Journal of Evolutionary Biology, 2025, **38**, 1548–1555
https://doi.org/10.1093/jeb/voaf103
Advance access publication 10 September 2025
Research Article

JOURNAL OF
Evolutionary Biology
eseb

On sperm length mean–variance relationships

David J. Hosken[1], John L. Fitzpatrick[2], Tommaso Pizzari[3], Dave J. Hodgson[1]

[1]Centre for Ecology and Conservation, University of Exeter, Penryn, United Kingdom
[2]Department of Zoology, University of Stockholm, Stockholm, Sweden
[3]Department of Biology, University of Oxford, Oxford, United Kingdom

Handling editor: Alejandro Gonzalez-Voyer, Associate editor: Masahito Tsuboi

Corresponding author: David J Hosken, Centre for Ecology and Conservation, University of Exeter, Penryn, TR109FE, United Kingdom. Email: d.j.hosken@exeter.ac.uk

Abstract

Sperm length is highly variable within ejaculates, between males, among populations, and across species. While theory makes strong predictions about expected mean sperm size, there is less clarity on variation in sperm, although studies have reported sperm-length variation consistent with some theoretical expectations. Typically, the coefficient of variation (CV) is used in these investigations to control for mean–variance scaling. However, a key assumption for this metric to be appropriate in controlling for mean sperm size is that the standard deviation in size scales linearly with the mean. Unfortunately, sperm-length mean–variation relationships are rarely reported making it hard to assess the validity of using CV as a way to compare mean-corrected sperm variation. Here, we investigate mean–variation relationships using 19,873 sperm length measures from 54 species and find little evidence of a consistent relationship between mean sperm-length and sperm-length variation among males within species, meaning CV is not appropriate for comparing relative (mean corrected) variation in sperm size at this level. We also find significant scaling of sperm-length variation with mean sperm-length across species, but the scaling exponent is consistently less than one, the exponent required by analyses using CV to control for sperm size. Our assessment shows that sperm mean–variation scaling relationships are rare within species and strong across species, but that neither supports the uncritical use of CV in studies of relative variation in sperm length.

Keywords: variation, sperm size, coefficient of variation, spermatozoa

Introduction

Sperm are highly specialized cells, characterized by extensive size variation at both inter- and intraspecific levels (Pitnick et al., 2008). Across species, sperm length varies four orders of magnitude (Kahrl et al., 2021), from the extremely short (e.g., ~6 μm in the Asian redtail catfish *Hemibagrus nemurus*, Muchlisin, 2004) to extraordinarily long (e.g., almost 6 cm in *Drosophila bifurca*, Pitnick et al., 1995). Within species, considerable variation in sperm length occurs both among males and within males, including within a single ejaculate (Birkhead et al., 2005; Fitzpatrick et al., 2010; Gage, 1998; Holt & van Look, 2004; Malo et al., 2006; Morrow & Gage, 2001; Pitnick et al., 2008; Sharma et al., 2013; Ward & Hauschteck-Jungen, 1993). However, despite obvious size variation and much research, a thorough understanding of variation in sperm length is lacking (Pitnick et al., 2008; but see Kahrl et al., 2021).

Postcopulatory sexual selection, and especially sperm competition, is a major source of selection on sperm length (Parker, 1984, 1993), and should deplete intraspecific length variation as sperm evolve towards an evolutionarily stable size (e.g., Calhim et al., 2007). Sperm competition theory suggests that selection should favour numerous small sperm unless there is some advantage to larger size (Parker, 1993), but sperm should evolve towards optimal size in either case. This should apply within and between species, and the high levels of intramale sperm variation in the Bullfinch are thought to be due to relaxed sperm competition, for example (Durrant et al., 2010). More generally, studies frequently investigate sperm-length variation relative to some measure of sperm competition risk while controlling for sperm size to test for adaptive explanations for size variation (e.g., Calhim et al., 2007; Immler et al., 2008; Sharma et al., 2013). Against this selective backdrop, another explanation for sperm length variation is that it is an inevitable consequence of imperfect sperm production, with developmental noise during spermatogenesis continually generating variation in sperm length (Cohen, 1967). This may explain some variation, but at least one study has found no evidence that developmental instability is associated with variation in sperm length (Hosken et al., 2003). Additionally, postcopulatory sexual selection should improve production quality—here meaning sperm length mapping to selective optima—during and after spermatogenesis (Hunter & Birkhead, 2002).

It is also possible that some variation in sperm length could be explained by haplo-diploid conflict or the production of multiple sperm types (Pitnick et al., 2008). Haplo-diploid conflict arises when the sperm phenotypic optima differ between the haploid sperm and the diploid males that produce them (Parker, 1993; Parker & Begon, 1993). However, once again sperm competition among genetically unrelated males should align the fitness interests of the haploid sperm with those of the diploid males, leading to convergent optimal sperm lengths (Parker & Begon, 1993; Pizzari & Foster, 2008). Additionally, sperm length variation exists in species that produce only one sperm type (Pitnick et al., 2008), so this cannot be a general cause of sperm length variation, but at least some sperm-length variation is expected to

Received March 17, 2025; revised June 11, 2025; accepted July 7, 2025

relate to the general phenomenon of mean–variance scaling (Lewontin, 1966; Sokal & Rohlf, 1981)—bigger things tend to be more variable. If so, understanding the natural scaling relationship between mean and variance is fundamentally important to adequately understand variation in sperm length, because without understanding this relationship, it is not clear how to control for sperm size in investigations of sperm-length variation.

