Molecular Evolution of Seminal Proteins in Field Crickets

José A. Andrés,* Luana S. Maroja,* Steven M. Bogdanowicz,* Willie J. Swanson,† and Richard G. Harrison*

*Department of Ecology and Evolutionary Biology, Cornell University; and †Department of Genome Sciences, University of Washington

In sexually reproducing organisms, male ejaculates are complex traits that are potentially subject to many different selection pressures. Recent experimental evidence supports the hypothesis that postmating sexual selection, and particularly sexual conflict, may play a key role in the evolution of the proteinaceous components of ejaculates. However, this evidence is based almost entirely on the study of *Drosophila*, a species with a mating system characterized by a high cost of mating for females. In this paper, we broaden our understanding of the role of selection on the evolution of seminal proteins by characterizing these proteins in field crickets, a group of insects in which females appear to benefit from mating multiply. We have used an experimental protocol that can be applied to other organisms for which complete genome sequences are not yet available. By combining an evolutionary expressed sequence tag screen of the male accessory gland in 2 focal species (*Gryllus firmus* and *Gryllus pennsylvanicus*) with a bioinformatics approach, we have been able to identify as many as 30 seminal proteins. Evolutionary analyses among 5 species of the genus *Gryllus* suggest that seminal protein genes evolve more rapidly than genes encoding proteins that are not involved with reproduction. The rates of synonymous substitution (d_S) are similar in genes encoding seminal proteins and genes encoding "housekeeping" proteins. For the same comparison, the rate of fixation of nonsynonymous substitutions (d_N) is 3 times higher in genes encoding seminal proteins, suggesting that the divergence of seminal proteins in field crickets has been accelerated by positive Darwinian selection. In spite of the contrasting characteristics of the *Drosophila* and *Gryllus* mating systems, the mean selection parameter ω and the proportion of loci estimated to be affected by positive selection are very similar.

Introduction

In many sexually reproducing animals, male ejaculates contain not only sperm but also chemically complex seminal fluids. Although natural selection likely has played a role in the evolution of male seminal secretions (e.g., in response to pathogens; Meister et al. 1997; Lung et al. 2001), current hypotheses stress the importance of postmating sexual selection in determining the composition of complex ejaculates (Cordero 1995; Eberhard 1996; Simmons 2001). Parker (1970) was the first to propose that in polyandrous species, in which male ejaculates compete for fertilization opportunities, sexual selection will continue beyond the struggle among males for access to females. This form of selection, mediated by sperm competition and cryptic female choice, has been shown to be responsible for many morphological, behavioral, and physiological adaptations (for a review see Simmons 2001). Indeed, postmating sexual selection may be the major evolutionary force driving the evolution of male ejaculates.

Comparative studies suggest that accessory gland size and ejaculate volume respond to changes in sexual selection pressures (Wedell 1993; Bissoondath and Wiklund 1995; Karlsson 1995; Reinhardt 2001; Garcia-Gonzalez and Gomendio 2004; Ramm et al. 2005). These studies highlight the role of selection acting on ejaculates as a whole. However, ejaculates are complex mixtures of many components, including proteins, sugars, amines, lipids, prostaglandins, and steroid-like hormones (see Gillott 2003). To understand the evolution of male ejaculates, it is essential to know how natural and sexual selection determine characteristics of these complex mixtures and of their individual components.

Key words: crickets, mating systems, positive selection, sexual conflict, sexual selection, reproductive proteins.

E-mail: jaa53@cornell.edu.

In insects, seminal proteins, produced by the accessory glands and/or other secretory glands of the male reproductive tract, are one of the major components of ejaculates, representing up to 90% of the seminal fluid (Heller et al. 1998). Some secreted proteins are major building blocks of the spermatophore, a proteinaceous capsule containing the ejaculate (Cheeseman et al. 1990; Paesen et al. 1992; Feng and Happ 1996). However, the majority of seminal proteins in insects are constituents of the male seminal fluid, and some of these proteins are known to influence female reproductive physiology and behavior (e.g., Neubaum and Wolfner 1999; Chapman et al. 2003; Liu and Kubli 2003).

In *Drosophila*, accessory gland proteins (*Acps*) evolve far more rapidly than do nonseminal proteins (Swanson et al. 2001; Kern et al. 2004), and molecular evolutionary studies have revealed a clear signature of positive selection (Aguadé et al. 1992; Tsaur and Wu 1997; Aguadé 1998, 1999; Tsaur et al. 1998, 2001; Begun et al. 2000; Swanson et al. 2001). Although the function of many of the positively selected *Acps* has not been identified, a subset of these proteins is known to influence female physiology and behavior, including oogenesis, ovulation, oviposition, sperm storage, and remating rate (e.g., Harshman and Prout 1994; Herndon and Wolfner 1995; Wolfner 1997; Neubaum and Wolfner 1999; Tram and Wolfner 1999). Some of these traits are intimately related to different sperm competition and/or cryptic female choice mechanisms. In fact, in *Drosophila melanogaster*, experimental evidence has demonstrated a link between sperm competition and molecular variation at 2 different *Acp* loci (Fiumera et al. 2005). Similarly, Clark and Swanson (2005) have recently found evidence for selection in primate seminal proteins, some of which are involved in the formation of copulatory plugs. Comparative evidence also suggests a positive causal relationship between the degree of polyandry and the strength of selection at the molecular level. Among primates and hominoids, the degree of polyandry seems to be associated with the strength of positive selection in

ejaculate coagulation factors (Kingan et al. 2003; Dorus et al. 2004). Finally, seminal protein–coding genes in both *Drosophila* (Begun and Linfords 2005) and primates (Clark and Swanson 2005) exhibit rapid evolutionary turnover, a pattern consistent with the predictions of some postmating sexual selection models (see Andrés and Arnqvist 2000).

Evidence of adaptive evolution in seminal proteins is based almost entirely on the study of *Drosophila Acps* and primate seminal fluids (Jensen-Seaman and Li 2003; Kingan et al. 2003; Dorus et al. 2004; Clark and Swanson 2005). As a result of this strong taxonomic bias, it is still not possible to draw general conclusions about the role of different selective forces in the evolution of seminal proteins. The *D. melanogaster* mating system is characterized by a high cost of mating to females, a consequence of the transfer of seminal proteins by the male (Chapman et al. 1995; Wigby and Chapman 2005; Stewart et al. 2005). The negative effect on female fitness, believed to be a pleiotropic effect of postmating sexual selection in males, results in interlocus genomic conflicts (Rice and Holland 1997; Holland and Rice 1998), which in turn can lead to antagonistic coevolution of signal/receptor systems. Thus, the observed signature of positive selection on *Drosophila Acps* may reflect the particular characteristics of sexually antagonistic coevolution in such a mating system.

In this paper, we aim to broaden our understanding of the role of selection on the evolution of seminal proteins by characterizing these proteins in a group of insects in which females appear to benefit from mating multiply. To achieve this aim, we have carried out an evolutionary expressed sequence tag (EST) screen of the male accessory gland in 2 closely related species of field crickets, *Gryllus firmus* and *Gryllus pennsylvanicus* (Alexander 1957; Harrison and Bogdanowicz 1997). The reproductive biology of field crickets is well known. Females are polyandrous (e.g., Solymar and Cade 1990) and store sperm in a single elastic spermatheca that expands to accomodate successive ejaculates (Simmons 1986; Sakaluk and Eggert 1996). Males transfer discrete packets of sperm (spermatophores), which also contain physiologically important substances (for a review see Stanley-Samuelson and Loher 1986). However, in contrast to *Drosophila*, field cricket females appear to benefit from multiple mating, and seminal fluids have a net positive effect on female fitness through both direct (increased lifetime fecundity) and indirect (i.e., genetic) benefits (Simmons 1988; Burpee and Sakaluk 1993; Wagner et al. 2001; Ivy and Sakaluk 2005).

