Collective sperm movements are shaped by post-copulatory sexual selection and phylogenetic history in *Peromyscus* mice

Kristin A. Hook, W. David Weber, Heidi S. Fisher*

Department of Biology, University of Maryland, 1200 Biology-Psychology Building, 4094 Campus Drive, College Park, MD 20742, U.S.A.

* Heidi S. Fisher
Email: hsfisher@umd.edu

Keywords:

mating systems; sexual selection; sperm competition; sperm conjugation; sperm motility

Author Contributions:

KAH and HSF conceived of the study, designed experiments, and interpreted results; KAH and WWD collected the data, KAH carried out the statistical analyses; all authors wrote the manuscript, gave final approval for publication, and agree to be held accountable for the work presented.

Funding:

This work was supported by a Eunice Kennedy Shriver National Institute of Child Health and Human Development K99/R00 Pathway to Independence Award to HSF [R00HD071972] and a National Science Foundation Postdoctoral Research Fellowship to KAH [1711817].

Abstract

Sperm of some species form motile, coordinated groups as they migrate through the female reproductive tract to the site of fertilization. This collective motion is predicted to improve sperm swimming performance and therefore may be beneficial in a competitive context, but limited evidence supports this theory. Here we examine sperm aggregates across closely-related species of *Peromyscus* mice that naturally vary by mating system, and thus sperm competition intensity. We find that phylogenetic history predicts the likelihood that sperm will aggregate, and that relative testis size is negatively associated with variation in number of aggregated cells, suggesting that sperm competition has a stabilizing effect on sperm group size. Moreover, we show that sperm aggregates are kinematically beneficial for some species but costly for others, and these differences are largely dependent on the orientation and composition of sperm within the groups. In addition, when we compared sperm of the two sister-species that aggregate most frequently, we find that sperm from the species that evolved under intense sperm competition forms aggregates with more efficient geometry more frequently than sperm from its monogamous congener. These results are consistent with the prediction that sperm aggregation evolved to improve motility in a competitive context; however, when monogamy evolved secondarily, relaxed sexual selection allowed for less efficient strategies to persist. Together, our findings in *Peromyscus* reveal that collective sperm behavior is likely to evolve rapidly and is shaped by changes in the selective regime.

Introduction

Sperm cells are one of the most diverse cell types in nature and exhibit striking variation both within and across taxa (Pitnick et al. 2009). In addition to being morphologically diverse, sperm may exhibit complex, emergent behaviors, including sperm conjugation, which occurs when two or more cells join together for motility or transport through the female reproductive tract before dissociating prior to fertilization (Pitnick et al. 2009; Higginson and Pitnick 2011). Although relatively rare, these sperm-sperm interactions have evolved multiple times across independent lineages of internally fertilizing species (Immler 2008; Pitnick et al. 2009; Higginson and Pitnick 2011), yet the adaptive significance of these gametic interactions remains unclear for many taxa.

One functional hypothesis posits that sperm aggregates may be advantageous if the combined force generated by multiple flagella enable sperm conjugates to swim faster than single cells (Moore et al. 2002; Immler et al. 2007). Improved sperm motility can be beneficial if it enables cells to move quickly through hostile environments within the female tract (Birkhead et al. 1993; Pizzari and Foster 2008), and in competitive environments in which females mate with multiple males, as sperm velocity is often correlated with competitive fertilization success (e.g., Birkhead et al. 1999; Gage et al. 2004; Pizzari and Foster 2008). While some studies that have quantified the motility of sperm aggregates have found empirical support for this hypothesis (e.g., Hayashi 1998; Moore et al. 2002; Woolley et al. 2009; Fisher and Hoekstra 2010; Pearcy et al. 2014), others have not (e.g., Ishijima et al. 1999; Tung et al. 2017). For example, sperm groups swim faster than solitary cells in the Norway rat (*Rattus norvegicus*), yet house mouse (*Mus musculus*) sperm swim slower as groups under identical experimental conditions (Immler et al. 2007). Another non-mutually exclusive hypothesis is that sperm aggregation facilitates migration through viscous or viscoelastic secretions of the female reproductive tract (Moore and Taggart 1995; Suarez 2016), including cervical mucus (Hanson and Overstreet 1981), which has received some empirical support (e.g., Moore and Taggart 1995; Hayashi 1998; Moore et al. 2002; but see Tung et al. 2017). In the grey short-tailed opossum (*Monodelphis domestica*), for example, sperm pairs swim

with greater motility than single sperm in viscous fluids; however, these sperm pairs were artificially induced to uncouple (Moore and Taggart 1995) and therefore the comparison may not be biologically informative (Pitnick et al. 2009). Additionally, sperm aggregation may protect sensitive regions of the sperm, such as the acrosome, from damage and preserve sperm functionality during passage through the male or female reproductive tracts (Phillips 1970; Tung et al. 1980; Bedford et al. 1984; Higginson and Henn 2012), or enhance egg penetration during fertilization (Mackie and Walker 1974; but see McGlinn et al. 1979). While it is often assumed that coordinated sperm movements are adaptive, inconsistent findings and a wide diversity of naturally variable sperm behavior have limited our understanding of the functional advantages of sperm aggregation.

Multiple independent origins of sperm conjugation suggests that the functional consequences, as well as mechanisms that regulate these cellular interactions, are likely to vary throughout nature (reviewed in Higginson and Pitnick 2011). Indeed, the formation of sperm aggregates and the number of grouped cells varies widely across taxa (Pizzari and Foster 2008). In mammals alone, sperm of some species assemble during epididymal transport and are molecularly "glued" to one another as bi-flagellate pairs in grey short-tailed opossums (Taggart et al. 1993), as bundles of roughly 100 cells in monotremes (Nixon et al. 2016), or as organized rouleaux of five or more cells in guinea pigs (*Cavia porcellus*; McGlinn et al. 1979). Conversely, mammalian sperm may also assemble after ejaculation to form variably-sized groups. For instance, sperm may form temporary clusters of up to sixteen cells in bulls (*Bos taurus*; Tung et al. 2017) or more fixed groups of up to 30 cells in house mice (Immler et al. 2007), 50 cells in the Norway rat (Immler et al. 2007), or thousands of cells in the wood mouse (*Apodemus sylvaticus*; Moore et al. 2002) whereby the hook-shaped heads interlock or attach to the flagella of other sperm. For these latter species in which sperm groups form after ejaculation, both single and aggregated sperm typically co-occur, thus allowing for direct comparisons between collective and solitary sperm movements within an ejaculate while controlling for within-male variability.

126 Closely-related species in the rodent genus, *Peromyscus,* produce sperm that naturally vary in
127 their collective behavior. Sperm of some species assemble temporary groups after ejaculation by
128 adhering to one another at their head region (Pearce et al. 2018) and disassemble prior to
129 fertilization (Fisher and Hoekstra 2010). In *P. maniculatus*, sperm selectively group with the most
130 closely-related cells to form motile groups of up to 30 cells (Fisher and Hoekstra 2010), but there
131 is a non-monotonic association between group size and swimming velocity, indicating that some
132 groups are faster than single cells but others are not (Fisher et al. 2014). Conversely, sperm
133 produced by their sister-species, *P. polionotus*, also form aggregates but do so indiscriminately
134 with related and unrelated cells (Fisher and Hoekstra 2010), are less likely to be optimally-sized,
135 and overall move in a less linear trajectory (Fisher et al. 2014). Intriguingly, sperm competition is
136 predicted to be more intense in *P. maniculatus* than in *P. polionotus* due to their different mating
137 systems. In *P. maniculatus*, both sexes mate with multiple partners, often in overlapping series just
138 minutes apart (Dewsbury 1985), and females frequently carry multiple-paternity litters in the wild
139 (Birdsall and Nash 1973), whereas both behavioral (Dewsbury 1981) and genetic data (Foltz 1981)
140 indicate that *P. polionotus* is strictly monogamous. Evidence suggests that monogamy has evolved
141 at least twice within the *Peromyscus* lineage (reviewed in Bedford and Hoekstra 2015; Turner et
142 al. 2010), thus enabling us to investigate if post-copulatory sexual selection, driven by female
143 mating behavior, has shaped the evolution of sperm aggregation (Moore et al. 2002) more broadly
144 across the *Peromyscus* lineage. In this study, we quantify intra- and inter-specific differences in
145 the size and performance of sperm aggregates under consistent, controlled conditions to examine
146 the evolution of collective sperm behavior and empirically test whether sperm aggregation
147 improves swimming performance.
148
149 **Materials and Methods**
150 (a) Sperm collection

We obtained captive *Peromyscus maniculatus bairdii, P. polionotus subgriseus, P. leucopus*, *P. eremicus*, and *P. californicus insignis* from the Peromyscus Genetic Stock Center at the University of South Carolina, and *P. gossypinus gossypinus* from Dr. Hopi Hoekstra at Harvard University and housed them in same-sex cages at 22ºC on a 16L:8D cycle in accordance with guidelines established by the Institutional Animal Care and Use Committee at the University of Maryland in College Park (protocol R-Jul-18-38). We sought samples from all available captive *Peromyscus* species and avoided wild-caught specimens to control for variation due to life experience. We obtained sperm samples from sexually mature males and accounted for relatedness among the focal males by assigning siblings a unique 'Family' ID. We euthanized males via isoflurane overdose and cervical dislocation, then weighed each male and both testes (Mettler Toledo, Switzerland). Next, we removed a single caudal epididymis, made several small incisions in the tissue, and submersed it in sperm medium (Modified Sperm Washing Medium, Irvine Scientific, USA) that was pre-warmed at 37ºC; to reduce differences in sperm density despite natural variation in epididymal sizes, we varied the volume (50µl - 1000µl) based on tissue size and accounted for these differences when estimating final sperm counts for each male. To collect sperm, we agitated the tissue at 300rpm (ThermoMixer F1.5, Eppendorf, Germany) at 37ºC for ten minutes, inverting the tube at the five- and ten-minute mark, then incubated the tissue undisturbed for two minutes. Using pipette tips cut to create a wider opening, we collected live sperm cells for analysis from just below the meniscus of the solution to enrich for the most motile sperm (Magdanz et al. 2019). Next we estimated sperm density using a computer-assisted sperm analysis (CASA) system (Ceros II Animal, Hamilton Thorne, USA) and verified with a Neubauer-improved hemocytometer (Marienfeld, Germany) then diluted samples with pre-warmed medium to reach a standard concentration of 300-400 cells summed across the five 5-second videos at 100X magnification for cell tracking optimality and efficiency.

