Comparative Proteomics Reveals Evidence for Evolutionary Diversification of Rodent Seminal Fluid and Its Functional Significance in Sperm Competition

Steven A. Ramm,†1 Lucy McDonald,‡1 Jane L. Hurst,* Robert J. Beynon,‡ and Paula Stockley*2

*Mammalian Behaviour and Evolution Group, Department of Veterinary Preclinical Science, University of Liverpool, Leahurst Campus, Neston, Cheshire, United Kingdom; and ‡Proteomics and Functional Genomics Group, Department of Veterinary Preclinical Science, University of Liverpool, Liverpool, United Kingdom

During insemination, males of internally fertilizing species transfer a complex array of seminal fluid proteins to the female reproductive tract. These proteins can have profound effects on female reproductive physiology and behavior and are thought to mediate postcopulatory sexual selection and intersexual conflict. Such selection may cause seminal fluid to evolve rapidly, with potentially important consequences for speciation. Here we investigate the evolution of seminal fluid proteins in a major mammalian radiation, the muroid rodents, by quantifying diversity in seminal fluid proteome composition for the first time across a broad range of closely related species. Using comparative proteomics to identify and cross-match proteins, we demonstrate that rodent seminal fluid is highly diverse at the level of both proteomes and individual proteins. The striking interspecific heterogeneity in seminal fluid composition revealed by our survey far exceeds that seen in a second proteome of comparable complexity, skeletal muscle, indicating that the complement of proteins expressed in seminal fluid may be subject to rapid diversification. We further show that orthologous seminal fluid proteins exhibit substantial interspecific variation in molecular mass. Because this variation cannot be attributed to differential glycosylation or radical differences in termination sites, it is strongly suggestive of rapid amino acid divergence. Sperm competition is implicated in generating such divergence for at least one major seminal fluid protein in our study, SVS II, which is responsible for copulatory plug formation via transglutaminase-catalyzed cross-linking after insemination. We show that the molecular mass of SVS II is positively correlated with copulatory plug anisodactylancy, and may thus conflict with female interests (Chapman 2000; Chapman et al. 2003a). The functional and evolutionary significance of seminal fluid have been most extensively explored in Drosophila, in which accessory gland protein (Acp)–induced changes in the mated female include stimulation of oviposition, mediation of sperm storage, and diminished female receptivity to remating (Wolfer 2002; Chapman and Davies 2004). Variation at Acp loci is associated with sperm competition success (Clark et al. 1995; Fiumera et al. 2005), and Acp deposition reduces female lifespan and may thus conflict with female interests (Chapman et al. 1995, 2003b; Wigby and Chapman 2005). Although less well studied, seminal fluid components in other taxa also have functions consistent with an influential role in postcopulatory sexual selection. For example, the accessory reproductive gland products of male mammals maintain sperm viability, enhance sperm motility and capacitation, stimulate the female reproductive tract, suppress female chemical and immunological challenges to sperm, and produce the copulatory plug (Shivaji et al. 1990). Additionally, seminal proteins from widely divergent taxa appear to be functionally conserved (Mueller et al. 2004; Braswell et al. 2006).

The role of sexual selection in shaping seminal fluid evolution has attracted interest because it could drive rapid divergence, which represents a potentially important source of postmating reproductive isolation during speciation (Price 1997; Rice 1998; Knowles and Markow 2001; Price et al. 2001; Andrés et al. 2008). Consistent with this idea, genes encoding many seminal proteins are subject to an unusually high rate of adaptive evolution (e.g., Swanson et al. 2001a; Clark and Swanson 2005; Andrés et al. 2006), and interlineage variation in evolutionary rates may be linked to the intensity of postcopulatory sexual selection (Dorus et al. 2004; Wagstaff and Begun 2005b; see also Hurle et al. 2007; Ramm et al. 2008). More broadly, the action of sexual selection may also lead to rapid divergence in seminal fluid proteomes. In support of this idea, there is some evidence for rapid diversification of genomic Acp complement in Drosophila (Begun and Lindfors 2005; Mueller et al. 2005; Wagstaff and Begun 2005a, 2007; Begun et al. 2006), as well as for wide interspecific differences between accessory gland protein extracts separated by electrophoresis (e.g., Thomas and Singh 1992). However, the evolutionary processes responsible for this latter finding cannot be inferred unless proteins are identified and cross-matched across species. Current proteomic methods permit precisely this, but no global survey of seminal fluid diversity at the protein level has yet been attempted in any animal group.