Although the reproductive biology of field crickets is relatively well known, there is little information on the composition of the seminal fluid. Prostaglandins or prostaglandin precursors are present in the seminal fluid of crickets in the genera *Acheta* and *Teleogryllus*, and when transferred to females, these compounds have been shown to trigger oviposition behavior (Destephano et al. 1974; Destephano and Brady 1977; Loher et al. 1981; Stanley-Samuelson and Loher 1983, 1986). However, these compounds are only a small fraction of the seminal fluid, and there is no available information on the proteinaceous components (but see Braswell et al. 2006).

Here we identify genes that encode seminal proteins, using functional and structural information on the genes expressed in the accessory gland, together with qualitative expression data. To examine the possible role of selection in the evolution of seminal proteins, we use codon-based maximum likelihood methods (Yang et al. 2000) to identify the signature of adaptive molecular evolution based on comparisons among 5 species of field crickets. We compare the patterns of molecular evolution that characterize genes encoding both seminal and nonseminal proteins. Furthermore, to begin to understand the impact of different mating systems on the evolution of seminal proteins, we also compare the pattern of evolution of field cricket seminal proteins with that seen for *Drosophila Acps*.

Materials and Methods
Isolation and Sequencing of Accessory Gland ESTs

The accessory glands of 11 field-collected adult male *G. pennsylvanicus* (Mt Pleasant, Ithaca, N.Y.) were dissected in DEPC-treated Ringer's solution under a dissecting microscope. Total RNA was extracted using the RNAeasy midi kit (QIAGEN, Valencia, CA). Poly(A) mRNA was isolated using an Oligotex (dT) binding kit (QIAGEN), and approximately 6 μg of poly(A) RNA was used to construct a size-fractionated (>400 bp) directional cDNA library in the plasmid vector pCMV.SPORT6 using a cDNA construction kit (Invitrogen, San Diego, CA). Recombinant plasmids were introduced into electrocompetent cells (ElectroMax DH5-α, Invitrogen) by electroporation (2.5 kV, 200 Ω, 25 μF) and plated on LB ampicilin agar with 50 μg/ml ampicilin. The resulting library contained 40 000 colony-forming units with an average insert size of ≈1.1 kb. Colonies were lifted onto Magna Graph membranes (Osmonics, Inc). To enrich for male-specific ESTs, we screened the library with ^{33}P-labeled *G. pennsylvanicus* female cDNA, using a RadPrime Labeling system (Invitrogen) to produce the radiolabeled probe. Prior to hybridization, membranes were washed with 2× standard saline citrate (SSC) buffer containing 0.1% sodium dodecyl sulfate (SDS). The membranes were then hybridized for 48 h at 65 °C in 7.5% SDS, 0.5 M phosphate (pH = 7.6) buffer. Following hybridization, the membranes were washed with 2× SSC, 0.1% SDS at 60 °C for 10 min. Nonhybridizing and weakly hybridizing recombinant colonies were picked manually and transferred to Luria–Bertani/ampicilin medium in 1.5 ml tubes. After incubation, plasmid DNA was extracted by boiling lysis (Sambrook et al. 2001). Isolated plasmid DNA was sequenced from the 5′ end using a Big Dye Terminator Cycle sequencing kit and an ABI-3730 XL automated sequencer (Applied Biosystems, Foster City, CA). Initial sequencing of ESTs (*n* = 96 clones) revealed 4 ESTs that constituted a large fraction of the library (see also Braswell et al. 2006). The library was then further screened with a random-primed probe generated from a mixture of polymerase chain reaction (PCR) products from these highly abundant ESTs. By excluding recombinant colonies that hybridized with this probe, we could minimize repeated sequencing of the same ESTs, increasing the rate of discovery of new transcripts. Primers and PCR conditions are described in the Supplementary Material online. For our study, we also used an accessory gland cDNA library from *G. firmus*, which was constructed,

screened, and sequenced in a similar fashion (see Braswell et al. 2006). Sequences were edited by eye, and any ambiguous calls were resolved by examining the chromatograms. High-quality EST sequences for both libraries (as determined by the Parkinson and Blaxter [2004] perl script Trace2dbest 3.0.1) have been deposited in GenBank (dbEST accession numbers 38387049–38388412).

Analyses of ESTs

To differentiate the coding and untranslated regions (UTRs) (i.e., 3′ and 5′ UTRs) of the individual transcripts represented in our library, we first clustered ESTs into contigs using the algorithms implemented by Seqman (DNASTAR). All alignments were visually checked, and consensus and singleton sequences were then blasted against the National Center for Biotechnology Information nonredundant databases using both nucleotide sequences (BlastN) and translated sequences (BlastX). For those sequences showing clear similarity (presumed homology) with previously annotated genes (E values $< 10^{-16}$), the coding region and open reading frame (ORF) were inferred directly from the alignment. For novel sequences (defined as those with no Blast hits with E values $< 10^{-16}$), we inferred the ORF using a combined strategy. First, we assessed the longest ORF possible and then we determined the likelihood of the sequences representing coding regions by combining Markov chain models of coding and noncoding regions with a Bayesian decision-making function. These latter analyses were performed using the Genemark.SPL software package (Borodovsky and McIninch 1993). The presence of a 3′ UTR was confirmed by searching for the AAUAAA polyadenylation signal.

Identification and functional characterization of the individual genes expressed in the accessory gland was carried out by sequence (BlastN and BlastX), structural, and functional homology. Structural and functional homology analyses based on hidden Markov model (HMM) profiles of conserved residues were run using the SMART package (Letunic et al. 2002). The presence and predicted cleavage sites of signal peptides were inferred using the HMM and neuronal network methods implemented by SignalIP 3.0 (Bendtsen et al. 2004).

Gene Expression Analyses

We used a PCR-based approach to obtain qualitative information on the expression patterns of the genes of interest (i.e., those identified by the homology analyses described above). Total RNA was extracted using Trizol (Invitrogen) from abdominal tissues (reproductive, secretory, and digestive) of 5 virgin *G. pennsylvanicus* females. Individual extractions of total RNA from the accessory glands of 3 last instar males and 4 adult males were carried out using the same protocol. Total RNA was used to synthesize single-stranded cDNA using SuperScript II and III reverse transcriptases (Invitrogen) and an oligo (dT) primer. Tissue-specific RNA/DNA heteroduplexes were then 20-fold diluted in water and used as PCR template. All cDNA extractions were screened for genomic DNA contamination using a diagnostic PCR assay designed to amplify a known region of the gene for elongation factor-1α. Because this gene region contains an intron, expected fragment sizes are different for amplifications from cDNA versus genomic DNA. Because our method of assessing expression patterns is sensitive to variation across amplifications, for each locus we independently replicated the assay at least twice, using 2 different males and 2 different females in each experiment. The results of the assay were highly repeatable, and none of the loci showed contradictory expression patterns across amplifications. Information on primers and PCR conditions can be found in the Supplementary Material online.

Evolutionary Analyses

We assessed DNA sequence divergence among 3–5 different species of the genus *Gryllus*. For each of the studied genes, we used single sequences from either all or a subset of the following species: *G. firmus*, *G. pennsylvanicus*, *Gryllus rubens*, *Gryllus veletis*, and *Gryllus bimaculatus*. *Gryllus bimaculatus* males were obtained from a large laboratory culture maintained by the Hoy laboratory at Cornell University. All other species were collected from natural populations: *G. firmus* from Guilford, Conn., *G. pennsylvanicus* from Ithaca, N.Y., *G. rubens* from Durham, N.C., and *G. veletis* from Ithaca, N.Y. For all our analyses, we used the following tree topology (*G. bimaculatus*, (*G. firmus*, *G. pennsylvanicus*), (*G. rubens*, *G. veletis*)). This tree topology reflects the phylogenetic relationships of the 5 *Gryllus* species based on mitochondrial gene cytochrome b sequences for these taxa (accession numbers: AY236340–47, AF248678–79, AF248675, AF248664, and AF248659–60; see also Huang et al. 2000).

