Ancient Pheromone Blend as an Alternative for Copulation in Internally Fertilizing Salamanders

Ines Van Bocxlaer[1*], Dag Treer[1*], Margo Maex[1*], Wim Vandebergh[1], Sunita Janssenswillen[1], Gwij Stegen[1], Philippe Kok[1], Bert Willaert[1], Severine Matthijs[1], Erik Martens[2], Anneleen Mortier[3], Henri de Greve[4,5], Paul Proost[3] and Franky Bossuyt[1]

1 Amphibian Evolution Lab, Biology Department, Vrije Universiteit Brussel (VUB), Pleinlaan 2, B-1050 Brussels, Belgium

2 Laboratory of Immunobiology, Department of Microbiology and Immunology, Rega Institute, Katholieke Universiteit Leuven (K.U. Leuven), Minderbroedersstraat 10 - box 1030, B-3000 Leuven, Belgium

3 Laboratory of Molecular Immunology, Department of Microbiology and Immunology, Rega Institute, Katholieke Universiteit Leuven (K.U. Leuven), Minderbroedersstraat 10 - box 1030, B-3000 Leuven, Belgium

4 Structural and Molecular Microbiology, VIB Department of Structural Biology, VIB, Pleinlaan 2, 1050 Brussels, Belgium.

5 Structural Biology Brussels, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium.

* These authors contributed equally to this work

Keywords: evolution, phylogeny, gene duplications, amphibians, protein pheromones

Author for Correspondence: Franky Bossuyt; phone: +32-2-6293648 e-mail: fbossuyt@vub.ac.be

Throughout the animal kingdom, internal fertilization - the merging of sperm and egg inside the female body - nearly invariably relies on the use of a copulatory organ. In contrast, males of advanced salamanders (Salamandroidea) attain internal fertilization by depositing a spermatophore on the substrate in the environment, which females subsequently take up with their cloaca. The aquatically reproducing modern Eurasian newts (Salamandridae) have taken this to extremes, since the majority does not display physical contact between the sexes and largely rely on females following the male track at spermatophore deposition. Although the use of pheromones has been widely assumed during their courtship, molecules able to induce the female following behaviour that culminates in insemination have not been identified. Here we show that uncleaved glycosylated SPF protein pheromones, secreted during courtship, are sufficient to elicit such behaviour in palmate newts (*Lissotriton h. helveticus*), indicating that these molecules obviate the need for copulation in these salamanders. Surprisingly, our finding of side-by-side secretion of Late Palaeozoic diverged proteins in a single species suggests that these molecules already had a courtship function in stem salamanders about 300 million years ago, rendering them one of the oldest vertebrate pheromone systems.

1. Introduction

Internal fertilization (i.e., the merging of sperm and egg inside the female body) is a widespread reproductive mode that is generally accomplished through copulation, i.e. the insertion of a copulatory organ into the female sex organ [1]. In contrast, males of most advanced salamanders (Salamandroidea, making up about 90 % of the more than 650 species of extant salamanders) reproduce by internal fertilization, but deposit a sperm package (spermatophore) on the substrate in the environment which females subsequently take up with their cloaca. In most families, an

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.457v1 | CC-BY 4.0 Open Access | received: 11 Aug 2014, published: 11 Aug

enhanced success rate of insemination is accomplished through contact, such as a coordinated tail-straddling walk, or amplexus in which the male sometimes drags the female over the spermatophore [2,3]. However, some male salamanders have abandoned close physical contact altogether and instead largely rely on tail-fanning courtship pheromones to the female [2,4,5]. These pheromones induce following behaviour, so that the female passes over the spermatophore and picks it up with her cloaca (Figure 1A, movie S1) [6].

In Caudata, the use of protein pheromones during male courtship rituals is known from terrestrial plethodontid salamanders, and a decapeptide attractant in Asian newts has been intensively studied [7-14]. However, no studies are available that have characterized pheromones that directly affect the female sexual following behaviour that is crucial for attaining insemination in aquatically reproducing newts. Here we purified courtship proteins that are tail-fanned by palmate newts (*Lissotriton helveticus*, Salamandridae) from water, experimentally tested them, and used transcriptomics and phylogenetics to estimate the age of the earliest divergence of present-day secreted proteins.

2. Material and methods

(a) Animals

The research was done with permission and according to the guidelines of Agentschap voor Natuur en Bos (permits ANB/BL-FF/V12-00050 and ANB/BL-FF/V13-00134). All experiments complied with EU and Belgian regulations concerning animal welfare. Animals were released back to the pond of their origin after the experiments were finished. We used 40 adult males and 40 females of each of the three species of newts (*Lissotriton helveticus*, *L. vulgaris* and *Ichthyosaura alpestris*). All species have an overlapping breeding season and were

collected in spring from ponds near Ternat, Belgium. The catching method and the housing conditions were the same to those described elsewhere [6].

(b) Behavioural experiments

Behavioural experiments were done in aged tap water with and without stimuli added. All experiments were performed on consecutive days, at the same time of the day, and under the same light and temperature conditions (see [6] for more details). To evaluate suitability of the animals to be used, all experiments were preceded by a receptivity test before the first experiment, and only receptive animals were selected [6]. For the species-specificity experiments, *L. vulgaris* and *I. alpestris* females were tested in parallel to confirm the potency of their courtship water used to test *L. helveticus* females. In all tests, female behaviour was recorded for twelve minutes, and the first two minutes of the experiment were discarded to allow acclimatization of the animals. Experiments were recorded using a digital camera connected directly to a computer, and the recordings were analysed for following and pointing. Following is female courtship behaviour similar to that under natural conditions with a courting male, where a female starts to closely follow his movements. Here we measured the cumulative amount of time in which at least one of the females incessantly shows interest towards the other one, including turning towards the other female and following her. More than 30 seconds of following was scored as positive. Pointing is a way to measure the change in behaviour without having to understand or observe specific types of female behaviour and is therefore more objective. Here we measured the cumulative amount of time (seconds) per couple that an imaginary straightforward line, perpendicular to the line connecting the eyes of the following female, intersects the other female's body. The differences in pointing between stimuli were tested with the Kruskal-Wallis test followed by post hoc two-tailed Mann-Whitney U test for pairwise comparisons [15]. The analyses were done using IBM SPSS Statistics for Windows [16].