In this investigation, we employ comparative proteomics to uncover broad-scale patterns in the evolution of seminal fluid expression and divergence and assess the potential functional significance of seminal fluid proteins in the context of sperm competition. We focus on the secretion of the major accessory reproductive gland—the seminal vesicle—from a cross-section of muroid rodents. This is an ideal model system for such an investigation for several

1 These authors contributed equally to this work.
2 These authors contributed equally as joint senior authors.

Key words: copulatory plug, proteomics, mammals, seminal fluid, sexual selection, sperm competition.

E-mail: p.stockley@liv.ac.uk.

doi:10.1093/molbev/msn237
Advance Access publication October 17, 2008

© The Author 2008. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org
reasons. Rodents are a speciose and diverse group (Steppan et al. 2004), with a wide range of mating systems and hence varying levels of sperm competition (Dewsbury 1984). They are amenable to study because representative species have been subject to intensive phenotypic characterization and genomic sequencing due to their importance as biomedical models. Moreover, in the context of the present study, it is already established for rodents that the size of the seminal vesicles correlates with sperm competition intensity across species (Ramm et al. 2005), as does the rate of nonsynonymous substitution at the locus that encodes one of its major products, SVS II (Ramm et al. 2008), suggesting that seminal vesicle proteins may well evolve under sexual selection. Here, by identifying and cross-matching these proteins across species, we aim to test for evidence of seminal fluid divergence both at the level of primary amino acid sequences of seminal fluid constituents and in the complement of seminal fluid proteins expressed in each species.

We use peptide mass fingerprinting (PMF, Henzel et al. 2003) and tandem mass spectrometry accompanied by manual de novo amino acid sequencing (MSMS, Perkins et al. 1999), in combination with genomic information available for Mus musculus and Rattus norvegicus, to identify the major proteins secreted by rodent seminal vesicles. In PMF, a purified protein is cleaved enzymically into fragments that, when mass-measured, constitute a diagnostic “fingerprint” that can be searched against databases of known or very similar proteins (Henzel et al. 2003). This approach is particularly effective if the protein exists in the database, although the method is tolerant to a degree of cross-species matching provided the sequences are greater than 70% identical (Lester and Hubbard 2002). By contrast, de novo sequencing, a characterization method that can be conducted without prior knowledge of the amino acid sequence, selects one peptide and by the application of more complex gas phase chemistry, elicits fragmentation of that peptide in a manner that permits deduction of the amino acid sequence of the unknown peptide. De novo sequencing is particularly valuable for characterization of proteins for which little or no genomic data exist and for proteins that show a high rate of evolution and hence protein sequence divergence.

Based on our comparative proteomics analyses, we provide 1) evidence of substantial interspecific diversity in seminal fluid proteome composition among rodents; 2) evidence that orthologous proteins exhibit extensive divergence, most likely due to changes in their amino acid sequence; and 3) evidence that divergence in the molecular mass of a major seminal fluid protein, the copulatory plug–producing SVS II, correlates with divergence in mating systems, suggesting that an SVS II molecule of greater molecular mass (containing more cross-linking sites for plug formation) could be favored under higher levels of sperm competition.

**Materials and Methods**

**Sample Collection**

Seminal vesicle secretions were expelled from the lumen of both seminal vesicles of dissected reproductive tracts of multiple adult males from each species: Acomys cahirinus (Ac), Apodemus sylvaticus (As), Clethrionomys [=Myodes] glareolus (Cg), Coelomys pahari (Cp), Meriones unguiculatus (Mu), Microtus agrestis (Ma), Mus musculus castaneus (Mc), Mus musculus domesticus (Md), Mus macedonicus (Mm), Mus spicilegus (Ms), Peromyscus californicus (Pc), Peromyscus leucopus (Pl), Peromyscus maniculatus (Pm), Peromyscus polionotus (Pp), Phodopus roborovskii (Pr), Phodopus sungorus (Ps), Pseudomys australis (Pa), and R. norvegicus (Rn). Samples were taken from animals confirmed to be in reproductive condition based on testis size. These animals were wild-derived and housed under stable environmental conditions in laboratory colonies prior to data collection (except A. sylvaticus, which were wild-caught). Subsequent separation and identification of proteins were based on independently analyzing samples obtained from single individuals (i.e., there was no pooling of samples within species), and species-specific patterns were confirmed by analyzing multiple (two to seven) individuals per species in almost all cases (except Ph. roborovskii and Ph. sungorus, where only one individual per species was analyzed). Samples were either analyzed immediately or stored at −20 °C. Prior to analysis, samples were placed in 0.5 ml 40 mM Tris pH 8.8, 4 M urea, 125 mM ethylenediaminetetraacetic acid and vortexed for 1 min, then centrifuged at 13,000 × g for 5 min at 4 °C. Supernatants were removed and stored at −20 °C.