The most frequently employed method to account for mean–variation associations is the coefficient of variation (CV) (or in some cases its multivariate equivalent which sums variation across traits) (Lande, 1977; Lewontin, 1966; Pélabon et al., 2020; Van Valen, 2005; and e.g., Fitzpatrick & Baer, 2011; Joly et al., 1989; Sharma et al., 2013). This statistic is generated by dividing the standard deviation of the trait in question by the trait mean (and in many cases multiplying by 100 to express CV as a percentage). CV is an attractive scaled measure of variation because it is dimensionless (mean and SD are measured in the same units of trait dimension) and hence it can be used to compare relative variation across traits, populations or species. Once researchers have generated CV values, they typically then assume that any size–variation relationship has been accounted for and interpret any difference in CV between groups as the signal of something "other than" natural scaling (e.g., Sharma et al., 2013; reviewed in Pélabon et al., 2020). One under-appreciated, but critical feature of CV is that it only equivalizes mean-scaled variation if trait dimensions scale linearly (Lewontin, 1966; Pélabon et al., 2020; Sokal & Rohlf, 1981) and this may not be always true. That is, if variation (the standard deviation: SD) does not scale to the mean linearly, then size correction can inflate or deflate relative variation, compromising comparisons.

Unfortunately, the underlying scaling relationships of variation to mean are rarely reported in studies of sperm length, and even when they are reported, CVs are still almost always used in these investigations (e.g., Calhim et al., 2007; Fitzpatrick & Baer, 2011; Immler et al., 2008; Joly et al., 1989; Kleven et al., 2008; Sharma et al., 2013). This is fine when there is interest in sperm variation as a proportion of mean size, but more problematic when *controlling* for mean-size. If the aim is to control for size, a key CV assumption (standard deviation of the trait scales linearly with the mean) needs to be met and this requires testing mean–variation scaling. Here, we estimate the natural sperm length mean–variance scaling relationships within and among species using a large number of invertebrate and vertebrate taxa. We then zoom in taxonomically to estimate how sperm length variation scales in 20 dungfly species and 27 species of Hymenoptera. This is the first widespread reporting of sperm mean–variation scaling and cautions against the use of CV in comparing *mean-controlled* variation in sperm length.

Methods

We use data sourced from an array of data-sets (Minder, 2002 (used in Sharma et al., 2013), unpublished data from Fitzpatrick et al., 2010 and Fitzpatrick & Baer, 2011, and from the published dataset in Simpson et al., 2014) in which measures of total sperm length (in microns) are taken for multiple sperm from each of multiple males. This dataset consisted of 19,873 sperm length measures, for 1,707 males, from 54 species. We aggregate these data to yield mean and standard deviation in sperm length per male.

We first examined how variance in sperm length is partitioned among the hierarchical scales within-male, among-male, and among-species, using a simple random effects model with no fixed predictors. We then examined within- and across-species associations between mean and variation in sperm length in a subset of taxa (9 species) deliberately chosen to reflect the taxonomic breadth of our dataset, while also capturing variation in total sperm length. This subsect comprised of two marine invertebrates (*Mytilus galloprovincialis* and *Heliocidaris erythrogramma*), two insects (*Bombus terrestris* and *Scathophaga stercoraria*), two fishes (*Poecilia reticulata* and *Melanotaenia australis*), one amphibian (*Crinia georgiana*), one bird (*Dromaius novaehollandiae*), and one mammal (*Homo sapiens*).

After log-transforming the means and dispersion measures for multiple males from multiple species, we infer within- and among-species mean–variation scaling using general linear models to describe associations between means and standard deviations. We do this initially using males as replicates, grouped by species, to assess within-species scaling, and then using species-average values of mean and standard deviation in sperm length to assess among-species scaling. We use log–log transformations in order to determine whether scaling relationships are effectively linear (slope of 1), additive (slope of zero), or power (slope greater than one, or less than one and different from zero). We then considered how within-male sperm-length variation scales within two distinct clades using 20 species of dungflies (Minder, 2002; Sharma et al., 2013) and 27 species of Hymenoptera (Fitzpatrick & Baer, 2011) chosen because of the large number of species within these groups.

We do not use phylogenetic control because for many species in these analyses there is no phylogeny (some species are unnamed) and most phylogenetic tools consider phylogenetic structuring in the residuals of the regression, which can cause a flattening of allometric relationships and hence a deflation of allometric exponents. Any inclusion of phylogenetic signal would require complex and very careful consideration of additional statistical artefacts, especially if the goal is to understand how to handle natural scaling relationships. Clearly, taxonomy is important in the structuring of sperm mean size and measures of dispersion, but here our point is simply to show that mean correcting variation is complicated.

