AN ABSTRACT OF THE THESIS OF

Martha M. von Reis for the degree of Master of Science in Zoology presented on May 24, 2007.
Title: Evolution of Sodefrin Precursor Pheromones in Salamandrid Newts

Abstract approved:

Frank L. Moore

An ancient pheromonal signal is found in aquatic courting salamandrids and terrestrial courting plethodontids, two highly divergent salamander families. In the aquatic courting salamandrids Cynops pyrrhogaster and C. ensicauda, the decapeptide sodefrin is cleaved from a larger 189 amino acid Sodefrin Precursor Factor (SPF) protein and released from a gland in the male’s cloaca during courtship. This decapeptide is capable of attracting females in a species-specific manner. In contrast, males of some terrestrial courting plethodontids apply uncleaved SPF directly to females to increase receptivity during courtship. In this study, we examine the presence, diversification, and evolution of this pheromone in two aquatic courting North American salamandrids, Notophthalmus viridescens and Taricha granulosa. We used degenerate primers, RT-PCR procedures, and cDNA obtained from sexually active males, in addition to genomic PCR methods. SPF was found to be expressed in abdominal gland region cDNA by males of both species and was additionally expressed in hedonic (genial) gland tissue of N. viridescens males. With respect to the presence of a cleaved pheromone product in these species, multiple alignments with C. pyrrhogaster sequences demonstrated a high degree of variability in the putative
decapeptide region. Cleavage sites were not identifiable. Site-specific tests reveal that positive selection has driven rapid evolutionary change at sites within *N. viridescens* and *C. pyrrhogaster* precursor proteins but not within *T. granulosa*. 
Evolution of Sodefrin Precursor Pheromones in Salamandrid Newts

by
Martha M. von Reis

A THESIS
submitted to
Oregon State University
in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented May 24, 2007
Commencement June 2008

APPROVED:

Major Professor, representing Zoology

Chair of the Department of Zoology

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Martha M. von Reis, Author
ACKNOWLEDGEMENTS

I would like to thank Frank Moore for acting as my major professor. In particular, his flexibility and guidance during the writing portion of my thesis are appreciated. I would also like to express my gratitude to past members of the Moore lab—Emma Coddington, Brian Searcy, Eliza Walthers, and Wendy Phillips—who have always kindly offered advice concerning graduate school and life in general. Their myriad perspectives were both insightful and humorous. And for Sam Bradford, I am most thankful. He has performed the work of 10 men by acting as mentor, friend, confidante, technical advisor, editor, coffee aficionado, day laborer, newt catcher, cat catcher, and Emily entertainer. I hope that my friendship has offered half as much in return.

While at OSU, I have been lucky enough to also have Stevan Arnold and Lynne Houck as mentors. I am grateful that they allowed me open access to their lab meetings and events and provided opportunity for travel to Highlands Biological Station. Thanks to Steve, who has offered critical feedback since the beginning of my graduate career (and taught a fantastic grant writing course) and to Lynne for her responsiveness and kindness. Several other members of their lab played a critical role in my graduate research. A big thank you to Richard Watts and Catherine Palmer for their advice and support during the grant writing process and beyond. Mostly, I am grateful for the enduring friendships I have made through my association with the Houck-Arnold lab—namely Amy Picard Hastings and Kevin Hastings, about whom I can’t say enough good things.
Lastly, I am grateful for my amazing family, Charles, Emily, and the dogs, who have (almost) without complaint endured and supported me (each in his or her own way) through my time as a graduate student. Thank you for your love and patience.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter 1- General Introduction</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 2- Evolution of Sodefrin Precursor Pheromones in Salamandrid newts</td>
<td>14</td>
</tr>
<tr>
<td>Abstract</td>
<td>15</td>
</tr>
<tr>
<td>Introduction</td>
<td>16</td>
</tr>
<tr>
<td>Materials and methods</td>
<td>20</td>
</tr>
<tr>
<td>Results</td>
<td>25</td>
</tr>
<tr>
<td>Discussion</td>
<td>42</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>49</td>
</tr>
<tr>
<td>References</td>
<td>49</td>
</tr>
<tr>
<td>Chapter 3- General Conclusions</td>
<td>53</td>
</tr>
<tr>
<td>Bibliography</td>
<td>56</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Deduced amino acid sequences of representative Sodefrin Precursor Factor (SPF) proteins</td>
<td>28</td>
</tr>
<tr>
<td>2.</td>
<td>Gene tree of expressed SPF sequences generated by parsimony analysis (PAUP v4.01b)</td>
<td>31</td>
</tr>
<tr>
<td>3.</td>
<td>Identification of sites under positive selection in SPF proteins from <em>C. pyrrhogaster</em> (Part A) and <em>N. viridescens</em> (Part B)</td>
<td>39</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Species sequence diversity from Sodefrin Precursor Factor (SPF) genes in five species</td>
<td>34</td>
</tr>
<tr>
<td>2.</td>
<td>Nonsynonymous to synonymous substitution rate ratios ($\omega$), prior probabilities ($p(\omega)$), and log-likelihood values ($\ln L$) for selection models fitted to newt Sodefrin Precursor Factor (SPF) genes</td>
<td>37</td>
</tr>
<tr>
<td>3.</td>
<td><em>C. pyrrhogaster</em> SPF variants from three populations</td>
<td>42</td>
</tr>
</tbody>
</table>
Evolution of Sodefrin Precursor Pheromones in Salamandrid Newts

Chapter 1

General Introduction

**Pheromones**

Pheromones are molecular signals, sociochemicals, used as a means of communication between animals of the same species. Vertebrate pheromones are most commonly described as signals released from one organism that initiate specific behavioral or developmental responses in a conspecific. Pheromones can act alone or in conjunction with acoustical, visual, and tactile cues. Organisms without acoustic capabilities may depend largely on chemical cues for communication. The potential benefits of using these chemical cues include their longevity over time and space, signal specificity, and low energy cost (Wyatt, 2003). Pheromonal signals have been widely described in insects and more recently in a number of vertebrates. The behavioral responses that can be elicited through the use of pheromones are wide ranging and include functions such as: aggregation of social groups for protection or mating, territorial marking, organization of social insects, predator avoidance via alarm signals, and mediation of sexual behaviors such as location of mates, mate choice, and female receptivity (see review by Wyatt, 2003). In salamanders, chemical signaling through the use of pheromones is perhaps the most significant factor in sex
recognition and courtship (Houck, 1986). My research has focused on male sex pheromones used to attract and/or increase sexual receptivity of female salamanders.

**Pheromonal signaling in salamanders**

Although auditory signals may be the predominant form of communication for anuran amphibians, it has become apparent that chemical signals play an important role in non-auditory urodele amphibians and some anurans. For many amphibians, the possible existence of pheromones has been demonstrated in behavioral tests, yet very few pheromones have been isolated and characterized. Among those that have been well characterized are both small peptides and large proteinaceous pheromones. In the aquatic breeding newts, *Cynops pyrrhogaster* and *C. ensicauda*, a water soluble decapeptide (sodefrin and silefrin, respectively) is produced by males and has the ability to attract receptive females when released into the water. The decapeptide is cleaved from a larger precursor protein consisting of ~189 amino acids (Kikuyama & Toyoda, 1999). The aquatic breeding frog *Littoria splendid* releases a 25 amino acid female-attracting pheromone called splendiferin (Wabnitz et al., 2000). In contrast, large proteinaceous pheromones have been found in terrestrial breeding salamanders. In *Plethodon shermani*, males produce a multi-component pheromone mixture which increases female receptivity, as assayed by a decreased time to complete sperm transfer. The primary components of this mixture consist of a 22kDa protein called Plethodontid Receptivity Factor (PRF) which alone can decrease courtship duration
(Rollmann et al., 1999) and a 7kDa component, Plethodontid Modulatory Factor (PMF), which has the opposite effect when applied alone (Houck et al., 2007). Also found in a different plethodontid salamander is another proteinaceous pheromone of ~25kDa which was behaviorally tested and shown to increase female receptivity in *Desmognathus ocoee* (Watts et al., 2007). Due to its molecular and biochemical similarity to the sodefrin precursor protein, this pheromone is termed Sodefrin Precursor Factor (SPF). PCR screening indicates that SPF is widely expressed throughout the plethodontid family. It is fascinating that this chemical signal has been retained since the divergence of these two families, between 45 and 80 million years ago (Wiens et al., 2006) yet has been significantly modified, apparently in conjunction with other changes in courtship behavior. In my research, I have used molecular techniques to explore the presence of SPF in additional aquatic courting newt species, *Taricha granulosa* and *Notophthalmus viridescens*. Using sequence similarities and evolutionary modeling, I have compared the sequences obtained from these North American newts to SPF from other aquatic courting newts and from terrestrial courting newts and have made inferences regarding the probability of a proteinacious versus peptide pheromone. In order to further describe the similarities and differences in these pheromones, it is useful to understand the behavioral contexts in which they are used.