For comparison with a nonreproductive proteome, we also collected skeletal muscle samples from a more limited range of these species (R. norvegicus, A. sylvaticus, the four Mus species, C. pahari, and M. agrestis) to assess the success of PMF and MSMS in identifying and cross-matching proteins from this tissue and hence, the relative degree of interspecific divergence seen within these two sampled proteomes. Skeletal muscle samples were homogenized in 10 ml of 20 mM sodium phosphate buffer, pH 8.0, and centrifuged for 45 min at 13,000 × g. The resultant supernatant fraction, predominantly comprising soluble glycolytic proteins, was used without further purification.

The protein concentration of all samples was determined using the Coomassie Plus protein assay (Perbio Science, Tattenhall, United Kingdom). The protein preparations (10 μg) were separated through a 12.5% polyacrylamide gel according to previously documented methods (Doherty et al. 2004) and visualized with Coomassie Brilliant Blue stain.

**In-Gel Trypsin Digestion**

Gel plugs (approximately 1 mm3) containing protein bands of interest were excised from 1D gels using a glass pipette and transferred to a 96-well plate. To each well, 25 μl of 50 mM ammonium bicarbonate, 50% (v/v) acetonitrile (ACN) was added and incubated at 37 °C for 20 min. This process was repeated until all the stain had been removed. The plugs were then washed in 50 mM ammonium bicarbonate, which was subsequently discarded. To break and then prevent reformation of disulfide bonds, dithiothreitol (25 μl 10 mM) was added to each plug and incubated for 30 min at 37 °C. The supernatant was discarded and replaced...
with iodoacetamide (25 μl 55 mM), and the incubation continued in the dark for 60 min. The gel was dehydrated using 25 μl of ACN, and incubation at 37 °C was resumed for 15 min. The supernatant was removed from the dehydrated plug, which was allowed to air-dry. Once dry, the gel was rehydrated in 50 mM ammonium bicarbonate (9 μl) containing trypsin (Roche Diagnostics, Basel, Switzerland) (1 μl of 100 ng/μl trypsin stock reconstituted in 50 mM acetate acid). After 30 min, 50 mM ammonium bicarbonate (10 μl) was added to each tube, and digestion was allowed to continue overnight at 30 °C; the digestion was then halted by the addition of 2 μl formic acid.

Peptide Mass Fingerprinting

Samples were analyzed using a matrix assisted laser desorption ionization time of flight (MALDI-ToF) mass spectrometer (M®LDI; Waters-Micromass, Manchester, UK). Data were collected over the range of 900–3500 thomsons (synonymous with mass to charge ratio, m/z). The mass spectrometer was calibrated using a mixture of four peptides: des-arg bradykinin (903.47DA), neurotensin (1,671.92DA), adrenocorticotropic hormone (2,464.20DA), and insulin β chain (3,493.65DA). Peptide mixtures were mixed in a 1:1 ratio with a saturated solution of α-cyano-4-hydroxycinnamic acid (Sigma, St. Louis, MO) in ACN:water:trifluoroacetic acid (50:49:1, v/v/v) and spotted directly onto a MALDI target for analysis. Data acquisition and processing were performed through the MassLynx software suite (version 4.0). Proteins were identified from their peptide mass fingerprint by manual searching using a locally implemented MASCOT (Perkins et al. 1999) server (version 1.9) against the MSDB database (ftp://ftp.ebi.ac.uk/pub/databases/MassSpecDB). Search parameters allowed a single missed tryptic cleavage, carboxymethyl modification of cysteine (fixed), oxidation of methionine (variable), and a peptide tolerance of ±150 ppm. The taxonomic space was restricted to rodents.