Selection estimates are based on coding regions of the transcripts isolated from male accessory gland cDNA. Total RNA from individual glands was extracted using Trizol Reagent (Invitrogen). Single-stranded poly (dT) cDNA was synthesized using SuperScript III reverse transcriptase (Invitrogen). Primers were designed to maximize the length of the coding region sequence for each of the selected genes, which were identified by the structural and functional homology analyses described above. PCR primers and conditions are available as Supplementary Material online. PCR products were diluted 1:2 in water and sequenced using either an ABI-3100 or ABI-3730 XL sequencer (Applied Biosystems). Sequences can be found in GenBank under accession numbers DQ520131–DQ520157 and DQ63065–DQ6300942.

Predicted coding sequences were aligned using the ClustalW algorithm (Thompson et al. 1994), as implemented in Megalign (DNASTAR). Individual gene alignments were used to estimate the rate of amino acid replacement substitution (d_N) relative to the rate of silent substitution (d_S), using the rate ratio $\omega = k_a^*/k_s^*$, where the number of nonsynonymous ($k_a^* = k_a + \varepsilon$) and synonymous ($k_s^* = k_s + \varepsilon$) substitutions have an associated error (ε) that reflects stochasticity in sequence determination (Barrier et al. 2003). For estimating d_N and d_S, we used the codon-based maximum likelihood method (CODEML) described by Goldman and Yang (1994) implemented in the program PAML 3.14 (Yang 2000). Estimates of the rate of evolutionary change (tree length) for each of the genes were also calculated using CODEML. In contrast

to parsimony-based methods, maximum likelihood analysis takes into account the base frequencies at the 3 codon positions and transition/transversion ratios. Because nonsynonymous mutations change the amino acid sequence of the protein, whereas synonymous mutations do not, the difference in their fixation rates provides an estimate of the selection pressures on the protein. For any set of amino acid residues, when $\omega = 1$, a neutral model of evolution cannot be rejected, whereas $\omega < 1$ indicates purifying selection, and $\omega > 1$ indicates positive selection. Nested models that allow different ω values were compared using a likelihood ratio test (LRT) as described in Yang et al. (2000) and Yang and Nielsen (2002). The LRT statistic was calculated as the negative of twice the difference in log likelihood ($-2[\log (L_0) - \log (L_1)]$) obtained from 2 different models, and its significance was estimated using a χ^2 distribution with degrees of freedom equal to the difference of estimated parameters. Variation in the d_N/d_S ratio between sites was modeled using both discrete and continuous distributions. To test for positive selection, we used 2 different comparisons, one discrete (M1a vs. M2a) and one continuous (M7 vs. M8). $L_0 = $ M1a (nearly neutral) and $L_1 = $ M2a (positive selection). M1a (nearly neutral) allows for 2 rate classes, one with $0 < \omega_0 < 1$ and the other with $\omega_0 = 1$. M2 includes an additional unconstrained rate class where ω can be greater than 1. The comparison $L_0 = $ M7 versus $L_1 = $ M8 is an alternative test for positive selection that allows for a continuous variation in substitution rates across sites.

Selection is likely to act along specific branches during the evolution of a lineage, rather than homogeneously across an entire phylogeny. To test if selection indeed varies among lineages, we compared the likelihood of a model assuming a single ω value for all lineages (i.e., one ratio model), with the likelihood of a model that estimates an independent ω for each of the branches (i.e., free-ratio model). Although variation in the ω ratio among lineages is a violation of the neutral model of molecular evolution, this comparison is not necessarily a test of adaptive evolution, and the free-ratio model is likely to produce inaccurate ω estimates (Yang and Bielawski 2000; Yang et al. 2000). Therefore, we only used this comparison for identifying lineages for which episodes of selection might have occurred.

Variation in the rate of evolutionary change and variation in selective constraints between different groups of genes were analyzed using random permutation tests (Manly 1991). This method of statistical inference has the advantage of not making any a priori assumptions about the distribution of the parameters (ω, d_N, d_S) estimated from the sampled populations and is particularly appropriate with unequal and small sample sizes. The significance value of each permutation test was calculated using 10,000 random reallocation data sets. Means and 95% confidence limits for the parameters of interest were obtained by bootstrapping across loci. All statistical analyses were carried out using R 2.1.1 (R Development Core Team 2004).

To compare the rate of divergence and selective constraints between genes encoding nonreproductive proteins and those encoding putative seminal proteins, we contrasted this latter subset of genes with a group of 6 accessory gland–expressed "housekeeping" genes, selected to represent different structural proteins and basic metabolic functions (β-actin, hexokinase, elongation factor-1α, fascilin, tubulin, and S9 ribosomal protein).

Results

Identification and Characterization of Male Accessory Gland Genes

The screened libraries of *G. firmus* ($n = 717$ ESTs) and *G. pennsylvanicus* ($n = 952$ ESTs) represent, respectively, 247 and 277 different gene transcripts or unigenes (here assumed to represent independent genes expressed in the accessory gland). Analysis of the EST frequency spectrum revealed that more than half of the recovered genes are represented by a single EST and that nearly 80% of the transcripts occur as either singletons or contigs that include only a small number of ESTs ($n \leq 3$). However, transcripts from a small group ($n = 11$–15) of highly expressed genes made up nearly 40% of all sequenced ESTs. Nucleotide, functional, and structural Blast analyses revealed that the predicted products of these highly expressed genes are shared between *G. firmus* and *G. pennsylvanicus* and that this group of genes is composed of ribosomal proteins or of novel sequences that showed no significant homology ($>10^{-16}$) to sequences in any of the databases we searched. Analysis of predicted products from accessory gland genes revealed a wide range of possible biological functions (see Braswell et al. 2006), but the function of nearly half of the accessory gland gene products remains unknown because the EST sequences showed no significant homologies with annotated genes.

Structural analyses for the subset of genes that showed no significant similarities using nucleotide-based alignments (BlastX, BlastN) revealed that ~30% of the transcripts included predicted signal peptides at their 5′ ends. Most of the predicted gene products in this group did not show any significant structural resemblance (e_{SMART} default parameters) to described protein profiles, and only a few of the predicted proteins show significant structural similarities with annotated domains. Most of the predicted products are structurally quite simple and are composed of a single domain. The most common domains found among the accessory gland proteins of both species are serine proteases (~5%), di- and carboxypeptidases (~3%), and c-type lectins (~3%). Our analyses also revealed that the most highly expressed genes in the accessory gland libraries (AG-0001F, AG-0076F, AG-0315F, AG-0317F) have a highly repetitive primary structure and a predicted helical and coiled-coil secondary structure. [Note that here and below we give the name of the EST from the *G. firmus* library, unless the EST was only found in the *G. pennsylvanicus* library. Many of the ESTs from one library have presumed homologues in the second library.]