**Courtship in salamandrid newts**
Salamandrid newts *C. pyrrhogaster, C. ensicaudea, T. granulosa,* and *N. viridescens* share several important courtship characteristics; courtship occurs in water, typically requires persuasion of a female to engage in and complete courtship, and concludes with a female’s insemination by one or more spermatophores that have been deposited onto the substrate. Details of courtship vary and are described for each species along with what has been previously detailed concerning pheromone secretion and delivery.

In *C. pyrrhogaster,* courtship consists of a tail-nudging walk in which a male approaches a female by nudging her with his snout and blocking her from moving forward (Arnold 1977). The male then moves in front of the female and performs a tail-waving display in which the decapeptide sodefrin (or other active variants) is released from the abdominal gland and wafted towards the female. The pair proceeds to walk forward with the female’s nares in close proximity to the male’s cloaca. The male deposits a spermatophore and leads the female so that her cloaca is positioned for insemination. The isolation and characterization of the decapeptide sodefrin represented the first known pheromone in an amphibian and the first peptide pheromone in a vertebrate. Studies have illustrated that sodefrin is a species-specific signal that is able to attract receptive conspecific females at low concentrations (Kikuyama & Toyoda, 1999; Yamamoto et al., 2000). Furthermore, the existence of regional variants has been demonstrated in *C. pyrrhogaster* populations (Iwata et al.,
2005), thus demonstrating the possibility that differences in sex pheromones may contribute to speciation.

Courtship in *N. viridescens* can occur rapidly with minimal effort by the male, if a receptive female is encountered. In this abbreviated courtship, males will deposit a spermatophore after a brief tail-fanning display (Houck & Arnold, 2003; Petranka, 1998). However, the majority of courtships require significant female persuasion, beginning with the male amplexing the female around her neck with his hind legs (Petranka, 1998). While holding the female, the male will alternate between rubbing her snout with his hedonic (cheek) glands and his forelimbs. At the same time, the male fans his tail and presumably wafts cloacal secretions containing pheromone toward the female (Houck & Arnold, 2003). This effort may last up to an hour and intensifies directly before a male dismounts to deposit a spermatophore. If the female is sufficiently receptive, she will nudge the male’s cloaca to stimulate spermatophore deposition. Then she will follow closely behind him as he leads her forward. The male then turns sideways to block her from moving beyond the spermatophore (Petranka, 1998). Products of the hedonic glands are known to be under hormonal control (Pool & Dent, 1977) and have been assumed to produce pheromones. However, molecular and biochemical characterizations of courtship pheromones have not previously been reported for this species.
In *T. granulosa*, courtship begins when a male captures a female by dorsal amplexus with all four limbs. During this prolonged stage of courtship, the male will periodically rub the female’s snout with his chin, strokes her abdomen with his hind legs, and rub his cloaca against her dorsum. These actions increase in frequency and intensity throughout the courtship. After hours or days a female will indicate that she is sexually receptive by lifting her head towards the male’s chin and releasing air bubbles (Propper, 1991). The male then dismounts and places his body perpendicularly in front of the female while exhibiting a tail-fanning display directed towards her (Rose & Moore, 2002). At this time, the male deposits a spermatophore. The female keeps her snout in close contact with the male’s cloaca as he leads her forward so that her cloaca is positioned above the spermatophore for insemination (for a more detailed account, see Propper, 1991). Although no male *T. granulosa* courtship pheromones have been described, pheromone delivery potentially occurs as the male rubs his snout against the female’s nares during the amplectic stage and/or during the tail-fanning display that occurs directly before spermatophore deposition.

**Courtship in terrestrial breeding plethodontids**

The family Plethodontidae is a very diverse group of terrestrial breeding salamanders containing over 300 species (Larson et al., 2003). Although the details differ between species, a typical courtship encounter begins when a male approaches a female and initiates head contact. If the female is willing, the pair will engage in a
tail-straddling walk in which the female straddles the male’s tail with her chin resting on his tail base. This stage of courtship can last for several hours. Eventually, if a male is able to maintain the female’s interest, he will deposit a spermatophore on the substrate and lead the female forward for sperm transfer into the female’s cloaca (see review by Houck and Arnold, 2003). Delivery of pheromone sometimes occurs upon initial encounter with a female and typically throughout the tail-straddling walk. Depending on the species, courtship pheromones are delivered either by ‘diffusion’ or ‘slapping’. In the diffusion mode of pheromone delivery, a male uses enlarged premaxillary teeth to scratch the female’s dorsum and then rubs secretions from his mental gland (glandular tissue on the chin that hypertrophies during the breeding season) into the abraded site. In the slapping method of pheromone delivery, a male will turn back during the tail-straddling walk and slap the female’s snout with his mental gland.

**Pheromone detection**

The vomeronasal organ is an accessory chemosensory system that acts to receive pheromonal and other signals and transmit them to the brain (Halpern & Martiniz-Marcos, 2003). In amphibians, the vomeronasal organ is a small microvillar lined epithelial layer located adjacent to the olfactory epithelium within the nasal cavity (Wirsig-Wiechmann et al., 2002). Pheromone is transmitted to the vomeronasal organ via capillary action at the nasolabial groove. Experiments on C.
pyrrhogaster and P. shermani indicate that application of pheromone stimulates cells of the vomeronasal organ (Toyoda & Kikuyama, 2000; Wirsig-Wiechmann et al., 2002; Wirsig-Wiechmann et al., 2006). Vomeronasal receptors belong to two families, V1R and V2R (Dulac & Axel, 1995; Hagino-Yamagishi et al., 2004). In mammals, pheromone detection can be mediated by stimulation of V1R receptors (Boschat et al., 2002) but no specific function has yet been assigned to V2R receptors. Both V1R and V2R families are diverse and show evidence of adaptive evolution via positive selection (Shi et al. 2005; Yang et al. 2005).

**Evolution of Amphibian Pheromones**

Evolution of a chemical signal by sexual selection can occur when a signal contributes to male mating success. In salamander mating systems, delivery of courtship pheromones can enhance male mating success. Thus, sexual selection is expected to play a role in the evolution of male courtship pheromones and rapid evolution can be a consequence (Arnold and Houck 1982). Since salamander courtship pheromones are chemical signals that act via female receptors, co-evolution (driven by sexual selection) may act to maintain concordance between the signal and the receptor. Indeed, co-evolution of pheromone and receptor is indicated by high levels of polymorphism, gene duplication, and positive selection observed in plethodontid SPF and PRF (Palmer et al., 2005; Palmer et al., 2007; Watts et al., 2007). These observations apparently are common to reproductive proteins that act as
isolating mechanisms (Swanson & Vacquier, 2002; Swanson & Vacquier, 2002; Swanson et al., 2001), by conferring information about species or population differences or genetic relatedness. Thus, rapidly evolving pheromonal signals may act as a reproductive barrier, thereby contributing to speciation. Rapid evolution associated with adaptive change (positive selection) has been identified in pheromones, pheromone associated proteins, and pheromone receptors (Emes et al., 2004; Shi et al., 2005; Willett, 2000). In contrast, other reproductive proteins that are not associated with isolating mechanisms (e.g. hormones and their receptors) are highly conserved and without evidence of rapid evolution by positive selection. Like most classes of proteins, these reproductive proteins have evolved by neutral divergence constrained by purifying selection (Endo et al., 1996).

Conclusions

In my graduate research, I have identified the presence of SPF, an ancient pheromonal signal, in two North American newts. In addition, I have investigated the evolutionary forces that have shaped the SPF pheromones in three aquatic breeding newts. The results increase our general understanding of molecular evolution of reproductive proteins and raise interesting questions regarding the use of SPF in aquatic breeding newts.

References


Chapter 2

Evolution of Sodefrin Precursor Pheromones in Salamandrid Newts

Martha M. von Reis
Abstract

An ancient pheromonal signal is found in aquatic courting salamandrids and terrestrial courting plethodontids, two highly divergent salamander families. In the aquatic courting salamandrids *Cynops pyrrhogaster* and *C. ensicauda*, the decapeptide sodefrin is cleaved from a larger 189 amino acid Sodefrin Precursor Factor (SPF) protein and released from a gland in the male’s cloaca during courtship. This decapeptide is capable of attracting females in a species-specific manner. In contrast, males of some terrestrial courting plethodontids apply uncleaved SPF directly to females to increase receptivity during courtship. In this study, we examine the presence, diversification, and evolution of this pheromone in two aquatic courting North American salamandrids, *Notophthalmus viridescens* and *Taricha granulosa*. We used degenerate primers, RT-PCR procedures, and cDNA obtained from sexually active males, in addition to genomic PCR methods. SPF was found to be expressed in abdominal gland region cDNA by males of both species and was additionally expressed in hedonic (genial) gland tissue of *N. viridescens* males. With respect to the presence of a cleaved pheromone product in these species, multiple alignments with *C. pyrrhogaster* sequences demonstrated a high degree of variability in the putative decapeptide region. Cleavage sites were not identifiable. Site-specific tests reveal
that positive selection has driven rapid evolutionary change at sites within *N. viridescens* and *C. pyrrhogaster* precursor proteins but not within *T. granulosa*.