(c) Collection and extraction of molecules from courtship water

Pheromone collection was done by placing a male and female for one hour in a plastic container (25x16x14 cm) filled with 600 ml of water (details are outlined in [6]). During sampling, couples were monitored for courtship behaviour and the amount of time a male fanned his tail was measured. For each condition, we sampled a minimum of 15 courting couples in which at least ten minutes of male tail-fanning occurred. Pheromones were extracted by applying non-courtship or courtship water of a single couple onto two separate solid phase extraction cartridges (300 ml per filter; RP-C8 and RP-C18 Sep-Pak plus cartridge, 400 mg sorbent, Waters, Milford, MA, USA) using a vacuum pump. Proteins were eluted from both cartridges with 7.5 ml of 90% (v/v) acetonitrile containing 0.1% (v/v) TFA. All acetonitrile was evaporated using a SpeedVac concentrator (SCV-100H, Savant instruments, Farmingdale, NY) for 1 h. After concentration, samples were pooled per condition and subjected to RP-HPLC.

(d) Purification of proteins

Peptides and proteins were partially separated using reversed-phase high-performance liquid chromatography (RP-HPLC). Pooled and concentrated samples were loaded onto a Source 5RPC column (4.6x150 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 a linear (from 0-100 % B in 80 minutes at 1 ml/min) or flattened gradient (30-65% B in 56 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).

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.457v1 | CC-BY 4.0 Open Access | received: 11 Aug 2014, published: 11 Aug

To further purify the candidate pheromones, HPLC fractions of interest were submitted to ion exchange chromatography. After evaporating the acetonitrile (SCV-100H, Savant instruments, Farmingdale, NY) samples were brought to pH of 7.5 by addition of buffer containing 20 mM bis-tris propane, 20 mM piperazine and 20 mM N-methyl piperazine (Sigma). Samples were loaded onto a 1 ml Hitrap DEAE Fast Flow (GE Healtcare Bio-sciences, flow rate 1ml/min) 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, Sigma) and washed for at least 10 minutes with the same buffer until all material in the effluent disappeared. Proteins were eluted with 15 mM bis-tris propane, 15 mM piperazine and 15 mM N-methyl piperazine (buffer B, pH 3, Sigma) by applying a linear gradient (from 0-100% B in 20 minutes). Detection of eluting proteins was performed at a wavelength of 280 nm and the eluate was collected in fractions of 1 ml. Purity of the fractions was assessed by mass spectrometry and non-reducing SDS-PAGE, using precast gels (Any kD Mini-PROTEAN TGX, Biorad, Hercules, CA, USA). After electrophoretic separation, proteins were visualized by silver staining (Silverquest Silver Staining kit, Invitrogen, Carlsbad, CA, USA).

(e) Mass spectrometry and amino acid sequence analyses

Mass analyses of the HPLC fractions were performed by electrospray ionization ion trap mass spectrometry on an ESQUIRE- LC MS (Bruker, Brussels, Belgium). In addition mass analyses of the desalted ion exchange fractions (Zip Tip C18, 10 µl, Millipore) were performed on an Amazon Speed ETD ion trap mass spectrometer. Characterization of the glycan moiety was done through in-source fragmentation on the Esquire ion trap mass spectrometer by gradually elevating the potential on skimmer 1 and the exit caps in the electrospray source. Peak fractions of courtship water collected during breeding season were subjected to a non-reducing SDS-PAGE using precast gels (Any kD Mini-PROTEAN TGX, BioRad, Hercules, CA, USA). After electrophoresis, proteins were transferred from the gel onto a PVDF membrane by semi-

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.457v1 | CC-BY 4.0 Open Access | received: 11 Aug 2014, published: 11 Aug

dry blotting (Trans Blot Turbo System, Bio-Rad) and stained with 0.1% Coomassie Brilliant Blue R-250 (Sigma, St. Louis, MO, USA; membrane not shown). All protein bands were excised from the blot for N-terminal sequencing on a 491 Procise cLC protein sequencer (Applied Biosystems, Foster City, CA, USA).

(f) Transcriptomics and gene expression estimates

Abdominal glands for RNA sequencing (RNA-seq) were sampled from a single male. Total RNA was extracted using TRI Reagent (Sigma-Aldrich) and the RNAeasy mini kit (Qiagen). Extracted RNA was sent to Baseclear (Leiden, The Netherlands) for RNA sequencing, de novo transcriptome assembly, and gene expression estimations. A pair-end cDNA sequencing library (PE50) was created with Illumina TruSeq RNA Library Preparation Kit and 52040842 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 Trinity [17] and transcript expression levels were estimated by mapping reads to the de novo assembled transcripts, on the RNA-seq module of the CLC Genomics Workbench (allowing two mismatches per read). Sodefrin precursor factor-like (SPF) sequences were identified through aligning assembled transcripts to a dataset containing SPF sequences from the Uniprot database using RAPsearch [18].

