

# Evolutionary Proteomics Reveals Distinct Patterns of Complexity and Divergence between Lepidopteran Sperm Morphs

Emma Whittington<sup>1</sup>, Timothy L. Karr<sup>2</sup>, Andrew J. Mongue<sup>3</sup>, Steve Dorus<sup>1,\*</sup>, and James R. Walters<sup>3,\*</sup>

<sup>1</sup>Center for Reproductive Evolution, Department of Biology, Syracuse University

<sup>2</sup>Center for Mechanisms of Evolution, The Biodesign Institute, Arizona State University

<sup>3</sup>Department of Ecology and Evolutionary Biology, University of Kansas

\*Corresponding authors: E-mails: sdorus@syr.edu; jrwalters@ku.edu.

Accepted: April 5, 2019

**Data deposition:** This project has been deposited at the ProteomeXchange Consortium via the PRIDE partner repository under the accession PXD010168.

## Abstract

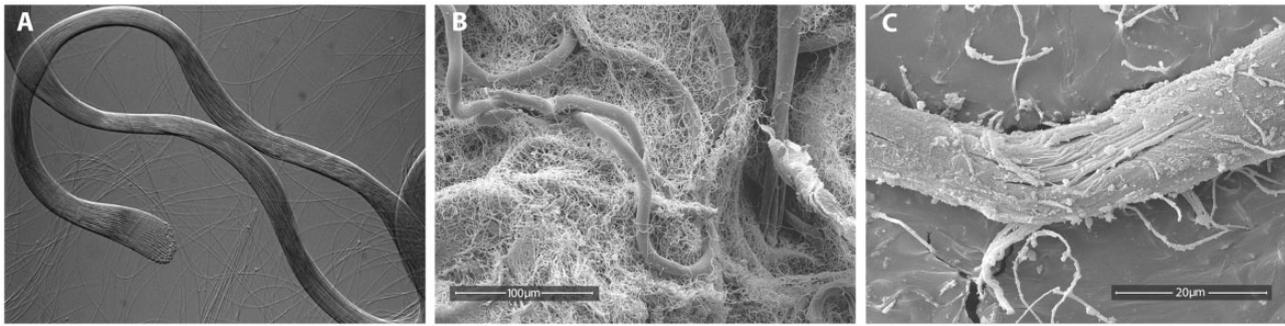
Spermatozoa are one of the most strikingly diverse animal cell types. One poorly understood example of this diversity is sperm heteromorphism, where males produce multiple distinct morphs of sperm in a single ejaculate. Typically, only one morph is capable of fertilization and the function of the nonfertilizing morph, called parasperm, remains to be elucidated. Sperm heteromorphism has multiple independent origins, including Lepidoptera (moths and butterflies), where males produce a fertilizing *eupyrene* sperm and an *apyrene* parasperm, which lacks a nucleus and nuclear DNA. Here we report a comparative proteomic analysis of eupyrene and apyrene sperm between two distantly related lepidopteran species, the monarch butterfly (*Danaus plexippus*) and Carolina sphinx moth (*Manduca sexta*). In both species, we identified ~700 sperm proteins, with half present in both morphs and the majority of the remainder observed only in eupyrene sperm. Apyrene sperm thus have a distinctly less complex proteome. Gene ontology (GO) analysis revealed proteins shared between morphs tend to be associated with canonical sperm cell structures (e.g., flagellum) and metabolism (e.g., ATP production). GO terms for morph-specific proteins broadly reflect known structural differences, but also suggest a role for apyrene sperm in modulating female neurobiology. Comparative analysis indicates that proteins shared between morphs are most conserved between species as components of sperm, whereas morph-specific proteins turn over more quickly, especially in apyrene sperm. The rapid divergence of apyrene sperm content is consistent with a relaxation of selective constraints associated with fertilization and karyogamy. On the other hand, parasperm generally exhibit greater evolutionary lability, and our observations may therefore reflect adaptive responses to shifting regimes of sexual selection.

**Key words:** spermatogenesis, fertility, sexual selection, parasperm, apyrene sperm, eupyrene sperm.

## Introduction: The Enigma of Sperm Heteromorphism

Sperm heteromorphism is a phenomenon in which males produce multiple distinct sperm morphs as a developmentally normal and regulated process during gametogenesis. This phenomenon has arisen across a broad range of taxa, including multiple independent origins in Insecta and Mollusca, and a few vertebrates (Swallow and Wilkinson 2002; Till-Bottraud et al. 2005; Hayakawa 2007). Sperm morphs are defined by

their fertilization capacity, with only one morph (eusperm) capable of successful fertilization, whereas the remaining morphs are not (parasperm) (Healy and Jamieson 1981). Although parasperm function has yet to be conclusively determined in any taxa, two adaptive hypotheses have been investigated: 1) Facilitation of eusperm or 2) mediation of sperm competition (reviewed in Swallow and Wilkinson 2002; Till-Bottraud et al. 2005). The former hypothesis is supported by observations that fertility can be strongly impacted



**FIG. 1.**—Microscopy of *Manduca* eupyrene sperm bundles and apyrene sperm dissected from the male reproductive tract. (A) Differential interference contrast image of a single bundle of eupyrene sperm, with unbundled individual apyrene sperm visible in the background. (B) Scanning electron microscopy (SEM) image of sheathed eupyrene bundles amongst an abundance of apyrene sperm. (C) SEM image of a eupyrene sperm bundle with a section of the sheath-matrix removed to reveal sperm tails. To our knowledge, these are the first SEM images of lepidopteran sperm in the public record; additional SEM images are provided in supplemental file S2, [Supplementary Material](#) online.

by parasperm absence or variety (Oppliger et al. 1998; Sahara and Kawamura 2002; Sahara and Takemura 2003). The latter is supported by the tailoring of parasperm investment in response to the intensity of sperm competition (He and Miyata 1997; Oppliger et al. 1998; Wedell and Cook 1999). As would be expected given their presumed functional diversification, parasperm and eusperm exhibit distinct evolutionary patterns, with parasperm diverging faster and showing greater predicted evolvability (Holman et al. 2008; Moore et al. 2013; Snook 1997).