**Introduction**

Pheromonal courtship signals have been described for members of the urodele families Salamandridae and Plethodontidae. In aquatic courting newts of the family Salamandridae, a female-attracting pheromone is produced as a ~190 amino acid precursor in the abdominal glands of sexually active males of two species; *Cynops pyrrhogaster* and *Cynops ensicauda* (Iwata et al., 1999; Kikuyama & Toyoda, 1999). The active region of the protein is a decapeptide that is cleaved from the carboxyl terminus, released into the water, and wafted, via a tail-fanning action, toward a female during courtship. In *C. ensicauda*, this decapeptide is called silefrin and differs by two amino acids from the *C. pyrrhogaster* pheromones (Iwata et al., 1999). In *C. pyrrhogaster*, two bioactive decapeptides have been identified, sodefrin and aonirin. Sodefrin is found ubiquitously in all populations tested. The aonirin variant, differing by only one amino acid from sodefrin, is found only in males from the Nara region of Japan (Nakada et al., 2007). Though these newts express both sodefrin and aonirin, females react more strongly to aonirin. Females from Chiba and Niigata, populations whose males do not express aonirin, were tested for attractiveness to this variant. Females from Niigata showed attraction under a very high dosage of aonirin and
females from Chiba showed no response at any dosage. Population specificity of pheromone variants may be a key factor leading to or enforcing reproductive isolation and speciation.

In contrast to the decapeptide pheromone found in *Cynops*, the major pheromone component found in the some plethodontid clades is a 22kD protein (Rollmann et al., 1999) called Plethodontid Receptivity Factor (PRF). This protein is synthesized in the male mental (chin) glands and is directly applied to a female, usually during the stage of courtship termed tail-straddling walk. Behavioral tests indicate that this proteinaceous pheromone increases female receptivity as measured by a decreased time to complete sperm transfer (Rollmann et al., 1999). Recently, another proteinaceous pheromone with no similarity to PRF was identified in the mental glands of *Desmognathus* salamanders (Watts et al., 2007). Due to its molecular similarity to the sodefrin precursor, it has been termed a Sodefrin Precursor Factor (SPF) protein. Further surveys for SPF have revealed that it is present, at various expression levels, in three major lineages of plethodontids (Palmer et al., 2007). Behavioral tests in *Desmognathus ocoee*, a species with very high expression of SPF, indicated that the full-length ~23kD protein, and not a cleaved peptide, acts to increase female receptivity when applied directly to a female during courtship (Watts et al., 2007).
It is remarkable that this ancient signal has been retained and used as a pheromonal signal throughout the 40-85 MY diversification (Wiens et al. 2006) of these salamander families. As described above, extreme modifications to courtship behavior, breeding environments, gland morphology, and pheromone delivery modes have taken place since their divergence. Other studies on the evolutionary history of amphibian courtship systems have revealed stasis at the behavioral and morphological level, but rapid and diversifying selection at the molecular level (Houck & Arnold, 2003; Palmer et al., 2007). Modifications to the SPF pheromone suggest a model in which a sexual signal has evolved and has been retained over millions of years with changes in utilization of the signal specific to the life history of each lineage.

Rapid evolution of reproductive proteins between closely related species has been observed in genera from fungi (for example see Brown & Casselton, 2001) to primates (Mundy & Cook, 2003). Such rapid evolution can lead to extreme variation and polymorphism at individual loci within populations, and may contribute to speciation when stable differences between populations act as reproductive barriers. Rapid evolution can be the result of relaxed functional constraint on a protein (neutral evolution), or alternatively, it can be due to positive selection acting at specific sites within a protein. Positive selection is typically identified by a nonsynonymous/synonymous substitution rate ratio of greater than one, which implies that nonsynonymous (amino acid changing) mutations offer an adaptive advantage
(Yang & Nielsen, 2002). Sites under positive selection are therefore likely to be functionally important and in many cases are directly involved in interaction with a receptor (Yang & Swanson, 2002; Swanson et al., 2001). While rapid evolution driven by positive selection is rare for most classes of proteins, communication molecules, in particular reproductive proteins that may act as sexual isolating mechanisms, are subject to rapid changes at sites in the molecule that are directly involved in interaction (Swanson & Vacquier, 2002). Because salamander courtship pheromones are responsible for increasing female receptivity, changes at receptor binding epitopes will likely have a direct affect on mating success. Thus, rapid changes to courtship pheromones are expected to be the result of sexual selection (Arnold & Houck, 1982). Consistent with other reproductive proteins, PRF and Plethodontid SPF have both been shaped by positive selection. Molecular modeling illustrates that some of the positively selected sites identified in PRF are located in areas of the protein that are likely to be receptor binding epitopes (Watts et al., 2004).

In this study, we ask whether modifications to production (abdominal versus mental gland) and application (water-borne versus direct) of the pheromone are correlated to differences between aquatic and terrestrial courting salamander families. The presence of SPF has been established for three major lineages of terrestrial courting plethodontids (Palmer et al., 2007). We have used molecular methods to determine the presence and expression patterns of SPF in two aquatic courting North
American newts, *Taricha granulosa* and *Notophthalmus viridescens*. In addition, *C. pyrrhogaster* SPF sequences were analyzed in order to compare levels of variability and patterns of positively selected sites to other aquatic courting newts and to terrestrial courting salamanders. In the aquatic courting newts, *C. pyrrhogaster* and *C. ensicauda*, the sodefrin decapeptide shows little variation within a species and the ~190 amino acid precursor has no known function and should therefore evolve in a neutral manner. In contrast, terrestrial courting plethodontid SPF is a highly variable multigene family that has evolved via rapid and incessant positive selection (Palmer et al., 2007).

**Materials and methods**

*Animals*

*T. granulosa* were collected from local breeding ponds in Benton County, Oregon. *N. viridescens* were obtained from a commercial breeder (Charles D. Sullivan Co. Inc.; http://researchamphibians.com/). A small sample of tail tissue was clipped from live animals of *Pleurodeles waltl* and *Neurergus kaiseri*, both permanently housed at Oregon State University. Animals were housed, fed, and handled in accordance with guidelines set forward by the Institutional Animal Use and Care Committee at Oregon State University. Tissue samples from other genera were obtained by loan from the University Of California Museum Of Vertebrate Zoology.
SPF cloning via RT-PCR

Tissues from the abdominal gland area and hedonic (genial) gland (*N. viridescens* only) were dissected from males during the courtship season, from male *N. viridescens* and *T. granulosa* outside of the courtship season (control animals), and from female *N. viridescens*. RNA was extracted using an RNAqueous™ kit (Ambion catalog no. 1912) and cDNA was reversed transcribed (Invitrogen catalog no. 11904-018) using random hexamer primers. PCR amplification was initially performed using degenerate primers (sense: 5’-TYC TTA CTC TMY TAG CAC CAT GAG-3’, antisense: 5’-TCC TCS TCA CAA GAY CAG AC-3’) designed against conserved portions of *Cynops pyrrhogaster* (GenBank accession no. [AJ245955](https://www.ncbi.nlm.nih.gov/nuccore/AJ245955), *Cynops ensicauda* (GenBank accession no. [AJ245956](https://www.ncbi.nlm.nih.gov/nuccore/AJ245956)) and *Triturus carnifex* (GenBank accession no. [AF446080](https://www.ncbi.nlm.nih.gov/nuccore/AF446080)) using *Taq* polymerase (Fermentas) with the following cycling profile: 94°C for 3 min; 4 cycles of 94°C for 45s, 50-58°C (gradient) for 45s, 72°C for 1 min 30s, 30 cycles of 94°C for 30s, 50-58°C (gradient) for 30s, 72°C for 1 min, and terminal extension at 72°C for 7 min. Specific primers were later designed for *Taricha granulosa* (sense: 5’-GGC ACA GAG ACT CCA TCA GC-3’, antisense: 5’-GCT GAC ATB AGT TCA TGT GCT-3’) and *N. viridescens* (sense: 5’-GCA CAG AGA CTC CAT CAC CTA C-3’, antisense: 5’-TTA CAA TGT TCC TCC TCT TCA C-3’). Full-length sequences obtained for these species were produced using specific primers, High Fidelity *Taq* polymerase (Invitrogen catalog no. 11304-011).
and the following cycling parameters; 94°C for 3 min; 4 cycles of 94° for 45s, 51°C for 45s, 68°C for 1 min 30s, 30 cycles of 94° for 30s, 51°C for 30s, 68°C for 1 min, and terminal extension at 68°C for 7 min. Products were separated on a 1% agarose gel, stained with ethidium bromide, and visualized with UV light. If a single band of the appropriate size was amplified, products were directly ligated into the TOPO T4 Vector (Invitrogen catalog no. K4530-20), transformed into TOP10 E. coli cells, spread on selective agar plates containing ampicillin (50 µg/ml), and grown at 37°C overnight. The resulting colonies were screened for inserts of the proper size using the vector primers T3 and T7. Sub-clones of colonies identified as positive for proper inserts were grown overnight in Luria Broth containing ampicillin (50 µg/ml), and plasmid DNA was isolated using the Qiagen Miniprep Kit (Qiagen catalog no.27104) according to the manufacturer’s instructions. Sequencing was carried out by the Central Services Laboratory at Oregon State University.