RACE (rapid amplification of cDNA ends) was performed to obtain complete protein sequences from different SPF precursors. Primers were designed on the 3'-untranslated region to amplify full-coding sequences of SPF transcripts as follows:

SPF_Primer_A, 5'-TTGTTAATAAWYATTCTGTAAAGARGCT-3'; SPF_Primer_B, 5'-GCCTTGTTGBCAAAAHKTCTTC-3'; SPF_Primer_C, 5'-ACAAYTWCTAAGCTGGHKTAG

GA-3'; SPF_Primer_D, 5'-GTGTGTATWTGRGGTATRAACAAAGGTC-3', SPF_Primer_E, 5'-CCAACAATTACTRRGMKGGAGTAGG-3'; SPF_Primer_F, 5'-CAACTACTAAGCTRRAGTM
RGAGTGC-3'; SPF_Primer_G, 5'-GGRTAGGATTGCGTCAGATGTT-3'; SPF_Primer_H, 5'-TAGGAATGTTTCTAYKGACKACTACTRAG-3'; SPF_Primer_I, 5'-CTATTGCTAAGCTG
KGGTG-3'; SPF_Primer_J, 5'-GCTGGCACATGGGCATGT-3'; SPF_Primer_K, 5'-GCCCAWA
CASKACTAAGCACATT-3'; SPF_Primer_L, 5'-GACTCTGVATTHCAGGTACTTGTAGAG-3'.

A total of 1 μg total RNA from the same extraction procedure as in RNA-seq was used to create RACE cDNA with the SMARTer-RACE cDNA amplification kit (Clontech). PCR products were amplified with FastStart High Fidelity Taq DNA polymerase (Roche). Amplification products were cloned into a pGEM-T Easy cloning vector (Promega) and vectors were transformed into TOP10 Competent Cells (Invitrogen). Colonies were picked randomly and inserts were amplified with Faststart Taq DNA polymerase. Amplification products were purified and sequenced by the VIB genetic service facility (Antwerp, Belgium). For comparison with protein masses found in courtship water, contiguous sequences (contigs) were assembled with CodonCode Aligner 3.7.1.1 (CodonCode Corporation) using a 99% similarity threshold, after quality trimming. Signal peptides, predicted using SignalP 4.0 [19], were removed and protein masses were calculated with the pI/Mw tool on Expasy (http://web.expasy.org/compute_pi/). Gene expression differences between SPF homologs were estimated using RNA-seq read counts (as described above) on the assembled homologs from our RACE procedure. Expression levels were determined using RPKM values [20].

(g) Phylogeny and divergence time estimates

We combined protein sequences of a representative set of 16 SPF precursors of *Lissotriton helveticus* found in this study with four plethodontid, three ambystomatid and six

other salamandrid sequences from Genbank. Two frog sequences were chosen as outgroup. Alignment of the protein sequences was done with MAFFT [21] using the L-INS-i method and resulted in a data matrix of 216 amino acids. Bayesian and likelihood analyses were performed using a LG amino-acid rate matrix. Maximum Likelihood (ML) analyses were run in PAUP [22], with empirical frequencies, estimated proportion of invariable sites (0.0527345) and distribution of rates at variable sites following a gamma distribution with four categories and estimated shape parameter 1.79548. This resulted in a single ML tree with likelihood score -Ln L = 8271.589. Bayesian analyses and Bayesian posterior probabilities were calculated in MrBayes [23]. Two runs of four Markov chain Monte Carlo (MCMC) chains each were executed in parallel for 5,000,000 generations, with a sampling interval of 500 generations and a burn-in corresponding to the first 1,000,000 generations. Speciation-duplication analyses were done using Notung [24]. To estimate the age of the earliest diversification in our SPF pheromones, we used a Bayesian relaxed molecular clock model implemented in Beast [25]. As a calibration point, we used the divergence of the (Salamandridae, Ambystomatidae) clade from Plethodontidae, a relationship that is widely accepted [26-30] and was also strongly supported in our ML tree. We used the mean (175.7 Mya) and standard deviation (14.8) of the last five studies [26-30] presented on Timetree [31] (version of 13 January 2014) to calibrate our tree with a central 95% range of 146.7-204.7 Mya. The sodefrin precursor sequence of *Cynops* was not included, because the end of the sequence (containing the sodefrin peptide) is not homologous with our full-length proteins.

3. Results and Discussion

We first optimized a behavioural assay in which female courtship responses can be measured in the absence of a male, thus giving experimental control over the application of candidate pheromones. A two-female experiment (modified from [6]) with palmate newts

(*Lissotriton h. helveticus*) was used in which females are exposed to male molecules emitted during tail-fanning (Figure 1a). Under natural conditions, females respond to male courtship pheromones by following the male closely for a prolonged period (Movie S1). The set-up of our two-female behavioural assay removes the secondary sexual morphological and chemical characteristics, as well as the visual cues of tail-fanning of a male, while retaining the required presence of another individual necessary to exhibit following behaviour. Using this assay, we first measured whether water in which a male had been tail-fanning (henceforth termed courtship water) was able to induce such following behaviour in females. As a first indication, we counted the number of couples in which a trained observer measured more than half a minute of following behaviour during ten minutes of observation (henceforth termed Nf). However, to have a fully objective way of measuring following behaviour, statistical comparisons (Kruskal-Wallis test followed by *post hoc* two-tailed Mann-Whitney U test) were performed with the time period that females faced each other during the experiment (which is largely caused by the following behaviour). Our behavioural assays indicate that palmate newt courtship water induced following of conspecific females (Figure 1b and see the electronic supplementary material, Table S1; Lh1: Nf = 9/12, compared to control 1: Nf = 1/11; $P < 0.01$). These tests confirm that courtship water is able to induce female courtship responses in palmate newts, even in the absence of the male secondary sexual characteristics and visual cues associated with tail-fanning. Females of palmate newts show a reduced response in courtship water of the congeneric species *L. vulgaris* (Figure 1b and Table S1; Lv: Nf = 4/12; $P = 0.065$), and no response in that of the more distantly related alpine newts (*Ichthyosaura alpestris*) (Figure 1b and Table S1; Ia: Nf = 0/12; $P < 0.01$), suggesting that tail-fanned courtship pheromones quickly evolve towards species-specificity.