One striking example of sperm heteromorphism occurs in Lepidoptera (moths and butterflies), where the parasperm morph, called *apyrene* sperm, lacks a nucleus and nuclear DNA (thoroughly reviewed by Friedlander, Seth and Reynolds 2005). Apyrene sperm are present in all studied species of Lepidoptera except for the most ancestrally diverging lineage. Thus, sperm heteromorphism appears to have arisen ancestrally and has been retained across taxa with diverse mating systems and sperm competition intensities. In most species, apyrene sperm vastly outnumber their nucleated eusperm counterparts (called *eupyrene* sperm), typically accounting for ~85–90% of sperm produced. Numerous microscopy studies contrasting apyrene and eupyrene sperm have revealed several structural differences beyond apyrene sperm lacking a nucleus, including 1) apyrene sperm lack an acrosome, 2) apyrene sperm are shorter, and 3) prior to ejaculation, eupyrene sperm remain bundled in an extracellular sheath (fig. 1). Developmentally, eupyrene sperm are produced before apyrene sperm, beginning in late larval instars and ceasing during pupal development. In contrast, the start of apyrene sperm production typically coincides with the initiation of pupation and lasts into adulthood. Apyrene and eupyrene spermatogenesis differs greatly, stemming from the improper pairing of homologous chromosomes in apyrene meiosis. Consequently, remaining nuclear fragments are ejected along with cytoplasmic debris during peristaltic squeezing in apyrene sperm (Friedlander et al. 2005). Other

meiotic differences include reduced endoplasmic reticulum (ER) and microtubule mass in apyrene sperm (Wolf 1992). Such structural and development differences between sperm morphs are likely associated with differences in the protein content of sperm morphs. However, very little is known about these differences at the molecular level.

In previous studies, we employed high-throughput liquid chromatography tandem mass spectrometry (LC–MS/MS) proteomics to analyze comixed apyrene and eupyrene samples from both the Carolina sphinx moth (*Manduca sexta*, henceforth *Manduca*) and monarch butterfly (*Danaus plexippus*) (Whittington et al. 2015, 2017). These studies revealed substantial divergence in proteome content between species but did not include direct comparisons between morphs. However, morph-specific proteome analysis using 2D gel electrophoresis in monarch indicated that the eupyrene sperm proteome was notably more complex (Karr and Walters 2015). Extending analyses to separately characterize apyrene and eupyrene sperm proteomes will establish their molecular differences and potentially improve our understanding of apyrene sperm function.

Here we report the results of LC–MS/MS proteomic analysis applied to isolated apyrene and eupyrene sperm samples. Doing so in both monarch and *Manduca* allowed us to contrast the overlap in protein content and function between morphs and between species. We found proteome composition has diverged more rapidly for morph-specific than shared proteins. Functional annotations broadly reflect known structural differences, and hint at a role for apyrene sperm in modulating female neurobiology.

## Materials and Methods

### Sperm Samples and Proteomic Analysis

Two species were used in this study, the monarch butterfly (*D. plexippus*, Monarch Watch, Lawrence, KS), and the Carolina sphinx moth (*M. sexta*, Carolina Biological,

Burlington, NC). Sperm samples were isolated from male seminal vesicles 5–10 days post eclosion via a small incision in the mid to distal region of the seminal vesicle. Apyrene and eupyrene sperm were isolated using the “panning” method described in Karr and Walters (2015). Briefly, total seminal vesicle contents were placed in a petri dish of phosphate buffered saline and separated via repeated bouts of “panning,” in which the petri dish is rotated in a circular motion resulting in the denser eupyrene bundles collecting in the center while apyrene sperm dissipate to the edge. Separation efficiency and sample purity were assessed by visual inspection under a microscope and was judged complete when no apyrene sperm could be visually detected among the eupyrene bundles, and vice versa, as represented in the images from Karr and Walters (2015). Samples from 3 to 5 males were pooled for each of 3 biological replicates in each species, resulting in a total of 12 samples.

Proteomic analyses followed the protocol reported in Whittington et al. (2017). Described briefly, each sample was size-separated on a poly-acrylamide gel and cut into four slices that were individually analyzed via LC–MS/MS. Resulting mass spectra were matched to predicted proteins from each species’ genome using the Trans-Proteomic Pipeline (Zhan and Reppert 2013; Deutsch et al. 2015; Kanost et al. 2016). Proteins included in the final sperm proteomes met the following criteria: 1) Identification in two or more biological replicates or 2) identification in a single replicate by two or more unique peptides. For quantitative analysis, relative abundance estimates were calculated using the normalized spectral factor method. All LC–MS/MS data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD010168 (Vizcaíno et al. 2016).

### Functional Annotation and Homology

Functional annotations and GO assignments were generated by PANNZER (Ashburner et al. 2000; Koskinen et al. 2015). GO-term enrichment tests were performed using the GOstats Bioconductor package, employing the “conditional =TRUE” setting to account for hierarchical redundancy in GO classifications (Falcon and Gentleman 2007). Hypergeometric tests used the union of apyrene and eupyrene proteins as the background “universe” of genes to identify terms enriched in morph-specific proteins or proteins shared between morphs.