SPF cloning via genomic PCR

In order to increase the number of species used in this study, a genomic PCR method was utilized for obtaining sequences from genomic DNA. DNA was extracted from a small portion of tail tissue, obtained from C. pyrrhogaster, Neurergus kaiseri and P. waltl (N=1 animal per species), using standard Proteinase K digestion, phenol-chloroform extraction, and ethanol precipitation methods. Additional tissue samples were obtained by loan from the University of California, Museum of Vertebrate
Zoology (MVZ): *Amphiuma means* (MVZ no.144884), *Amphiuma tridactylum* (MVZ no. 241480), *Siren intermedia* (MVZ no.200518), *Salamandra salamandra* (MVZ no. 236840), *Mertensiella luschani* (MVZ no. 230154), *Pachytriton labiatus* (MVZ no. 230355), *Paramesotriton hongkongensis* (MVZ no. 230366), *Tylototriton shanjing* (MVZ no. 219761), and *Cynops orientalis* (MVZ no. 231161). Genomic DNA was extracted from these samples by heating for two hours at 56°C in 200µL of a 5% Chelex 100 (Sigma no. C7901) and Proteinase K (200 µg/ml) mixture suspended in sterile water. Following the 56°C incubation, samples were boiled for 8 minutes and then vortexed and centrifuged before use in PCR. All genomic DNA was partially digested with the restriction enzyme Spe I in order to facilitate amplification. Nested PCR was performed using degenerate primers (pair 1 [outer reaction] sense: 5’-TYC TTA CTC TMY TAG CAC CAT GAG-3’, antisense: 5’-GCT GAC ATC AGT TCA TRT GYT-3, pair 2 [inner, nested reaction] sense: 5’-TGY CTA TTW TGC GAG CAR TG-3’, antisense: 5’-TCC TCS TCA CAA GAY CAG AC-3’). Cycling parameters for both reactions using High Fidelity *Taq* polymerase were as follows: 94°C for 3 min; 4 cycles of 94° for 45s, 51°C for 45s, 68°C for 1 min 30s, 30 cycles of 94° for 30s, 51°C for 30s, 68°C for 1 min, and terminal extension at 68°C for 7 min. Positive and negative controls (autoclaved purified water) were used during each step of the amplification process. The resulting products were processed for sequencing as outlined above.
Cynops SPF sequences

Iwata et al. obtained five *C. pyrrhogaster* males from each of three regions of Japan (Chiba, Niigata, and Nara) and extracted total RNA from abdominal glands in order to PCR amplify the decapeptide sodefrin (see Iwata et al., 2004 for complete details). The resulting dataset contained 135 SPF sequences, of which only two contained a full open reading frame. An additional 100 sequences contained enough data for further analysis of regions of interest, as determined by the site-specific positive selection analyses.

Construction of SPF gene tree

The following four sequences and their accession numbers were obtained using PSI-BLAST; *C. pyrrhogaster* (AJ245955, AB280579), *C. ensicauda* (AJ245956) and *Triturus carnifex* (AF446080). As described previously, additional sequences were obtained from *Taricha granulosa* (N=14) and *N. viridescens* (N=31), and *C. pyrrhogaster* (N=7). Sequences were aligned using ClustalX (v1.8) with gap opening penalties of 30 and gap extension penalties of 0.6 for pairwise comparisons and gap opening of 15.0 and gap extension penalties of 0.5 for multiple comparisons. Delaying of divergent sequences was reduced to 20%. Other optional parameters remained at default settings. Bayesian analyses were performed using Mr. Bayes (v3.1.2). Four Monte-Carlo Markov chains were run simultaneously for 1,000,000 generations with sampling of every 100 trees. The first 2000 trees were discarded as
‘burn-in’ to ensure that the log likelihood values for trees had converged to a stable value. The remaining 8,000 trees were used to create a majority rule consensus tree. Additionally, maximum parsimony analyses using the same ClustalX alignments were executed in PAUP (v4.0 Beta) with a minimum of 100 bootstrapping events per tree. Gapped positions were excluded from analysis.

Selection analyses

Selection at specific amino acid sites was characterized from estimates of the ratio of nonsynonymous to synonymous substitution rates ($\omega$). In this test, $\omega=1$ for sites that have evolved neutrally, while $\omega<1$ indicates purifying selection and $\omega>1$ indicates positive selection ($\omega^+$). The ratio $\omega$ was estimated using the program CODEML as executed in the PAML (http://abacus.gene.ucl.ac.uk/software/paml.html) package (v3.15). These likelihood ratio tests are derived from different codon substitution models described in detail by Yang et al. (2000). The five site-specific codon substitution models fit to each dataset in these studies are briefly described here. Model 0 (M0, one ratio) allows one freely estimated $\omega$ category at each site, the Model 1a (M1a, NearlyNeutral) allows two $\omega$ categories estimating $0<\omega_0<1$ from the data and fixing $\omega_1=1$, thereby excluding the possibility of positive selection. Model 2a (M2a, PositiveSelection) allows for positive selection ($\omega>1$) by allowing three $\omega$ categories; $\omega_0<1$ and $\omega>1$ estimated from the data and $\omega_1=1$ fixed. Model 7 (M7, beta) assumes a beta distribution of $\omega$ sites limited between 0 and 1. Model 8 (M8,
beta + omega) adds an extra site class to M7 which is estimated from the data. Positive selection was determined by comparing M1a vs. M2a and M7 vs. M8 using the Likelihood Ratio Test (LRT) statistic against a chi-square distribution with two degrees of freedom (Yang, 1997; Yang & Nielsen, 2002). The Bayes Empirical Bayes (BEB) approach was used to identify posterior probabilities for site-classes and to determine which sites have experienced positive selection (ω>1) with greater than 90% probability.

**Results**

*SPF is present in two North American newts*

The degenerate PCR approach using cDNA preparations from *T. granulosa* and *N. viridescens* resulted in sequences with a high degree of similarity to the SPF gene produced by *C. pyrrhogastor* and *C. ensicauda*. These sequences were expressed only in males collected during the breeding season. *N. viridescens* males were found to express SPF-like sequences in the hedonic (genial) gland in addition to the abdominal gland region. PCR tests of glandular areas on the face of male *T. granulosa* in breeding condition were negative, as were PCR tests on tissue from non-breeding condition males from both species. Additionally, PCR tests of cDNA derived from *N. viridescens* females were negative for SPF expression. BLASTx analyses of *T. granulosa* and *N. viridescens* sequences returned *C. pyrrhogastor*, *Triturus carnifex*, and *C. ensicauda* sodefrin precursor sequences as the top hits, with identity
levels between 62% and 68%. The open reading frame for *Taricha granulosa* and *N. viridescens* is 199 and 200 amino acids, respectively. This size is comparable to the 189 amino acid open reading frame of the SPF protein from *C. pyrrhogaster*. An alignment of the deduced amino acid sequences (Figure 1) shows a high degree of identity throughout the protein with strong conservation of cysteine residues, essential for tertiary protein structure. These observations support the idea that sequences amplified from *T. granulosa* and *N. viridescens* are orthologous to the SPF protein expressed in males of *C. pyrrhogaster*, which acts as a female-attracting pheromone in its cleaved form.

*North American newts lack a decapeptide*

Despite similarities to *C. pyrrhogaster* SPF protein, the region of the SPF sequences of *T. granulosa* and *N. viridescens* where an active decapeptide would be expected is highly dissimilar to the sodefrin decapeptide (Figure 1). Divergence might have lead to a very different peptide pheromone, but these sequences also lack identifiable cleavage sites that would lead to excision of a peptide.