We optimized a protocol for sampling proteins emitted during male tail-fanning directly from water and compared reversed-phase high-performance liquid chromatography (RP-HPLC) elution profiles of courtship water with that of water in which a non-courting male and female

had been held. These analyses show a recurrent pattern of elution profiles showing a peak that is present in water in which a male has been tail-fanning [>20 profiles, one shown; Figure 1c, courtship peak, CP (orange)], but absent in (*i*) water in which non-courting males and females were held for the same amount of time [5 profiles, one shown; Figure 1c, male-female water, MF (blue)] and (*ii*) courtship water sampled at the end of the breeding season (see the electronic supplementary material, figure S1). A behavioural assay with the fractions of the courtship peak indicated that they induced following (Figure 1b, CP: Nf = 9/10, compared to water, control 2: Nf = 1/12; $P < 0.05$) in a way that is not significantly different from that observed in courtship water (Lh2: Nf = 11/13, $P = 0.385$). Conversely, non-courting male-female water resulted in female reactions that were not significantly different from those in control water (MF: Nf = 2/11; control 2: Nf = 1/12; $P = 0.166$). These observations indicate that the RP-HPLC fractions of the courtship peak effectively contain the male courtship pheromones that induce female following behaviour. N-terminal amino acid sequencing (Edman sequencing) of these pooled fractions indicated the presence of multiple isoforms of the Sodefrin Precursor-like Factor (SPF) family (see the electronic supplementary material, table S2). These proteins were considered ideal pheromone candidates, because a full-length protein (i.e., not cleaved, except for the signal peptide) of this family identified from the mental gland of the terrestrially reproducing plethodontid salamander *Desmognathus ocoee* was shown to increase female receptivity [14].

We further characterized the diversity of SPF proteins by combining transcriptome analyses of the pheromone-producing abdominal gland of a single male with proteome analyses of RP-HPLC fractions of the courtship peak. Whole transcriptome sequencing (RNAseq) and de novo assembly of nearly 52 million (Mio) reads revealed 4.1 Mio reads (7.9 %) belonging to this SPF family of molecules. RACE-PCR sequencing revealed 32 different cDNA precursor sequences (GenBank numbers KJ402326 - KJ402357) encoding for 31 unique mature proteins. RNAseq expression analyses indicated five isoforms as most abundant, together making up 94.0% of the SPF transcripts identified in the transcriptome (Figure 2a, 2b). Interestingly, the

294 pairwise amino acid divergences between these sequences were between 19.2 and 78.8 %,
295 indicating that these proteins do not only result from allelic variation. To determine the presence
296 of post-translational modifications and to confirm that these precursors are also effectively
297 translated and tail-fanned to the female, we performed an RP-HPLC with a prolonged gradient
298 (Figure 2c) to obtain a better separation of the SPF proteins, and combined mass spectrometry
299 analyses and Edman sequencing of individual fractions. Mass spectrometry analyses indicated
300 the presence of an oligosaccharide with 2 N-acetylglucosamine units (GlcNAc) and multiple
301 hexoses (see the electronic supplementary material, Figure S2) attached to the available
302 glycosylation sites of the proteins. Individual RP-HPLC fractions revealed the presence of
303 multiple proteins for which the glycosylated masses (up to eight hexoses) match the theoretically
304 predicted masses derived from the cDNA precursors (see the electronic supplementary material,
305 table S3). Additionally, several of these predictions could be confirmed by N- terminal amino
306 acid sequencing (see the electronic supplementary material, table S2). This indicates that SPF is
307 effectively present as multiple uncleaved proteins with different levels of glycosylation
308 (glycoforms) in the courtship peak.

309 Next we performed ion exchange chromatography to purify SPF from the courtship peak
310 fractions. SDS-PAGE, mass spectrometry and Edman sequencing all indicated that this led to
311 removal of non-SPF as well as some of the SPF proteins, and resulted in a sample containing two
312 SPF proteins (SPF 1: Mr = 21036.8; SPF 3: Mr = 20326.9) with multiple glycoforms (Figure
313 2d). A two-female behavioural experiment with these proteins resulted in a significant increase
314 of the female following behaviour (Nf = 8/10) compared to control water (Nf = 0/11, $P < 0.01$)
315 (Figure 2d), and confirms that SPF proteins alone are able to induce female following, even in
316 the absence of visual stimuli of a courting male. Finally, we used the same techniques to purify a
317 single SPF isoform (SPF 3: Mr = 20326.9) with its glycoforms (Figure 2e). A behavioural test
318 with this protein induced following in half of the couples (Nf = 5/10, compared to none in
319 control water; $P < 0.05$) (Figure 2e). This experiment indicates that a single isoform is able to

elicit female courtship responses. Future investigations on the relative ability of various individual SPF proteins to induce female courtship responses in the palmate newt and related species could give important insights in the evolution of species-specificity of protein pheromones.

