Individual phenotypic plasticity explains seasonal variation in sperm morphology in a passerine bird

Tim Schmoll¹,², Oddmund Kleven³ and Maria Rusche¹

¹Evolutionary Biology, Bielefeld University, Bielefeld, Germany, ²Institute of Evolutionary Ecology and Conservation Genomics, Ulm University, Ulm, Germany and ³Norwegian Institute for Nature Research (NINA), Trondheim, Norway

ABSTRACT

Background: Spermatozoa display marked variation in size and form among and within animal species. In birds, comparative evidence suggests that post-copulatory sexual selection resulting from extra-pair copulations is a major driver of interspecific variation in sperm traits. However, little is known about the extent, determinants, and dynamics of intraspecific variation in avian sperm traits.

Goal: Characterize and analyse variation in sperm morphology within and among two natural populations of great tits, Parus major, a socially monogamous passerine with frequent extra-pair matings.

Methods: We studied both a German and a Norwegian population of P. major. In the German population, we sampled spermatozoa during both the first clutch egg-laying and the nestling period (partly from the same individual males). In the Norwegian population, we sampled spermatozoa during the pre-laying/egg-laying period. We determined the overall size of spermatozoa as well as making separate measurements of sperm head, midpiece, and tail length.

Results: In the German population, spermatozoa were significantly shorter during the nestling period than during the egg-laying period. Individual phenotypic plasticity was responsible for the seasonal dynamics in sperm morphology. Changes in flagellum length (sum of midpiece and tail length) rather than changes in head length accounted for the change observed in total length. We found that changes in flagellum length were attributable to both midpiece and, in particular, tail shortening. Consequently, the ratio ‘midpiece/total length’ increased over the breeding cycle. Controlling statistically for seasonal variation, sperm total length was significantly repeatable across sperm samples from the same males. Furthermore, spermatozoa sampled in a Norwegian population early in the season differed from those obtained from the German population during egg-laying, but not from those obtained from the German population during the nestling period.

Conclusions: Individual phenotypic plasticity across the breeding season may contribute to intraspecific variation in avian sperm morphology. Our comparison across populations illustrates that seasonal variation in sperm dimensions within populations may confound between-population comparisons unless sampling date in relation to reproductive phenology is controlled for.
Keywords: great tit, individual phenotypic plasticity, *Parus major*, population divergence, repeatability, seasonality, sperm competition, sperm morphology.

INTRODUCTION

Spermatozoa share a function across all animal taxa: that of fertilizing eggs. Despite this common task, spermatozoa display enormous variation in size, form, and motility at all taxonomic levels (Pitnick et al., 2009). Studies of comparative and experimental evolution in birds, fishes, insects, and mammals suggest that post-copulatory sexual selection constitutes a powerful evolutionary force contributing to this variation (e.g. Fitzpatrick et al., 2009; Pitnick et al., 2009; Tourmente et al., 2011; Higginson et al., 2012; Rowe et al., 2015; Godwin et al., 2017).

In socially monogamous bird species, female extra-pair mating behaviour is widespread (Griffith et al., 2002; Westneat and Stewart, 2003), increases the opportunity for sexual selection (e.g. Webster et al., 1995; Vedder et al., 2011), and represents a major source of post-copulatory sexual selection. For example, both sperm length and sperm velocity have been shown to be positively correlated with the frequency of extra-pair paternity across passerine birds (Briskie et al., 1997; Kleven et al., 2009). Thus progress has recently been made in understanding the selective pressures in the evolutionary past that have contributed to the marked interspecific variation of avian sperm traits that we see today (Jamieson, 2007).

In a within-species context, sperm trait variation in general and in birds in particular has received much less attention. This is in marked contrast to other reproductive traits, such as secondary sexual characters, although analysing the ecological and evolutionary dynamics of intraspecific sperm trait variation in natural populations is of major importance for understanding the adaptive function and microevolution of sperm traits under post-copulatory sexual selection. Few studies, for example, have addressed such fundamental topics as geographical (e.g. Lüpold et al., 2011; Schmoll and Kleven, 2011; Lifjeld et al., 2012; Hogner et al., 2013; Laskemoen et al., 2013a; Stenstad et al., 2016) or seasonal (Lüpold et al., 2012; Cramer et al., 2013; Laskemoen et al., 2013a; Edme et al., 2018) variation in avian sperm morphology. Furthermore, the potential of individual phenotypic plasticity to contribute to intraspecific variation in sperm traits has been largely neglected and evidence for gamete plasticity is scarce in general (Marshall, 2015). In birds, less than a handful of studies have established individual phenotypic plasticity in sperm morphology in relation to, for example, sperm competition risk (Immler et al., 2010), season and harem size (Lüpold et al., 2012), or social dominance rank (Rojas Mora et al., 2017).