(b) Live sperm analysis

177 To conduct live sperm observations, we gently reverse pipetted 4µl of the sperm solution into
178 12µl of pre-warmed medium on a plastic slide within a 9mm x 0.12mm imaging spacer (Grace
179 Bio-Labs, USA) and covered by a plastic cover slip. This set-up served as a control and represents
180 a 'low-viscosity' environment. To test sperm motility in a 'high-viscosity' environment, we
181 followed the same procedures except that we mixed 4µl of sperm solution with 12µl of pre-
182 warmed medium enriched with methylcellulose (Sigma Aldrich M 7140; 15cP, 2% in water;
183 Suarez and Dai 1992). We then recorded 5-second videos at 60 frames/sec on the CASA system,
184 capturing at least five videos per male but recorded additional videos for samples with lower
185 sperm density (n = 57) to ensure an adequate number of observed cells per male. Videos were
186 recorded at 59±16 minutes post-harvest from the epididymal tissue, dependent on dilutions. The
187 number of cells analyzed are reported in Table 1.
188 We characterized sperm aggregation by scoring CASA videos using direct observations
189 because the system tracks particles, and thus each track may represent a single cell or an
190 aggregate. We counted the number of cells represented by each track on at least three different
191 frames/track. From these data, we calculated the proportion of cells that aggregated for each male
192 by dividing the total number of aggregates by the total number of motile cells across all tracks
193 (Crawley 2013). Then we calculated the mean number of cells in aggregate (i.e., aggregate size)
194 by dividing the sum of cells in aggregate by the sum of aggregates, both across all tracks for each
195 male and across all males for each species. Finally, we calculated the coefficient of variation (CV)
196 for both the proportion and number of cells that aggregated within each species using the
197 following formula: (standard deviation/mean) x 100. In addition, to further characterize
198 differences in sperm aggregation for the two species whose sperm were observed to aggregate the
199 most extensively, *P. maniculatus* and *P. polionotus*, we qualitatively scored the composition and
200 orientation of cells within the sperm aggregates. For males within these species, we calculated the
201 proportion of aggregates that were: 'aligned' (all sperm adhered to one another in a head-to-
202 flagella orientation and included no immotile, morphologically abnormal or damaged cells),

203 'defective' (one or more sperm were abnormal, immotile, or stuck to the slide), or 'opposed' (all
204 sperm are normal and motile but were not oriented in the same direction), the latter of which
205 included star-shaped aggregates (Fisher et al. 2014).
206 We recorded the following metrics for each recorded track (i.e., single sperm and aggregates):
207 *straight-line velocity* (VSL; calculated using the straight-line distance between the first and last
208 detected positions of the sperm head, divided by the time taken to swim the track; also known as
209 average velocity), *curvilinear velocity* (VCL; calculated using the summed distance between the
210 sperm head positions in each frame divided by the time taken to travel the track; also known as
211 speed), *average path velocity* (VAP; the time-averaged velocity of the sperm head along its
212 average path), and *linearity* (LIN; the ratio of VSL to VCL to measure the straightness of the
213 trajectory; World Health Organization 2010). We calculated the mean of each kinematic parameter
214 for both single cells and sperm aggregates separately for each of three populations of sperm cells:
215 all cells, motile cells (i.e., devoid of visually inspected tracks in which cells were unmoving, stuck,
216 or featured an obvious morphological abnormality such as a kinked midpiece), and progressively
217 motile cells (i.e., motile cells with a VSL $\geq$ 25μm/sec). We used these data to calculate the
218 proportion of motile aggregates by dividing the sum of motile aggregates by the total number of
219 sperm aggregates, and the proportion of progressively motile aggregates by dividing the sum of
220 progressive aggregates by the total number of sperm aggregates for each male (Crawley 2013). For
221 our kinematic analyses, we focused on the motile sperm dataset to eliminate artifacts from
222 damaged or dead cells, and the total sperm population to quantify frequency of aggregation and
223 aggregate size in an effort to capture natural aggregation rates (results from other cell populations
224 are reported in Table S1, Figure S1).
225
226 (c) Statistical Analyses
227 We performed all statistical analyses using R version 3.4.2 (R Core Team 2016) and visually
228 inspected diagnostic plots (qqplots and plots of the distribution of the residuals against fitted

values) to validate model normality. Only the best fitting models are reported here. We created all figures using the 'ggplot2' package with R (Wickham 2016). One *P. californicus* male was excluded from the aggregate analysis dataset because their measurements represented clear outliers. All means are presented ± 1 standard error.

To compare species differences in the proportion of aggregated cells, we used the mean values for each male and a generalized linear mixed model (GLMM) using the glmer function from the "lme4" R package and a logit link function (Bates et al. 2015). The binomial response was the number of sperm cells in aggregate, and the total number of sperm cells was the binomial denominator. In the initial statistical model, we observed the residual deviance to be larger than the residual degrees of freedom, which is an indication of overdispersion (Crawley 2013). We thus used an observation-level random effect (OLRE) as a random factor in all subsequent analyses to control for overdispersion (Harrison 2014). We considered family ID as a random factor in the initial model and both random factors were then used in bivariate analyses for predictors of interest that could potentially explain differences in the proportion of aggregated cells. These predictors included male age, pairing status, the timing of video recordings relative to harvest of the epididymal tissue, and the number of videos recorded. Only predictors that had a *p*-value at or below 0.20 were considered for the final model. We further screened these predictors for collinearity with other significant predictors using linear models and removed collinear predictors, so that only the one with greater relative significance was included in the final GLMM. The remaining model included pairing status and species as fixed factors. We calculated the variance inflation factor (VIF) values and found evidence of collinearity due to two VIF values above the recommended threshold of three (Zuur et al. 2010). Thus, we removed pairing status as a fixed factor and family ID as a random factor, leaving only species as a fixed factor within our final model. Post-hoc pairwise comparisons were performed using Tukey HSD adjustments for multiple comparisons from the "LSmeans" R package (Lenth 2016). We additionally analyzed whether the proportion of aggregated cells correlates with sperm swimming speed using the same methods

outlined above with VCL as the main explanatory variable within both inter- and intra-species GLMMs. Our inter-species model included pairing status as a fixed factor and both family ID and an OLRE as random factors, whereas the intra-species models reported here included only an OLRE as a random factor.

To compare species differences in the number of aggregated cells, we used the mean values for each male and initially used a linear mixed model (LMM) using the lmer function from the "lme4" R package, but eventually reverted to using a linear model (LM) because the family ID random factor did not significantly contribute to the residual variability in the response variable. Predictors that were considered for the initial LMM included male age, pairing status, the timing of video recordings relative to harvest of the epididymal tissue, total sperm cells, number of recorded videos, an interaction between the latter two variables, and the ratio of total sperm cells to the number of videos recorded. We considered predictors with p-values < 0.20 for the final model, but first screened each for collinearity with other significant predictors using a linear model. Whenever collinearity was present, only the predictor with the greater relative significance was included in the LM. We dropped non-significant explanatory variables one at a time based on model comparisons using an analyses of variance test to determine the minimal adequate model, but were unable to meet the normality assumptions for this model. We also assessed species differences in the proportion of aggregates that were motile or progressively motile from the total population of aggregates using binomial GLMM. The binomial response was the number of motile sperm aggregates, and the total number of sperm aggregates was the binomial denominator. Our final model contained both an OLRE due to detected overdispersion and family ID as random factors.

To determine if sperm aggregates have motility or force benefits over single cells, we initially performed a principal component analysis (PCA) using three related swimming performance measures (VSL, VCL, and VAP) to reduce dimensionality and obtain a composite measure for motile solitary and aggregated sperm cells in both low- and high-viscosity media. Because we

281 found that species distributions overlapped for this composite measure for both single sperm and
282 aggregates in low- and high-viscosity media (Figures S2 and S3), we focused on individual
283 kinematic parameters (VSL, VCL, VAP, LIN) within each species separately. Using our dataset of
284 mean values per male, we used a paired student's t-test to compare each kinematic parameter
285 between solitary sperm cells and sperm aggregates within males for each separate species to
286 determine if sperm aggregates have motility benefit over single cells. To determine if aggregates
287 have a force benefit over single sperm, we conducted these same analyses in high-viscosity media.
288 We then combined these two datasets in low- and high- viscosity media and compared sperm
289 aggregate kinematics in both low- and high- viscosity media at the intra-male level using a paired
290 student's t-test within each species.