*Structural notes*

The number of cysteine residues varies across species but most residues are conserved throughout the protein. *N. viridescens* and *T. granulosa* share 19 cysteine residues located at amino acid sites 20, 23, 26, 34, 41, 48, 69, 78, 97, 98, 103, 122,
125, 134, 142, 149, 171, 177, and 195 (Figure 1). Results from DISULFIND (A. Vullo and P. Frasconi; PredictProtein) indicate with high probability values that all cysteines form disulfide bonds except number 19 at position 195. The SPF protein from *C. pyrrhogaster* contains 14 cysteine residues but none in the C-terminal end of the protein after position 134 (Figure 1). All are predicted to participate in disulfide bond formation except the 15th cysteine present only in the aonirin variant. This added cysteine is located in the penultimate amino acid position and is speculated to be the location of a disulfide-linked dimer of stored (inactive) precursor protein (Nakada et al., 2007). Thirteen of 14 cysteine residues in *C. pyrrhogaster* and 13 of 13 cysteine residues of *C. ensicauda* align perfectly with *T. granulosa* and *N. viridescens*. Ten cysteine residues are conserved across the six species compared including the
Figure 1 - Deduced amino acid sequences of representative Sodefrin Precursor Factor (SPF) proteins. The sequence labeled C. pyrrhogaster is an SPF protein found in the Chiba area of Japan (accession no. AJ245955). C. pyrrhogaster* refers to the variant (aonirin, accession no. AB280579) found in the Nara region of Japan. Yellow highlighted areas show the position of cysteine residues. Red highlighted areas show amino acid sites that have experienced strong positive selection with >90% probability (>60% in D. ocoee; from (Palmer et al., 2007)). The orange box highlights the decapeptide that must be cleaved from the SPF protein before acting as a female.
attractant. The total number of residues is listed at the end of each sequence. plethodontid D. ocoee. Among N. viridescens, T. granulosa and D. ocoee, there are an additional four conserved cysteine residues at positions 125, 142, 149, and 171.

Attempts to predict secondary and tertiary structural elements resulted only in very weak models with low confidence measures. However, given the predicted cysteine bonding, SPF proteins are expected to be tightly folded.

*Sequences are highly variable*

Sequencing of 60 full-length clones derived from cDNA preparations (N=5 species) resulted in a total of 39 unique deduced amino acid sequences. Phylogenetic analysis indicates species-specific grouping of expressed SPF sequences (Figure 2). Average interspecific nucleotide dissimilarity for all sequences was 17.2%, with 206 of 697 sites conserved across all sequences. Interspecific protein dissimilarity averaged 27.4%. Within each species, nucleotide dissimilarity ranged from 1.7% to 16.8% while protein dissimilarity was higher with a range from 5.3% to 28.8%, with 63 of 202 sites conserved across all five species. Conserved amino acids include 13 cysteine residues, some of which are likely participants in disulfide bond formation and are therefore essential for proper tertiary protein structure. Conservation was much higher toward the amino terminal end of the protein and was very low in the carboxyl decapptide region of sodefrin (Figure 1).
Within lineages, sequence diversity was highest in *N. viridescens*. Of the 31 sequences obtained from 9 individuals of *N. viridescens*, there were a total of 28 unique nucleotide sequences and 22 unique deduced amino acid sequences (Table 1). These sequences had an average nucleotide dissimilarity of 8%. The highest sequence diversity of 23.2% was observed within the seven *N. viridescens* deduced amino acid sequences (N=5 individuals) obtained from abdominal glands. In contrast, hedonic (genial) gland sequences and genomic sequences both showed approximately 12% sequence dissimilarity at the protein level (Table 1). The SPF gene tree suggests a gene duplication produced two distinct groups of *N. viridescens* SPF genes, both of which contain hedonic and abdominal gland region sequences (Figure 2).

The 16 *T. granulosa* sequences obtained from 8 individuals had an overall nucleotide sequence dissimilarity of 3.5% with 11 unique nucleotide sequences and 7 unique protein translations (Table 1). The two genomic sequences represent the majority of the diversity seen in *T. granulosa*. The 14 sequences obtained from abdominal gland region cDNA had a nucleotide sequence dissimilarity of 3.5%, while the two genomic sequences were approximately 19% dissimilar from one another and from abdominal gland sequences. The *T. granulosa* sequences form a sister group to the *N. viridescens* sequences (Figure 2).

Diversity levels were also determined for nine full-length sequences derived from abdominal gland total RNA of *C. pyrrhogaster*. Two sequences are available
from the GenBank database (accession nos. \textbf{AJ245955} and \textbf{AB280579}), five sequences were amplified and cloned at Oregon State University, and two additional sequences (Iwata et al., 2004) were provided by the Kikuyama laboratory (Waseda University, Japan).
**Figure 2**- Gene tree of expressed SPF sequences generated by parsimony analysis (PAUP v4.01b). Phylogeny is rooted using *D. ocoee* SPF as an outgroup. Bootstrap support (>50%) is shown. Sequences obtained from GenBank are indicated by accession numbers. *N. viridescens* indicates sequence was derived from abdominal gland region cDNA. *N. viridescens* indicates sequence was derived from hedonic cheek gland cDNA.

These nine sequences are derived from between 3-6 individual animals and are from three separate populations. Of the nine sequences, there are seven unique nucleotide sequences and seven unique translations (Table 1). The degree of sequence diversity in *C. pyrrhogaster* is similar to that observed in *T. granulosa* with 2.9% nucleotide sequence dissimilarity and 6.6% amino acid dissimilarity. Approximately half of the observed diversity is due to the presence of one sequence (GenBank accession no. AB280579). This sequence contains a newly identified regional sodefrin variant termed aonirin (Nakada et al., 2007). The precursor to this variant shares 86.2% amino acid identity with the SPF protein first identified in newts of the Chiba area of Japan (GenBank accession no. AJ245955).

**SPF was amplified from the genome of diverse salamander families**

PCR using species-specific primers was performed on genomic DNA preparations from *T. granulosa* and *N. viridescens*. Due to the lack of introns within the genomic structure of the SPF gene, we were able to amplify the full coding region in a single PCR reaction. Thus, this approach was expanded to amplify SPF in distantly related genera and families. Using degenerate primers designed from the conserved untranslated regions of *C. ensicauda* (GenBank accession no. AJ245956.1)
and Plethodontid *D. ocoee* SPF (GenBank accession no. **DQ097065.1**), we were able to amplify full-length genomic sequences from *Neurergus kaiseri* and *P. waltl*. Additional degenerate primers and a nested PCR approach were used to amplify partial sequences (lacking the 5’ end) from the genome of the following species: *Amphiuma means*, *Amphiuma tridactylum*, *Siren intermedia*, *Mertensiella luschani*, *Pachytriton labiatus*, *Paramesotrition hongkongensis*, *Tylototriton shanjing* and *Cynops orientalis*.

Genomic SPF sequences varied in the level of diversity observed. Two full-length genomic sequences obtained from *Taricha granulosa* were 19.2% dissimilar at the nucleotide level and 31.4% dissimilar at the amino acid level. This dissimilarity is striking compared to the 0.5% nucleotide dissimilarity and 5.9% amino acid dissimilarity observed in the 14 *T. granulosa* sequences derived from abdominal gland region cDNA (Table 1). In contrast, the five genomic sequences obtained from two *Notophthalmus viridescens* individuals showed a more modest amount of diversity with 6.1% dissimilarity at the nucleotide level and 12.1% at the amino acid level. This dissimilarity is comparable to the level of diversity observed in the 19 (N=9 animals) sequences cloned from *N. viridescens* hedonic gland cDNAs and less than 12% nucleotide dissimilarity and 23.2% amino acid dissimilarity observed in the 7 *N. viridescens* sequences derived from abdominal gland region cDNAs. Genomic sequences derived from *Neurergus kaiseri* and *Pleurodeles waltl* both showed
relatively high sequence dissimilarity at the protein level with 28.8% and 21.1%, respectively (Table 1). Genomic SPF sequences obtained by nested degenerate PCR resulted in sequences that clearly belong to the SPF gene family. However, the variability of these genomic sequences within a single species resulted in a gene tree with no support for monophyly at the species level (data not shown).
Table 1- Species sequence diversity from Sodefrin Precursor Factors (SPF) genes in five species. The sample size ‘N’ refers to the number of animals used to obtain sequences. Nucleotide dissimilarity was calculated as the number of changes per site. Amino acid translations were determined using standard computer methods.