Using the structural analyses, we identified a subset of accessory gland–expressed genes ($n = 70$) that, based on 2 different criteria, are likely to encode seminal proteins (i.e., candidate seminal proteins). First, we selected genes potentially encoding secreted extracellular proteins. In this category, we included both secreted proteins (i.e., those with a predicted signal peptide) and novel transcripts with incomplete 5′ ends (that may or may not include a signal peptide). An obvious concern with including this latter group is
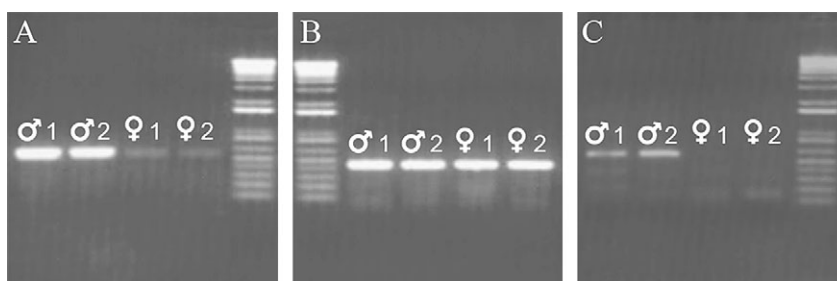
FIG. 1.—Qualitative gene expression patterns of accessory gland genes. PCR amplification of 3 genes in 4 different individuals, 2 males and 2 females. Left to right: (*A*) Gene showing a male-biased expression pattern. (*B*) Accessory gland housekeeping gene (elongation factor-1α) showing a qualitatively similar pattern of expression in both sexes. (*C*) Gene showing male specific expression.

that we may misidentify nonsecreted proteins as putative seminal proteins. However, this "inclusive" strategy compensates for the inability to generate cDNA libraries comprised only of full-length transcripts. Second, because major functional classes appear to be conserved among seminal proteins (Mueller et al. 2004), we also targeted those genes predicted to encode products belonging to the protein classes represented in seminal fluids of *Drosophila* and mammals (i.e., regulators of proteolysis: proteases and serpins, lipases, lectins, and cysteine-rich proteins of the CRISP family).

Gene Expression Patterns

All candidate genes were examined for male (i.e., accessory gland specific) expression. Using a PCR-based approach, we searched for sex- and stage-specific gene expression patterns among the candidate seminal protein genes described above. Approximately 27% (*n* = 19) of these genes were expressed in the accessory gland of adult males, but no PCR product was obtained using cDNA from female abdominal tissue as template. Another ~14% (*n* = 10) of the genes gave strong PCR products using male accessory gland cDNA as template, but only faint bands appeared when the template was cDNA from abdominal tissues of females (fig. 1). These 29 "male-expressed" genes did not vary in the temporal pattern of gene expression; they were all expressed in the accessory gland from last instar juvenile and adult males. The vast majority of these male-expressed genes fall into 3 categories. Five of the genes (AG-0001F, AG-0076F, AG-0315F, AG-0316F, AG-0317F) code for peptides containing one or several internal repeats. This class of genes includes those that are very highly expressed (see above). The inferred products of these genes are rich in aliphatic and neutral amino acids, with alanine, asparagine, and leucine constituting 40%–50% of the residues of the predicted proteins. These proteins are all strikingly similar to each other, in some cases sharing the same repeat motifs. A second class of male-expressed genes (AG-0159F, AG-0253P, AG-0308F, AG-0508F) encode inactive serine protease precursors (i.e., zymogens), each containing a signal peptide, a short activation peptide, and a single serine protease domain with significant similarities to other trypsin- or chymotrypsin-like precursors. All but one (AG-0253P) of these serine protease domains contain the Asp-His-Ser catalytic triad characteristic of this protein family. Finally,

~60% (*n* = 17) of the male-expressed genes (AG-0005F, AG-0015F, AG-0020F, AG-0024F, AG-0032F, AG-0042F, AG-0055F, AG-0064F, AG-0068F, AG-0143F, AG-00177F, AG-0211F, AG-0271F, AG-0227, AG-0197P, AG-0286P, AG-0334P) showed no significant functional or structural homology with previously described genes.

Expression levels of genes encoding putative seminal proteins appear to vary widely. An EST frequency spectrum analysis on the subset of genes with no detectable homology with housekeeping genes (and therefore putative seminal proteins) revealed that a relatively small set of genes (*n* = 27) account for the majority of the recovered ESTs (~70%) and that genes with internal repeats represent over 30% of all ESTs (fig. 2). Within this group of highly expressed genes, there is a nearly significant association between the presence of a predicted signal peptide and a male-biased expression pattern (χ^2 = 5.25, Monte-Carlo simulated *P*-value = 0.053).

Variation in Selective Constrains Among Male Accessory Gland Genes

Using sequence data from putative homologs in 5 species of field crickets, we estimated the distribution of synonymous (d_S) and nonsynonymous (d_N) substitutions for a subset of the accessory gland genes considered to represent seminal proteins. In this group of genes, the distribution of the number of synonymous substitutions differs significantly from that observed for nonsynonymous substitutions (permutation test *P* = 0.0001). The nonsynonymous distance distribution has a mean (0.026 ± 0.0004) smaller than that for the distribution of synonymous substitutions (0.070 ± 0.0009, fig. 3). The estimates of the selection parameter ω in this set of genes range from 0 to 1.96. The mean value of ω obtained by resampling (Bustamante et al. 2002) is 0.382 (95% confidence interval [CI]: 0.283–0.499), and although only 3 of the candidate genes showed a clear signature of adaptive evolution (i.e., ω estimate exceeding 1), there are another 6 genes with inferred d_N/d_S ratios greater than 0.5 (see fig. 4). [Note that we were unable to estimate ω for all the putative seminal protein genes because in a few cases sequence data could not be obtained for more than the 2 focal cricket species.]

The mean rate of evolutionary change for accessory gland genes that encode putative seminal proteins (mean tree length 0.12 ± 0.016) is almost 3 times higher than that
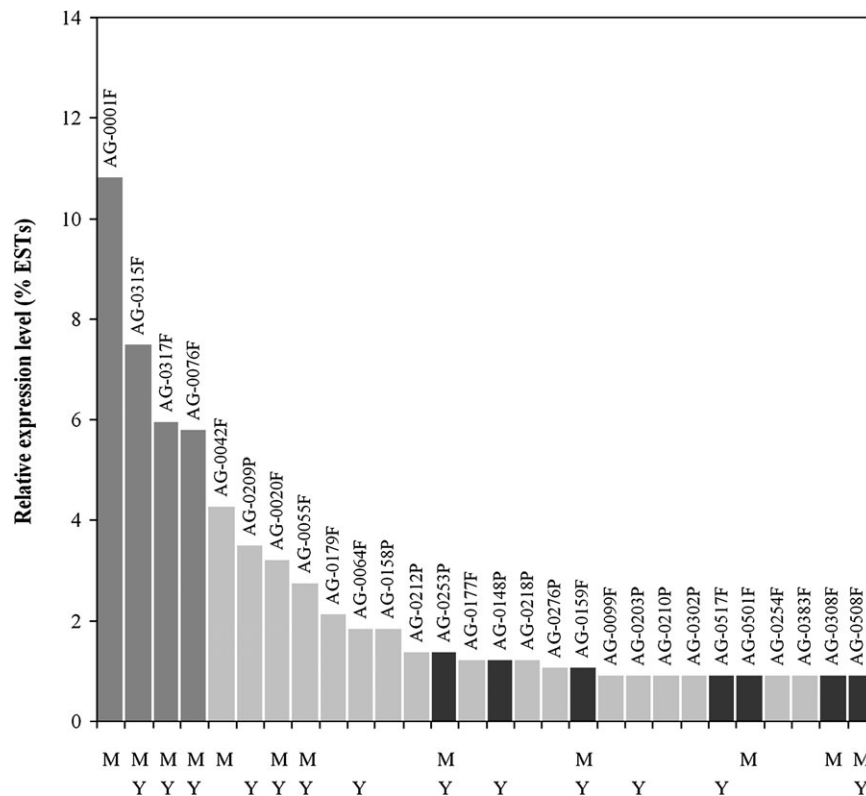
FIG. 2.—Relative gene expression levels of the most abundant putative seminal proteins in the seminal fluid of *Gryllus firmus* and *Gryllus pennsylvanicus*. Expression levels were estimated as the percentage of ESTs considered to represent the same gene transcript relative to the total number of ESTs present in the group of candidate genes. M: Male-biased expression pattern. Y: Signal peptide present. Light gray bars represent genes showing no structural or functional similarity to annotated genes. Dark gray bars represent genes encoding repetitive proteins. Black bars represent genes encoding serine proteases or c-type lectins, classes of proteins that are known to be present in the seminal fluid of mammals and *Drosophila*.