Tandem Mass Spectrometry

Electrospray ionization tandem mass spectrometry was performed on a Q-ToF Micro instrument (Waters) coupled online to reversed-phase high-performance liquid chromatography from a Dionex Ultimate system (Dionex Ltd., Camberley, UK) with a PepMap C18 column (LC packings), 15 cm × 75 m, bead size 3 m, and pore size 100 Å. Prior to separation, aliquots (10–20 μl) were desalted in-line using a Dionex Switchos unit, fitted with a 1-mm × 300-μm C18 precolumn. The precolumn was initially equilibrated in 0.2% (v/v) formic acid at 30 μl/min. Peptides were then loaded and washed for 3 min at the same flow rate, after which the trap and downstream PepMap column were developed with 90% ACN/0.2% formic acid, introduced at a linear gradient of 0–50% in 50 min at 0.2 μl/min. The quadrupole mass analyzer was set to allow the passage of selected precursor ions into the gas cell, where they were fragmented by collision with argon. The masses of the resulting fragment ions were then determined by the ToF analyzer. MSMS data were processed using MassLynx software suite (version 4.0) and interpreted manually to generate de novo sequence information. Peptide sequences were searched against the protein database using NCBI protein–protein Blast (Altschul et al. 1990) using the option to search against short nearly exact matches. The taxonomic space was limited to rodents.

Deglycosylation

To investigate the degree of posttranslational modification of common seminal fluid proteins and to test for differential posttranslational modification among species, we repeated the sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) analysis for N = 14 species but with the addition of a deglycosylation step prior to separation. To achieve this, samples were treated with an enzymatic deglycosylation kit (Prozyme, San Leandro, CA). A portion (100 μg protein) of each sample was made up to 30 μl using deionized water. Samples were processed according to the manufacturer’s instructions. The control glycoprotein, bovine fetuin (included in kit), was also subjected to the same deglycosylation treatment as the seminal vesicle samples. Once incubated, 10 μg of each treated sample was separated on SDS-PAGE along with 10 μg of the corresponding untreated sample.

Phylogenetic Analysis

To further explore the potential selection pressures driving seminal fluid evolution, we conducted phylogenetically controlled comparative analyses using relative testis size as an index of the intensity of postcopulatory sexual selection. The relationship between sperm competition level and relative testis size is a well-established evolutionary trend in many animal groups (Parker et al. 1997), including rodents (Ramm et al. 2005). Analyses were conducted to compare interspecific variation in the size of one seminal fluid protein, SVS II, and overall seminal fluid complexity with relative testis size. To control for potential phylogenetic inertia in the data (nonindependence of species traits), an implementation of the phylogenetic general linear model approach of Gage and Freckleton (2003) was employed, which is equivalent to generalized least squares (Freckleton et al. 2002). Phylogenetic relationships among species in the data set were inferred from the literature (e.g., Steppan et al. 2004) and unit branch lengths assumed (Freckleton et al. 2002). Independent variables entered into the model were male body mass and testis mass, a significant effect of the latter being taken as evidence for an influence of postcopulatory sexual selection. In the first analysis, SVS II molecular mass estimates were entered as the dependent variable. These were obtained by plotting mobility (= distance traveled on gel) as a function of log_{10}(molecular weight) for each of the protein standards and then interpolating from the mobility of proteins of interest to estimate unknown masses for presumed orthologs. In the second analysis, we investigated variation in seminal fluid complexity by automatically counting the number of bands present at greater than 1% intensity in each lane using the Phoretix 1D Quantifier program (Version 4.01; Non-Linear Dynamics Ltd., Newcastle, UK) and entering this as the dependent variable.
Results
Identification and Cross-Matching of Seminal Fluid Proteins