<table>
<thead>
<tr>
<th>Species</th>
<th>N</th>
<th>Source</th>
<th>Cloned Sequences</th>
<th>Unique Nucleotide Sequences</th>
<th>% Nucleotide Sequence Dissimilarity</th>
<th>Unique Translations</th>
<th>% Amino Acid Dissimilarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>T. granulosa</td>
<td>8</td>
<td>combined</td>
<td>16</td>
<td>11</td>
<td>3.5 ± 0.3</td>
<td>7</td>
<td>5.3 ± 0.7</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>abdominal gland cDNA</td>
<td>14</td>
<td>7</td>
<td>0.5 ± 0.1</td>
<td>5</td>
<td>5.9 ± 2.2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>genomic DNA</td>
<td>2</td>
<td>2</td>
<td>19.2 ± 2.0</td>
<td>2</td>
<td>31.4 ± 4.3</td>
</tr>
<tr>
<td>N. viridescens</td>
<td>9</td>
<td>combined</td>
<td>31</td>
<td>28</td>
<td>8 ± 0.8</td>
<td>22</td>
<td>15.1 ± 1.8</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>abdominal gland cDNA</td>
<td>7</td>
<td>7</td>
<td>12 ± 0.9</td>
<td>7</td>
<td>23.2 ± 2.4</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>hedonic cDNA</td>
<td>19</td>
<td>17</td>
<td>6.1 ± 0.6</td>
<td>14</td>
<td>12.0 ± 1.4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>genomic DNA</td>
<td>5</td>
<td>5</td>
<td>6.1 ± 0.6</td>
<td>4</td>
<td>12.1 ± 1.7</td>
</tr>
<tr>
<td>N. kaiseri</td>
<td>1</td>
<td>genomic DNA</td>
<td>2</td>
<td>2</td>
<td>16.8 ± 1.7</td>
<td>2</td>
<td>28.8 ± 3.8</td>
</tr>
<tr>
<td>P. waltl</td>
<td>1</td>
<td>genomic DNA</td>
<td>3</td>
<td>3</td>
<td>9.5 ± 0.9</td>
<td>3</td>
<td>21.1 ± 2.9</td>
</tr>
<tr>
<td>C. pyrrhogaster</td>
<td>3-6</td>
<td>abdominal gland cDNA</td>
<td>9</td>
<td>7</td>
<td>2.9 ± 0.4</td>
<td>7</td>
<td>6.6 ± 1.2</td>
</tr>
</tbody>
</table>
Salamandridae SPF sequences have evolved in response to positive selection

Site-specific models of selection were fit to *N. viridescens*, *T. granulosa*, *C. pyrrhogaster*, and a combined dataset and compared using Likelihood-Ratio Tests (LRT). Five codon-substitution models were used to characterize the modes of selection acting at each amino acid site along the SPF gene. Models of selection allowing only for neutral and purifying selection (M0, M1a, and M7) could be excluded in favor of models allowing for positive selection (M2a and M8) in *N. viridescens* and *C. pyrrhogaster* but not in *T. granulosa* (Table 2). The combined dataset contained 60 sequences from five species. Parameters from models M2a and M8 indicate that approximately 4.6% of amino acid sites within the SPF gene have experienced positive selection with $\omega=2.6$ (Table 2). The dataset of *N. viridescens* contained 31 sequences (9 animals) from mRNA derived from two gland-secreting areas as well as genomic DNA preparations. Models M2a and M8, both allowing for positive selection ($\omega^+$), provided the best fit to the data, indicating that approximately 4.5% of sites have experienced positive selection (>90% probability; Table 2) with $\omega=3.66$. *N. viridescens* sequences were also divided by tissue type (hedonic cheek gland or abdominal gland) for separate site-specific positive selection tests. The dataset containing sequences obtained from abdominal gland region cDNA (N=7 sequences) were best fit by models allowing for positive selection. These models (M2a and M8) indicate that approximately 3% of sites within the gene have evolved
under positive selection with $\omega=5.04$. The following nine amino acid sites throughout the *N. viridescens* SPF protein were identified to have experienced strong positive selection: 88, 105, 131, 136, 139, 156, 164, 168, and 169 (Figure 1). Separate analyses of hedonic cheek gland sequences ($N=19$) showed no significant differences between models allowing for positive selection (M2a and M8) and those models constraining the data to neutral and purifying selection. Similarly, the *T. granulosa* dataset, comprised of sixteen sequences (14 abdominal gland cDNA and 2 genomic DNA), did not support positive selection as a strong mechanism in the populations studied; models allowing for positive selection did not provide a significantly better fit to the data ($LRT=-2.25$ with 2 d.f., $0.5<\chi^2<0.1$). Analysis of the *C. pyrrhogaster* dataset ($N=9$ sequences) indicated that 5.8% (>90% probability) of sites are under very strong positive selection with $\omega=22.61$. The twelve sites showing strong and very rapid levels of positive selection occur in the second half of the protein at sites 110, 118, 156, 158-163, 181, 182, and 185 (Figure 1). The pattern of site-specific positive selection is illustrated in Figure 3. Here we can see that the majority of sites have evolved under neutrality constrained by purifying selection. Interspersed throughout the precursor protein are sites that show an elevated probability of positive selection having occurred. Most apparent is the group of positively selected sites ($\omega^+$), -NWFDV-, located approximately 30 amino acids upstream of the decapeptide sodefrin (Figure 3).
**Table 2** - Nonsynonymous to synonymous substitution rate ratios \((\omega)\), prior probabilities \((p(\omega))\), and log-likelihood values \((\ln L)\) for selection models fitted to newt Sodefrin Precursor Factor (SPF) genes. Site specific models for positive selection were implemented using CODEML within the program PAML (v3.15). Positive selection is diagnosed using the omega ratio and the likelihood-ratio test with two degrees of freedom for the comparison of Model 1a to Model 2a. Model 7 and Model 8 were also implemented with similar results and are not shown. ‘p’ denotes the proportion of sites falling within the specified omega \((\omega)\) categories. Percent positive selection indicates the percent of sites across SPF gene estimated to have experienced positive selection with greater than 90% probability. Combined dataset includes sequences from *N. viridescens*, *T. granulosa*, *P. waltl*, *Neurergus kaiseri* and *C.pyrrhogaster* with ‘N’ referring to the number of sequences in each analysis.

<table>
<thead>
<tr>
<th>Species</th>
<th>N</th>
<th>Model 1a</th>
<th>Model 2a</th>
<th>(\chi^2)</th>
<th>% positive selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined Dataset (N=5 species)</td>
<td>60</td>
<td>(p(\omega=0.07) = 0.39)</td>
<td>(p(\omega=0.07) = 0.36)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(p(\omega=1.00) = 0.61)</td>
<td>(p(\omega=1.00) = 0.47)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\ln L = -3549.04)</td>
<td>(\ln L = -3534.00)</td>
<td>&lt;0.001</td>
<td>4.1%</td>
</tr>
<tr>
<td><em>N. viridescens</em> abdominal, hedonic, and genomic</td>
<td>31</td>
<td>(p(\omega=0) = 0.40)</td>
<td>(p(\omega=0.00) = 0.36)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(p(\omega=1) = 0.60)</td>
<td>(p(\omega=1.00) = 0.48)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\ln L = -2608.47)</td>
<td>(\ln L = -2589.01)</td>
<td>&lt;0.001</td>
<td>4.5%</td>
</tr>
<tr>
<td><em>N. viridescens</em> abdominal gland</td>
<td>7</td>
<td>(p(\omega=0) = 0.40)</td>
<td>(p(\omega=0.00) = 0.34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(p(\omega=1) = 0.60)</td>
<td>(p(\omega=1.00) = 0.55)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\ln L = -1978.51)</td>
<td>(\ln L = -1961.04)</td>
<td>&lt;0.001</td>
<td>3.0%</td>
</tr>
<tr>
<td><em>N. viridescens</em> hedonic gland</td>
<td>19</td>
<td>(p(\omega=0) = 0.32)</td>
<td>(p(\omega=0.06) = 0.36)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(p(\omega=1) = 0.68)</td>
<td>(p(\omega=1.00) = 0.54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\ln L = -1512.15)</td>
<td>(\ln L = -1510.79)</td>
<td>NS</td>
<td>N/A</td>
</tr>
<tr>
<td><em>T. granulosa</em></td>
<td>16</td>
<td>(p(\omega=0) = 0.28)</td>
<td>(p(\omega=0.00) = 0.39)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(p(\omega=1) = 0.72)</td>
<td>(p(\omega=1.00) = 0.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\ln L = -1416.46)</td>
<td>(\ln L = -1415.33)</td>
<td>NS</td>
<td>N/A</td>
</tr>
<tr>
<td><em>C. pyrrhogaster</em></td>
<td>9</td>
<td>(p(\omega=0) = 0.24)</td>
<td>(p(\omega=0.00) = 0.86)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(p(\omega=1) = 0.76)</td>
<td>(p(\omega=22.61) = 0.14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\ln L = -1127.30)</td>
<td>(\ln L = -1109.74)</td>
<td>&lt;0.001</td>
<td>5.8%</td>
</tr>
</tbody>
</table>
Analysis of *C. pyrrhogaster* partial sodefrin precursor sequences