of a group comprised of basic metabolic enzymes and structural genes expressed in the same tissue (mean tree length 0.038 ± 0.012, $n = 6$, permutation test $P = 0.047$). This difference in the rate of evolutionary change is driven by an increased fixation rate of nonsynonymous mutations (d_N) in the set of putative seminal protein genes (0.029 ± 0.005) relatively to that of the control genes (0.01 ± 0.002, permutation test $P = 0.027$). In contrast, the estimated rate of synonymous substitutions is similar between these 2 groups (putative seminal proteins: 0.046 ± 0.02, accessory gland control genes: 0.074 ± 0.01, permutation test $P = 0.317$). The difference in the rate of evolutionary change is more obvious for the subset of putative seminal proteins that show a male-biased expression pattern (0.149 ± 0.02 vs. 0.038 ± 0.012, permutation test $P = 0.007$). As in the previous comparison, this higher mean rate of change does not arise as a result of differences in the rate of change at synonymous sites (d_S permutation test $P = 0.220$) but because of an accelerated fixation rate of nonsynonymous mutations (d_N: 0.050 ± 0.02 vs. 0.01 ± 0.002, permutation test $P = 0.006$).

The value of the selection parameter ω for the set of control genes (0.039 ± 0.02, $n = 6$) is also significantly smaller than the estimated ω value for the genes that encode putative seminal proteins (0.38 ± 0.07, $P < 0.001$, $n = 35$). Moreover, the mean ω value for the subset of putative seminal proteins that showed a male-biased expression pattern is significantly higher (0.63 ± 0.14) than for those that do

not show qualitative differences in gene expression between the sexes (0.27 ± 0.04, permutation test $P = 0.047$). The action of selection seems to be homogeneous across lineages, and only one of the LRT comparisons between one- and free-ratio ω models was statistically significant (table 1).

Selection at the molecular level is likely to target only a small subset of amino acid residues within a functional protein. Therefore, we have also analyzed the selective constraints of candidate genes using 2 different models of codon substitution that allow ω to vary among different classes of sites (M1a vs. M2a and M7 vs. M8). The LRT results using a discrete model of variation (M1a vs. M2a) were consistent with those obtained using a continuous β distribution of ω values among sites (M7 vs. M8), and 2 genes showed statistically significant evidence for adaptive evolution.

Finally, comparisons between field cricket and *Drosophila Acp* evolution revealed that the selective constraints on seminal proteins are quite similar across species with contrasting mating systems. Thus, the mean of the selection parameter ω in putative seminal proteins in field crickets, is similar to that observed in *Drosophila Acps* ($\omega = 0.5$, bootstrap 95% CI of 0.344–0.646, permutation test $P = 0.319$). Likewise, the inferred proportion of seminal proteins with $\omega \geq 0.5$ (i.e., those that are likely to be under positive selection—see below) is similar in *Drosophila* and *Gryllus* ($\chi^2 = 0.656$, Monte-Carlo simulated P-value $= 0.443$).
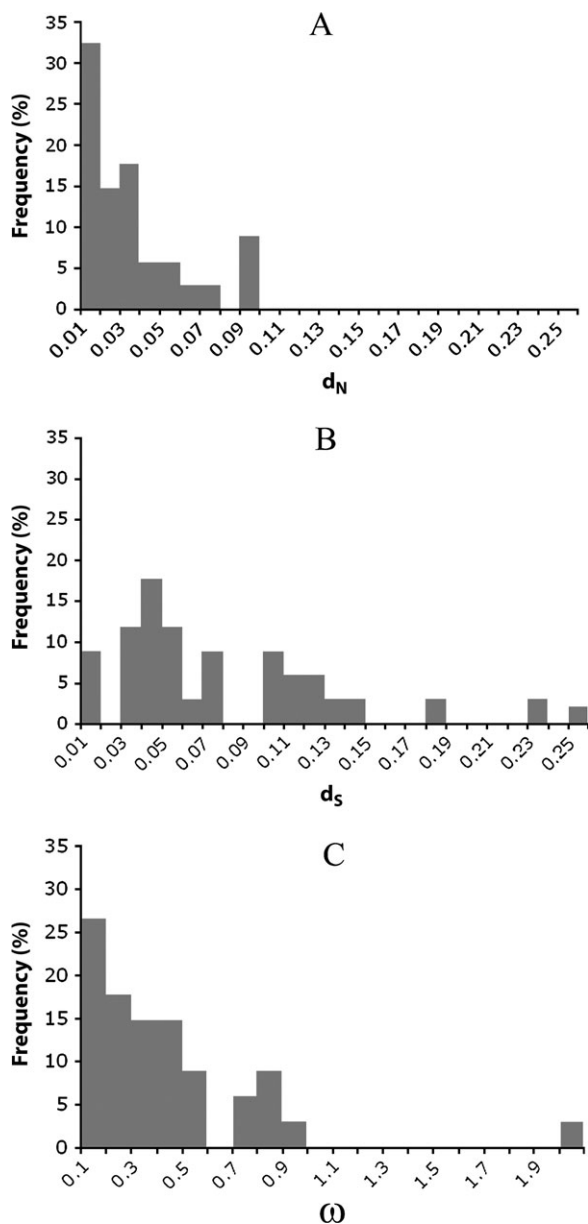
FIG. 3.—Distribution of (*A*) nonsynonymous (d_N), (*B*) synonymous (d_S) distances, and (*C*) estimated ω values among putative seminal protein orthologs in *Gryllus bimaculatus*, *Gryllus firmus*, *Gryllus pennsylvanicus*, *Gryllus rubens*, and *Gryllus veletis*.
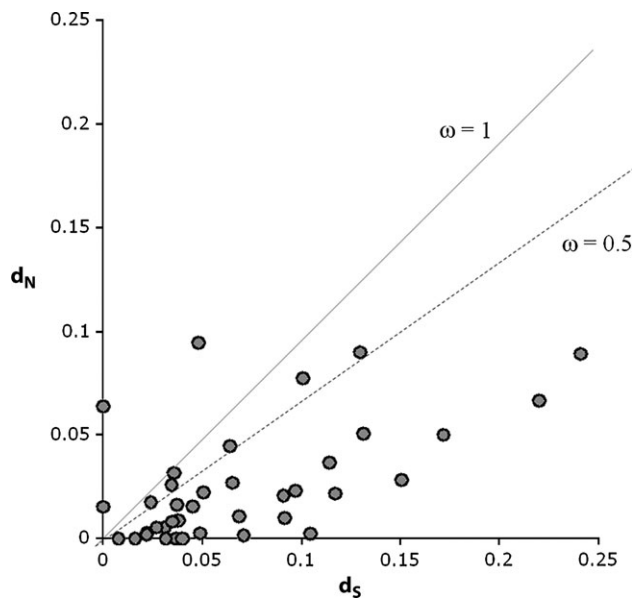

FIG. 4.—Plot of estimated d_N versus d_S values for 35 genes predicted to encode seminal proteins in field crickets, assuming no rate heterogeneity among the included lineages (*Gryllus bimaculatus*, (*Gryllus firmus*, *Gryllus pennsylvanicus*), (*Gryllus rubens*, *Gryllus veletis*))—see table 1. The solid line represents the neutral expectation of $d_N/d_S = 1$. The dashed line represents the threshold value of d_N/d_S considered to identify genes that may have undergone episodes of positive selection (see text for more details).