Throughout, our analyses use simple, weighted general linear models of log–log associations between standard deviation and mean sperm length. Noting that estimates of within-male variation in sperm size will be sensitive to sample size, we weight each male's contribution by the sample size of sperm cells used to measure mean and dispersion. We report slopes (±SEs), *t*-tests and *F*-tests of significance, and correlation coefficients where relevant. We do not regress CV against mean in our analyses because the use of "mean sperm length," as both the explanatory variable and the denominator of the response variable, risks the introduction of spurious statistical artefacts. In Supplementary Material, we provide more synthetic analyses of within- and among-species mean-dispersion scaling, using mixed effects models that regress standard deviation against mean, and CV against mean, with species as a random effect.

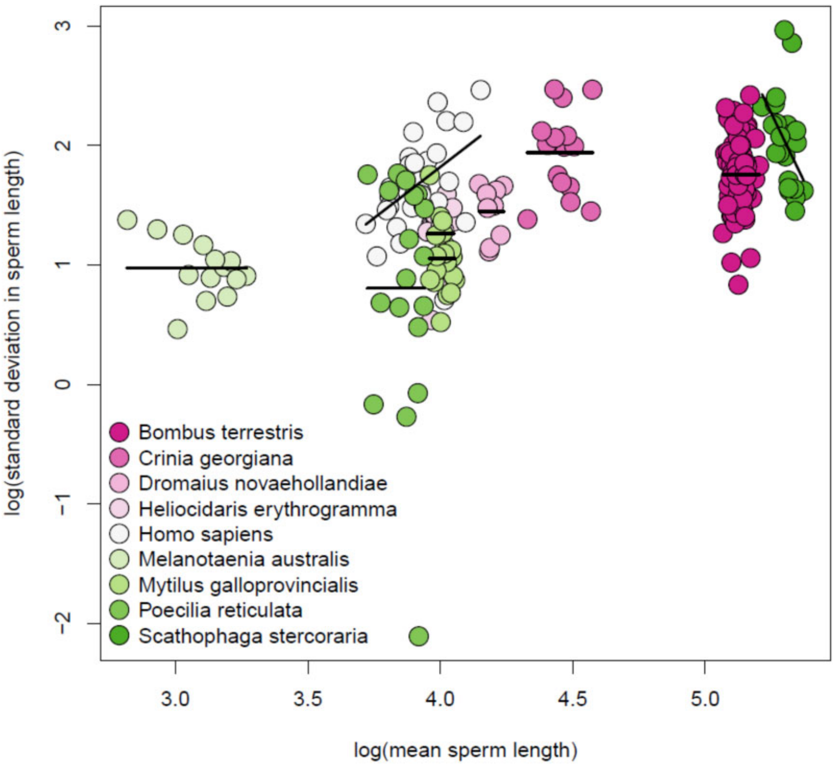

Figure 1. Log–log relationships between sperm length standard deviation and mean sperm length among-males, within-species, in a broad range of taxa. In most cases, there was no significant association and only in humans does sperm-length variation increase with mean sperm length. This cautions against the use of the CV in comparing relative (= mean corrected) variation in sperm size.

Results

Sources of variation in sperm length

Variance-partitioning of the complete dataset, using a simple random effects model, showed that the vast majority (99.8%) of sperm length variation occurs among-species. The remaining variation is mainly within-males (68.4%) compared to among-males (31.6%), demonstrating that within-male variation is very small compared to taxonomic variation, but large compared to variation among males within species.

Sperm mean–variation associations: broad taxonomy

From our restricted (9 species) within-species survey, only two taxa show significant within-species mean–variation associations (Figure 1), with the standard deviation in the sperm of *Homo sapiens* increasing with mean length (±SE), with a scaling exponent of 1.71 (±0.63) (significantly greater than zero, $t_{27} = 2.72$, $P = 0.011$; not significantly different from one, $t_{27} = 1.13$, $P = 0.208$), while dispersion in the dungfly *Scathophaga stercoraria* scales negatively with mean, with an exponent of −4.60 (±2.10) (significantly less than zero, $t_{21} = -2.20$, $P = 0.040$). The remaining seven species show no significant within-species scaling of dispersion with mean sperm length. Thus, the only species for which use of CV might be deemed appropriate is *Homo sapiens*, because its scaling is more than additive and cannot be distinguished from linear. For all other species, the use of CV is inappropriate to compare relative (mean controlled) sperm-length variation.

Comparing the mean–variation scaling relationship among these 9 species used the mean of the per-male standard deviations and the mean of the per-male means. Across these species, there was a significant mean-standard deviation relationship, and the slope of this association was 0.57 (±0.11) (Figure 2), significantly greater than 0 ($t_7 = 3.50$, $P = 0.001$) and smaller than the CV-expectation of 1 ($t_7 = -3.94$, $P = 0.004$), but not significantly different from 0.5 ($t_7 = 0.68$, $P = 0.679$). This means that comparisons of CV would not differentiate between apparent changes in dispersion caused by natural scaling vs. actual changes in the scatter around the mean. The slope of ∼0.5 tells us that variance scales near-linearly with the mean (because the scaling exponent for variance will be twice the scaling exponent for standard deviation), suggesting that comparisons of the coefficient of *variance* (i.e., variance scaled by the mean, CVar, rather than standard deviation scaled by mean, CV) might be adequate for use in comparisons between species.