291 To assess the structure of *P. maniculatus* and *P. polionotus* aggregates, we used generalized
292 linear models (GLM) to compare the proportions of 'defective' and 'opposed' aggregates to
293 'aligned' aggregates. For the composition GLM, the binomial response was the number of
294 'defective' sperm aggregates, and the total number of sperm aggregates that were either 'defective'
295 or 'aligned' was the binomial denominator. We used a paired student's t-test to compare the VCL
296 of these aggregates within males. For the orientation GLM, the binomial response was the number
297 of 'opposed' sperm aggregates, and the total number of sperm aggregates that were either
298 'opposed' or 'aligned' was the binomial denominator. We used a paired student's t-test to compare
299 the VCL of these aggregates within males.

300 Finally, to account for variation in phylogenetic relationships among of the species used in this
301 study, we adopted a phylogenetic generalized least squares approach (Pagel 1999; Freckleton et al.
302 2002) using the "caper" (Orme et al. 2013) and "APE" (Paradis et al. 2004) packages in R and
303 using an ultra-metric phylogenetic tree of *Peromyscus* (provided by Dr. Roy Neal Platt II, Texas
304 Biomedical Research Institute), based on sequence variation in the mitochondrial gene,
305 cytochrome B. The species' relationships within this tree matched those from other previously
306 established phylogenies of *Peromyscus* (Bradley et al. 2007; Turner et al. 2010). We used this

307 phylogeny as a covariate in regression analyses to investigate the effect of relative testis weight on
308 sperm aggregation, including the proportion of aggregated cells and aggregate size, and the within-
309 species CV for each of these parameters. To control for differences in male body when examining
310 testis mass, we included body mass as a separate fixed factor within our analyses, a method better
311 suited to estimating relative testis weight is size rather than using the ratio of testis to body mass or
312 residuals (García-Berthou 2001; Lüpold et al. 2009).
313
314 **Results**
315 We investigated the frequency of sperm aggregation in each species and found significant
316 differences in the proportion of aggregated cells among species (binomial GLMM: $n = 134$, $P <$
317 0.001; Table 1, 2; Figure 1), with more variance across species than within species (variance
318 across species = 0.10; variance within each species < 0.10). Specifically, sperm from *P.*
319 *maniculatus* and *P. polionotus* similarly aggregate the most, whereas *P. leucopus* and *P.*
320 *gossypinus* sperm similarly aggregate the least among all the species. Pairwise comparisons
321 adjusted for multiple comparisons using LSmeans revealed that species within these pairs do not
322 significantly differ from one another, whereas all other pairwise species comparisons do (Table 1,
323 2). The coefficient of variation (CV) for the proportion of cells aggregated within each species are
324 as follows: 68.8% for *P. californicus*, 50.6% for *P. eremicus*, 14.4% for *P. polionotus*, 10.8% for
325 *P. maniculatus*, 125.7% for *P. leucopus*, and 60.5% for *P. gossypinus*. We found the sperm of
326 males with greater VCL were less likely to form aggregates (binomial GLMM: $n = 130$, $p =$
327 0.1435), specifically within *P. gossypinus* (binomial GLMM: $n = 21$, $p = 0.0292$), *P. leucopus*
328 (binomial GLMM: $n = 22$, $p = 0.000193$), and *P. polionotus* (binomial GLMM: $n = 24$, $p =$ 1.46E-
329 05). Controlling for phylogenic relationships and body mass, we found no effect of testis size ($F_{2,3}$
330 $= 2.606$, $P = 0.15816$) or the within-species CV ($F_{2,3} = 0.3603$, $P = 0.5604$) on the proportion of
331 aggregated cells.

Moreover, we found significant differences in the mean number of cells aggregated among species (LM: $F_{6,126} = 56.37$, $P < 0.001$; Table 1, 2; Figure 2A), with more variance observed across species ($s^2 = 1.96$) than within species ($s^2 < 1.00$, except for *P. polionotus* [$s^2 = 2.15$]). Post-hoc pairwise comparisons revealed that both *P. maniculatus* and *P. polionotus* produce the largest sperm aggregates ($P < 0.05$ for all pairwise comparisons), whereas *P. gossypinus*, *P. leucopus*, and *P. californicus* produce the smallest aggregates ($P < 0.05$ for pairwise comparisons), the latter of which produces sperm cell aggregates that are statistically similar in size to those produced by *P. eremicus* ($P = 0.9203$; Figure 2A). Controlling for phylogenic relationships and body mass, we found a significant effect of testis weight on the within-species CV of aggregate size ($F_{2,3} = 8.398$, $P = 0.02655$, Figure 2B), but found less of an effect on the direct measure of aggregate size ($F_{2,3} = 3.058$, $P = 0.1522$). The CV for the number of cells aggregated within each species were as follows: 26.5% for *P. californicus*, 19.2% for *P. eremicus*, 27.5% for *P. polionotus*, 18.4% for *P. maniculatus*, 13.3% for *P. leucopus*, and 7.2% for *P. gossypinus*.

By comparing sperm aggregate composition and orientation within the species that produce the largest and most frequent sperm aggregates, we found that there are significantly more sperm aggregates in *P. maniculatus* in which all sperm are aligned in a head-to-flagella orientation (99.1%, 731/738) than in *P. polionotus* (87.5%, 720/826; GLM: $n = 42$, $P = 3.92e\text{-}12$) and that these aggregates have significantly faster VCL compared to aggregates with unaligned cell orientations in both species (Figure 3; paired t-tests: *P. maniculatus* $t = 5.9627$, $df = 4$, $P = 0.003972$; *P. polionotus* $t = 11.247$, $df = 11$, $P = 2.257e\text{-}07$). We also found that there are significantly fewer sperm aggregates in *P. maniculatus* with immotile, stuck, or morphologically abnormal cells (8.1%, 64/795) than in *P. polionotus* (11.5%, 94/814; binomial GLM: $n = 40$, $P = 0.00503$). Importantly, aggregates with these defects had significantly lower VCL compared to aggregates without in both species (*P. maniculatus* $VCL_{aligned} = 179.60 \pm 6.88$ μm/sec, $VCL_{defective} = 132.47 \pm 6.18$ μm/sec, paired t-test: $t = 20.627$, $df = 14$, $P = 7.075e\text{-}12$; *P. polionotus* $VCL_{aligned} =$

357 135.48 ± 3.51 μm/sec, $VCL_{defective}$ = 115.62 ± 5.22 μm/sec, paired t-test: t =16.312, df = 20, P =
358 5.079e-13).

359 When comparing the proportion of motile and progressively motile aggregates across species,
360 our pairwise comparisons revealed that *P. eremicus* produced a significantly smaller proportion of
361 motile aggregates than all other species ($P < 0.05$ for all *P. eremicus* pairwise comparisons; $P >$
362 0.05 for all other pairwise comparisons). Fitted values of the proportion of motile aggregates using
363 LSmeans were 0.77 ± 0.04 for *P. eremicus*, 0.92 ± 0.02 for *P. polionotus*, 0.91 ± 0.02 for *P.*
364 *gossypinus*, 0.95 ± 0.01 for *P. maniculatus*, 0.93 ± 0.02 for *P. leucopus*, and 0.91 ± 0.02 for *P.*
365 *californicus*. Moreover, post-hoc comparisons revealed that *P. eremicus, P. polionotus,* and *P.*
366 *gossypinus* all had the smallest proportions of progressively motile aggregates ($P < 0.05$ for all
367 pairwise comparisons; fitted values using LSmeans were 0.64 ± 0.06, 0.65 ± 0.05, and 0.75 ± 0.05,
368 respectively), the latter species of which did not significantly differ from *P. californicus* ($P =$
369 0.1981; 0.87 ± 0.02 for LSmeans fitted values). Conversely, *P. maniculatus* and *P. leucopus* had
370 the largest proportions of progressively motile aggregates ($P = 0.3417$; fitted values using
371 LSmeans were 0.94 ± 0.02 and 0.91 ± 0.02, respectively), the latter of which did not differ from *P.*
372 *californicus* ($P = 0.9109$).

373 Overall, we found species-specific differences in the effect of sperm aggregation on motility,
374 regardless of environmental complexity (Table 3, Figure 4). In low-viscosity medium, we found
375 that sperm aggregates have a significantly greater VCL in *P. maniculatus*, VSL in *P. maniculatus*,
376 *P. leucopus*, and *P. californicus*, LIN in *P. leucopus*, and VAP in *P. californicus* compared to
377 single cells. Conversely, sperm aggregates had a significantly lower VCL, VSL, and VAP velocity
378 in *P. polionotus* and *P. gossypinus* than single sperm in the low-viscosity medium (Figure 4). In
379 the high-viscosity medium, we found that sperm aggregates have a significantly greater VCL in *P.*
380 *maniculatus* and *P. californicus* and a higher VSL and VAP in *P. californicus* (Figure 4) compared
381 to single sperm. Conversely, sperm aggregates in the high viscosity medium had a significantly

lower LIN in *P. californicus*, *P. eremicus*, *P. polionotus*, and *P. maniculatus* as well as a reduced VSL and VAP in *P. polionotus* than single cells (Figure 4).