Initial analysis by 1D gel electrophoresis revealed substantial species differences in the complexity and mass distribution of the seminal vesicle secretion among the 18 species of muroid rodents surveyed (fig. 1A). In all species, the protein profile was dominated by a small number of highly expressed proteins. PMF identified these proteins in *M. musculus* and *R. norvegicus* (supplementary table 1). However, for the remaining species, a further 143 peptides from 75 bands on the gel could only be cross-matched to known *M. musculus* and *R. norvegicus* proteins after MSMS and de novo amino acid sequencing (supplementary table 2), consistent with a high degree of sequence variation between samples. As a consequence of these analyses, 86 out of 89 (97%) major protein bands for which identification was attempted were identified by either PMF or, more commonly, de novo sequencing conducted manually. As further confirmation of the success of manual de novo sequencing, after this study was completed, cDNA sequences were obtained independently for Svs2 in six species common to this study (Ramm et al. 2008) and in all cases confirmed that our de novo sequencing data and subsequent identifications were 100% accurate (other than the caveat that MS-based sequencing cannot readily discriminate the isobaric pair leucine and isoleucine). This comprehensive proteome profiling is summarized in table 1 and fig. 1B. The majority of these major seminal fluid proteins are products of the rapidly evolving substrate for transglutaminase (REST) gene family (Lundwall and Lazure 1995; Lin et al. 2005).

Comparative Diversity of the Seminal Vescile Proteome

To assess the relative rate of diversification in seminal fluid proteomes, we compared levels of diversity from our survey with those seen in a second tissue proteome, derived from soluble proteins of skeletal muscle, which largely comprises enzymes responsible for glycolysis. This sample has a similar degree of complexity to the seminal vesicle proteins, being dominated by relatively few, high-abundance proteins. In contrast to seminal fluid proteomes, though, the protein profiles in the 1D muscle gel (fig. 2A) show remarkable similarity across all species tested—a pattern of similarity that is preserved throughout an extended group of more distantly related mammalian species (McDonald L, Beynon RJ, unpublished results). For example, beta enolase and creatine kinase could be identified in every rodent species of muroid rodents surveyed (fig. 1A). Bands labeled with the same number/letter are orthologous proteins (1: SVS I, 2: SVS II, 3: SVS III, 4: SVS IV, 5: SVS V, 6: SVS VI, 7: SVS VII, A: albumin, C: creatine kinase, F: FAD-dependent sulfhydryl oxidase, and Y: cytochrome c; see supplementary tables 1 and 2 for identifications). Bands lacking a label were either not successfully cross-matched or cross-matching was not attempted; bands with multiple labels produced hits to different proteins, most likely because they contain multiple proteins of very similar molecular mass.

Diversity in the Molecular Mass of Seminal Vescile Proteins

The gel analysis of the seminal vesicle proteins demonstrated that for some of these, there was considerable species-to-species variation in migration through the gel. This
could be attributable to differences in the length of the primary protein sequence or to size modifying posttranslational modification, particularly glycosylation. The only proteins for which there was good evidence for glycosylation were the CEACAM proteins that were present in a minority of species (supplementary table 2). When seminal vesicle proteins were treated with endoglycosaminidases to remove carbohydrate, the protein bands corresponding to the CEACAM protein in Mi. agrestis and Ph. sungorus both showed a substantial decrease in mass. This is not surprising as CEACAM proteins are known to be heavily N-glycosylated in vivo (Williams et al. 1991). Other than the CEACAM proteins, no other proteins changed in mobility after deglycosylation (supplementary fig. 1). Hence, there is no evidence that a major form of posttranslational modification (glycosylation) is responsible for the observed pattern of seminal fluid protein mass divergence associated with either SVS II (see below) or any of the other SVS proteins identified across the species set.

The fact that de novo sequencing was required to identify seminal vesicle proteins in all species other than M. musculus and R. norvegicus, and that in each case, only a minority of sequenced peptide fragments retained sufficient sequence conservation to match to orthologous sequences in these model species (supplementary table 2; e.g., SVS II, supplementary fig. 2), strongly suggests that these proteins exhibit amino acid sequence divergence. There was no evidence for divergence caused by mutations in initiation or termination codons, because peptides that could be cross-matched to model species were dispersed throughout the length of the protein, including at the N and C termini (e.g., SVS II in supplementary fig. 2). We therefore infer that the size divergence in proteins that were identified across multiple species is due to amino acid variance throughout the entire sequence. In contrast to wide variation between species (fig. 1), we found no evidence for polymorphism within individual samples (N = 2–7) from any one species, whether assessed in terms of gel banding patterns or MALDI-ToF fingerprints (results not shown).