Orthologs between monarch and *Manduca* gene sets were predicted via the proteinortho pipeline (Lechner et al. 2011) with default settings, using the longest isoform per gene. Predicted orthologs identified in the sperm proteome of both species were further classified as *sperm homologs*, and those found in the same subset between species were *subset homologs*. In a few cases, paralogy resulted in small gene groups with a one-to-many or many-to-many relationship between species. Sperm proteins identified within such

paralogous groups in both species were also classified as a sperm or subset homologs. Proportions of homologous proteins were compared between the three subsets of proteins, with significant differences assessed as nonoverlapping 95% confidence intervals, generated by 1,000 bootstrap replicates.

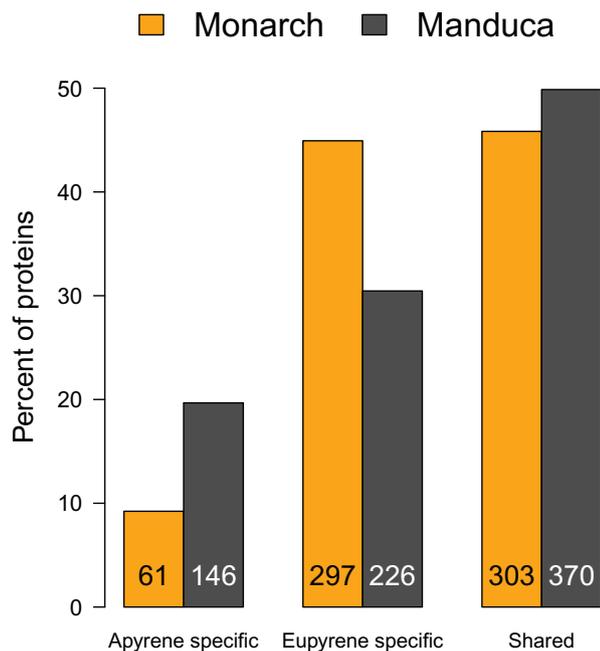
### Microscopy

For light microscopy, sperm were dissected from *Manduca* seminal vesicles and imaged using differential interference contrast on an Olympus BX60 microscope.

For SEM images, an adult male *Manduca* or monarch was dissected to obtain intact seminal vesicles containing eupyrene and apyrene sperm. This tissue was fixed in 10% formalin for 3.5 h, and then washed four times with 70% ethanol, with 5 min in between each wash. Samples were rested for 1 h and washed twice with distilled water with 30 min in between washes. All liquid was removed, and samples were soaked overnight in 1% OsO<sub>4</sub> in the dark. The next day, samples were washed two times with distilled water then dried with a series of increasing ethanol washes (70%, 95%, 100%) with 10 min rest in between each wash. Seminal vesicles were removed from ethanol, placed on a depressed slide, and treated with hexamethyldisilazane as a final drying step. After 30 min, excess liquid was removed, and the samples were air dried for an additional 10 min before being placed on a prepared specimen mount stub. After mounting to the stub, seminal vesicles were ruptured with a needle and the contents were spread across the stub using a dissecting pin. The stub was then sputter coated with 35 nm of gold and imaged on an FEI Versa 3D Dual Beam microscope at the University of Kansas’s Microscopy and Analytical Imaging Lab.

## Results I: Proteome Composition, Complexity, and Function

LC–MS/MS analysis of isolated apyrene and eupyrene sperm samples identified a combined total of 742 proteins in *Manduca* and 661 proteins in monarch (fig. 2; [supplementary tables S1 and S2, Supplementary Material](#) online). Our analysis confirms the previous observation in monarch that apyrene sperm is less complex in protein composition relative to eupyrene sperm (Karr and Walters 2015), and extends this observation to *Manduca*. Intersecting results from the two sperm morphs yields three protein “subsets,” those that are detected only in eupyrene sperm, those detected only in apyrene sperm, and proteins detected in both sperm morphs (“shared”). For convenience, we refer to proteins detected only in one sperm morph as being “specific” to that morph. However, it is crucial to note that in LC–MS/MS studies, failure to detect a protein is not necessarily a robust indication of absence, as the protein may well be present at abundances



**Fig. 2.**—Portion of proteins found in each subset of the sperm proteome. Proteins identified only in apyrene or eupyrene sperm are “specific,” whereas proteins identified in both morphs are “shared.” Bar heights represent percent of total proteins. Numbers at the base of bars are the counts of proteins identified in each subset.

too low to be readily detected. Proteins that are differentially abundant, where that difference spans the detection threshold, may be erroneously classified as “specific.” We have sought to assess the impact of this potential artifact in our analysis by examining the proportion of morph-specific proteins across a range of minimum abundance thresholds. Details of this thresholding analysis are given in [supplementary file 3, Supplementary Material](#) online, and results do not indicate that such differential detection biases are strongly influencing our assessment of morph-specific proteins.

In both species, approximately half of identified proteins were shared between sperm types (fig. 2). Given that these two species diverged over 100 Ma (Heikkilä et al. 2012), it seems likely that reduced apyrene proteome complexity and substantial overlap between morphs is generally representative of Lepidoptera. Nonetheless, the proportions of proteins specific to or shared between morphs differed substantially between species (fig. 2;  $\chi^2 = 47.1$ ,  $df = 2$ ,  $P < 0.0001$ ), primarily reflecting monarch’s greater disparity in complexity between morph-specific proteomes.

In both species, approximately three-quarters of sperm proteins were successfully annotated with gene ontology (GO) terms, with no significant difference in the proportion of annotated proteins between the morph-specific or shared subsets ( $\chi^2 < 4.2$  in both species,  $df = 1$ ,  $P > 0.05$ ). Consequently, it is unlikely that our results are impacted by any biases in annotation quality and coverage between

subsets. GO-term enrichments highlight broad functional distinctions among morph-specific and shared proteins (tables 1 and 2; [supplementary material S3, Supplementary Material](#) online). Although apyrene and eupyrene sperm necessarily play distinct (though yet unresolved) roles in fertilization, these discrete morphs have many similarities in morphology (e.g., axonemal-based flagellum), physiology (e.g., ATP production) and behavior (e.g., motility) (Friedlander et al. 2005). Thus shared proteins are expected to be enriched for GO-terms related to these functions, as well as others also associated with spermatozoa. Consistent with this prediction, in both species the shared set of proteins tend to have broad associations with the cytoskeletal structure, mitochondria, and cilia (the sperm flagellum is a modified cilium [Dallai 2014]).