Phylogenetic analyses combining our palmate newt SPF cDNA precursors with available sequences [11] on Genbank confirm that SPF diversification goes beyond allelic variation by indicating multiple gene duplication events (Figure 3). Speciation-duplication analyses identified speciation events that conform to established higher-level phylogenetic relationships of salamanders (Figure 3a), but also recognized two duplications that occurred before the Plethodontidae-Salamandridae divergence (Figure 3b, nodes 1 and 2). The strongly supported relationship of SPF proteins from lungless salamanders (Plethodontidae) with a clade uniting an *Ambystoma* and our salamandrid SPF3 precursors (Figure 3b, indicated with an asterisk) reveals the orthologs corresponding to the Salamandridae-Plethodontidae divergence [15-20], and defines a split (Figure 3b, node 3) that had remained unidentified in previous studies [11, 12]. We used the mean and standard deviation (175.7 +/-14.8 Mya) of the last five studies presented on Timetree [26-31] (version of 13 January 2014) to calibrate this node with a central 95% range of 146.7-204.7 Mya and to estimate precursor divergence times with a Bayesian relaxed molecular clock model implemented in Beast [25]. Our results reveal a Late Palaeozoic duplication event (Figure 3b, node 1) that denotes the early onset of SPF diversification and protein secretion (the latter as indicated by our protein characterization from courtship water) at about 288.4 Mya (95% HPD = 200.6-385.1 mya) (see the electronic supplementary material, table S4). Our time estimates for salamander speciation nodes in the gene tree (Table S4) are close to the mean of the last five studies that estimated the divergence times of the Ambystomatidae-Salamandridae split (146.8 MYA +/- 33.1) and the onset of diversification of Plethodontidae (72.1 +/- 23.9) [25-30]. Additionally, the two nodes that represent the same speciation event in *Lissotriton* in our gene tree (Figure 3b, nodes 9 and 10) have similar age

346 estimates. All these results together strengthen confidence in our divergence time estimates,
347 including the duplication events of SPF genes. The timing of the earliest SPF divergence
348 therefore is close to the origin of stem salamanders (Anura-Caudata divergence, estimated at
349 295.5 +/- 21 Mya, Figure 3) [26] and considerably predates the currently known use of this
350 protein system (i.e., in crowngroup plethodontids, Figure 3, green circle and branches).

351 Our study not only characterizes the pheromones behind the intriguing female following
352 behaviour in salamandrid newts, but also expands the evidence for the use of uncleaved SPF
353 pheromones from a single family (Plethodontidae) to potentially all salamanders. Uncleaved SPF
354 proteins until now were shown to be functional as a pheromone in a single plethodontid species
355 *Desmognathus ocoee* (Figure 3b, green circle) [14]. Although an SPF-derived pheromone was
356 initially discovered in a salamandrid, the cleaved active decapeptide (sodefrin, an attractant in
357 *Cynops*) originated through a translational frame shift and as a consequence shows no homology
358 with uncleaved SPF protein pheromones. Additionally, the short peptide obtained its pheromone
359 function in the genus *Cynops*, and therefore independently from uncleaved proteins
360 (Janssenswillen et al., submitted). To our knowledge, our study of an aquatic salamandrid is the
361 first to expand the effective behavioural evidence for a courtship pheromone function of
362 uncleaved SPF proteins outside the family of plethodontids. Additionally, the side-by-side
363 secretion of anciently diverged proteins (Figure 3b, red circles) in our newt species suggests that
364 the courtship function for these proteins considerably predates the Salamandridae-Plethodontidae
365 divergence. Although cDNA studies in individual species already indicated the presence of
366 multiple isoforms [11,12], the known diversity of SPF precursors in each of these families
367 resulted from family-specific gene duplications and/or polymorphisms [11,12]. In contrast, our
368 palmate newts tail-fan proteins of which the estimated divergence dates back to the Late
369 Palaeozoic (Figure 3b, node 1) and our results therefore strongly suggest a pheromone function
370 for these molecules already in the earliest salamanders, about 300 mya.

371 Our combined evidence indicates that, although very different courtship behaviours can be

observed across the evolutionary tree of salamanders [3,10,14], the function of uncleaved SPF proteins to regulate female sexual receptiveness originated early in salamander evolution and has been conserved with various observable effects in multiple salamander lineages ever since. In palmate newts, and likely also in related species with female following behaviour, these pheromones obviate the copulatory organ by ensuring that eggs and sperm can merge in the safe environment of the female body.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.457v1 | CC-BY 4.0 Open Access | received: 11 Aug 2014, published: 11 Aug

References

1. Beckett BS. 1986 *Biology: a modern introduction* Oxford: Oxford University Press.
2. Halliday TR. 1990 *The Evolution of Courtship Behavior in Newts and Salamanders.* In Advances in the Study of Behavior, vol. 19, JSR Peter, JB Slater and B Colin, eds. San Diego: Academic Press Inc., pp. 137-169.
3. Houck LD, Arnold S. 2003. *Courtship and mating behavior. In Reproductive biology and phylogeny of Urodela*, DM Sever, ed. Enfield: Science Publishers, Inc., pp. 383-424.
4. Verrell P, Halliday T, and Griffiths M. 1986 The annual reproductive cycle of the smooth newt *Triturus vulgaris* in England. *J. Zool.* **210**, 101-119. (doi: 10.1111/j.1469-7998.1986.tb03623.x)
5. Sever DM, Verrell PA, Halliday TR, Griffiths M, 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.
6. Treer D, Van Bocxlaer I, Matthijs S, Du Four D, Janssenswillen S, Willaert B, Bossuyt F. 2013 Love is blind: indiscriminate female mating responses to male courtship pheromones in newts (Salamandridae). *PloS One* **8**, e56538. (doi: 10.1371/journal.pone.0056538)
7. Cummins SF, Bowie JH. 2012 Pheromones, attractants and other chemical cues of aquatic organisms and amphibians. *Nat. Prod. Rep*. **29**, 642-658. (doi: 10.1039/c2np00102k)
8. Kikuyama S, Toyoda F. 1999 Sodefrin: a novel sex pheromone in a newt. *Rev. Reprod.* **4**, 1-4. (doi: 10.1530/ror.0.0040001)
9. Woodley SK. 2010 Pheromonal communication in amphibians. *J. Comp. Physiol. A* **196**, 713-727. (doi: 10.1007/s00359-010-0540-6)
10. Kikuyama S, Toyoda F, Ohmiya Y, Matsuda K, Tanaka S, Hayashi H. 1995 Sodefrin: a female-attracting peptide pheromone in newt cloacal glands. *Science* **267**, 1643-1645. (DOI: 10.1126/science.7886452)
11. Osikowski A, Babik W, Grzmil P, Szymura JM. 2008. Multiple sex pheromone genes are expressed in the abdominal glands of the smooth newt *Lissotriton vulgaris* and montandon's newt *L. montandoni* Salamandridae. *Zool. Sci.* **25**, 587-592. (doi: 10.2108/zsj.25.587)
12. Palmer CA, Watts RA, Houck LD, Picard AL, Arnold SJ. 2007 Evolutionary replacement of components in a salamander pheromone signaling complex: more evidence for phenotypic-molecular decoupling. *Evolution* **61**, 202-215. (doi: 10.1111/j.1558-5646.2007.00017.x)
13. Iwata T, Conlon JM, Nakada T, Toyoda F, Yamamoto K, Kikuyama S. 2004 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. (doi: 10.1016/j.peptides.2003.10.027)