Postcopulatory Sexual Selection and Seminal Vesicle Proteins

Next, we investigated the functional significance of molecular mass diversity found in the most taxonomically widespread of the seminal vesicle proteins identified from our survey, SVS II (fig. 3A). This protein cross-links after ejaculation in the presence of a prostate-derived transglutaminase, causing the ejaculate to coagulate and form a copulatory plug occluding the female reproductive tract. The cross-linking region is characterized by the presence of short (typically six to seven amino acids) repeat motifs containing a glutamine (Q) residue at the first position and a lysine (K) residue at the third position, and it is these that serve as a substrate for transglutaminase and thus form the cross-links involved in plug formation (Lundwall 1996). Previous reports in both rodents (Lundwall 1996) and primates (Hurle et al. 2007) have noted species differences in the primary amino acid sequence of coagulation proteins affecting the number of sites available for cross-linking. We therefore hypothesized that the diversity in the molecular mass of rodent SVS II revealed by our proteomics survey could arise through divergence in the
number of cross-linking sites and that this divergence could be driven by sperm competition.

Two lines of evidence suggest that the majority of molecular mass divergence we report for SVS II does indeed occur through insertions and deletions in the cross-linking region. First, our identification procedure predominantly cross-matched peptides for SVS II at the less variable N- and C-terminal regions (supplementary fig. 2), as predicted if the amino acid sequence of the central repetitive region is less well preserved. Second, when we analyzed molecular mass variation in the six rodent species for which full cDNA sequence data are available (Mus musculus domesticus, M. m. castaneus, M. m. macedonicus, M. m. spicilegus, C. pahari, A. sylvaticus, and R. norvegicus; data from Ramm et al. 2008), we found substantial interspecific variation in repeat number (26–35 Q–K repeats, supplementary fig. 3) and a significant positive correlation between the repeat number found in each species and the corresponding molecular mass of SVS II estimated from our data (Pearson correlation: \( N = 6; r = 0.91; P = 0.01 \)). Molecular masses for these proteins calculated by SAPS (Brendel et al. 1992, implemented at http://www.ncbi.nlm.nih.gov/Tools/saps) based on the corresponding cDNA sequence from Ramm et al. (2008) confirm the same positive correlation with repeat number (\( r = 0.98, P < 0.001 \)) and also confirm that our estimated molecular masses correlate very closely with those expected based on the full cDNA sequence (\( r = 0.88, P = 0.02 \)).

To assess whether sperm competition could drive divergence in SVS II, we next asked whether species with larger relative testis size tend to have an SVS II molecule of greater molecular mass (and presumably more sites available for cross-linking during plug formation). As predicted, we found a significant positive correlation between relative testis size and the molecular mass of the SVS II protein (\( N = 17, r = 0.61, P < 0.01 \); fig. 3B) and, after control for body mass, a significant effect of testis mass on SVS II molecular mass in a phylogenetically controlled general linear model (\( t = 2.76, P = 0.015 \); table 2, a).

In addition to affecting individual proteins, postcopulatory sexual selection could also influence overall levels of protein complexity in seminal fluid (Poiani 2006). However, in contrast to the likely influence of sperm competition detected on SVS II, relative testis size did not correlate with overall levels of seminal fluid complexity, as measured by variation in the number of major protein bands present (table 2, b). Given that we focused on major protein bands in each species—and in reality, there will be many more minor proteins present (e.g., Pilch and Mann 2006)—this is perhaps unsurprising (see also Poiani 2006), because there is relatively little variation between species in the number of abundant proteins present (notwithstanding substantial interspecific variation in their identity).

Discussion

Evolutionary Dynamics of Seminal Fluid Proteomes

By identifying and cross-matching seminal fluid proteins for the first time across a broad range of closely related species, our study reveals substantial interspecific diversity in rodent seminal fluid proteomes, with most species expressing a unique complement of major seminal vesicle proteins. This diversity is evident despite apparent conservation in the number of major proteins expressed, because most species typically express only a few (three to seven) major proteins. Although "missing" proteins in a particular species could in some cases still be expressed at low levels—and therefore not be included in our comparison—the level of diversity in seminal fluid revealed by our study is nevertheless striking. Differences in seminal fluid composition between domesticated mammals have been described previously (Shivaji et al. 1990), but by extending our knowledge to a much greater number but taxonomically more restricted range of species, we are able to show here that seminal fluid diversification can occur over relatively short evolutionary timescales. Our study also reveals that interspecific diversity in the seminal vesicle proteome far exceeds diversity seen in the second proteome we investigated, that of skeletal muscle. Given that we did not observe any polymorphism within species, such results are consistent with rapid evolution as a result of directional selection, although a more comprehensive investigation of within-species polymorphism awaits future study.
The proteomics-based pattern we have documented for rodents complements genomic evidence that genes encoding Drosophila seminal fluid proteins may be subject to rapid evolutionary turnover (Begun and Lindfors 2005; Mueller et al. 2005; Wagstaff and Begun 2005a, 2007; Begun et al. 2006) and evidence in primates that certain seminal fluid proteins may be subject to rapid evolutionary dynamics observed are due to radical substitutions in the repetitive central cross-linking region of SVS II is likely to have occurred through insertion/deletion across species. Because the majority of divergence in SVS II is likely to have occurred through insertion/deletion substitutions in the repetitive central cross-linking region of the protein, there could be an obvious link here to the dynamics of plug formation. Hence, the most plausible explanation for an increased molecular mass of SVS II under sperm competition might be that a protein of greater molecular mass contains more sites for cross-linking and thus.

