein superfamily. *FEBS J.* **274**, 2300–2310 (2007).
- Palmer, C. A., Watts, R. A., Hastings, A. P., Houck, L. D. & Arnold, S. J. Rapid evolution of plethodontid modulating factor, a hypervariable salamander courtship pheromone, is driven by positive selection. *J. Mol. Evol.* **70**, 427–440 (2010).
- Wilburn, D. B. *et al.* Proteomic and UTR analyses of a rapidly evolving hypervariable family of vertebrate pheromones. *Evolution*. **66**, 2227–2239 (2012).
- Wilburn, D. B. *et al.* Structural insights into the evolution of a sexy protein: novel topology and restricted backbone flexibility in a hypervariable pheromone from the red-legged salamander, *Plethodon shermani*. *PLoS ONE* **9**, e96975 (2014).
- Rollmann, S. M., Houck, L. & Feldhoff, R. C. Proteinaceous pheromone affecting female receptivity in a terrestrial salamander. *Science*. **285**, 1907–1909 (1999).
- Houck, L. D. *et al.* A recombinant courtship pheromone affects sexual receptivity in a plethodontid salamander. *Chem. Senses*. **33**, 623–631 (2008).
- Houck, L. D. *et al.* A candidate vertebrate pheromone, SPF, increases female receptivity in a salamander in *Chemical signals in vertebrates 11* (eds Hurst J. L., Beynon R. J., Roberts S. C. & Wyatt T. D.) 213–221 (Springer, New York, 2008).
- Van Bocxlaer, I. *et al.* Side-by-side secretion of late Palaeozoic diverged courtship pheromones in an aquatic salamander. *P. Roy. Soc. B-Biol. Sci.* **282**, 20142960 (2015).
- Janssenswillen, S. *et al.* Origin and diversification of a salamander sex pheromone system. *Mol. Biol. Evol.* **32**, 472–480 (2015).
- Maki, N. *et al.* Expression profiles during dedifferentiation in newt lens regeneration revealed by expressed sequence tags. *Mol. Vis.* **16**, 72–78 (2010).
- Putta, S. *et al.* From biomedicine to natural history research: EST resources for ambystomatid salamanders. *BMC Genomics*. **5**, 54 (2004).
- Sal-site: the Ambystoma gene and EST database. Ambystoma mexicanum version 4.0. (grant no. R24OD010435)* [database on the Internet], Available at: www.ambystoma.org. (Accessed: June 2014).
- Palmer, C. A., Watts, R. A., Houck, L., Picard, A. L. & Arnold, S. J. Evolutionary replacement of components in a salamander pheromone signalling complex: more evidence for phenotypic-molecular decoupling. *Evolution*. **61**, 202–215 (2007).
- Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*. **21**, 2104–2105 (2005).
- Janssenswillen, S. *et al.* High pheromone diversity in the male cheek gland of the red-spotted newt *Notophthalmus viridescens* (Salamandridae). *BMC Evol. Biol.* **15**, 54 (2015).
- Zhang, P. & Wake, D. B. Higher-level salamander relationships and divergence dates inferred from complete mitochondrial genomes. *Mol. Phylogenet. Evol.* **53**, 492–508 (2009).
- Wyatt, T. D. Proteins and peptides as pheromone signals and chemical signatures. *Anim. Behav.* **97**, 273–280 (2014).

31. Raes, J. & Van de Peer, Y. Functional divergence of proteins through frameshift mutations. *Trends Genet.* **21**, 428–431 (2005).
32. Woodley, S. K. Chemical Signaling in Amphibians In *Neurobiology of Chemical Communication* (ed. Mucignat-Caretta C.) Ch. 8, 255–285 (CRC Press, Boca Raton, FL, 2014).
33. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
34. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
35. Swofford, D. L. PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. (Sinauer Associates, Sunderland, Massachusetts, 1998).
36. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
37. Rambaut, A., Suchard, M., Xie, D. & Drummond, A. Tracer v1.6. Available at <http://beast.bio.ed.ac.uk/Tracer> (Accessed: 24th September 2015), (2014).
38. Stamatakis, A., Hoover, P. & Rougemont, J. A rapid bootstrap algorithm for the RAxML Web Servers. *Syst. Biol.* **57**, 758–771 (2008).
39. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE)*, 1–8 (New Orleans, LA 14 November 2010).
40. Chen, K., Durand, D. & Farach-Colton, M. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J. Comput. Biol.* **7**, 429–447 (2000).

Acknowledgements

This research was supported by a European Research Council starting grant (ERC 204509, project TAPAS), the Fonds voor Wetenschappelijk Onderzoek (FWO) Vlaanderen (grant no. G.0133.08 and G.026715N) and the Strategic Research Program program of the Vrije Universiteit Brussel (SRP30). IVB is supported by a post-doctoral fellowship from FWO-Vlaanderen. MM is supported by a doctoral fellowship from FWO-Vlaanderen. The Hercules foundation of the Flemish government provided funding to PP to purchase LC-MS/MS equipment (contract AKUL/11/31). We thank Kim Roelants for the *Cynops pyrrhogaster* drawing.

Author Contributions

All authors conceived and designed the experiments; I.V.B., M.M., S.J.W. and W.V. performed the transcriptome experiments; I.V.B., M.M., D.T., S.J.W., R.J. and P.P. performed the proteome experiments; I.V.B., W.V. and F.B. analysed the transcriptome data; I.V.B., M.M., D.T., R.J. and P.P. analysed the proteome data; I.V.B. and F.B. wrote the paper; I.V.B. prepared figures and tables and I.V.B., P.P. and F.B. contributed reagents, materials and analysis tools. All authors reviewed drafts of the paper.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Van Bocxlaer, I. *et al.* Beyond sodefrin: evidence for a multi-component pheromone system in the model newt *Cynops pyrrhogaster* (Salamandridae). *Sci. Rep.* **6**, 21880; doi: 10.1038/srep21880 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>