Similarly, structural differences between apyrene and eupyrene sperm are observed in morph-specific GO-term enrichments. Most prominently, apyrene sperm lack a nucleus and nuclear DNA. Additionally, during meiosis, apyrene sperm lack thick layers of perinuclear ER typically observed in eupyrene sperm, suggesting the ER is greatly reduced or missing in mature apyrene sperm (Wolf 1992). Accordingly, among eupyrene-specific proteins, GO-terms associated with the ER and nuclear membrane are enriched, particularly in *Manduca*. Also, in *Manduca*, terms associated with protein–DNA packaging are among the most significantly enriched in eupyrene-specific proteins, reflecting chromatin-related proteins absent from apyrene sperm. Apyrene sperm also lack an acrosome, a vesicle/vacuole organelle typically located in the head of sperm; a corresponding enrichment for vacuole-related GO-terms was identified among eupyrene-specific proteins in both species. Finally, an enrichment of terms associated with extracellular structures is unique to eupyrene-specific proteins. In contrast to apyrene sperm, eupyrene sperm from an individual cyst are packaged and transferred to the female in bundles, sheathed in a proteinaceous extracellular matrix that is subsequently degraded in the female (fig. 1; additional SEM figures are given in [supplemental file 2, Supplementary Material](#) online). Eupyrene sperm samples in this study were isolated from seminal vesicles and thus were still bundled, hence these “extracellular” eupyrene proteins likely comprise this sheathing.

Although GO-term analysis at the molecular level broadly mirrors previously known structural differences between sperm morphs, it does not yield much insight into apyrene sperm function. We find few commonalities between species among GO-terms enriched in apyrene-specific proteins ([supplementary material S3, Supplementary Material](#) online) and this may reflect distinct functions of apyrene sperm between our study species. Parasperm potentially act as vehicles transporting molecules to the female reproductive tract, as is proposed in some mollusk species. In *Littorina obtusa*, a sperm protein LOSP (*Littorina* Sperm Protein) has been found exclusively in granules within parasperm, which presumably will be

**Table 1**

The Gene Ontology (GO) Cellular Component Terms Significantly Enriched in the Three Subsets of Lepidopteran Sperm

Manduca			Monarch		
GO: ID (CC)	P-Value	Description	GO: ID (CC)	P-Value	Description
<b>Shared</b>					
GO: 0031514	0.00010	Motile cilium	GO: 0031514	2.30E-07	Motile cilium
GO: 0042995	0.00028	Cell projection	GO: 0120038	0.00022	Plasma membrane bounded cell projection part
GO: 0099081	0.00094	Supramolecular polymer	GO: 0043209	0.00031	Myelin sheath
GO: 0015630	0.00205	Microtubule cytoskeleton	GO: 0097014	0.00052	Ciliary plasm
GO: 0120038	0.00285	Plasma membrane bounded cell projection part	GO: 0005739	0.00083	Mitochondrion
GO: 0043209	0.00446	Myelin sheath	GO: 0005622	0.00131	Intracellular
GO: 0005739	0.00558	Mitochondrion	GO: 0120025	0.00141	Plasma membrane bounded cell projection
GO: 1990204	0.00728	Oxidoreductase complex	GO: 0099568	0.00292	Cytoplasmic region
GO: 0005929	0.00958	Cilium	GO: 0044430	0.00398	Cytoskeletal part
GO: 0005622	0.01137	Intracellular	GO: 0015630	0.00440	Microtubule cytoskeleton
<b>Apyrene-specific</b>					
GO: 0005839	0.00732	Proteasome core complex	GO: 0016021	0.01110	Integral component of membrane
GO: 1905369	0.00893	Endopeptidase complex	GO: 0044451	0.02725	Nucleoplasm part
GO: 0005881	0.02507	Cytoplasmic microtubule	GO: 0005912	0.03632	Adherens junction
GO: 0034702	0.03798	Ion channel complex	GO: 0150034	0.03639	Distal axon
GO: 1905354	0.03798	Exoribonuclease complex	GO: 0044440	0.03639	Endosomal part
GO: 0005868	0.03798	Cytoplasmic dynein complex	GO: 0044459	0.03839	Plasma membrane part
GO: 0001534	0.03798	Radial spoke			
<b>Eupyrene-specific</b>					
GO: 0005783	3.25E-07	Endoplasmic reticulum (ER)	GO: 0098588	0.00109	Bounding membrane of organelle
GO: 0044815	0.00013	DNA packaging complex	GO: 0016471	0.00335	Vacuolar proton-transporting V-type ATPase complex
GO: 0032993	0.00040	Protein-DNA complex	GO: 0005773	0.00611	Vacuole
GO: 0044391	0.00116	Ribosomal subunit	GO: 0033180	0.00763	Proton-transporting V-type ATPase, V1 domain
GO: 0005578	0.00124	Proteinaceous extracellular matrix	GO: 0031984	0.01306	Organelle subcompartment
GO: 0042175	0.00269	Nuclear outer membrane-ER	GO: 0016469	0.01524	Proton-transporting two-sector ATPase complex
GO: 0098588	0.00293	Bounding membrane of organelle	GO: 0005783	0.03806	ER
GO: 0098827	0.00294	ER subcompartment	GO: 0030659	0.03926	Cytoplasmic vesicle membrane
GO: 0016471	0.00382	Vacuolar proton-transporting V-type ATPase complex	GO: 0033181	0.03926	Plasma membrane proton-transporting V-type ATPase complex
GO: 0005918	0.00382	Septate junction	GO: 0005794	0.04726	Golgi apparatus