Discussion
Searching for Seminal Proteins in the Transcriptome Haystack: Identification of Candidate Genes

Our analysis of the transcriptome of accessory glands of adult male field crickets revealed the presence of a large number of proteins with a wide range of functions and levels of expression. Therefore, the first goal of our study was to filter out genes encoding structural proteins and proteins involved in basic metabolic functions, in order to focus on genes encoding seminal proteins. An obvious first step in identifying genes encoding seminal proteins is to find in the transcriptome those sequences that are homologues of seminal protein genes that have been described in other insects. Thus, we used a bioinformatics approach to compare the sequences of the accessory gland of field crickets to *Acps* of *Drosophila*, the only other insect in which these proteins have been described. However, this analysis failed to reveal any homology between candidate genes in crickets and *Drosophila Acps*. This result is not surprising, given that seminal proteins are known to be rapidly evolving in *Drosophila* (Swanson et al. 2001).

An alternative approach for identifying seminal protein genes is to select candidates based on our knowledge of the functions of proteins that are secreted by the accessory glands and/or other secretory glands of the male reproductive tract. Adopting this definition, we used an approach analogous to that described by Mueller et al. (2005) to identify a set of putative seminal protein genes in field crickets. Thus, we consider seminal protein genes to be genes expressed in the accessory gland that: (1) do not show any significant homology with housekeeping genes that have no known function related to reproduction, (2) encode a protein with a predicted signal peptide, and (3) show predominantly male (i.e., accessory gland) expression. In addition, because functional classes of seminal proteins seem to be conserved (Mueller et al. 2004), we also looked for genes showing functional and/or structural similarities with previously described seminal proteins or *Acps*.

Using these criteria, we have been able to identify as many as 30 different genes that we consider to represent bona fide seminal proteins. As a group, these genes are highly expressed in the accessory gland and show significant association between secretion and male-biased expression patterns, as expected if this set of genes includes genes that encode seminal proteins.

Table 1
Detection of Selection by Maximum Likelihood Analyses

Gene	LRT Free Ratio	M1a–M2a	M7–M8
AG-0254F	2.9101	2.1044	2.1046
AG-0383F	2.3648	0.1992	0.2457
AG-0501F	2.3759	0.0000	0.0000
AG-0508F	**14.667**	0.0000	0.0003
AG-0517F	3.0012	0.4959	0.4960
AG-0141F	2.9369	0.0000	0.0000
AG-0143F	0.0001	**6.0308**	**6.0308**
AG-0207F	1.5262	5.6505	5.7481
AG-0241F	5.7434	0.0000	0.0001
AG-0032F	0.9534	0.0000	0.0000
AG-0055F	4.5967	3.9321	4.1435
AG-0005F	5.5503	2.2518	2.5830
AG-0064F	3.6186	1.2325	1.2800
AG-0065F	1.3425	0.0000	0.0000
AG-0085F	6.8638	0.0000	0.0001
AG-0105F	2.9736	0.0000	0.0000
AG-0148P	6.5523	0.2324	0.3856
AG-0014P	4.5657	**18.304**	**19.434**
AG-0197P	7.3394	5.4323	5.4381
AG-0198P	5.5729	0.2813	0.3357
AG-0201P	0.0002	0.0011	0.0002
AG-0203P	7.4018	0.0363	0.0568
AG-0220P	4.5201	0.7681	0.7682
AG-0248P	2.6993	0.0000	0.0001
AG-0253P	1.0855	3.0409	3.0705
AG-0285P	8.9288	0.0000	0.0000
AG-0302P	6.3206	1.1588	1.2558
AG-0310P	2.8441	0.0000	0.0000
AG-0316P	8.1318	0.8504	0.9612
AG-0345P	1.6875	0.0000	0.0000
AG-0368P	0.0004	1.2464	1.2464
AG-0376P	2.7587	0.0000	0.0003
AG-0404P	0.0003	0.0005	0.0002

Note.—Likelihood ratio statistic ($2\Delta l$) values between nested models of molecular evolution. The free-ratio column represents the difference in likelihood between a free ω ratio model and a one ω ratio model. The other 2 comparisons represent likelihood differences between models that allow ω to vary among sites including (M2a–M8) or excluding (M1a and M7) positive selection. Significant ($P < 0.05$) LRT values are printed in bold.

Structural and functional analyses of this group of genes revealed several distinct classes of proteins that may play specific roles in sperm transfer: Several highly expressed genes encode repetitive peptides that are very similar to each other and have characteristic secondary structure. We suggest that these genes encode structural (i.e., spermatophore building) seminal proteins, both because of the high levels of gene expression and because products of these genes are similar to proteins that constitute the spermatophore of other insects (*Coleoptera*: Paesen et al. 1992; Feng and Happ 1996; *Orthoptera*: Cheeseman et al. 1990). Many of the other male-expressed and secreted proteins belong to the same functional classes as do previously described seminal proteins. Thus, we have found serine proteases as well as c-type lectins, 2 of the major functional classes represented in the seminal proteins of other orthopterans (Cheeseman et al. 1990), as well as in *Drosophila* and mammals (Mueller et al. 2004, 2005). The presence of the same functional classes in animals with very different life histories and mating systems may represent a case of functional convergence in response to similar selection pressures. Obviously, further studies are needed to test this hypothesis.

Divergence Studies Reveal the Footprints of Adaptive Evolution

To gain insight into the evolutionary forces affecting the divergence of seminal proteins in field crickets, we estimated and compared rates of synonymous and nonsynonymous substitutions in 2 groups of genes, both expressed in the accessory gland of sexually mature males. One group of genes encodes candidate seminal proteins, and a second group of genes are considered to be housekeeping genes that encode structural proteins and enzymes.

The estimated rate of divergence of seminal protein genes was significantly higher than that of the control set of genes, supporting the hypothesis that seminal protein genes tend to evolve more rapidly than genes with products presumably not involved in reproduction (Swanson et al. 2001, Kern et al. 2004). Because rates of synonymous substitution (d_S) were similar in genes encoding seminal proteins and genes not related to reproduction, significant differences in mutation rate or codon bias between seminal proteins and the set of housekeeping genes cannot be invoked as an explanation. In contrast, the rate of fixation of nonsynonymous substitutions (d_N) is 3 times higher in seminal proteins than in other accessory gland–expressed genes. These observed differences in the rates of substitution of synonymous and nonsynonymous mutations can be explained under 2 alternative hypotheses: (1) in field crickets, positive selection is relatively more important in the evolution of seminal proteins than in proteins not related to reproduction; or (2) purifying selection resulting from functional constraints is relatively more important in housekeeping genes than in the genes encoding seminal proteins. The second explanation argues that housekeeping and seminal protein genes are basically evolving under strong purifying selection but that seminal proteins are significantly less constrained than housekeeping genes.

The estimated mean for the selection parameter ω was an order of magnitude larger in genes encoding seminal proteins than in housekeeping genes. However, most of the genes encoding putative seminal proteins have d_N/d_S ratios less than 1, suggesting that many of the proteins of the seminal fluid are under purifying selection associated with functional constraint. However, within the class of genes thought to encode seminal proteins, there is a wide range of ω values. For 3 genes, the d_N/d_S ratio is greater than 1, suggesting strong positive selection acting at these loci. Moreover, another 6 proteins have d_N/d_S ratios greater than 0.5, whereas none of the control genes show an overall ω greater than 0.5. $\omega > 0.5$ has been proposed as a threshold value to identify candidate genes that may have experienced episodes of adaptive evolution (Swanson et al. 2001, 2004; Clark and Swanson 2005). In fact, in a recent study, Swanson et al. (2004) have shown that in a wide variety of taxa a very high proportion (95%) of the genes with an estimated overall ω value greater than 0.5 showed statistical evidence for adaptive evolution upon closer examination using analyses that allow for variation in the d_N/d_S between sites (Yang et al. 2000). Thus, our results suggest that a significant proportion of the seminal proteins in field crickets are indeed under positive selection and support a scenario in which the relative importance of positive selection in the

evolution of the genes expressed in the accessory gland is related to their biological function. On average, positive selection seems to be more important in the evolution of genes encoding putative seminal proteins than in the evolution of nonreproductive genes.