Sperm mean–variation associations

A survey of sperm measurements in 20 species of dungfly yielded sufficient information (number of males >4) for 14 within-species regressions of sperm variation against mean sperm length (Figure 3). In only one of these 14 species (*Scathophaga stercoraria*: see above) did we find a significant (but negative) association between (log) standard deviation and (log) mean (slope = −4.60 (±2.10), $t_{19} = -2.20$, $P = 0.040$): all other regressions were non-significant ($|t| < 1.27$; $P > 0.21$), and even the significant one does not survive Bonferroni correction for multiple testing and Type I errors. For none of these species would the use of CV be appropriate to compare relative sperm variation across males.

The across-species regression of standard deviation against mean length reveals a significant positive association, with a

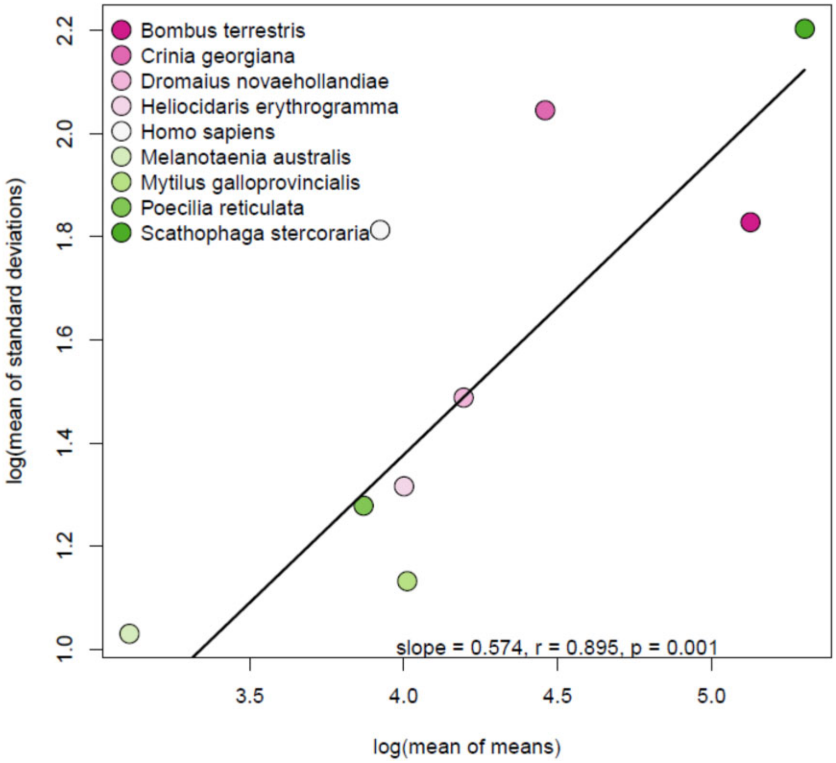

Figure 2. The relationship between sperm length standard deviation and mean size among species across broad taxonomic space. This relationship suggests the use of CV might generate false inferences when the aim is to compare scatter around means across species.

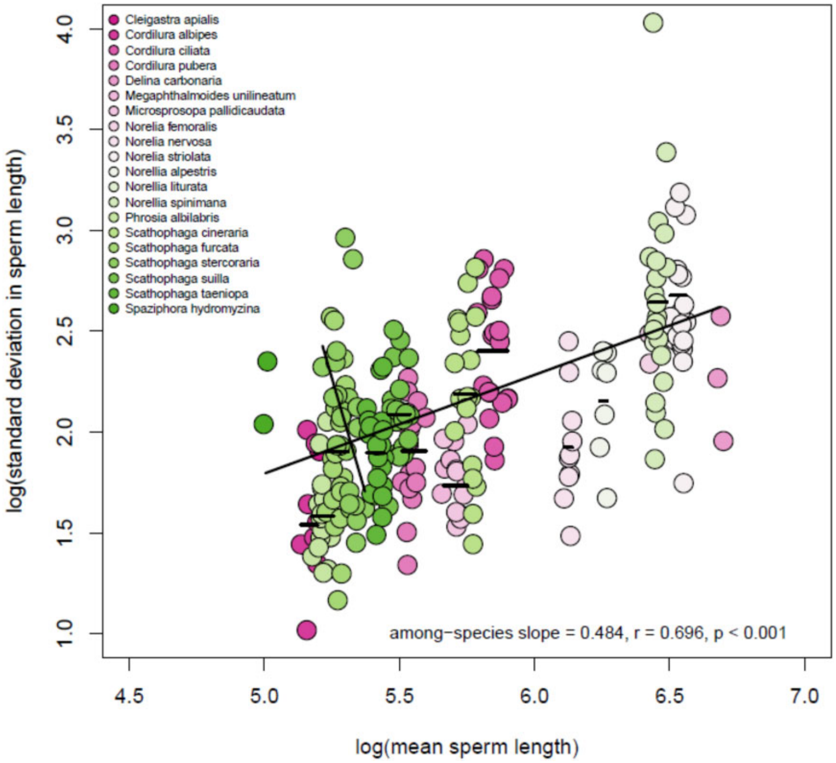

Figure 3. Within- and among-species associations between standard deviation and mean sperm length, in 20 species of dungfly. Regression lines per species have zero slope if non-significant. None of the within-species associations recommend use of CV. The regression of average standard deviation against average mean length, per species, has a log–log slope of 0.453 and hence use of CV would not be appropriate in comparing data scatter around the mean.