delivered to the female via exocytosis in the female reproductive tract (Lobov et al. 2018). Alternatively, proteins may bind to sperm, similar to some seminal fluid proteins in *Drosophila melanogaster*, to aid transport into the female. Along these lines, *neuron development* is the most significantly enriched "Biological Process" term among monarch apyrene-specific proteins (table 2). Other terms associated with neuronal-development are also enriched in *Manduca* (though with less significance; table 2; supplementary table S3, Supplementary Material online). It is well-known in *Drosophila* that components of the male ejaculate impact

female neurobiology, modulating postmating shifts in behavior and physiology (Chow et al. 2013). In *Helicoverpa armigera* moths, male accessory gland extracts produce a strong postmating response in females (Fan et al. 1999), mediated by a receptor specifically expressed in both female neural and reproductive structures (Hanin et al. 2011, 2012). Thus, there is precedent for male-derived proteins to modulate female neuro-endocrinology and reproductive physiology. It is therefore plausible that apyrene sperm deliver neuro-endocrine active proteins that modulate female postmating responses. This is in contrast to previously proposed functions for apyrene

**Table 2**

The Gene Ontology (GO) Biological Process Terms Significantly Enriched in the Three Subsets of Lepidopteran Sperm

Manduca			Monarch		
GO: ID (BP)	P-Value	Description	GO: ID (BP)	P-Value	Description
<b>Shared</b>					
GO: 0120031	0.00002	Plasma membrane bounded cell projection assembly	GO: 0120031	0.00004	Plasma membrane bounded cell projection assembly
GO: 0044782	0.00003	Cilium organization	GO: 0044782	0.00005	Cilium organization
GO: 0015980	0.00018	Energy derivation by oxidation of organic compounds	GO: 0060285	0.00009	Cilium-dependent cell motility
GO: 0007018	0.00039	Microtubule-based movement	GO: 0006163	0.00112	Purine nucleotide metabolic process
GO: 0060285	0.00098	Cilium-dependent cell motility	GO: 0006165	0.00128	Nucleoside diphosphate phosphorylation
GO: 0070925	0.00117	Organelle assembly	GO: 0048515	0.00128	Spermatid differentiation
GO: 0009150	0.00273	Purine ribonucleotide metabolic process	GO: 0006091	0.00176	Generation of precursor metabolites and energy
GO: 0009144	0.00273	Purine nucleoside triphosphate metabolic process	GO: 0006753	0.00176	Nucleoside phosphate metabolic process
GO: 0009142	0.00394	Nucleoside triphosphate biosynthetic process	GO: 0000226	0.00220	Microtubule cytoskeleton organization
GO: 0006753	0.00412	Nucleoside phosphate metabolic process	GO: 0003341	0.00262	Cilium movement
<b>Apyrene</b>					
GO: 0042326	0.00433	Negative regulation of phosphorylation	GO: 0048666	0.00093	Neuron development
GO: 0043171	0.00638	Peptide catabolic process	GO: 0048699	0.00193	Generation of neurons
GO: 0010259	0.01062	Multicellular organism aging	GO: 0051252	0.00373	Regulation of RNA metabolic process
GO: 0010563	0.01068	Negative regulation of phosphorus metabolic process	GO: 1903506	0.00379	Regulation of nucleic acid-templated transcription
GO: 0018208	0.01619	Peptidyl-proline modification	GO: 0045892	0.00381	Negative regulation of transcription, DNA templated
GO: 0006749	0.01619	Glutathione metabolic process	GO: 0030154	0.00408	Cell differentiation
GO: 0009056	0.01874	Catabolic process	GO: 0051052	0.00479	Regulation of DNA metabolic process
GO: 0051014	0.03943	Actin filament severing	GO: 1902679	0.00599	Negative regulation of RNA biosynthetic process
GO: 0021859	0.03943	Pyramidal neuron differentiation	GO: 0007399	0.00881	Nervous system development
GO: 0021884	0.03943	Forebrain neuron development	GO: 0045934	0.00893	Negative regulation of nucleobase-containing compound
<b>Eupyrene</b>					
GO: 0099131	0.00004	ATP hydrolysis coupled ion transmembrane transport	GO: 0099131	0.00072	ATP hydrolysis coupled ion transmembrane transport
GO: 0034976	0.00009	Response to endoplasmic reticulum stress	GO: 0015988	0.00072	Energy coupled proton transmembrane transport, against electrochemical gradient
GO: 0015988	0.00011	Energy coupled proton transmembrane transport, against electrochemical gradient	GO: 0044092	0.01014	Negative regulation of molecular function
GO: 0007424	0.00025	Open tracheal system development	GO: 0043062	0.01800	Extracellular structure organization
GO: 0071705	0.00072	Nitrogen compound transport	GO: 0042176	0.01853	Regulation of protein catabolic process
GO: 0034660	0.00112	ncRNA metabolic process	GO: 0032269	0.02023	Negative regulation of cellular protein metabolic process

(continued)

Table 2 Continued

Manduca			Monarch		
GO: ID (BP)	P-Value	Description	GO: ID (BP)	P-Value	Description
GO: 0008104	0.00123	Protein localization	GO: 0044265	0.02473	Cellular macromolecule catabolic process
GO: 0009058	0.00136	Biosynthetic process	GO: 0071702	0.02565	Organic substance transport
GO: 0009059	0.00296	Macromolecule biosynthetic process	GO: 1905114	0.02715	Cell surface receptor signaling pathway involved in cell-cell signaling
GO: 0001763	0.00349	Morphogenesis of a branching structure	GO: 0050790	0.03536	Regulation of catalytic activity

sperm in Lepidoptera, including to aid eusperm transport in the female reproductive tract, and to bias paternity (Silberglied et al. 1984; Cook and Wedell 1999; Watanabe et al. 2000). Additional parasperm functions have been proposed in other taxa. Of note, parasperm in *Drosophila pseudoobscura* have been shown not to affect female remating, in opposition to findings in the green-veined white butterfly *Pieris napi* (Snook 1998; Cook and Wedell 1999). In combination, these results refute a single function for parasperm. It seems more likely that parasperm vary in function among taxa.