Positive selection is apparently an important evolutionary force driving the evolution of seminal proteins in organisms with polyandrous mating systems (*Drosophila* and primates). There is ample evidence that a subset of *Drosophila Acps*, intimately related to mechanisms of sperm precedence, are (or have been) under strong positive selection (e.g., Aguadé 1998, 1999). This selection pressure may arise primarily as the result of the underlying conflict between the different reproductive interests of males and females. In fact, in the *Drosophila* mating system, molecular adaptations to sperm competition in males seem to have a negative effect on female fitness (Wigby and Chapman 2005). Thus, it has been hypothesized that selection acting on seminal proteins directly related to sperm competition mechanisms fuel the antagonistic coevolution of reproductive signal/receptor systems, which in turn may be responsible for both rates of divergence and patterns of codon substitutions observed in seminal proteins.

Although such models of seminal protein evolution are reasonable and appear to explain patterns of evolution for at least one specific *Acp* locus (Wigby and Chapman 2005), our results suggest that observed patterns of divergence and nonsynonymous substitutions in seminal proteins in polyandrous species are not exclusively a consequence of conflict. The comparison between *Drosophila Acps* and *Gryllus* seminal proteins reveals that, in spite of the contrasting characteristics of their mating systems, the mean selection parameter ω and the proportion of loci assumed to be affected by positive selection are very similar in these 2 polyandrous taxa. In field crickets, experimental evidence suggest that seminal proteins have a positive rather than a negative effect on female fitness. In fact, during copulation, *Gryllus* males transfer seminal fluid products to females that increase female life span and lifetime fecundity (Wagner et al. 2001). So, postmating sexual selection driven by sperm competition and/or by a process of cryptic female choice analogous to conventional female choice (Eberhard 1996) are likely candidates responsible for the pattern of rapid evolution of seminal proteins in these taxa. However, other deterministic evolutionary forces such as natural selection cannot yet be ruled out. Further comparisons including both monandrous and polyandrous lineages would help to clarify the role of the different types of selection in the evolution of seminal proteins, the major component of seminal fluids.

Supplementary Material

Primers and PCR conditions are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

Literature Cited

Aguadé M. 1998. Different forces drive the evolution of the *Acp*26Aa and *Acp*26Ab accessory gland genes in the *Drosophila melanogaster* species. Genetics 150:1079–89.

Aguadé M. 1999. Positive selection drives the evolution of the *Acp*29AB accessory gland protein in *Drosophila*. Genetics 152:543–51.

Aguadé M, Miyashita N, Langley CH. 1992. Polymorphism and divergence in the *Mst*26A male accessory gland region in *Drosophila*. Genetics 132:755–70.

Alexander RD. 1957. The taxonomy of the field crickets of the Eastern United States (Orthoptera: Gryllidae: Acheta). Ann Entomol Soc Am 50:584–602.

Andrés JA, Arnqvist G. 2000. Genetic divergence of seminal signal-receptor system in houseflies: the footprints of sexually antagonistic coevolution? Proc R Soc Lond B Biol Sci 268:399–405.

Barrier M, Bustamante CD, Yu J, Purugganan MD. 2003. Selection on rapidly evolving proteins in the *Arabidopsis* genome. Genetics 163:723–33.

Begun DJ, Lindfords HA. 2005. Rapid evolution of genomic *Acps* in the melanogaster subgroup of *Drosophila*. Mol Biol Evol 22:2010–21.

Begun DJ, Whitley P, Todd BL, Waldrip-Dail HM, Clark AG. 2000. Molecular population genetics of male accessory gland proteins in *Drosophila*. Genetics 156:1879–88.

Bendtsen JD, Nielsen H, Heijne GV, Brunak S. 2004. Improved prediction of signal peptides: SignalIP 3.0. J Mol Evol 340:783–95.

Bissoondath CJ, Wiklund C. 1995. Protein content of spermatophores in relation to monandry/polyandry in butterflies. Behav Ecol Sociobiol 37:365–71.

Borodovsky M, McIninch J. 1993. GeneMark: parallel gene recognition for both DNA strands. Comput Chem 17:123–33.

Braswell WE, Andres JA, Maroja L, Harrison RG, Howard DJ, Swanson WJ. 2006. Identification and comparative analysis of accessory gland proteins in Orthoptera. Genome. Forthcoming.

Burpee DM, Sakaluk SK. 1993. Repeated matings offset the cost of reproduction in female crickets. Evol Ecol 7:240–50.

Bustamante CD, Nielsen R, Hartl DL. 2002. A maximum likelihood method for analyzing pseudogene evolution: implications for silent site evolution in humans and rodents. Mol Biol Evol 19:110–7.

Chapman T, Bangham J, Vinti G, Seifried B, Lung O, Wolfner MF, Smith HK, Partridge L. 2003. The sex peptide of *Drosophila melanogaster*: female post-mating responses analyzed by using RNA interference. Proc Natl Acad Sci USA 100:9923–28.

Chapman T, Liddle LF, Kalb JM, Wolfner MF, Partridge L. 1995. Cost of mating in *Drosophila melanogaster* females is mediated by accessory gland products. Nature 373:242–4.

Cheeseman MT, Gillot C, Ahmed I. 1990. Structural spermatophore proteins and trypsin-like enzyme from the accessory reproductive glands of the male grasshopper, *Melanoplus sanguinipes*. J Exp Zool 255:193–204.

Clark NJ, Swanson WJ. 2005. Pervasive adaptive evolution in primate seminal proteins. PLoS Genet 1:335–42.

Cordero C. 1995. Ejaculate substances that affect female reproductive physiology and behavior-honest or arbitrary traits. J Theor Biol 174:453–61.

Destephano DB, Brady UE. 1977. Prostaglandin and prostaglandin-synthetase in the cricket *Acheta domesticus*. J Insect Physiol 23:905–11.

Destephano DB, Brady UE, Lovins RE. 1974. Synthesis of prostaglandin by reproductive tissue of the male house cricket, *Acheta domesticus*. Prostaglandins 6:71–9.

Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT. 2004. Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. Nat Genet 36:1326–9.

Eberhard WG. 1996. Female control: sexual selection by cryptic female choice. Princeton, NJ: Princeton University Press.

Feng X, Happ GM. 1996. Isolation and sequencing of the gene encoding *Sp23*, a structural protein of the spermatophore of the mealworm beetle, *Tenebrio mollitor*. Gene 179:257–62.

Fiumera AC, Dumont BL, Clark AG. 2005. Sperm competitive ability in *Drosophila melanogaster* is associated with variation in male reproductive proteins. Genetics 169:243–57.

Garcia-Gonzalez F, Gomendio M. 2004. Adjustment of copula duration and ejaculate size according to the risk of sperm competition in the golden egg bug (*Phyllomorpha laciniata*). Behav Ecol 15:23–30.

Gillott C. 2003. Male accessory gland secretions: modulators of female reproductive physiology and behavior. Annu Rev Entomol 48:163–84.

Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–36.

Harrison RG, Bogdanowicz SM. 1997. Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. Evolution 51:493–505.

Harshman LG, Prout T. 1994. Sperm displacement without sperm transfer in *Drosophila melanogaster*. Evolution 48:758–66.