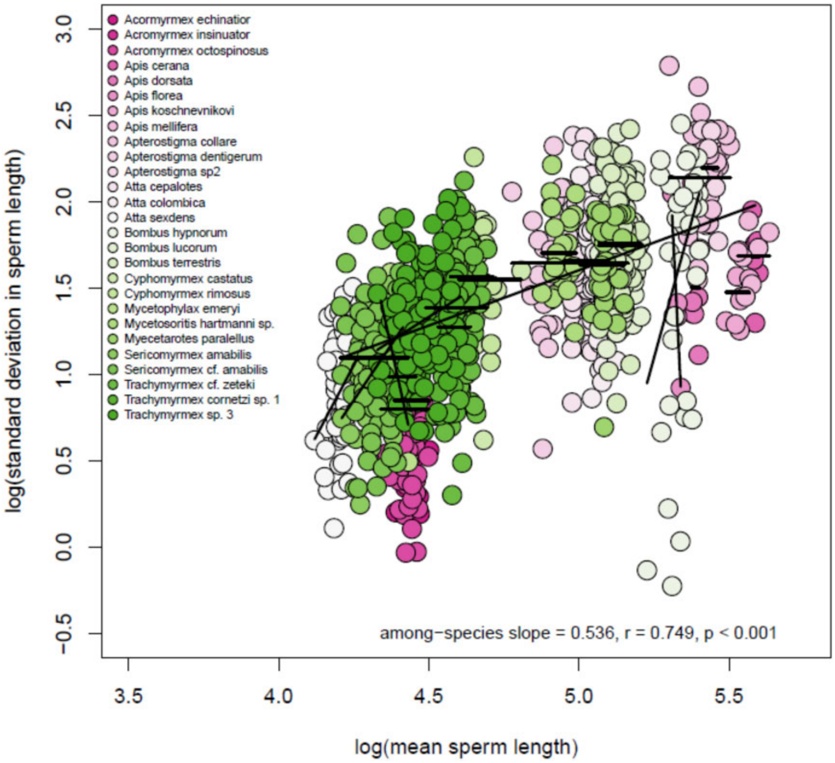

Figure 4. Within- and among-species associations between standard deviation and mean sperm length, in 27 species of Hymenoptera. Regression lines per species have zero slope if non-significant. None of the within-species associations recommend use of CV except perhaps *Trachymyrmex* sp. 3. The regression of average standard deviation against average mean length, per species, has a log–log slope of 0.64 and hence use of CV would not be appropriate in comparing data scatter around the mean.

log–log slope of 0.484 (±0.12), which is significantly greater than 0 ($t_{18} = 3.99$, $P < 0.001$), significantly lower than the CV-requirement of 1 ($t_{18} = -4.25$, $P < 0.001$) and not significantly different from the slope of 0.5 found in our broader taxonomic survey ($t_{18} = -0.13$, $P = 0.390$). Again this provides tantalizing evidence that the linear scaling of *variance* (not standard deviation) with mean sperm length, among species, might hold across taxa. This requires further research.

Sperm mean–variation associations: Hymenoptera

Now looking at 27 species of Hymenoptera (Figure 4), it appears there were six significant associations between (log) standard deviation and (log) mean sperm length: in *Apis florea* (slope = −34.48 (±8.96), t_4 = -3.85, $P = 0.018$); *Atta sexdens* (slope = 3.28 (±1.57), $t_{54} = 2.08$, $P = 0.042$); *Bombus hypnorum* (slope = 6.31 (±2.38), $t_{35} = 2.98$, $P = 0.005$); *Mycetophylax emeryi* (slope = −7.86 (±2.75), $t_9 = -2.86$, $P = 0.019$); *Sericomyrmex amabilis* (slope = 2.58 (±0.95), $t_{38} = 2.73$, $P = 0.010$); and *Trachymyrmex* sp. 3 (slope = 1.11 (±0.46), $t_{100} = 2.39$, $P = 0.019$). All other regressions were non-significant, and of the significant slopes, only two would fit the requirements of CV (slope not significantly different from 1; *Atta sexdens* $t_{54} = 1.45$, $P = 0.139$; *Trachymyrmex* sp. 3 $t_{100} = 0.23$, $P = 0.387$), but significant slopes also go in the "wrong" direction and none survive Bonferroni correction for multiple testing and Type I errors. Therefore, for none of these species would the use of CV be appropriate to compare sperm variation across males.

The among species regression of standard deviation against mean length reveals a significant positive association, with a log–log slope of 0.54 (±0.095), which is significantly greater than 0 ($t_{25} = 5.66$, $P < 0.001$), significantly lower than the CV-requirement of 1 ($t_{25} = -4.89$, $P < 0.001$) and not significantly different from the slope of 0.5, as found in our broader taxonomic and dungfly surveys ($t_{25} = 0.38$, $P = 0.366$). Again, this contravenes requirements for use of CV in among-species comparisons of sperm length dispersion.