## Results II: Sperm Protein Homology

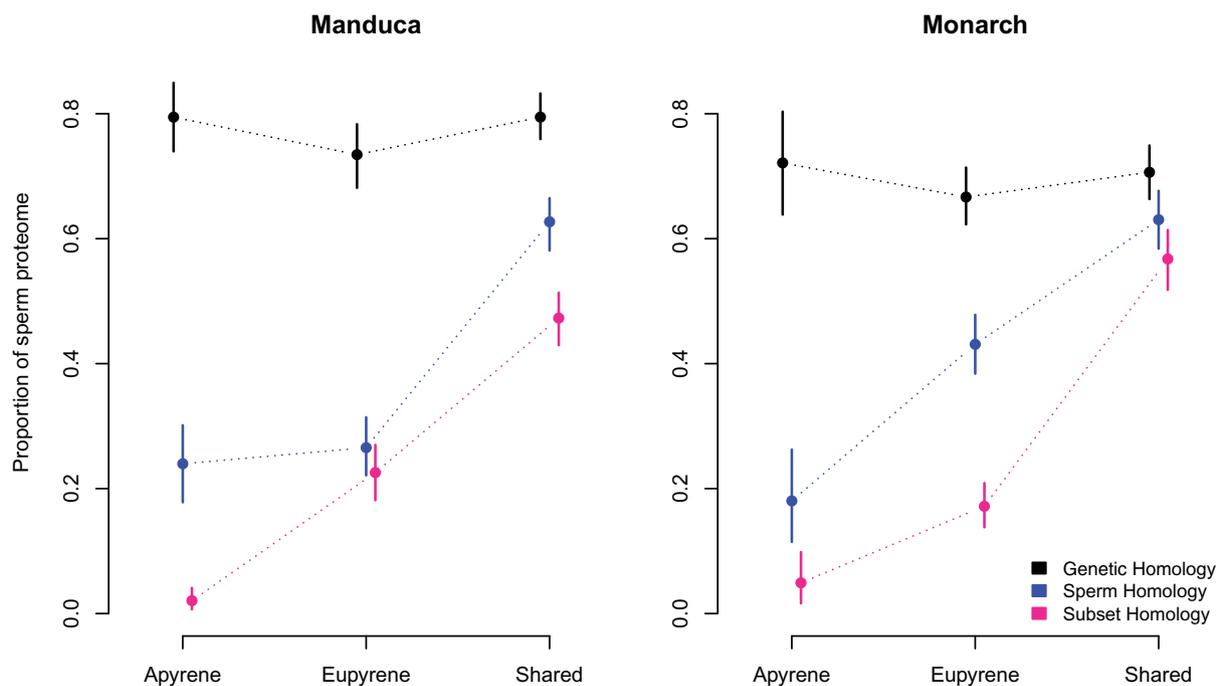
Spermatozoan morphology is strikingly diverse across animals, an observation seemingly at odds with their fundamental role in reproduction, which should arguably result in evolutionary constraint. This diversity is often explained as the outcome of sexual selection, particularly sperm competition, which can drive the rapid evolution of male reproductive characters (Pitnick et al. 2009). It has been suggested that nonfertilizing parasperm allow the resolution of these potentially conflicting selective pressures via division of labor. For instance, eusperm may shoulder the selective constraints of fertilization, whereas parasperm primarily function in sperm competition and experience reduced evolutionary constraint and increased adaptive selection (Kura and Nakashima 2000). Consistent with this, quantitative genetic analyses indicate greater evolvability in *D. pseudoobscura* parasperm (Moore et al. 2013). Together, these hypotheses offer a plausible explanation for the contrasting patterns of genetic homology versus sperm homology we observe among lepidopteran sperm morphs (fig. 3). We analyzed homology between monarch and *Manduca* at three hierarchical levels. Firstly, at the level of the whole genome, defined as *genetic homology*, requiring a sperm protein gene in one species to have a predicted ortholog in the genome of the other species. Secondly, at the level of the sperm proteome, termed *sperm homology*, requiring a sperm protein in one species have a predicted ortholog present in the sperm proteome of the other. Finally, sperm homologs can be further classified as occurring within the same subset (e.g., apyrene-specific) in both species, termed *subset homology*.

In the genome of each species, proportions of predicted orthologs (i.e., genetic homology) are indistinguishable across the three subsets of proteins (fig. 3, black). However, the pattern is strikingly different for rates of sperm homology (fig. 3, blue), which is greatly reduced for morph-specific proteins relative to genetic homology. For shared proteins, this effect is much less pronounced. Thus, proteins shared between sperm morphs are also relatively conserved as sperm proteins across species, likely reflecting morphological and physiological characteristics broadly common to sperm. In contrast, morph-specific protein content turns over more rapidly than gene gain/loss occurring between species at the whole genome level.

Comparing morph-specific subsets, our results suggest protein turn-over is faster in apyrene than eupyrene sperm. This is most prominent in monarch, where sperm homology is significantly lower among apyrene-specific than eupyrene-specific proteins (based on nonoverlapping 95% confidence intervals). This difference is not apparent in *Manduca* until examining subset homology (fig. 3, magenta). Subset homology shows similar patterns between species and strongly indicates that apyrene-specific proteins are the least conserved among the subsets examined. Only three proteins were found to be unique to apyrene sperm in both species.