Heller KS, Faltin S, Fleischmann P, Helversen OV. 1998. The chemical composition of the spermatophore in some species of phanaropterid bushcrickets (Orthoptera: Tettigonioidea). J Insect Physiol 44:1001–8.

Herndon LA, Wolfner MF. 1995. A *Drosophila* seminal protein, *Acp26Aa*, stimulates egg laying in females for 1 day after mating. Proc Natl Acad Sci USA 92:10114–8.

Holland B, Rice W. 1998. Chase-away sexual selection: antagonistic seduction versus resistance. Evolution 52:1–7.

Huang Y, Orti G, Sutherlin M, Duhachek A, Zera A. 2000. Phylogenetic relationships of North American crickets inferred from mitochondrial DNA data. Mol Phylogenet Evol 17:48–57.

Ivy TM, Sakaluk SK. 2005. Polyandry promotes enhanced offspring survival in decorated crickets. Evolution 59:152–9.

Jensen-Seaman MI, Li WH. 2003. Evolution of the hominid semenogelin genes, the major proteins of ejaculated semen. J Mol Evol 57:261–70.

Karlsson B. 1995. Resource allocation and mating systems in butterflies: a comparative study. Evolution 49:955–61.

Kern AD, Jones CD, Begun DJ. 2004. Molecular genetics of male accessory gland in the *Drosophila simulans* complex. Genetics 167:725–35.

Kingan SB, Tatar M, Rand DM. 2003. Reduced polymorphism in the chimpanzee semen coagulating protein, semenogelin I. J Mol Evol 57:159–69.

Letunic I, Goodstadt L, Dickens NJ, Doerks T, Schultz J, Mott R, Ciccarelli F, Copley RR, Ponting CP, Bork P. 2002. Recent improvements to the SMART domain-based sequence annotation resource. Nucleic Acids Res 30:242–4.

Liu HF, Kubli E. 2003. Sex-peptide is the molecular basis of the sperm effect in *Drosophila melanogaster*. Proc Natl Acad Sci USA 100:9929–33.

Loher W, Ganjian WI, Kubo I, Stanley-Samuelson D, Tobe S. 1981. Prostaglandins: their role in egg laying in the cricket *Teleogryllus commodus*. Proc Natl Acad Sci USA 78:7835–8.

Lung O, Kuo L, Wolfner MF. 2001. Males transfer antibacterial proteins from their accessory gland and ejaculatory duct to their mates. J Insect Physiol 47:617–22.

Manly BFJ. 1991. Randomization and Monte Carlo methods in biology. New York: Chapman and Hall.

Meister M, Lemaitre B, Hoffman JA. 1997. Antimicrobial peptide defense in *Drosophila*. Bioassays 19:1019–26.

Mueller JL, Ravi Ram K, McGraw LA, Bloch Qazi C, Siggia ED, Clark AG, Aquadro CF, Wolfner MF. 2005. Cross-species comparison of *Drosophila* male accessory gland protein genes. Genetics 171:131–43.

Mueller JL, Ripoll DR, Aquadro CF, Wolfner MF. 2004. Comparative structural modeling and inference of conserved protein classes in *Drosophila* seminal fluid. Proc Natl Acad Sci USA 101:13542–7.

Neubaum DM, Wolfner MF. 1999. Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. Genetics 153:845–57.

Paesen GC, Schwartz MB, Peferoen M, Weyda F, Happ GM. 1992. Amino acid sequence of *Sp23*, a structural protein of the spermatophore of the mealworm beetle, *Tenebrio mollitor*. J Biol Chem 267:18852–7.

Parker GA. 1970. Sperm competition and its evolutionary consequences in the insects. Biol Rev 45:525–67.

Parkinson J, Blaxter M. 2004. Expressed sequence tags: analysis and annotation. Methods Mol Biol 270:93–126.

Ramm S, Parker GA, Stockley P. 2005. Sperm competition and the evolution of male reproductive anatomy in rodents. Proc R Soc Lond B Biol Sci 272:949–55.

Reinhardt K. 2001. Determinants of the ejaculate size in a grasshopper. Behav Ecol Sociobiol 50:503–10.

Rice WR, Holland B. 1997. The enemies within: intergenomic conflict, interlocus contest evolution (ICE) and the intraspecific red queen. Behav Ecol Sociobiol 41:1–10.

Sakaluk SK, Eggert AK. 1996. Female control of sperm transfer and intra-specific variation in sperm precedence: antecedents to the evolution of a courtship food gift. Evolution 50:694–703.

Sambrook J, Russel DW, Sambrook J. 2001. Molecular cloning: a laboratory manual. New York: Cold Spring Harbor Laboratory Press.

Simmons LW. 1986. Female choice in the field cricket, *Gryllus bimaculatus* (De Geer). Anim Behav 34:1463–470.

Simmons LW. 1988. The contribution of multiple mating and spermatophore consumption on the lifetime reproductive success of female field crickets (*Gryllus bimaculatus*). Ecol Entomol 13:57–69.

Simmons LW. 2001. Sperm competition and its evolutionary consequences in the insects. Princeton, NJ: Princeton University Press.

Solymar BD, Cade WH. 1990. Heritable variation for female mating frequency in field crickets, *Gryllus integer*. Behav Ecol Sociobiol 26:73–6.

Stanley-Samuelson DW, Loher W. 1983. Arachidonic and other long-chain polyunsaturated fatty acids in spermatophores and spermathecae of *Teleogryllus commodus*: significance in prostaglandin-mediated reproductive behaviour. J Insect Physiol 29:41–5.

Stanley-Samuelson DW, Loher W. 1986. Prostaglandins in insect reproduction. Ann Entomol Soc Am 79:841–53.

Stewart AD, Morrow EH, Rice WR. 2005. Assessing putative interlocus sexual conflict in *Drosophila melanogaster* using experimental evolution. Proc R Soc Lond B Biol Sci 272:2029–935.

Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc Natl Acad Sci USA 95:4051–4.

Swanson WJ, Wong A, Wolfner MF, Aquadro CF. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* identifies genes subjected to positive selection. Genetics 168: 1457–65.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–80.

Tram U, Wolfner MF. 1999. Male seminal fluid proteins are essential for sperm storage in *Drosophila melanogaster*. Genetics 153:837–44.

Tsaur SC, Ting CT, Wu CI. 1998. Positive selection drives the evolution of a gene of male reproduction, *Acp*26Aa, of *Drosophila* II. Divergence versus polymorphism. Mol Biol Evol 15:1040–6.

Tsaur SC, Ting CT, Wu CI. 2001. Sex in *Drosophila mauritania*: a very high level of amino acid polymorphism in a male reproductive protein gene, *Acp*26Aa. Mol Biol Evol 18:22–6.

Tsaur SC, Wu CI. 1997. Positive selection and the molecular evolution of a gene of male reproduction, *Acp*26Aa of *Drosophila*. Mol Biol Evol 14:544–9.

Wagner WE, Kelley RJ, Tucker KR, Haper CJ. 2001. Females receive a life-span benefit from male ejaculates in a field cricket. Evolution 55:994–1001.

Wedell N. 1993. Spermatophore size in bushcrickets: comparative evidence for nuptial gifts as sperm protection devices. Evolution 47:1203–12.

Wigby S, Chapman T. 2005. Sex peptide causes mating costs in female *Drosophila melanogaster*. Curr Biol 15:316–21.

Wolfner MF. 1997. Tokens of love: functions and regulations of *Drosophila* accessory male products. Insect Biochem Mol Biol 27:179–92.

Yang Z. 2000. Phylogenetic analysis by maximum likelihood (PAML). London: University College London.

Yang Z, Bielawski B. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol Biol 15:496–503.

Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908–17.

Yang Z, Nielsen R, Goldman N, Pedersen AMK. 2000. Codon substitution models for heterogenous selection pressure at amino acid sites. Genetics 155:431–49.