Discussion

It is common in evolutionary ecology to compare trait variation, between groups or treatments, following correction by mean trait size (reviewed in Pélabon et al., 2020) and this is certainly the case in studies of comparing sperm-length variation (e.g., Calhim et al., 2007; Immler et al., 2008; Kleven et al., 2008; Joly et al., 1989; Fitzpatrick & Baer, 2011; Sharma et al., 2013). However, the presumption that this removes any size–variance signal is only true when the standard deviation of the trait scales linearly with the mean (i.e., with slope of 1 on a log–log plot and intercept of 0), a point that has been made in other contexts (Pélabon et al., 2020; Tomkins & Simmons, 2002; and see e.g., Eberhard et al., 1998; Hosken, 2008).

We have shown here that CV should not be used in comparisons of mean correct variation in sperm length without prior testing of the underlying assumptions, a point made more generally recently (Pélabon et al., 2020). When such comparisons are made within-species, our analyses suggest there should be no expectation of size–variation associations (but a minority of species defy this pattern and deserve further scrutiny). Our broad survey across species provides some evidence of linear scaling of *variance* rather than of *standard deviation*. If this is the case, the CVar, where variance is divided by mean, could be used as a simple scaled measure of variation. This requires further investigation. However, this coefficient loses the benefit of being dimensionless (it has units of trait dimension), and endorsing the use of CVar risks generating an assumption that sperm have variance that scales linearly with the mean, which may not always be the case. Investigators still need to understand how mean and variance scale.

While trait variation frequently scales with the mean (Pélabon et al., 2020), and this is true of body mass and clutch-size, for example, the relationship is frequently not linear (reviewed in Pélabon et al., 2020). This is also what we found for some sperm estimates. However, for many of the associations we investigated here, sperm-length variation does not scale with mean sperm-size, or even scales negatively with sperm length, and although there is no reason to expect within species relationships to reflect across species relationship, our findings reveal vast differences across comparisons and taxa. All this means that when studies wish to compare variation while controlling for sperm size, care is needed. As noted above, if the aim is to compare variation as a *proportion* of mean size, then CV can still be employed; however, just do not assume it *controls* for (removes the effects of) sperm size in an unbiased way. That is, the fundamental question being addressed is critical in assessing the validity of employing CV regardless of mean–variation scaling (discussed in Houle et al., 2011), but when controlling for mean sperm size, CV does not appear appropriate for the species we investigated here.

Collectively, our findings beg the question, what should actually be used to make valid comparisons of samples of mean corrected variation in sperm length? Comparisons are often made between two treatment groups or two populations that are predicted to differ in mean trait value. One solution is to assume that the size–variation association displayed in the "control" group can be extrapolated to the "treatment" group (e.g., in experimental comparisons), hence we recommend general linear modelling of log–log associations between variance and mean, which allow the analyst to test for differences in slope and intercept between groups (and see Pélabon et al., 2020). We nonetheless caution that these analyses come with an important warning, because extrapolation of relationships beyond natural boundaries (e.g., from control to treatment) can be risky. We demonstrate this by showing that males within-species generally show no scaling of variance with mean, while linear scaling of variance (but not standard deviation) commonly occurs among species.

One way to compare the mean-controlled variation (scatter) in sperm length that is not due to mean–variance scaling is to establishing the actual scaling of variance to the mean (regress variance on mean) for each taxon and take the residual variance (that not explained by the scaling relationship: the least squares errors) and compare this across taxa. This testing of residual variation tells us something about the relative tightness of the relationship in mean–variance scaling (how much noise there is in sperm-length) across taxa, and for example, if one were interested in whether sperm competition risk impacted variation in sperm length, then the variance residuals could be regressed against testes size with the expectation that species with larger testes have smaller variance residuals. However, these types of analyses will not reveal why scaling relationships vary across taxa. We also add a serious cautionary note that including residuals to more complex models (as outlined here) could introduce a new set of issues, and lead to biased parameter estimation, a point that has been made elsewhere (Freckleton, 2002). Caution is needed. Other approaches could also potentially be used, possibly even data transformations, but thought is also required when following this route (e.g., Houle et al., 2011).

Two other cautions emerge from this data set, one related to measurement error and one to trait scaling. For a number of species, very few sperm (≤ 10) are measured per male to estimate means and variances. This may be reasonable to obtain an average sperm size, but to obtain more precise estimates of variation, larger sample sizes are required (we suggest >30). Additionally, the repeatability of sperm size measures need to be assessed. While for some taxa means and variances seem to be consistent across measurers, studies, and time (e.g., *Scathophaga stercoraria*: Ward & Hauschteck-Jungen (1993); Dobler & Hosken (2010)), assessment of measurement error is needed in studies of sperm variation. Finally, many have noted that differences in trait scaling can generate differences in CVs that are independent of the natural dispersion of the data, even though interest is usually focused on data scatter around the mean (e.g., Eberhard et al., 1998; Fitzpatrick & Baer, 2011; Hosken et al., 2005; Hosken, 2008). That is, CV can vary because of dispersion or scaling and these may reflect very different underlying biological phenomena. For example, a greater CV in sperm size may be detected in taxon A compared to taxon B either because of the underlying scaling exponent of sperm size to testis size, for example (and e.g., sperm size scales with testis size in *Drosophila*: Pitnick, 1996), or because the scatter of data around the mean is different, but this fundamental difference in causation is typically ignored.