## Discussion: Apyrene Sperm Function and Evolution

Apyrene sperm, along with other independently evolved instances of nonfertilizing parasperm, have long presented an evolutionary and functional enigma. Our comparison of protein content between apyrene and eupyrene sperm does not appear to immediately favor, nor exclude, any of the myriad hypothesized explanations for their existence (Swallow and Wilkinson 2002); indeed parasperm function likely varies among taxa. Nonetheless, it is an unexpected result to see in both species the apyrene-specific enrichment of GO-terms related to neuronal development, raising novel and intriguing possibilities for mechanisms by which apyrene sperm may mediate sperm competition, fertilization, or delayed female remating, among other hypothesized functions.



**Fig. 3.**—Proportions of proteins homologous between monarch and *Manduca* in each subset of the sperm proteome. The proportion of homologs are plotted for each subset of the sperm proteome: apyrene-specific, eupyrene-specific, and shared. Three different, increasingly stringent, criteria for homology are displayed. Genetic homology (black) indicates predicted orthologs found in the genomes of both species. Sperm homology (blue) indicates predicted orthologs found in the sperm of both species, regardless of sperm morph. Subset homology (magenta) indicates sperm homologs found in the same subset of the sperm proteome in both species.

The relatively rapid turnover of the apyrene-specific proteome is consistent with parasperm experiencing distinct selective pressures compared with eusperm. It is unclear whether this reflects relaxed constraint, greater adaptation, or some combination thereof. As the nonfertilizing sperm morph, apyrene sperm are free of constraints associated with egg interactions, karyogamy, and embryogenesis (Snook and Karr 1998). Consequently, proteins involved in these processes become superfluous to apyrene sperm function. The reduced complexity of the apyrene sperm proteome therefore likely reflects streamlining of a eupyrene “ancestor” present at the root of Lepidoptera. This process might be expected to be random, causing differential protein loss between lineages. Additionally, freed from selective constraints apyrene sperm may undergo lineage specific and adaptive functional specialization, thus compounding the pattern of increased divergence among apyrene-specific proteins. Mating system and the intensity of sperm competition are likely to affect the rate and direction of specialization. Notably, females are far more promiscuous in monarch than *Manduca* (Snow et al. 1974; Drummond 1984), which may explain some of the differences between species observed here. Discerning the relative contributions of relaxed constraint and adaptation to the rapid turnover of apyrene-specific proteins presented here is a novel goal for sperm heteromorphism research. In generating a substantial list of morph-specific proteins for

further research, our LC–MS/MS analysis of lepidopteran sperm represents an important step toward better understanding the still-enigmatic role of apyrene sperm, and parasperm more broadly.

### Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

### Acknowledgments

We thank Monarch Watch for support in rearing monarch butterflies, Sheri Skerget for expert technical assistance, and Desiree Forsythe for preliminary analyses. We thank the University of Cambridge Proteomics Facility, including Mike Deery, Renata Feret and Kathryn Lilley for excellent proteomic support, and Eric Sedore and Larne Pekowsky for computational support (reflecting National Science Foundation award OAC-1541396/ACI-1541396). Kirsten Jensen and Kaylee Herzog generously assisted with SEM imaging. Computing for this project was performed on the Syracuse University Crush Virtual Research Cloud. We thank the Cambridge Proteomics Facility, including Mike Deery, Renata Feret and Kathryn Lilley for excellent proteomic support, computational support from Eric Sedore and Larne Pekowsky (supported by

NSF award OAC-1541396/ACI-1541396). Funding for this research was provided by Syracuse University to S.D., by University of Kansas (KU) to J.R.W., by the KU Gould Fellowship and National Science Foundation award DEB-1701931 to A.J.M and J.R.W., and by Syracuse University and Marilyn Kerr Fellowships to E.W.