Discussion

While it is known that collective sperm behaviors have evolved independently in a number of taxa (Higginson and Pitnick 2011), it remains unclear how sperm aggregation evolves among closely related species. Our comparative study reveals that sperm aggregating behaviors vary across mice in the genus *Peromyscus*. We observed an effect of phylogenetic history on the frequency of sperm aggregation, indicating that collective sperm behavior likely evolved prior to the divergence of present-day species. Additionally, we find a negative association between relative testis weight, a robust proxy for intensity of sperm competition in rodents (Ramm et al. 2005), and the coefficient of variation for the number of cells in aggregate, suggesting that sexual selection has a stabilizing effect on sperm aggregate size. We then compared the motility of single sperm and sperm aggregates across all species under low- and high-viscosity conditions, and show that aggregation is kinematically beneficial for some species yet costly for others, regardless of environmental complexity. When examining sperm from the species that aggregate the most profoundly, we find that the formation of the aggregates and the orientation of the cells within the group are critical to kinematics. Moreover, we observe more aggregates with efficient geometry in the species that has evolved under strong post-copulatory sexual selection compared to its monogamous sister-species. These findings support the prediction that sperm aggregation evolved in *Peromyscus* to improve motility in a competitive context but reveal that relaxed selection may have enabled less efficient strategies to persist, thereby generating diversity in collective sperm behaviors within these closely-related species.

Our results reveal distinct species-specific differences across *Peromyscus* mice in the frequency of sperm aggregation and the average size of these cellular groups. Multiple *Peromyscus* species produce sperm that aggregate more extensively than other studied muroid

408 rodents (Tourmente et al. 2016), with the notable exception of the wood mouse (*Apodemus*
409 *sylvaticus*; Moore et al. 2002). In general, the proportion of sperm cells that aggregate is most
410 similar within each sister-species pair. Specifically, *P. maniculatus* and *P. polionotus* produce
411 sperm that aggregate the most (>80% of sperm), but the pair they are most closely-related to, *P.*
412 *gossypinus* and *P. leucopus*, are the species least likely to produce sperm that aggregate (<10% of
413 sperm); the most distantly-related species pair we assessed, *P. californicus* and *P. eremicus*, both
414 produce sperm with a moderate propensity to aggregate (~25% of sperm). These findings support
415 one of two possibilities for the evolution of sperm aggregates within these species: (a) a genus-
416 wide ancestral trait of moderate sperm aggregation with subsequent diversification leading to an
417 increase in *P. maniculatus* and *P. polionotus* and a decrease in *P. gossypinus* and *P. leucopus*, or
418 (b) the independent evolution of aggregation in the ancestors of the *P. californicus*-species pair
419 and the *P. maniculatus*-species pair. Such complex evolutionary histories with losses and
420 recurrences of sperm conjugation, and subsequent species divergence, have also been
421 demonstrated in diving beetles (Dytiscidae; Higginson et al. 2012a), consistent with the
422 evolutionary lability that we observe. Our experimental results may explain selection against
423 sperm aggregation if forming groups reduces sperm swimming performance, which we find in at
424 least one species, *P. gossypinus*. However, we found that sperm produced by their sister-species,
425 *P. leucopus*, also rarely aggregate, despite our observation that these rare collective groups have a
426 greater average velocity and are more linear than single cells. Interestingly, we observed that these
427 two promiscuous species that rarely aggregate, *P. gossypinus* and *P. leucopus*, have the largest
428 relative testes of the species studied, suggesting that they may have evolved increased sperm
429 production to improve competitive fertilization success (sensu Parker 1982), rather than
430 adaptations that influence motility (Snook 2005). Moreover, within these species (and *P.*
431 *polionotus*) we found a negative association between sperm speed and aggregation, even after
432 controlling for sperm density, indicating that differences in aggregate formation are not simply a

433 by-product of increased encounter rates. Together these results suggest different strategies
434 employed by divergent species across the *Peromyscus* genus in response to sperm competition.
435 In contrast to our results on the frequency of sperm aggregation within species, we found that
436 the average size of sperm groups does not align as closely with phylogenetic relationships. In
437 addition, we found that species with relatively larger testes, which is positively associated with
438 increased sperm competition (Ramm et al. 2005), exhibit less variation (CV) in aggregate size.
439 This result supports the prediction that relaxed sperm competition allows for greater intermale
440 variation to persist in a population (Calhim et al. 2007) and suggests that this post-copulatory
441 sexual selection may be stabilizing sperm aggregate size for a species-specific 'optimum' (Fisher
442 et al. 2014). Similarly, other studies have shown that the strength of sexual selection regulates
443 variance in sperm morphology across taxa and at multiple levels of organization, including within-
444 and between-males (Immler et al. 2008; Fitzpatrick and Baer 2011; Carballo et al. 2019) and
445 within- and between-species (Calhim et al. 2007; Rowley et al. 2019). A study on sperm bundles
446 across ten *Carabus* ground beetles also found intense selection on bundle size, which are
447 dimorphic and either small or large; the large, but not small, sperm bundles are positively
448 correlated with measures of sperm competition risk, including copulatory piece length and mate
449 guarding, suggesting that diversity of large sperm bundles is associated with sexual selection
450 (Takami and Sota 2007). In contrast to these findings that competition drives sperm-sperm
451 interactions, a study on the evolution of such sperm traits in diving beetles found that variation in
452 sperm conjugation is more associated with female reproductive tract architecture (Higginson et al.
453 2012b,a). Therefore, while our results suggest that stabilizing selection on sperm aggregate size is
454 associated with an increase in sperm competition given the correlation with relative testis weight,
455 mechanisms of female control (Eberhard 1996) may also play an important evolutionary role.
456 We compared the motility of single and aggregated sperm sampled from the same male to test
457 whether sperm aggregates swim faster or more efficiently than single sperm, which is predicted if
458 the combined force of multiple flagella enhances their motility (Moore et al. 2002). We found

459 improved kinematic measures in half of the species studied (*P. maniculatus, P. californicus*, and
460 *P. leucopus*), thus supporting this functional hypothesis; however, in several species we found
461 aggregation had some negative (*P. polionotus* and *P. gossypinus*) or no (*P. eremicus*) impact on
462 motility relative to single sperm cells. While we found support for a theoretical prediction that
463 sperm aggregates achieve greater straight-line velocity because they move in a more linear path of
464 travel rather than at a faster speed (i.e., curvilinear velocity; Fisher et al. 2014) in two of our six
465 focal species (*P. californicus* and *P. leucopus*), we did not find this kinematic benefit for sperm
466 aggregation in all *Peromyscus* species. These results corroborate other studies in more disparate
467 taxonomic groups that have quantified sperm aggregation motility and found inconsistent results.
468 For example, sperm trains exhibit greater swimming progressive motility in the wood mouse
469 (Moore et al. 2002), and greater velocity than individual sperm in the Norway rat, but not the
470 house mouse (Immler et al. 2007). In invertebrates, the swimming velocity of fishfly sperm
471 increases with number of sperm in a bundle (Hayashi 1998), but in a marine snail, there is no
472 differences in swimming speed between paired and single sperm (Ishijima et al. 1999). One
473 possible explanation for these differences across taxa is that cell orientation within an aggregate is
474 critical for its collective motility. Sperm cells are predicted to be faster if they generate increased
475 force with proportionally less drag (see Higginson and Pitnick 2011 and references therein); such
476 effects may be true for sperm aggregations as well in which cells conjoin head-to-tail, thereby
477 increasing the length of the collective unit, or in which flagella within the group beat
478 synchronously (Higginson and Pitnick 2011). Our results support that even sperm aggregates that
479 conjoin head-to-head and are thus wider, as they are in *Peromyscus,* can offer a motility
480 advantage.

481 Our results suggest that relaxed sexual selection may allow the persistence of less optimal
482 strategies based on the quantitative and qualitative differences we observed among sperm
483 aggregates of different species. The most interesting motility results are those of the sister-species
484 pair with divergent mating systems that both form the most frequent and largest sperm aggregates.

485 In *P. maniculatus*, a promiscuous species, sperm aggregates exhibit greater straight-line and
486 curvilinear velocity compared to single cells, but the opposite was true for its monogamous
487 congener, *P. polionotus*. We find that these kinematic difference are associated with differences in
488 aggregate geometry; when sperm heads and flagella are not oriented in the same direction, the
489 cells within an aggregate exert opposing forces on one another, thereby reducing the overall
490 motility of the group (Fisher et al. 2014; Pearce et al. 2018). Indeed, we found that sperm from the
491 monogamous *P. polionotus* males are less likely to form aggregates with all sperm aligned and
492 more likely to include immotile or morphologically abnormal sperm, consequently resulting in
493 slower aggregates than those of *P. maniculatus* (Pearce et al. 2018). This finding is consistent with
494 previous reports that *P. polionotus* sperm tend to form optimal-sized aggregates less often than in
495 *P. maniculatus* (Fisher et al. 2014). Together, these observations further support the hypothesis
496 that sperm aggregation evolved prior to the divergence of the species pair (Fisher and Hoekstra
497 2010), and when monogamy evolved secondarily in *P. polionotus* (Greenbaum et al. 1978; Turner
498 et al. 2010), relaxed sexual selection allowed for the persistence of less motile sperm traits. In line
499 with this prediction, we observed the smallest proportion of motile and progressively motile sperm
500 aggregates in another monogamous species, *P. eremicus,* but the largest proportion of
501 progressively motile aggregates in two promiscuous species, *P. maniculatus* and *P. leucopus*.
502 Similar results have been reported in house mice, in which males evolving under intense sperm
503 competition produced a greater proportion of motile sperm, compared to males from lineages
504 subject to relaxed selection (Firman and Simmons 2011). Together these findings support that the
505 motility benefits conferred by sperm aggregation are associated with variation in post-copulatory
506 sexual selection.