Postcopulatory Sexual Selection and Seminal Fluid Protein Evolution

Our proteomics survey also reveals substantial interspecific variation in the molecular mass of rodent seminal vesicle proteins, most likely due to amino acid sequence divergence. Although proteomic- and genomic-derived data cannot be directly compared, our results are consistent with rapid divergence seen in several genes encoding seminal proteins in Drosophila (Swanson et al. 2001a; Mueller et al. 2005), crickets (Andrés et al. 2006), and mammals (Dorus et al. 2004; Clark and Swanson 2005; Ramm et al. 2008). Thus, rapid evolution of its constituent proteins is emerging as a general feature of seminal fluid.

We also report that interspecific divergence in the molecular mass of SVS II could be functionally significant in the context of sperm competition. This protein is the major component of the rodent copulatory plug, which forms following ejaculation when SVS II molecules are cross-linked by a prostate-derived transglutaminase (Shivaji et al. 1990). The plug is implicated in sperm competition through its effects on chastity enforcement and/or sperm transport (Ramm et al. 2005), and the gene encoding SVS II shows a heightened rate of nonsynonymous substitution in more promiscuous rodent lineages (Ramm et al. 2008). In this study, we found that the molecular mass of the SVS II protein correlates positively with sperm competition level across species. Because the majority of divergence in SVS II is likely to have occurred through insertion/deletion substitutions in the repetitive central cross-linking region of the protein, there could be an obvious link here to the dynamics of plug formation. Hence, the most plausible explanation for an increased molecular mass of SVS II under sperm competition might be that a protein of greater molecular mass contains more sites for cross-linking and thus.

Table 2
Phylogenetic Analyses of Interspecific Variation in (a) SVS II Molecular Mass and (b) Number of Major Seminal Fluid Proteins with Respect to Sperm Competition Level

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>n</th>
<th>λ²</th>
<th>Intercept</th>
<th>Body Mass Slope</th>
<th>Testis Mass Slope</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) SVS II mass</td>
<td>17</td>
<td>0</td>
<td>3.919 ± 0.143 (P &lt; 0.001)</td>
<td>−0.022 ± 0.034 (P = 0.53)</td>
<td>0.087 ± 0.032 (P = 0.015)</td>
<td>0.39</td>
</tr>
<tr>
<td>(b) Number of proteins</td>
<td>18</td>
<td>0</td>
<td>7.556 ± 2.214 (P &lt; 0.005)</td>
<td>−0.462 ± 0.517 (P = 0.39)</td>
<td>0.091 ± 0.464 (P = 0.85)</td>
<td>0.05</td>
</tr>
</tbody>
</table>

* Note that the maximum likelihood λ estimates of 0 imply that these traits are not strongly constrained by phylogeny.
undergoes a higher rate or greater extent of coagulation or forms a more stable copulatory plug (see also Lundwall 1996; Jensen-Seaman and Li 2003; Hurle et al. 2007). Consistent with this suggestion, the only species in which SVS II was not detected in our survey was Ph. californicus; genetic evidence suggests that the P. californicus mating system involves strict monogamy (Ribble 1991), and males have extremely small testes relative to their body size (Linzey and Layne 1969). There is also evidence that interspecific variation in the cross-linking region of coagulation proteins could extend beyond rodents, because Hurle et al. (2007) reported that both semenogelins SEMG1 and SEMG2 exhibit wide variation in the number of transglutaminase domains found across 13 and 14 primate species, respectively. Moreover, sexual selection could again provide a plausible explanation, with a “general (albeit imperfect) trend between the number of repeated transglutaminase domains and the relative amount of both female promiscuity and semen coagulation” (Hurle et al. 2007, p. 280), corroborating our inferences for rodents.