In summary, we have found wide variation in sperm-length mean–variance scaling, with no evidence that sperm length standard deviation generally scales linearly with mean length. This renders the most common way to standardize variation to mean sperm size, the CV, dangerous and we do not recommend its use unless accompanied by careful testing of the validity of the assumption of linear scaling of variation (SD) to mean. Why sperm show such variable mean–variance relationships remains to be established, adding another question to those that need addressing if we are to fully understand sperm length variation.

Supplementary material

Supplementary material is available at *Journal of Evolutionary Biology* online.

Data availability

Data have been uploaded to Dryad along with R-code used in the analysis. See https://doi.org/10.5061/dryad.r4xgxd2s6

Author contributions

David J. Hosken (Conceptualization [equal], Investigation [equal], Writing – original draft [equal], Writing – review & editing [equal]), John L. Fitzpatrick (Conceptualization [equal], Investigation [equal], Writing – original draft [equal],

Writing – review & editing [equal]), Tommaso Pizzari (Conceptualization [equal], Writing – review & editing [equal]), and Dave J. Hodgson (Conceptualization [equal], Data curation [equal], Formal Analysis [equal], Writing – original draft [equal], Writing – review & editing [equal])

Funding

None declared.

Acknowledgments

We thank Boris Baer for providing data on Hymenopteran sperm, Jonathan Evans and Paco Garcia-Gonzalez for their assistance in collecting the sea urchin data used in these analyses, and Aria Minder for collecting all the dungfly data. We also thank the referees and Editor for their careful reviews which greatly helped us clarify the manuscript. We finally thank everyone at the BoS meetings over the years for helping us understand the issues we highlight here, and we dedicate the paper to our great friend Matt Gage.

Conflict of interest

None declared.

References

Birkhead, T. R., Pellattt, E. J., Brekke, P., & Castillo-Juarez, H. (2005). Genetic effects on sperm design in the zebra finch. *Nature*, *434*, 383–387. https://doi.org/10.1038/nature03374

Calhim, S., Immler, S., & Birkhead, T. R. (2007). Postcopulatory sexual selection is associated with reduced variation in sperm morphology. *PLoS One*, *2*, e413. https://doi.org/10.1371/journal.pone.0000413

Cohen, J. (1967). Correlations between chiasma frequency and sperm redundancy. *Nature*, *215*, 862–863. https://doi.org/10.1038/215862a0

Dobler, R., & Hosken, D. J. (2010). Response to selection and realized heritability of sperm length in the yellow dung fly (*Scathophaga stercoraria*). *Heredity*, *104*, 61–66. https://doi.org/10.1038/hdy.2009.93

Durrant, K. L., Dawson, D. A., Burke, T., & Birkhead, T. R. (2010). The unusual sperm morphology of the Eurasian Bullfinch (*Pyrrhula pyrrhula*) is not due to the phenotypic result of genetic reduction. *The Auk*, *127*, 832–840. https://doi.org/10.1525/auk.2010.10070

Eberhard, W. G., Huber, B. A., Rodriguez, R. L., & Rodriguez, V. (1998). One size fits all? Relationships between the size and degree of variation in genitalia and other body parts in twenty species of insects and spiders. *Evolution*, *52*, 415–431. https://doi.org/10.2307/2411078

Fitzpatrick, J. L, Garcia-Gonzales, F., & Evans, J. P. (2010). Linking sperm length and velocity: the importance of intra-male variation. *Biology Letters*, *6*, 797–799. https://doi.org/10.1098/rsbl.2010.0231

Fitzpatrick, J. L., & Baer, B. (2011). Polyandry reduces sperm length variation in social insects. *Evolution*, *65*, 3006–3012. https://doi.org/10.1111/j.1558-5646.2011.01343.x

Freckleton, R. P. (2002). On the misuse of residuals in ecology: regression of residuals vs. multiple regression. *Journal of Animal Ecology*, *71*, 542–545. https://doi.org/10.1046/j.1365-2656.2002.00618.x

Gage, M. J. G. (1998). Mammalian sperm morphometry. *Proceedings of the Royal Society of London*, *265*, 97–103. https://doi.org/10.1098/rspb.1998.0269

Holt, W. V., & van Look, K. J. W. (2004). Concept in sperm heterogeneity, sperm selection and sperm competition as biological foundations for laboratory tests of semen quality. *Reproduction*, *127*, 527–535. https://doi.org/10.1530/rep.1.00134

Hosken, D. J. (2008). Clitoral variation says nothing about female orgasm. *Evolution & Development*, *10*, 393–395. https://doi.org/10.1111/j.1525-142X.2008.00247.x

Hosken, D. J., Garner, T. W. J., & Blanckenhorn, W. U. (2003). Asymmetry, testis and sperm size in yellow dung flies. *Functional Ecology*, *17*, 231–236. https://doi.org/10.1046/j.1365-2435.2003.00723.x

Hosken, D. J., Minder, A. M., & Ward, P. I. (2005). Male genital allometry in Scathophagidae (Diptera). *Evolutionary Ecology*, *19*, 501–515. https://doi.org/10.1007/s10682-005-1023-z

Houle, D., Pélabon, C., Wagner, G. P., & Hansen, T. F. (2011). Measurement and meaning in biology. *The Quarterly Review of Biology*, *86*, 3–34. https://doi.org/10.1086/658408

Hunter, F. M., & Birkhead, T. R. (2002). Sperm viability and sperm competition in insects. *Current Biology*, *12*, 121–123. https://doi.org/10.1016/S0960-9822(01)00647-9