507 When we compared single and aggregated sperm in a viscous environment to test if
508 aggregation improves motility in more complex fluids, similar to the secretions or cervical mucus
509 (Hanson and Overstreet 1981) of the female reproductive tract (Suarez 2016; Simons and Olson
510 2018), we found that collective motion is beneficial for some species but costly for others.

511 Compared to single sperm, aggregates swim with greater curvilinear velocity in *P. maniculatus*
512 and *P. californicus*, and greater straight-line and average path velocities in *P. californicus,* but
513 aggregates were less linear than single cells in all species except for *P. leucopus* and *P.*
514 *gossypinus*. While other studies have found kinematic benefits for sperm conjugates compared to
515 single cells in higher viscosities in the gray short-tailed opossum (Moore and Taggart 1995), the
516 wood mouse (Moore et al. 2002), and the fishfly (Hayashi 1998), a study in bulls found that sperm
517 were slower, exhibited less organized swimming patterns, and were less likely to cluster in viscous
518 fluids (Tung et al. 2017), both of which are consistent with our analysis of *Peromyscus* aggregates.
519 Ultimately, the benefit of sperm aggregation depends on the relative importance of each kinematic
520 parameter during sperm migration *in vivo*. Although beyond the scope of this study, we predict
521 that improved linearity afforded by collective motion may help to direct the sperm through the
522 dynamic fluids of the female reproductive tract (Ishikawa et al. 2016) and that increased velocity
523 will reduce the time it takes for the sperm to arrive at the fertilization site.

524 In conclusion, our study highlights the diversity of sperm aggregation within a single
525 taxonomic lineage and how selection has shaped the formation and performance of these cellular
526 groups. We show that both evolutionary history and varying levels of post-copulatory sexual
527 selection influence the frequency and size of sperm groups. Moreover, we find that sperm
528 aggregation can improve sperm motility in both simple and complex fluids, but this is not
529 consistent across all species. Theoretical predictions (Fisher et al. 2014; Pearce et al. 2018) and
530 emerging empirical evidence suggests that motility benefits may only be realized if cells maintain
531 optimal alignment within the groups and, if achieved, may provide these sperm with a competitive
532 advantage in the female reproductive tract (Higginson et al. 2012b). Future work investigating
533 sperm aggregates *in vivo* (e.g., Ishikawa et al. 2016; Wang and Larina 2018) will shed light on the
534 co-evolution of these unique gametic behaviors and the enormously variable and dynamic female
535 reproductive tracts through which sperm must navigate.
536

Acknowledgments:

We are grateful to Erica Glasper for providing *P. californicus* males and Hopi Hoekstra for providing *P. gossypinus* males. Thanks to Harrison Arsis, Madeline-Sophie Dang, Catherine Liu, and Audrey Mvemba for their assistance with video analysis and to Philip Johnson, Danielle Adams, Sam Church, and Shelby Wilson for statistical advice. We thank Neal Platt for providing the ultra-metric *Peromyscus* tree for use in our statistical analysis.

References:

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67:1–48.

Bedford, J. M., J. C. Rodger, and W. G. Breed. 1984. Why so many mammalian spermatozoa - a clue from marsupials? Proceedings of the Royal Society of London B: Biological Sciences 221:221–233.

Bedford, N. L., and H. E. Hoekstra. 2015. The natural history of model organisms: Peromyscus mice as a model for studying natural variation. eLife Sciences 4:e06813.

Birdsall, D. A., and D. Nash. 1973. Occurrence of successful multiple insemination of females in natural populations of deer mice (*Peromyscus maniculatus*). Evolution 27:106–110.

Birkhead, T. R., J. G. Martinez, T. Burke, and D. P. Froman. 1999. Sperm mobility determines the outcome of sperm competition in the domestic fowl. Proceedings of the Royal Society of London. Series B: Biological Sciences 266:1759–1764.

Birkhead, T. R., A. P. Møller, and W. J. Sutherland. 1993. Why do females make it so difficult for males to fertilize their eggs? Journal of Theoretical Biology 161:51–60.

Bradley, R. D., N. D. Durish, D. S. Rogers, J. R. Miller, M. D. Engstrom, and C. W. Kilpatrick. 2007. Toward a molecular phylogeny for *Peromyscus*: evidence from mitochondrial cytochrome-*b* sequences. Journal of Mammalogy 88:1146–1159.

562 Calhim, S., S. Immler, and T. R. Birkhead. 2007. Postcopulatory sexual selection is associated
563 with reduced variation in sperm morphology. PLoS ONE 2:e413.
564 Carballo, L., A. Battistotti, K. Teltscher, M. Lierz, A. Bublat, M. Valcu, and B. Kempenaers.
565 2019. Sperm morphology and evidence for sperm competition among parrots. Journal of
566 Evolutionary Biology 32:856–867.
567 Crawley, M. J. 2013. The R Book Second Edition. John Wiley & Sons, Ltd., England.
568 Dewsbury, D. A. 1981. An exercise in the prediction of monogamy in the field from laboratory
569 data on 42 species of muroid rodents. Biologist 63:138–162.
570 Dewsbury, D. A. 1985. Interactions between males and their sperm during multi-male copulatory
571 episodes of deer mice (*Peromyscus maniculatus*). Animal Behaviour 33:1266–1274.
572 Eberhard, W. 1996. Female Control: Sexual Selection by Cryptic Female Choice. Princeton
573 University Press.
574 Firman, R. C., and L. W. Simmons. 2011. Experimental evolution of sperm competitiveness in a
575 mammal. BMC Evolutionary Biology 11:19.
576 Fisher, H. S., L. Giomi, H. E. Hoekstra, and L. Mahadevan. 2014. The dynamics of sperm
577 cooperation in a competitive environment. Proceedings of the Royal Society of London B:
578 Biological Sciences 281:20140296.
579 Fisher, H. S., and H. E. Hoekstra. 2010. Competition drives cooperation among closely related
580 sperm of deer mice. Nature 463:801.
581 Fitzpatrick, J. L., and B. Baer. 2011. Polyandry reduces sperm length variation in social insects.
582 Evolution 65:3006–3012.
583 Foltz, D. W. 1981. Genetic evidence for long-term monogamy in a small rodent, *Peromyscus*
584 *polionotus*. The American Naturalist 117:665–675.
585 Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data:
586 a test and review of evidence. The American Naturalist 160:712–726.

587 Gage, M. J. G., C. P. Macfarlane, S. Yeates, R. G. Ward, J. B. Searle, and G. A. Parker. 2004.
588 Spermatozoal traits and sperm competition in Atlantic salmon: relative sperm velocity is
589 the primary determinant of fertilization success. Current Biology 14:44–47.
590 García-Berthou, E. 2001. On the misuse of residuals in ecology: testing regression residuals vs. the
591 analysis of covariance. Journal of Animal Ecology 70:708–711.
592 Greenbaum, I. F., R. J. Baker, and P. R. Ramsey. 1978. Chromosomal evolution and the mode of
593 speciation in three species of *Peromyscus*. Evolution 32:646–654.
594 Hanson, F. W., and J. W. Overstreet. 1981. The interaction of human spermatozoa with cervical
595 mucus *in vivo*. American Journal of Obstetrics and Gynecology 140:173–178.
596 Harrison, X. A. 2014. Using observation-level random effects to model overdispersion in count
597 data in ecology and evolution. PeerJ 2:e616.
598 Hayashi, F. 1998. Sperm co-operation in the Fishfly, *Parachauliodes japonicus*. Functional
599 Ecology 12:347–350.
600 Higginson, D. M., and K. R. H. Henn. 2012. Effects of sperm conjugation and dissociation on
601 sperm viability *in vitro*. PloS one 7:e34190.
602 Higginson, D. M., K. B. Miller, K. A. Segraves, and S. Pitnick. 2012a. Convergence, recurrence
603 and diversification of complex sperm traits in diving beetles (Dytiscidae). Evolution
604 66:1650–1661.
605 Higginson, D. M., K. B. Miller, K. A. Segraves, and S. Pitnick. 2012b. Female reproductive tract
606 form drives the evolution of complex sperm morphology. Proceedings of the National
607 Academy of Sciences 109:4538–4543.
608 Higginson, D. M., and S. Pitnick. 2011. Evolution of intra-ejaculate sperm interactions: do sperm
609 cooperate? Biological Reviews 86:249–270.
610 Immler, S. 2008. Sperm competition and sperm cooperation: the potential role of diploid and
611 haploid expression. Reproduction 135:275–283.