An additional 100 sequences from the *C. pyrrhogaster* dataset containing the region exhibiting high levels of positive selection were analyzed further to determine if changes at that location are associated with the aonirin variant or show other regional specificity. For the purposes of this analysis, we looked at a five amino acid region that showed the highest posterior probability levels and variability. Extreme variation was present in this region of the protein in all three populations and was not specifically linked to a particular sodefrin decapeptide variant (see Table 3). From the Niigata region of Japan (N=43 sequences) there were five unique variants in this portion of the protein, two of which were unique to this population. The Nara region of Japan (N=26 sequences) also had five variants at this location, three of which were unique to this population. Both sequences coding for the aonirin variant shared -KKLNA- at this region as did one additional sequence that coded for sodefrin. The –KKLNA- variant was also shared by three sequences obtained from the Niigata population. Sequences from the Chiba area of Japan (N=31) only had two variants in this part of the protein, neither unique to the region. The first variant –KMLSA- is shared with sequences from the Niigata population and was present in five sequences from Chiba, all of which coded for sodefrin. The other 26 sequences contained variant –FWSYF-. Seventeen of these sequences coded for sodefrin, eight coded for the
biologically inactive decapeptide, [Asn$^{10}$]sodefrin (see Nakada et al., 2007), and one sequence coded for [Asp$^{10}$]sodefrin, which has not been tested for biological activity.
Figure 3 - Identification of sites under positive selection in SPF proteins from *C. pyrrhogaster* (Part A) and *N. viridescens* (Part B). Deduced amino acids were derived from sequences amplified from abdominal gland cDNA. Colored bars indicate the probability that a site falls within three possible site categories. White bars indicate purifying selection with an omega category of 0.00. Gray bars indicate neutral selection with an omega category of 1.00 and orange bars indicate positive selection (omega=22.61 for *C. pyrrhogaster* and 5.04 for *N. viridescens*). *C. pyrrhogaster* sites show no white bars since the model collapsed to two categories. Posterior probabilities were calculated using codeml NsSites Model 2a as implemented in PAML (v3.15).
Figure 3 (continued)
Comparison to Plethodontid SPF

Lineage-specific differences in the strength and presence of positive selection are shared between plethodontid and salamandrid SPF and PRF genes. An analysis of SPF across plethodontid salamander lineages indicates that as few as 3.2% and as many as 28% of sites may have been shaped by positive selection ($\omega^+$). Positive selection in salamandrid SPF falls within this range. SPF has been most well characterized in the plethodontid salamander *D. ocoee*. In this species, between 6.6% and 14% ($\omega=10.53$) of sites are predicted to have evolved under positive selection (Palmer et al., 2007). The 25 sites predicted to have evolved under positive selection (identified under CODEML model M3 with >60% probability; (Palmer et al., 2007)) are indicated in Figure 1. Since Palmer et al. (2007) use a different model (M3 instead of M2a or M8, all in PAML v3.14-v3.15) and lower probability cut-off levels, it is difficult to make a direct comparison of $\omega^+$ sites between plethodontids and salamandrids. In general, it appears that for *D. ocoee* (data not shown) and *N. viridescens* (Figure 3), individual sites dispersed throughout the SPF protein have been affected and posterior probability levels are variable throughout the remainder of sites. In contrast, $\omega^+$ sites within *C. pyrrhogaster* occur in clusters exclusively in the second half of the SPF protein. Furthermore, posterior probability levels, indicating neutral evolution constrained by purifying selection, are extremely uniform across the remaining sites within the precursor protein (Figure 3).
Table 3- *C. pyrrhogaster* SPF variants from three populations. Amino acids 159-164 (site numbers as in Figure 1) are highly variable and have evolved under positive selection, but have no known function. Variants at this region are listed along with the sodefrin decapeptide variant they are associated with. Sample size ‘N’ refers to the number of sequences.

<table>
<thead>
<tr>
<th>Amino acids 159-164</th>
<th>Niigata (N=43)</th>
<th>Nara (N=26)</th>
<th>Chiba (N=31)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-KMLSA-</td>
<td>sodefrin (N=1)</td>
<td></td>
<td>sodefrin (N=5)</td>
</tr>
<tr>
<td>-FWSYF-</td>
<td>sodefrin (N=4)</td>
<td>[Asn&lt;sup&gt;10&lt;/sup&gt;]sodefrin (N=2)</td>
<td>sodefrin (N=17)</td>
</tr>
<tr>
<td>[Asn&lt;sup&gt;10&lt;/sup&gt;]sodefrin (N=2)</td>
<td></td>
<td>[Asn&lt;sup&gt;10&lt;/sup&gt;]sodefrin (N=9)</td>
<td>[Asn&lt;sup&gt;10&lt;/sup&gt;]sodefrin (N=8)</td>
</tr>
<tr>
<td>-KKLNA-</td>
<td>sodefrin (N=3)</td>
<td>sodefrin (N=1)</td>
<td>[Asp&lt;sup&gt;10&lt;/sup&gt;]sodefrin (N=1)</td>
</tr>
<tr>
<td>-NWFDV-</td>
<td>sodefrin (N=27)</td>
<td>aonirin (N=2)</td>
<td></td>
</tr>
<tr>
<td>-FWFHV-</td>
<td>sodefrin (N=7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-FWFHA-</td>
<td>sodefrin (N=8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-FWFSV-</td>
<td>sodefrin (N=1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-NWYCA-</td>
<td>sodefrin (N=5)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discussion

The molecular, morphological, and seasonal similarities between *C. pyrrhogaster* SPF and the *N. viridescens* and *T. granulosa* sequences provide strong evidence that we have identified genes coding for pheromones. Together with the identification of SPF in Plethodontid salamanders, these results illustrate how a family of genes coding for a chemical signal has been retained as a pheromone for millions of years yet has undergone drastic changes in the location of its glandular production, post-translational processing, and delivery method. North American and Asian newts exhibit similar aquatic courtship behavior, wherein pheromone is presumably released.
from a male’s cloaca and wafted toward a female via a tale-fanning action (Houck and Arnold 2003). The similarities between these newt families suggest that the SPF gene products in *T. granulosa* and *N. viridescens* may be active as a peptide (as in the genus *Cynops*) rather than a large protein (as in plethodontids). However, despite close molecular affinities to the *Cynops* spp. SPF genes and similar pheromone delivery behavior (tail-fanning), our data suggests that the pheromone is likely active in its full-length uncleaved form in both *T. granulosa* and *N. viridescens*. Positive selection analyses reveal that different selection mechanisms are acting in each species, indicating that the underlying evolutionary forces are varied and complex.

The three species-specific decapeptides, sodefrin, aonirin, and silefrin, found in the genus *Cynops*, require cleavage from precursor proteins before acquiring their ability to act as female-attracting pheromones. Although *N. viridescens* and *T. granulosa* SPF belong to the same gene family as *Cynops* spp. SPF, there is no evidence of an active peptide for these two species. The bioactive decapeptides present in the genus *Cynops* require cleavage at monobasic amino acid processing sites. To generate sodefrin and silefrin, Arg-Ser and Lys-Ile bonds must be cleaved, probably by serine proteases (Ishizuka et al., 2004). An alignment of the precursor proteins reveals a lack of identity in the region containing the decapeptide pheromones and there are no similar cleavage sites present in *N. viridescens* or *T. granulosa* SPF. The lack of cysteine residues in the carboxyl terminal of the *Cynops* SPF protein
leaves this non disulfide bonded portion of the protein open to enzymatic cleavage. The SPF sequences of *N. viridescens* and *T. granulosa* both contain an additional five cysteine residues in this region of the protein and all except the last cysteine residue are predicted to participate in the formation of disulfide bonds. The presence of the nine disulfide bonds dispersed throughout the SPF protein in these species indicates that the protein has a compact tertiary structure, thereby providing little room for the activity of bulky cleavage enzymes. Given the lack of an identifiable peptide, the lack of cleavage sites, and the presence of cysteine bonds throughout the terminal end of the protein, it is unlikely that a bioactive peptide is cleaved from the SPF protein in *T. granulosa* or *N. viridescens*. Bioactivity of a full-length uncleaved form of the SPF protein has been demonstrated in *D. ocoee* (Watts et al., in press). Based on these observations, it appears that the ancestral state of this pheromone is likely to be active in its uncleaved form and that the genus *Cynops* has evolved use of a cleaved decapeptide as its bioactive pheromone. We predict that future behavioral assays will show activity of the full-length uncleaved SPF proteins in *N. viridescens* and *T. granulosa*.