Several additional rodent seminal vesicle proteins identified in fewer species from our survey could also have roles in postcopulatory sexual selection. For example, other common proteins match to SVS I (9 species), SVS III (11 species), and SVS VI (13 species). The function of these proteins is not well characterized, but SVS I and SVS III could act as additional plug components because they also cross-link in the presence of transglutaminase (Shivaji et al. 1990). SVS IV (nine species) has immune-modulating and anti-inflammatory properties in the female reproductive tract (Ialenti et al. 2001), suggesting a possible role in mediating sexual conflicts over sperm survival (Fedorka and Zuk 2005). The sperm motility–enhancing protein Ceacam 10 (of the carcinoembryonic antigen–related cell–cell adhesion molecule, or CEACAM, gene family) is present in Mus (see also Li et al. 2005), and other CEACAM proteins are found in Ph. roborovskii, Ph. sungorus, and Mi. agrestis. The functional properties of all these proteins suggest that they may be relevant to postcopulatory sexual selection; our data now provide a basis from which to investigate the adaptive significance of species differences in their occurrence.

Conclusions

The evolutionary diversification of male seminal fluid proteomes revealed by our survey fits into a wider pattern of rapid evolution seen among reproduction-related genes (Swanson and Vacquier 2002). This applies to genes expressed in other male reproductive tissues (e.g., Good and Nachman 2005; Nadeau et al. 2007) and also to several female reproductive genes (e.g., Swanson et al. 2001b, 2004; Pröschel et al. 2006; Turner and Hoekstra 2006). Understanding the selective pressures and evolutionary processes responsible for this pattern—as well as integrating molecular and organismal perspectives—remains a significant challenge, though the action of sexual selection is now clearly implicated as a likely factor in some cases (e.g., Dorus et al. 2004; Wagstaff and Begun 2005b; Nadeau et al. 2007; Ramm et al. 2008; this study).

Our study also illustrates the potential utility of proteomics as a powerful tool in evolutionary biology. By augmenting emerging genomic data, the combination of separation science with high-sensitivity mass spectrometry–based analysis permits the identification of proteins from complex samples in nonmodel species, dramatically increasing the range of organisms that can be studied and consequently the types of questions that can be addressed. However, these data are only won by direct interpretation of product ion fragmentation data to infer the protein sequence; global or “shotgun” proteomics approaches are unlikely to yield equivalently confident identifications, and so these global methods are most efficient with known genomes. There are no methods for automated high-quality de novo sequence interpretation, and this is largely done manually. Thus, such approaches are best suited to detailed comparative analysis of a few proteins, as in this paper, derived from multiple species. Such data not only permit unambiguous identification of proteins that have been mass–measured by an independent method but also identify relatively conserved protein sequences that can be used to define primers for cDNA amplification, leading to rapid acquisition of the remainder of the protein sequence through the cognate cDNA. In employing these methods to produce the first comprehensive survey of seminal fluid proteome diversity, we provide a basis for elucidating the complex evolutionary dynamics and adaptive significance of this key ejaculate component in a model vertebrate group.

Supplementary Material

Supplementary tables and figures are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We are grateful to M. Bennett, K. Blasdell, W. Breed, J. Crossland, A. Loudon, A. Orth, G.A. Parker, S. Siah, and S. Unwin for supplying biological samples; to C. Ponting, R. Emes, and P. Oliver for collaboration to produce the gene sequence data reanalyzed in this study; to R. Freckleton for providing the code to conduct the comparative analyses; and to M. Nachman and three anonymous reviewers for their helpful comments on the manuscript. Experimental assistance from L. Burgess, R. Humphries, S. Jopson, and J. Waters and instrumentation support from D. Robertson are also gratefully acknowledged. This work was funded by a grant from the Natural Environment Research Council (to J.R.B., J.L.H., and P.S.). S.A.R was supported by a studentship from the University of Liverpool (to P.S. and G.A. Parker).

Literature Cited


Michael Nachman, Associate Editor

Accepted October 13, 2008