612 Immler, S., S. Calhim, and T. R. Birkhead. 2008. Increased postcopulatory sexual selection
613 reduces the intramale variation in sperm design. Evolution 62:1538–1543.
614 Immler, S., H. D. Moore, W. Breed, and T. Birkhead. 2007. By hook or by crook? Morphometry,
615 competition and cooperation in rodent sperm. PloS one 2:e170.
616 Ishijima, S., S. A. Ishijima, and B. A. Afzelius. 1999. Movement of Turritella spermatozoa:
617 Direction of propagation and chirality of flagellar bends. Cell Motility 44:85–95.
618 Ishikawa, Y., T. Usui, M. Yamashita, Y. Kanemori, and T. Baba. 2016. Surfing and swimming of
619 ejaculated sperm in the mouse oviduct. Biology of Reproduction 94:89, 1–9.
620 Lenth, R. V. 2016. Least-squares means: The R package 'lsmeans'. Journal of Statistical Software
621 69:1–33.
622 Lüpold, S., G. M. Linz, J. W. Rivers, D. F. Westneat, and T. R. Birkhead. 2009. Sperm
623 competition selects beyond relative testes size in birds. Evolution 63:391–402.
624 Mackie, J. B., and M. H. Walker. 1974. A study of the conjugate sperm of the dytiscid water
625 beetles *Dytiscus marginalis* and *Colymbetes fuscus*. Cell and Tissue Research 148:505–
626 519.
627 Magdanz, V., S. Boryshpolets, C. Ridzewski, B. Eckel, and K. Reinhardt. 2019. The swim-up
628 technique separates bovine sperm by metabolic rates, motility and tail length. bioRxiv
629 624502.
630 McGlinn, S. M., B. A. Shepherd, and J. Martan. 1979. Rouleau Formation and fertility of
631 spermatozoa in guinea pigs treated with cyproterone acetate. Archives of Andrology
632 2:197–202.
633 Moore, H. D., K. Dvoráková, N. Jenkins, and W. Breed. 2002. Exceptional sperm cooperation in
634 the wood mouse. Nature 418:174–177.
635 Moore, H. D. M., and D. A. Taggart. 1995. Sperm pairing in the opossum increases the efficiency
636 of sperm movement in a viscous environment. Biology of Reproduction 52:947–953.

637 Nixon, B., H. Ecroyd, J.-L. Dacheux, F. Dacheux, V. Labas, S. D. Johnston, and R. C. Jones.
638 2016. Formation and dissociation of sperm bundles in monotremes. Biol Reprod 95:1–11.
639 Orme, D., R. Freckleton, G. Thomas, and T. Petzoldt. 2013. The caper package: comparative
640 analysis of phylogenetics and evolution in R. R package version 5:1–36.
641 Pagel, M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.
642 Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of phylogenetics and evolution in R
643 language. Bioinformatics 20:289–290.
644 Parker, G. A. 1982. Why are there so many tiny sperm? Sperm competition and the maintenance
645 of two sexes. Journal of Theoretical Biology 96:281–294.
646 Pearce, D. J. G., L. A. Hoogerbrugge, K. A. Hook, H. S. Fisher, and L. Giomi. 2018. Cellular
647 geometry controls the efficiency of motile sperm aggregates. Journal of The Royal Society
648 Interface 15:20180702.
649 Pearcy, M., N. Delescaille, P. Lybaert, and S. Aron. 2014. Team swimming in ant spermatozoa.
650 Biology Letters 10:20140308.
651 Phillips, D. M. 1970. Ultrastructure of spermatozoa of the wooly opossum *Caluromys philander*.
652 Journal of Ultrastructure Research 33:381–397.
653 Pitnick, S., D. J. Hosken, and T. R. Birkhead. 2009. Sperm Morphological Diversity. Pp. 69–149
654 *in* T. R. Birkhead, D. J. Hosken, and S. Pitnick, eds. Sperm Biology: An Evolutionary
655 Perspective. Academic Press, London.
656 Pizzari, T., and K. R. Foster. 2008. Sperm sociality: cooperation, altruism, and spite. PLOS
657 Biology 6.
658 R Core Team. 2016. R: A Language and Environment for Statistical Computing. R Foundation for
659 Statistical Computing, Vienna, Austria.
660 Ramm, S. A., G. A. Parker, and P. Stockley. 2005. Sperm competition and the evolution of male
661 reproductive anatomy in rodents. Proceedings of the Royal Society B: Biological Sciences
662 272:949–955.

663 Rowley, A., L. Locatello, A. Kahrl, M. Rego, A. Boussard, E. Garza-Gisholt, R. M. Kempster, S.
664 P. Collin, E. Giacomello, M. C. Follesa, C. Porcu, J. P. Evans, F. Hazin, F. Garcia-
665 Gonzalez, T. Daly-Engel, C. Mazzoldi, and J. L. Fitzpatrick. 2019. Sexual selection and
666 the evolution of sperm morphology in sharks. Journal of Evolutionary Biology 32:1027–
667 1035.
668 Simons, J. E., and S. D. Olson. 2018. Sperm motility: models for dynamic behavior in complex
669 environments. Pp. 169–209 *in* M. Stolarska and N. Tarfulea, eds. Cell Movement.
670 Springer International Publishing, Cham.
671 Snook, R. R. 2005. Sperm in competition: not playing by the numbers. Trends in Ecology &
672 Evolution 20:46–53.
673 Suarez, S. S. 2016. Mammalian sperm interactions with the female reproductive tract. Cell and
674 Tissue Research 363:185–194.
675 Suarez, S. S., and X. Dai. 1992. Hyperactivation enhances mouse sperm capacity for penetrating
676 viscoelastic media. Biology of Reproduction 46:686–691.
677 Taggart, D. A., J. L. Johnson, H. P. O'Brien, and H. D. M. Moore. 1993. Why do spermatozoa of
678 the American marsupial form pairs? A clue from the analysis of sperm-pairing in the
679 epididymis of the grey short-tailed opossum, *Monodelphis domestica*. Anatomical Record
680 236:465–478.
681 Takami, Y., and T. Sota. 2007. Sperm competition promotes diversity of sperm bundles in
682 *Ohomopterus* ground beetles. Naturwissenschaften 94:543–550.
683 Tourmente, M., D. Zarka-Trigo, and E. R. S. Roldan. 2016. Is the hook of muroid rodent's sperm
684 related to sperm train formation? Journal of Evolutionary Biology 29:1168–1177.
685 Tung, C., C. Lin, B. Harvey, A. G. Fiore, F. Ardon, M. Wu, and S. S. Suarez. 2017. Fluid
686 viscoelasticity promotes collective swimming of sperm. Scientific Reports 7:1–9.

687 Tung, K. S. K., A. Okada, and R. Anagimachi. 1980. Sperm autoantigens and fertilization. I.
688 effects of antisperm autoantibodies on rouleaux formation, viability, and acrosome
689 reaction of guinea pig spermatozoa. Biology of Reproduction 23:877–886.
690 Turner, L. M., A. R. Young, H. Römpler, T. Schöneberg, S. M. Phelps, and H. E. Hoekstra. 2010.
691 Monogamy evolves through multiple mechanisms: evidence from V1aR in deer mice.
692 Molecular Biology and Evolution 27:1269–1278.
693 Wang, S., and I. V. Larina. 2018. *In vivo* three-dimensional tracking of sperm behaviors in the
694 mouse oviduct. Development 145:dev157685.
695 Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer.
696 Woolley, D. M., R. F. Crockett, W. D. I. Groom, and S. G. Revell. 2009. A study of
697 synchronisation between the flagella of bull spermatozoa, with related observations.
698 Journal of Experimental Biology 212:2215–2223.
699 World Health Organization. 2010. WHO laboratory manual for the examination and processing of
700 human semen.
701 Zuur, A. F., E. N. Ieno, and C. S. Elphick. 2010. A protocol for data exploration to avoid common
702 statistical problems. Methods in Ecology and Evolution 1:3–14.
703

FIGURES AND TABLES

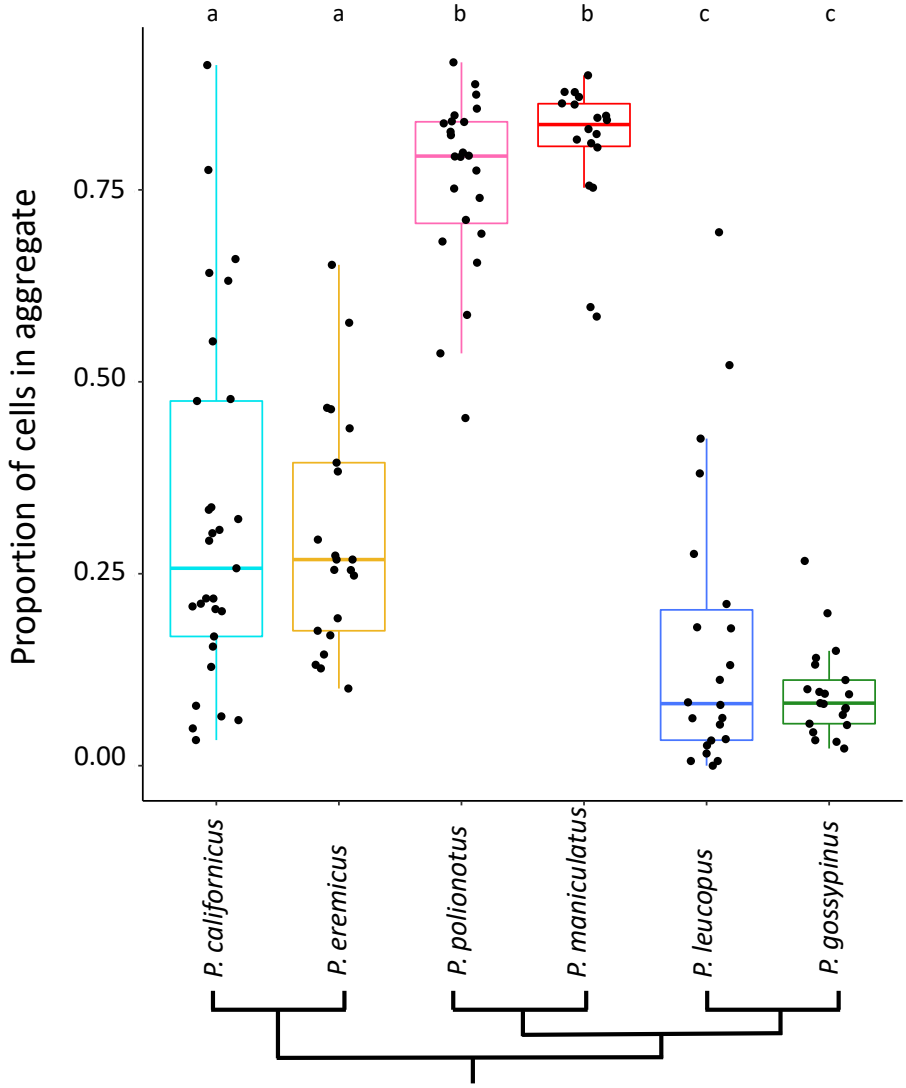

Figure 1.