14. Houck LD, Watts RA, Mead LM, Palmer CA, Arnold SJ, Feldhoff PW, Feldhoff RC. 2008 *A candidate vertebrate pheromone, SPF, increases female receptivity in a salamander.* In Chemical Signals in Vertebrates 11, JL Hurst, RJ Beynon, SC Roberts and TD Wyatt, eds. New York: Springer, pp. 213-221.
15. Dytham C. 2011 *Choosing and Using Statistics: a Biologist's Guide*, Third Edition. Wiley-Blackwell, Chichester.
16. IBM SPSS Statistics for Windows Version 22.0. IBM Corp, Armonk, NY, 2013.
17. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, 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. Biotech.* **29**, 644-652. (doi: 10.1038/nbt.1883)
18. Ye Y, Choi J-H, Tang H 2011 RAPSearch: a fast protein similarity search tool for short reads. *BMC Bioinformatics* **12**,159. (doi: 10.1186/1471-2105-12-159)
19. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011 SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **810**, 785-786. (doi: 10.1038/nmeth.1701)
20. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008 Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **57**, 621-628. (doi: 10.1038/nmeth.1226)
21. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **304**, 772-780. (doi: 10.1093/molbev/mst010)
22. Swofford DL. 1998 PAUP*. Phylogenetic Analysis Using Parsimony *and Other Methods. Version 4. Sinauer Associates, Sunderland, Massachusetts.
23. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012 MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **613**, 539-542. (doi: 10.1093/sysbio/sys029)
24. Chen K, Durand D, Farach-Colton M. 2000 NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J. Comput. Biol.* **7**, 429-447. (doi:10.1089/106652700750050871)
25. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012 Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969-1973. (doi: 10.1093/molbev/mss075)

26. Wiens JJ. 2007 Global patterns of diversification and species richness in amphibians. *Am. Nat. Suppl*. **170**, S86-106. (doi:10.1086/519396)
27. Roelants K, Gower DJ, Wilkinson M, Loader SP, Biju SD, Guillaume K, Moriau L, Bossuyt F. 2007 Global patterns of diversification in the history of modern amphibians. *Proc. Natl. Acad. Sci. U.S.A*. **104**, 887-892. (doi: 10.1073/pnas.0608378104)
28. Vieites DR, Min MS, Wake DB. 2007 Rapid diversification and dispersal during periods of global warming by plethodontid salamanders. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19903-19907. (doi: 10.1073/pnas.0705056104)
29. Zhang P, Wake DB. 2009 Higher-level salamander relationships and divergence dates inferred from complete mitochondrial genomes. *Mol. Phylogenet. Evol.* **53**, 492-508. (doi: 10.1016/j.ympev.2009.07.010)
30. Pyron RA. 2010 A likelihood method for assessing molecular divergence time estimates and the placement of fossil calibrations. *Syst. Biol.* **59**, 185-194. (doi: 10.1093/sysbio/syp090)
31. Hedges SB, Dudley J, Kumar S. 2006 TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**, 2971-2972. (doi:10.1093/bioinformatics/btl505. PMID 17021158)

Acknowledgments

We thank Dirk Cieters and Luc Platteau from Natuurpunt Ternat and the Vermeir family for their collaboration and help in obtaining the newts for the research. We thank Han Remaut and Ghislain Opdenakker for use of lab equipment, Rik Janssens and Mieke Gouwy for lab assistance, and Willem Meilink for drawing salamander figures. We also thank the Agentschap voor Natuur en Bos for giving permits. 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 the Hercules foundation (purchase of LC-MS equipment, AKUL/11/31). SJ receives a grant from the IWT, MM receives an aspirant fellowship from FWO, IVB, PK and AM are supported by a postdoctoral fellowship from FWO.