## Literature Cited

- Ashburner M, et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 25(1):25–29.
- Chow CY, Wolfner MF, Clark AG. 2013. Large neurological component to genetic differences underlying biased sperm use in *Drosophila*. *Genetics* 193(1):177–185.
- Cook PA, Wedell N. 1999. Non-fertile sperm delay female remating. *Nature* 397(6719):486–486.
- Dallai R. 2014. Overview on spermatogenesis and sperm structure of Hexapoda. *Arthropod Struct Dev.* 43(4):257–290.
- Deutsch EW, et al. 2015. Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Prot Clin Appl.* 9(7–8):745–754.
- Drummond BA. 1984. Multiple mating and sperm competition in the Lepidoptera. In: Smith RL, editor. *Sperm Competition and the Evolution of Animal Mating Systems*. New York: Academic Press. p. 291–370.
- Falcon S, Gentleman R. 2007. Using GOstats to test gene lists for GO term association. *Bioinformatics* 23(2):257–258.
- Fan Y, Rafaeli A, Gileadi C, Kubli E, Applebaum SW. 1999. *Drosophila melanogaster* sex peptide stimulates juvenile hormone synthesis and depresses sex pheromone production in *Helicoverpa armigera*. *J Insect Physiol.* 45(2):127–133.
- Friedlander M, Seth R, Reynolds S. 2005. Eupyrene and apyrene sperm: dichotomous spermatogenesis in Lepidoptera. *Adv Insect Physiol.* 32:206–308.
- Hanin O, Azrielli A, Applebaum SW, Rafaeli A. 2012. Functional impact of silencing the *Helicoverpa armigera* sex-peptide receptor on female reproductive behaviour. *Insect Mol Biol.* 21(2):161–167.
- Hanin O, Azrielli A, Zakin V, Applebaum S, Rafaeli A. 2011. Identification and differential expression of a sex-peptide receptor in *Helicoverpa armigera*. *Insect Biochem Mol Biol.* 41(8):537–544.
- Hayakawa Y. 2007. Parasperms: morphological and functional studies on nonfertile sperm. *Ichthyol Res.* 54(2):111–130.
- He Y, Miyata T. 1997. Variations in sperm number in relation to larval crowding and spermatophore size in the armyworm, *Pseudaletia separata*. *Ecol Entomol.* 22(1):41–46.
- Healy JM, Jamieson B. 1981. An ultrastructural examination of developing and mature paraspermatozoa in *Pyrazus ebeninus* (Mollusca, Gastropoda, Potamididae). *Zoomorphology.* 98(2):101–119.
- Heikkilä M, Kaila L, Mutanen M, Peña C, Wahlberg N. 2012. Cretaceous origin and repeated tertiary diversification of the redefined butterflies. *Proc Biol Sci.* 279:1093–1099.
- Holman L, Freckleton RP, Snook RR. 2008. What use is an infertile sperm? A comparative study of sperm-heteromorphic *Drosophila*. *Evolution* 62(2):374–385.
- Kanost MR, et al. 2016. Multifaceted biological insights from a draft genome sequence of the tobacco hornworm moth, *Manduca sexta*. *Insect Biochem Mol Biol.* 76:118–147.
- Karr TL, Walters JR. 2015. Panning for sperm gold: isolation and purification of apyrene and eupyrene sperm from lepidopterans. *Insect Biochem Mol Biol.* 63:152–158.
- Koskinen P, Törönen P, Nokso-Koivisto J, Holm L. 2015. PANNZER: high-throughput functional annotation of uncharacterized proteins in an error-prone environment. *Bioinformatics* 31:1544–1552.
- Kura T, Nakashima Y. 2000. Conditions for the evolution of soldier sperm classes. *Evolution* 54(1):72–80.
- Lechner M, et al. 2011. Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* 12:124.
- Lobov AA, et al. 2018. LOSP: a putative marker of parasperm lineage in male reproductive system of the prosobranch mollusk *Littorina obtusata*. *J Exp Zool. (Mol Dev Evol.)* 330(4):193–201.
- Moore AJ, Bacigalupe LD, Snook RR. 2013. Integrated and independent evolution of heteromorphic sperm types. *Proc Roy Soc B: Biol Sci.* 280(1769):20131647.
- Oppliger A, Hosken DJ, Ribi G. 1998. Snail sperm production characteristics vary with sperm competition risk. *Proc Biol Sci.* 265(1405):1527–1534.
- Pitnick SS, Hosken DJ, Birkhead TR. 2009. Sperm morphological diversity. In: Pitnick SS, Hosken DJ, Birkhead TR (eds). *Sperm Biology: An Evolutionary Perspective*. New York: Academic Press. p. 69–149.
- Sahara K, Kawamura N. 2002. Double copulation of a female with sterile diploid and polyploid males recovers fertility in *Bombyx mori*. *Zygote* 10(1):23–29.
- Sahara K, Takemura Y. 2003. Application of artificial insemination technique to eupyrene and/or apyrene sperm in *Bombyx mori*. *J Exp Zool.* 297A(2):196–200.
- Silberglied RE, Shepherd JG, Dickinson JL. 1984. Eunuchs: the role of apyrene sperm in Lepidoptera? *Am Nat.* 123(2):255–265.
- Snook RR. 1997. Is the production of multiple sperm types adaptive? *Evolution* 51(3):797–808.
- Snook RR. 1998. The risk of sperm competition and the evolution of sperm heteromorphism. *Anim Behav.* 56(6):1497–1507.
- Snook RR, Karr TL. 1998. Only long sperm are fertilization-competent in six sperm-heteromorphic *Drosophila* species. *Curr Biol.* 8(5):291–294.
- Snow JW, Copeland WWW, Goodenough JL. 1974. The tobacco hornworm: notes on morphology and mating habits. *J Georgia Entomol Soc.* 9:36–41.
- Swallow JG, Wilkinson GS. 2002. The long and short of sperm polymorphisms in insects. *Biol Rev Camb Philos Soc.* 77(2):153–182.
- Till-Bottraud I, Joly D, Lachaise D, Snook RR. 2005. Pollen and sperm heteromorphism: convergence across kingdoms? *J Evolution Biol.* 18(1):1–18.
- Vizcaíno JA, et al. 2016. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 44(D1):D447–D456.
- Watanabe M, Bon'no M, Hachisuka A. 2000. Eupyrene sperm migrates to spermatheca after apyrene sperm in the swallowtail butterfly, *Papilio xuthus* L. (Lepidoptera: Papilionidae). *J Ethol.* 18(2):91–99.
- Wedell N, Cook PA. 1999. Butterflies tailor their ejaculate in response to sperm competition risk and intensity. *Proc Biol Sci.* 266(1423):1033–1039.
- Whittington E, et al. 2017. Contrasting patterns of evolutionary constraint and novelty revealed by comparative sperm proteomic analysis in Lepidoptera. *BMC Genomics* 18(1):931.
- Whittington E, Zhao Q, Borziak K, Walters JR, Dorus S. 2015. Characterisation of the *Manduca sexta* sperm proteome: genetic novelty underlying sperm composition in Lepidoptera. *Insect Biochem Mol Biol.* 62:183–193.
- Wolf KW. 1992. Spindle membranes and microtubules are coordinately reduced in apyrene relative to eupyrene spermatocytes of *Inachis io* (Lepidoptera, Nymphalidae). *J Submicrosc Cytol Pathol.* 24:381–394.
- Zhan S, Reppert SM. 2012. MonarchBase: the monarch butterfly genome database. *Nucleic Acids Res.* 41(D1):D758–D763.

Associate editor: Adam Eyre-Walker