Immler, S., Calhim, S., & Birkhead, T. R. (2008). Increased postcopulatory sexual selection reduces the intramale variation in sperm design. *Evolution*, *62*, 1538–1543. https://doi.org/10.1111/j.1558-5646.2008.00393.x

Joly, D., Cariou, M. -L., Lachaise, D., & David, J. R. (1989). Variation of sperm length and heteromorphism in Drosophilid species. *Genetics Selection Evolution*, *21*, 283–293. https://doi.org/10.1186/1297-9686-21-3-283

Kahrl, A. F., Snook, R. R., & Fitzpatrick, J. L. (2021). Fertilization mode drives sperm length evolution across the animal tree of life. *Nature Ecology & Evolution*, *5*, 1153–1164. https://doi.org/10.1038/s41559-021-01488-y

Kleven, O., Laskemoen, T., Fossoy, F., & Lifjeld, J. T. (2008). Intraspecific variation in sperm length is negatively related to sperm competition in passerine birds. *Evolution*, *62*, 494–499. https://doi.org/10.1111/j.1558-5646.2007.00287.x

Lande, R. (1977). On comparing coefficients of variation. *Systematic Zoology*, *26*, 214–217. https://doi.org/10.2307/2412845

Lewontin, R. C. (1966). On the measurement of relative variability. *Systematic Zoology*, *15*, 141–142. https://doi.org/10.2307/2411632

Malo, A. F., Gomendio, M., Garde, J., & Roldan, E. R. S. (2006). Sperm design and sperm function. *Biology Letters*, *2*, 246–249.

Minder, A. (2002). *Coevolution of sperm, testis and the female reproductive tract morphology in Scathophagidae. Diploma thesis*, University of Zurich.

Morrow, E. H., & Gage, M. J. G. (2001). Consistent significant variation between individuals in spermatozoal morphometry. *Journal of Zoology*, *254*, 147–153. https://doi.org/10.1017/S0952836901000656

Muchlisin, Z. A. (2004). Ultra structure of baung (*Mystus nemurus*) spermatozoa. *Berkala Penelitian Hayati*, *9*, 75–77.

Parker, G. A. (1984). Sperm competition and the evolution of animal mating strategies. In R.L. Smith (Ed.), *Sperm competition and the evolution of animal mating systems*(pp. 2–61). Academic Press, London.

Parker, G. A. (1993). Sperm competition games: sperm size and number under adult control. *Proceedings of the Royal Society B*, *253*, 245–254.

Parker, G. A., & Begon, M. E. (1993). Sperm competition games: sperm size and number under gametic control. *Proceedings of the Royal Society B*, *253*, 255–262.

Pélabon, C., Hilde, C. H., Einum, S., & Gamelon, M. (2020). On the use of the coefficient of variation to quantify and compare trait variation. *Evolution Letters*, *4*, 180–188.

Pitnick, S. (1996). Investment in testes and the cost of making long sperm in Drosophila. *The American Naturalist*, *148*, 57–80. https://doi.org/10.1086/285911

Pitnick, S., Hosken, D. J., & Birkhead, T. R. (2008). Sperm morphological diversity. In T.R. Birkhead, D.J. Hosken, & S. Pitnick (Eds.), *Sperm biology: an evolutionary perspective*(pp. 69–149). Academic Press, London.

Pitnick, S., Spicer, G. S., & Markow, T. A. (1995). How long is a giant sperm. *Nature*, *375*, 109. https://doi.org/10.1038/375109a0

Pizzari, T., & Foster, K. R. (2008). Sperm sociality: cooperation, altruism and spite. *PLoS Biology*, *6*, e130. https://doi.org/10.1371/journal.pbio.0060130

Sharma, M. D., Minder, A. M., & Hosken, D. J. (2013). No association between sperm competition and sperm length variation across dung flies (Scathophaga). *Journal of Evolutionary Biology*, *26*, 2341–2349. https://doi.org/10.1111/jeb.12232

Simpson, J. L., Humphries, S., Evans, J. P., & Fitzpatrick, J. L. (2014). Relationships between sperm length and speed differ among three internally and three externally fertilizing species. *Evolution*, *68*, 92–104. https://doi.org/10.1111/evo.12199

Sokal, R. R., & Rohlf, J. (1981). *Biometry: the principles and practice of statistics in biological research*. WH Freeman and Co, London.

Tomkins, J. L., & Simmons, L. W. (2002). Measuring relative investment: a case study of testes investment in species with alternative male reproductive tactics. *Animal Behaviour*, *63*, 1009–1016. https://doi.org/10.1006/anbe.2001.1994

Van Valen, L. (2005). The statistics of variation. In B. Hallgrimsson, & B.K. Hall (Eds.), *Variation: a central concept in biology*(pp. 29–48). Elsevier Academic Press, Burlington, MA.

Ward, P. I., & Hauschteck-Jungen, E. (1993). Variation in sperm length in the yellow dungfly *Scathophaga stercoraria* (L.). *Journal of Insect Physiology*, *39*, 545–547. https://doi.org/10.1016/0022-1910(93)90035-P

Received March 17, 2025; revised June 11, 2025; accepted July 7, 2025