The observed natural variation in the proportion of aggregated sperm cells for six closely related species of *Peromyscus* mice (phylogeny adapted from Bradley et al. 2007). Box-plots represent median and interquartile ranges with raw data overlaid. Statistically significant differences at the $P = 0.05$ level are denoted by differing letters; shared letters denote no statistical difference.

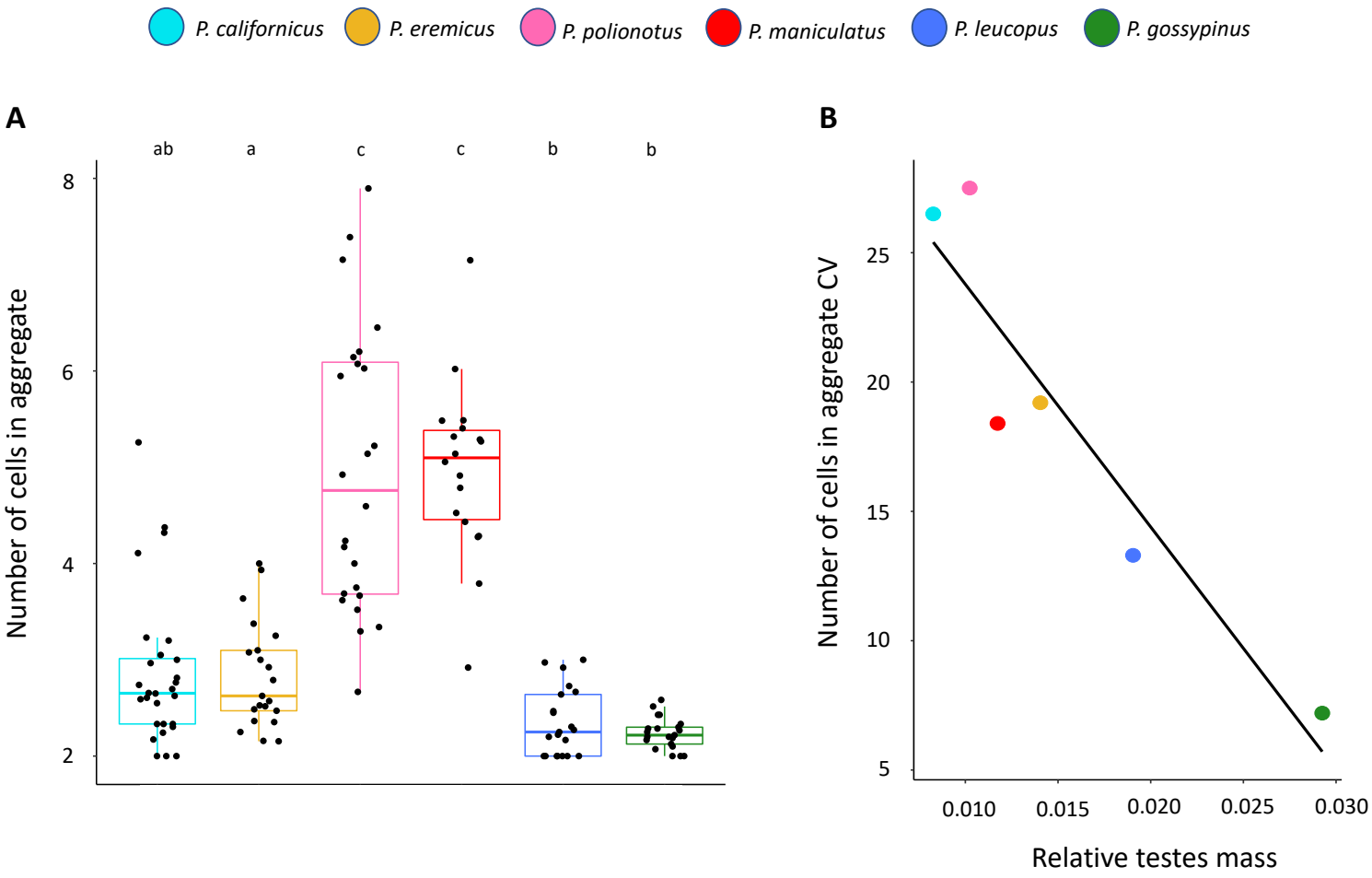

Figure 2.

Sperm aggregate size varies among species and is negatively associated with relative testes size. (A) The observed natural variation in the number of aggregated sperm cells for six closely related species of *Peromyscus* mice. Box-plots represent median and interquartile ranges with raw data overlaid. Statistically significant differences at the $P = 0.05$ level are denoted by differing letters; shared letters denote no statistical difference. (B) When controlling for phylogenetic relationships, the coefficient of variation (CV) for the number of aggregated sperm cells negatively correlates with relative testis mass across these species. Note truncated y-axes.

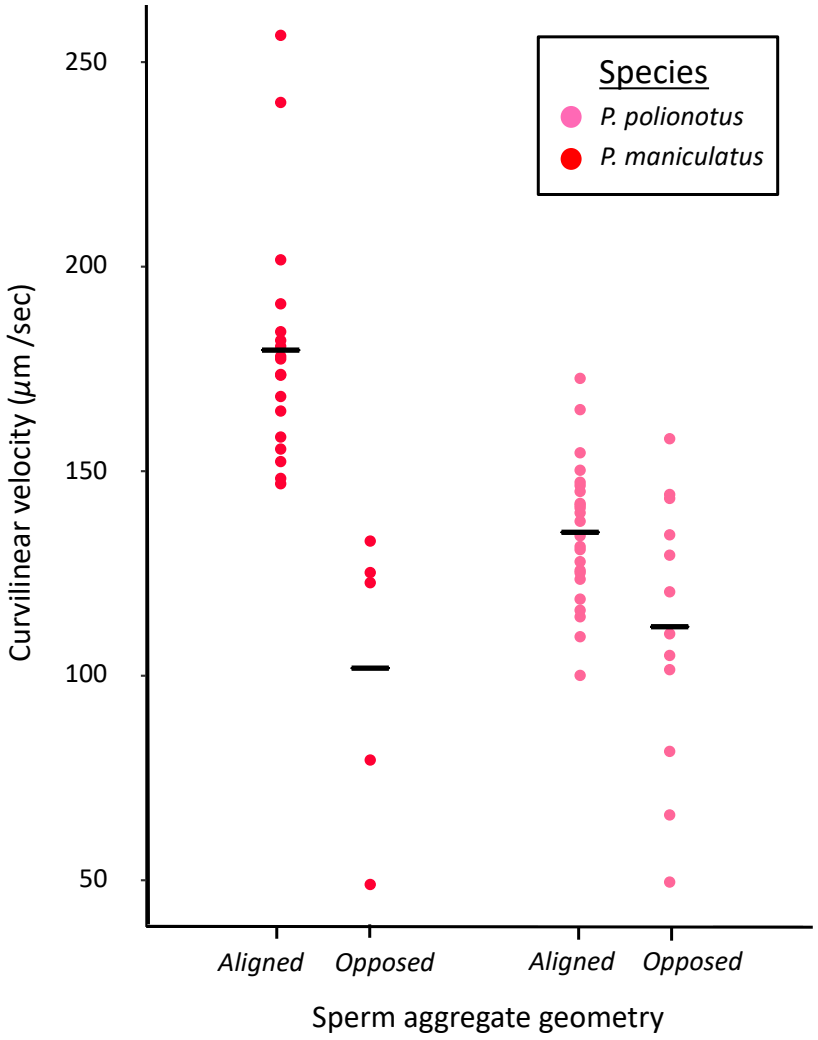

Figure 3.

The effect of sperm aggregate geometry on the curvilinear velocity (μm/sec) of sperm aggregates for two species that aggregated most – *Peromyscus maniculatus* and *Peromyscus polionotus*. Circles represent mean values per male within each species, and black lines represent the mean value within each category. Note truncated y-axis.