Prior work on three populations of *C. pyrrhogaster* has focused on the sodefrin decapeptide region and has revealed the presence of multiple variants, some active and others apparently not (Nakada et al., 2007; Iwata et al., 2004). Changes to amino acid sites within the protein could be due to neutral divergence or to adaptive evolution. In
this study, we focus on the entire precursor protein and describe for the first time how positive selection has acted to adaptively modify the *Cynops* SPF protein. The majority of the *C. pyrrhogaster* SPF gene is strikingly conserved. However, selection analyses reveal that positive selection has acted strongly to modify specific sites, resulting in a regionally specific bioactive variant found in *C. pyrrhogaster* males from the Nara region of Japan. Interestingly, site specific positive selection is very high in several of the amino acids located in the 30 amino acid sites directly upstream of the decapeptide. In particular, there is a five amino acid stretch which is not at all conserved. Analysis of partial sequences reveals that there is high diversity at this location of the protein in all three populations of *C. pyrrhogaster*. Additionally, particular amino acids at this location are not perfectly correlated to either the aonirin or the sodefrin pheromone variant (though a larger dataset of aonirin variants may reveal some correlation). The reason for such strong and rapid diversification is unclear. Since only the cleaved decapeptide region of the precursor gene is reported to be active, we would expect the remainder of the gene to evolve under neutral selection. Typically, if sites in a chemical signal have a particularly strong indication of positive selection, one would predict they are important for recognition by a receptor. This observation provides evidence that the precursor protein retains some yet unidentified ability to act as a pheromone in its uncleaved form or that other cleavage products may play an important role in communication. This example of
rapid diversification of a pheromonal component illustrates how positive selection may play a key role in speciation.

In this study, we sought to identify the pattern of $\omega^+$ sites in *T. granulosa* and *N. viridescens* and to compare that pattern to SPF from plethodontids and *C. pyrrhogaster*. The resulting picture of evolution was different for these two closely related North American newts. *N. viridescens* has a particularly interesting pheromonal system consisting of two pheromone-producing glands, with pheromones derived from each gland apparently evolving under different selective pressures. In addition to abdominal gland production of SPF, males of *N. viridescens* produce SPF proteins in their hedonic cheek glands which are rubbed over the female’s nares during courtship. Thus far, this is the only species found to produce SPF in more than one gland. While positive selection has acted to increase diversity in sequences obtained from the abdominal gland region, SPF sequences obtained from the hedonic glands have evolved, like most other classes of proteins, under neutral divergence restrained by purifying selection. Positive selection observed within the abdominal gland form of the protein indicates that rapid change at sites within these proteins has been beneficial to courting males. In contrast, the lack of positive selection on hedonic cheek glands SPF sequences implies that pheromones secreted from these glands may not have a direct affect on male mating success but rather play an accessory role in courtship. Instead, female stimulation by pheromones may be predominantly carried
out via the release of pheromones from the abdominal glands of males. If so, behavioral experiments would show that abdominal gland forms of the protein are more efficacious at stimulating females.

In contrast to the rapid evolution observed within *C. pyrrhogaster* and within abdominal gland sequences of *N. viridescens*, SPF sequences obtained from abdominal glands of *T. granulosa* have low diversity and no evidence of site-specific positive selection. However, this lack of detection could be due to a small sample size with little variation or to the small geographical area (a single pond) from which animals were sampled. When only weak levels of positive selection are acting at sites within a gene, the power of the likelihood-ratio test used to detect positive selection is low (Anisimova et al., 2002). In such cases, a larger dataset (i.e., more sequences) are required to identify positively selected sites. The *T. granulosa* datasets contained only a small number of sequences with little variation. Additional analyses including more sequences may lead to the detection of sites that have evolved under positive selection. If the lack of positive selection is a true representation of the evolution of *T. granulosa* SPF, it could imply that these pheromonal cues play a different role in *T. granulosa* courtship and that another pheromone component is acting to increase mating success. Many, if not most, pheromones are found in multi-component mixtures and the various components may play different roles. In the family Plethodontidae, PRF is produced in a multi-component mixture that also includes an unrelated protein called
Plethodontid Modulatory Factor (PMF). Behavioral tests with both components and PRF alone facilitate courtships by decreasing the time to sperm transfer. PMF, on the other hand, has the opposite effect when delivered alone, though may contribute to mating success by inducing a quiescent state during courtship (Houck et al., 2007). If a particular pheromonal component is not related to mating success and instead affects some other aspect of courtship (such as fertility of mates or survivorship of the male or female) evolution is expected to be controlled by natural selection (Arnold & Houck, 1982) and the rapid change of amino acids is not predicted. In plethodontid species that produce both SPF and PRF, only one or the other pheromone is subject to strong positive selection. These differences in positive selection are predicted to be due to the replacement of one pheromone by the other, since both SPF and PRF elicit the same behavioral response (Palmer et al., 2007).

Genomic PCR reveals the widespread occurrence of a family of SPF-like proteins in urodeles. The high diversity and lack of monophyly observed in genomic sequences suggests that gene duplication may have played a role in the evolution of at least some of the observed genomic sequences. Thus, these sequences may represent a family of related proteins with a separate function. Analysis of cDNA libraries from abdominal glands and PCR from other organs will be useful in determining the presence and potential role of a more diverse family of SPF-like proteins.
We expected to find that modifications to this pheromonal signal occurred in accordance with the major life-history transition from aquatic to terrestrial courtship. However, instead it appears that the selective forces shaping these changes are more complex. Despite similarities in gland production regions according to terrestrial or aquatic courtship, structural and positive selection data lead us to conclude that the use of a full-length protein rather than a decapeptide did not necessarily occur in conjunction with terrestrial courtship. Further biochemical characterizations of SPF and behavioral tests with aquatic courting newts will help to resolve these questions.

Acknowledgements

We thank C. Samuel Bradford for comments on multiple drafts of the manuscript. Thanks to Melissa Haye for laboratory assistance. The University of California, Museum of Vertebrate Zoology, provided salamander tissue for genomic PCR. Partial length *Cynops pyrrhogaster* SPF sequences were provided by the Kikuyama Lab (Waseda University, Japan) This research was funded by National Science Foundation grants 0237083-DGE and IRC DEB-0078343.

References


Arnold, S.J. 1977. The evolution of courtship behavior in New World salamanders


Contributions to the field

Molecular studies of salamander courtship pheromones have been the subject of many recent papers (Palmer et al., 2005; Palmer et al., 2007; Watts et al., 2007; Iwata et al., 2005). The data presented in this thesis contribute novel findings to a growing body of work in this field. In particular, I have focused on increasing our knowledge of pheromones in aquatic courting salamandrids in order to compare what is known about SPF proteins acting as pheromones on both sides of the major life-history transition from aquatic to terrestrial courtship.

Many of my results highlight the similarities between pheromone components from a wide variety of species. All salamander pheromones studied to date show high levels of diversity, polymorphism, gene duplication, and positive selection in some lineages and not others. However, I also present several novel and unexpected results. This is the first study to demonstrate multiple pheromone producing glands in a single species, Notophthalmus viridescens. Furthermore, it is of interest to note that the pheromones produced in each gland are apparently subject to different selective pressures. Additionally, this is the first study to demonstrate our ability to amplify SPF sequences using genomic DNA and PCR techniques. The lack of introns in the SPF gene and success with degenerate primers allow for detection of pheromone
sequences in highly diverse salamander families. Lastly, it is interesting to find intense positive selection in areas of the *C. pyrrhogaster* SPF protein reported to lack biological activity. These findings suggest that pheromonal signaling in *Cynops spp.* deserves further study to elucidate possible pheromonal functions of either the full-length precursor protein or other cleaved peptide products.

**Future directions**

These studies provide the framework for additional work on pheromones in aquatic courting newts. In order to better understand this complex signaling system, biochemical characterization of the SPF proteins coupled with behavioral tests are necessary. Additionally, it is important that behavioral tests are designed to clarify the difference between an attractant (as described for *C. pyrrhogaster*) and a pheromone that modifies the behavior of a female during courtship (as with Plethodontid SPF and PRF). While studies in *C. pyrrhogaster* and *C. ensicauda* have revealed the ability of sodefrin and its variants and silefrin to attract a female’s attention, the presentation of pheromone occurred in the absence of a courtship encounter when a female would normally receive such a signal. Thus, other tactile and visual stimuli were missing and there was no opportunity to observe how a female might respond in the context of a courtship encounter. Since pheromones are delivered after a male has encountered and initiated courtship with a female, I predict that the primary function of SPF
pheromones in *T. granulosa* and *N. viridescens* is to modify a female’s behavior during courtship.

Lastly, in order to gain a better understanding of the evolutionary forces shaping pheromones, it is important to better characterize female reception. The identification of female pheromone receptors in the vomeronasal organ will allow for studies focused on the co-evolutionary interactions between male signal and female response.

**References**


BIBLIOGRAPHY