FIGURES

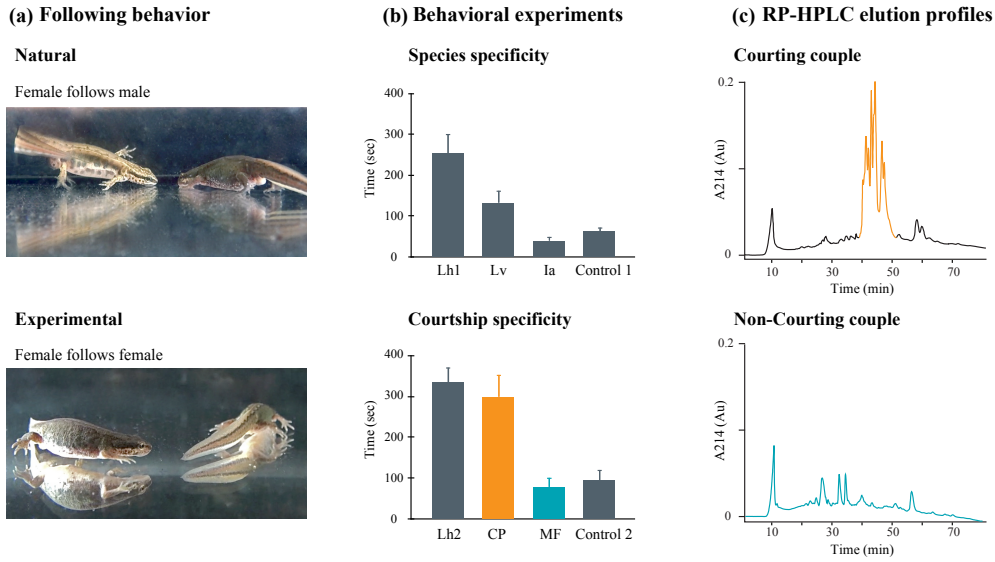

Figure 1. Identification and isolation of male courtship pheromones. (**a**) up: Tail-fanning of the pheromones towards the nose of the female persuades her to follow the path of the retreating male; down: analogous response of a female in pheromone-containing water during a two-female behavioural bio-assay. (**b**) Behavioural assays and species-specificity. up: The time that females of *L. helveticus* showed following behaviour in courtship water of their own species (Lh1), *L. vulgaris* (Lv), *I. alpestris* (Ia) and control (Control 1, H_2O); down: The fractions composing the courtship peak (CP, orange) induced female responses that are similar to those in full courtship water (Lh2). Male-female water (MF) resulted in few female responses, and was similar to the control (Control 2, H_2O) (**c**) Comparison of RP-HPLC profiles of courtship water and male-female (non-courtship) water. Courtship water shows a courtship peak (orange) that is absent in MF water (blue), indicating that males largely release courtship pheromones (that induce female following) during tail-fanning.