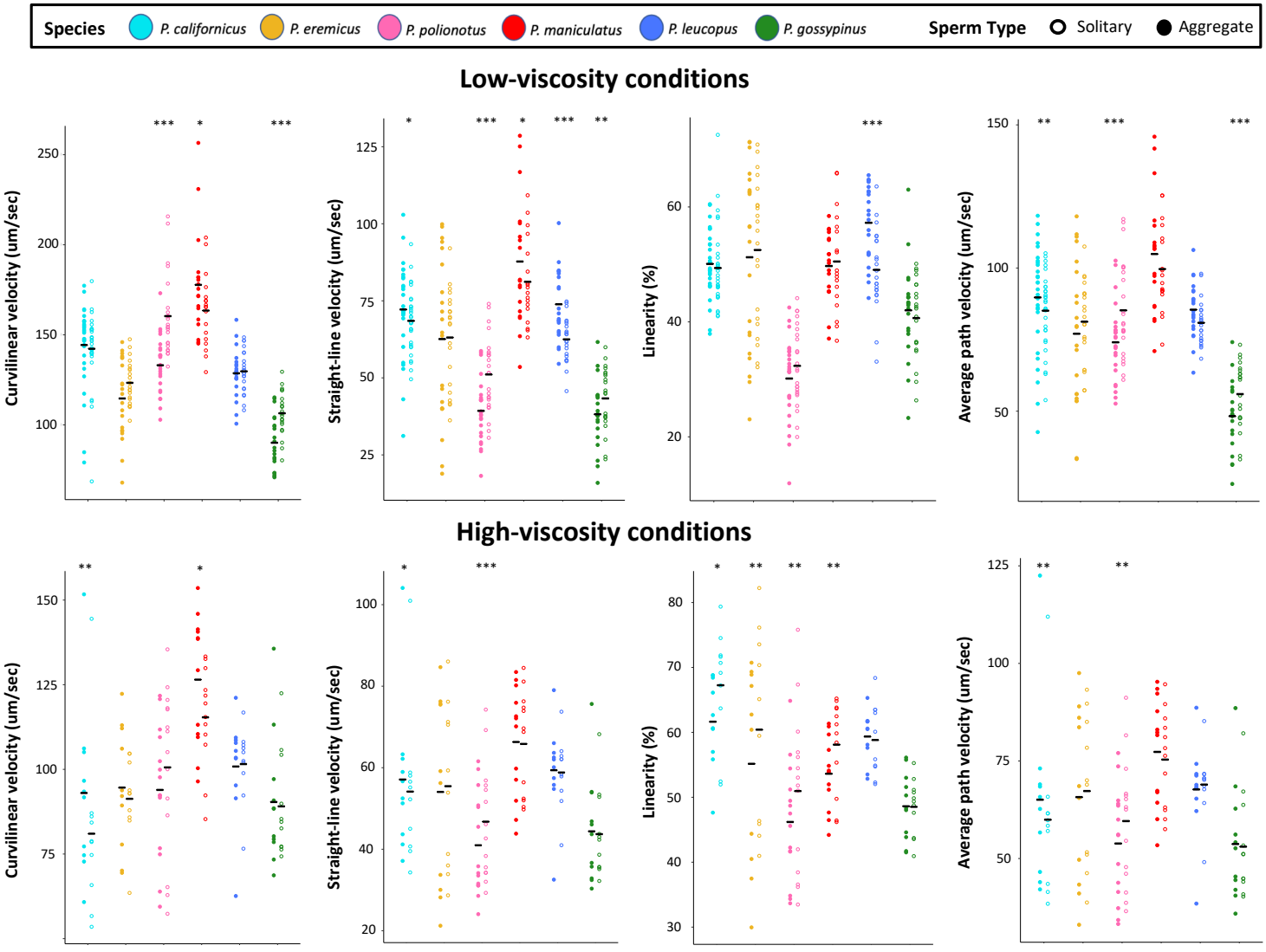

Figure 4.

Kinematic parameters of sperm aggregates (closed circles) and solitary sperm cells (open circles) for six species of *Peromyscus* mice in low- and high-viscosity conditions. Circles represent mean values per male, and black lines represent species means. Statistical significance levels comparing aggregated and solitary cells within each species are indicated by * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Note truncated y-axes.

TABLE 1.

Summary of live sperm aggregate results for mice in the genus *Peromyscus*

Peromyscus Species	No. males	No. sperm	% motile sperm	% progressive sperm	No. sperm aggregates	% cells in aggregate	% motile aggregates	% progressive aggregates
californicus	28	8235	81.2%	78.9%	823	29.7% (2446/8235)	90.5% (745/823)	86.6% (713/823)
eremicus	21	4906	67.4%	58.2%	513	30.2% (1481/4906)	77.8% (399/513)	66.7% (342/513)
polionotus	24	6360	82.4%	57.3%	949	79.5% (5059/6360)	90.9% (863/949)	66.1% (627/949)
maniculatus	18	4991	85.0%	84.3%	822	81.3% (4059/4991)	93.4% (768/822)	92.3% (759/822)
leucopus	22	6341	87.6%	84.5%	361	15.0% (949/6341)	90.3% (326/361)	88.1% (318/361)
gossypinus	21	5970	82.2%	70.3%	254	9.6% (576/5970)	88.6% (225/254)	70.1% (178/254)

Motile sperm cells include only those that exhibited movement patterns. Progressive sperm cells are motile cells with a straight-line velocity $\geq$ 25 μm/sec.

TABLE 2.

Fixed effects from a binomial generalized linear mixed model examining differences in the proportion of aggregated sperm cells and a linear model examining differences in the number of aggregated sperm cells across six species of *Peromyscus* mice

GLMM: PROPORTION OF CELLS IN AGGREGATE					
Model Term	Beta (SE)	Exp (beta)	95% CI	z	Pr(>|z|)
Intercept	1.52 (0.22)				
eremicus	-2.47 (0.30)	0.07	(0.04, 0.13)	-8.14	**< 0.001**
gossypinus	-3.94 (0.31)	0.02	(0.01, 0.03)	-12.89	**< 0.001**
californicus	-2.58 (0.29)	0.07	(0.04, 0.12)	-9.06	**< 0.001**
leucopus	-3.79 (0.30)	0.02	(0.01, 0.04)	-12.46	**< 0.001**
polionotus	-0.25 (0.29)	0.44	(0.30, 0.58)	-0.85	0.393
LM: NUMBER OF CELLS IN AGGREGATE					
Model Term	Beta (SE)	Exp (beta)	95% CI	t	Pr(>|z|)
Intercept	3.90 (0.26)				
Total Sperm Cells	0.00 (0.00)	0.50	(0.50, 0.50)	5.83	**< 0.001**
eremicus	-1.97 (0.24)	0.12	(0.08, 0.18)	-8.15	**< 0.001**
gossypinus	-2.77 (0.24)	0.06	(0.04, 0.09)	-11.53	**< 0.001**
californicus	-2.19 (0.23)	0.10	(0.07, 0.15)	-9.68	**< 0.001**
leucopus	-2.68 (0.24)	0.06	(0.04, 0.10)	-11.18	**< 0.001**
polionotus	0.04 (0.23)	0.51	(0.40, 0.62)	0.16	0.877
For both models, all rows are being compared with the intercept – *Peromyscus maniculatus*. 95% confidence intervals (CI) were calculated for each effect size.					

TABLE 3.

Results from an intra-male analysis comparing motile solitary and aggregated sperm kinematics in low- and high-viscosity conditions for six species of *Peromyscus* mice to test whether sperm aggregates confer kinematic benefits (shaded in gray)

PEROMYSCUS SPECIES	df	KINEMATIC VARIABLE			
		Curvilinear Velocity (μm/sec)	Linearity (VSL/VCL)	Straight-Line Velocity (μm/sec)	Average Path Velocity (μm/sec)
LOW-VISCOSITY CONDITIONS					
californicus	*28*	t = -1.0545, p = 0.3007	t = -0.56153, p = 0.5789	t = -2.2982, ***p* = 0.02923**	t = - 2.8162, ***p* = 0.008805**
eremicus	*20*	t = 1.6225, p = 0.1204	t = 0.73517, p = 0.4708	t = 0.1312, p = 0.8969	t = 1.0464, p = 0.3079
polionotus	*23*	t = 9.4575, ***p* = 2.1566e-09**	t = 1.699, p = 0.1028	t = 5.8355, ***p* = 6.026e-06**	t = 6.0729, ***p* = 3.408e-06**
maniculatus	*17*	t = 2.2482, ***p* = 0.03812**	t = 0.48075, p = 0.6368	t = -2.2335, ***p* = 0.03924**	t = -1.9206, p = 0.07172
leucopus	*20*	t = 0.23337, p = 0.8178	t = -4.8385, ***p* = 9.973e-05**	t = -5.5521, ***p* = 1.959e-05**	t = -1.838, p = 0.08096
gossypinus	*20*	t = 5.4048, ***p* = 2.73e-05**	t = -1.0247, p = 0.3177	t = 3.715, ***p* = 0.001369**	t = 4.479, ***p* = 0.0002298**
HIGH-VISCOSITY CONDITIONS					
californicus	*9*	t = -3.7465, ***p* = 0.003357**	t = 2.682, ***p* = 0.02512**	t = -2.7743, ***p* = 0.0216**	t = -4.1106, ***p* = 0.002634**
eremicus	*9*	t = -0.71739, p = 0.4913	t = 4.2087, ***p* = 0.002277**	t = 0.62918, p = 0.5449	t = 0.56505, p = 0.5858
polionotus	*13*	t = 2.0349, p = 0.06278	t = 3.9242, ***p* = 0.001745**	t = 4.8314, ***p* = 0.0003279**	t = 4.18, ***p* = 0.001079**
maniculatus	*11*	t = -2.9397, ***p* = 0.01345**	t = 3.9259, ***p* = 0.002369**	t = -0.30835, p = 0.7636	t = -1.0543, p = 0.3114
leucopus	*9*	t = 0.32429, p = 0.7531	t = -0.31635, p = 0.759	t = -0.34159, p = 0.7405	t = 0.95055, p = 0.3667
gossypinus	*10*	t = -0.63835, p = 0.95376	t = -0.076451, p = 0.9406	t = -0.60402, p = 0.5593	t = -0.71643, p = 0.4901

Statistical results are based on paired-student t-tests.