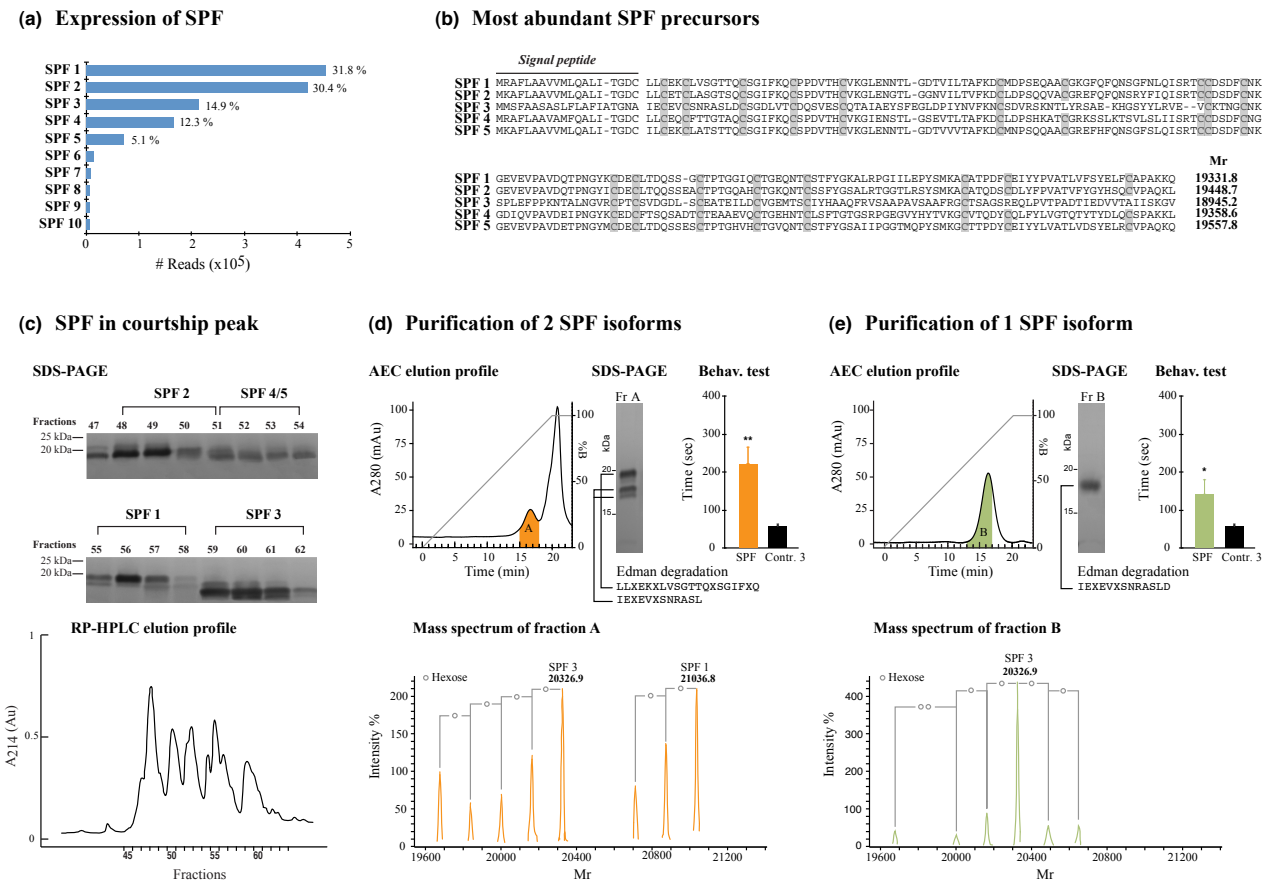

Figure 2. Transcriptomic and proteomic analyses of SPF proteins. (**a**) RNAseq expression level (percentage of total SPF) of the ten most abundant SPF precursors in the abdominal gland of a male (Reads Per Kilobase per Million mapped reads, RPKM). (**b**) MAFFT alignment and theoretical masses of the five most abundant SPF proteins. Cysteins are indicated in grey. (**c**) SDS-PAGE (silver staining) and RP-HPLC elution profile of SPF proteins in a courtship peak. See the electronic supplementary material, table S2 for Edman sequencing and table S3 for mass spectrometry analysis of individual fractions. (**d, e**) Anion exchange chromatography (AEC) elution profile, silver stained SDS-PAGE, mass spectrometry (deconvoluted mass spectra), Edman sequencing, and behavioural tests of two SPF pheromones (**d**) and a single SPF pheromone (**e**). Asterisks indicate significance levels: *P < 0.05, **P < 0.01.

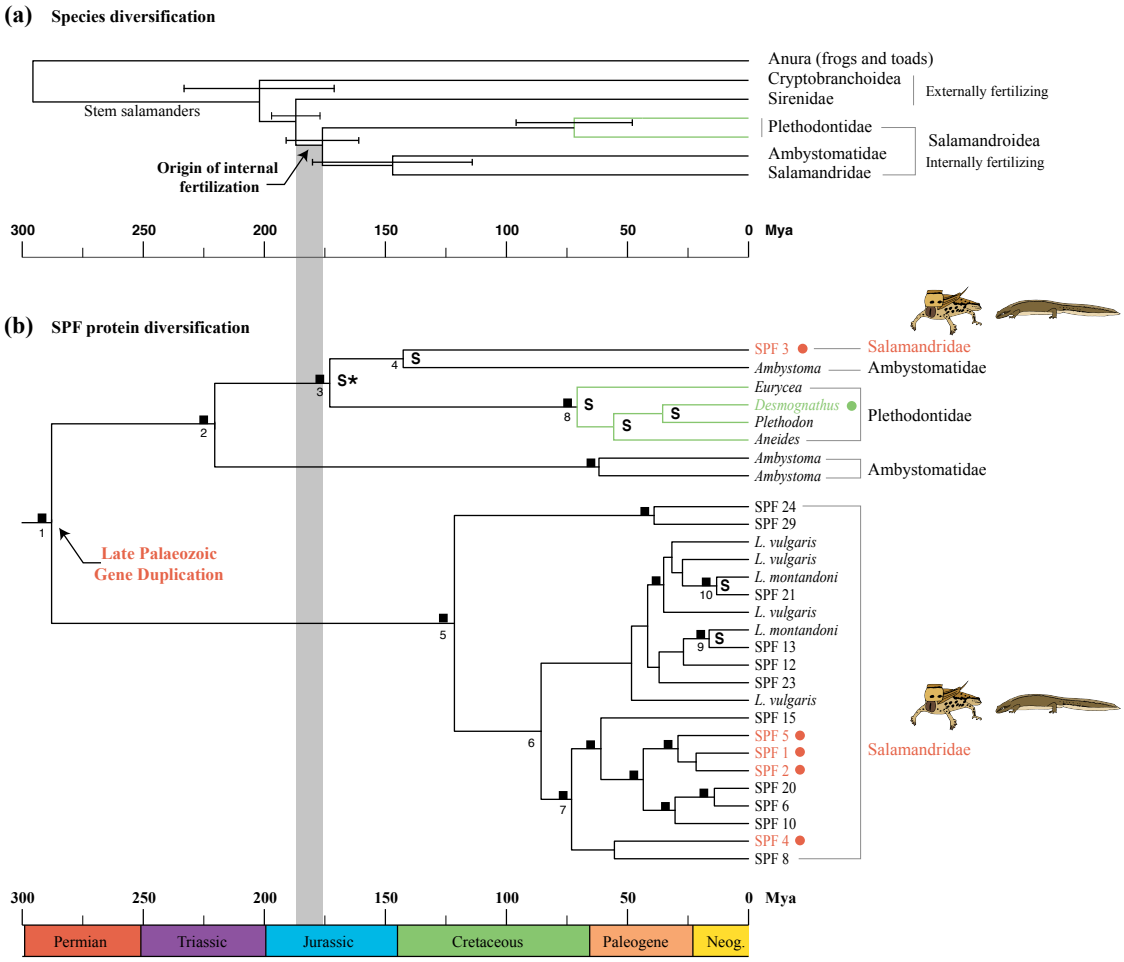

Figure 3. Time estimates (a) species diversification. The mean and standard deviations for species diversifications were calculated from the last five published estimates [26-30]. The origin of internal fertilization in the ancestor of Salamandroidea is indicated in gray. **(b) SPF protein diversification.** The tree shows Bayesian dating estimates, the asterisk denotes the calibration point. The diversity of plethodontid precursors was chosen to reflect the largest known SPF divergences in this family. The fact that our gene tree of plethodontid SPF's corresponds to the higher taxonomic level relationships of these species therefore indicates that the known SPF variation is the result of family-specific variation and/or gene duplications (in agreement with [12]). Bayesian Posterior Probabilities >95 are indicated with black squares. Speciation nodes are indicated with S, all other nodes are considered duplication nodes [by Notung analyses, see Materials and methods]. Node numbers refer to age estimates in table S4.

Numbered SPF's all indicate sequences from the abdominal gland of the palmate newt. The top five expressed proteins, which were also confirmed in courtship water, are indicated with red circles. The green circle denotes the species in which the pheromone function was demonstrated in plethodontids.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.457v1 | CC-BY 4.0 Open Access | received: 11 Aug 2014, published: 11 Aug