Here we analyse variation in sperm morphology within and among two natural populations of great tits, *Parus major*, a socially monogamous passerine bird with frequent extra-pair mating behaviour (see overview in table 1 of Lubjuhn et al., 2007). We determined the overall size of spermatozoa as well as making separate measurements of sperm head, midpiece, and tail length. Furthermore, we include in our analysis two sperm proportions that either showed seasonal variation in other species – flagellum/head length ratio (Lüpold et al., 2012; Cramer et al., 2013) – or predicted competitive fertilization success in a passerine bird species – midpiece/total length ratio (Laskemoen et al., 2010). We demonstrate individual phenotypic plasticity (*sensu* Nussey et al., 2007) in sperm morphology in relation to the reproductive cycle and discuss possible non-adaptive and adaptive explanations for this in a sperm competition context and its implications for comparisons of sperm traits across populations.
METHODS

Study populations and field methods

We sampled sperm of territorial male great tits between 17 April and 20 May 2010 in northwest Germany (near Lingen/Ems, 52°27′N, 7°15′E) during egg-laying of the focal males’ social females (N = 20) and later during the nestling feeding period (N = 24); and between 17 and 30 April 2010 in southern Norway (near Kråkstad, 59°41′N, 10°55′E) during the pre-laying and/or egg-laying period (N = 10; no detailed information on reproductive phenology is available, but egg-laying in a nearby nest box area began on 25 April). We sampled ten males from the German population during both periods, resulting in a total of 54 sperm samples from 44 males. Bird were caught with mist-nets or by hand and sampled non-invasively by cloacal massage (Wolfson, 1952; Laskemoen et al., 2013b). This typically took between just a few and approximately 30 seconds with birds showing no visible signs of stress. Sperm samples were diluted and mixed well in approximately 3 µL standard phosphate buffered saline and immediately transferred into 250 µL of an approximately 5% formaldehyde solution (equivalent to an approximately 12.5% formalin solution assuming a stock solution of 40% formaldehyde). Samples were stored at room temperature until sperm morphometric analysis in the laboratory [differential storage duration has been shown not to affect avian sperm length (Schnoll et al., 2016)].

Sperm morphometry

Sperm morphometric analysis was conducted by a single observer (M.R.). A droplet of approximately 3.6 µL sperm solution was transferred onto a microscope slide, covered with a coverslip, and examined immediately by digital light microscopy at 400× magnification under bright-field conditions using an Olympus BX50 microscope. A micrometre scale was pictured for each sperm sample before slides were screened for intact spermatozoa with no obviously abnormal morphology. Pictures of approximately 30 spermatozoa with clearly addressable sperm sections (head, midpiece, tail) were taken per sample with a Canon PowerShot A95 AiAF digital camera, of which 20 were selected for further analysis. M.R. measured the sperm head, midpiece, and tail length of 19.8 ± 0.5 (mean ± SD) spermatozoa per sample to a precision of 0.01 µm during a continuous measuring period using ImageJ 1.43 (Rasband, 1997–2012). We calculated sperm total length as the sum of these components and flagellum length as the sum of midpiece and tail length. To enforce blind measurement with respect to sperm sample identity, T.S. anonymized all samples before analysis, including an additional make-believe sample containing an identical number of pictures, but composed of spermatozoa from three different males to be measured twice to determine measurement error via repeatability analysis. Repeatability of measurements was ≥ 0.91 for all components (all F_{19,20} ≥ 23.7, all P < 0.001; see Appendix Table A1 at evolutionary-ecology.com/data/3131Appendix.pdf).

Statistical analysis

We used R v.3.1.1 (R Development Core Team, 2014) and linear mixed effects models (LME) to test for seasonal effects and for differences between populations in sperm traits. We included sperm sample identity (for German samples only) and male identity as nested random
effects to account for the non-independence of measurements obtained from the same sperm sample and the same individual and to estimate between-male variance in sperm total length. We determined the significance of fixed effects by removing the focal term from a maximum likelihood (ML) fit of the model and the significance of random intercept effects by removing the focal term from a restricted maximum likelihood (REML) re-fit of our model. Thus \( P \)-values in the context of LME analysis refer to the increase in model deviance (compared against a \( \chi^2 \) distribution) when a term is removed from the model. We used the R package ‘rptR’ (Stoffel et al., 2017) to calculate between-sperm-sample repeatability of sperm total length with 95% confidence intervals (parametric bootstrapping, \( N = 10,000 \)) based on mixed effects models. We report both repeatability of sperm length based on measurements of individual spermatozoa and repeatability of mean sperm length per sperm sample (the latter to allow direct comparisons with other studies).

For the German samples, we used sperm sampling date relative to the day the first egg was laid during the first brood period by a focal male’s social female (hereafter laying date) to analyse seasonal effects on sperm traits in regard to individual reproductive phenology. We used within-subject centring of this covariate in additional regression models to tease apart within- and between-individual effects (van de Pol and Wright, 2009). As detailed data on reproductive phenology were not available for the Norwegian population (see above), we used a three-level factorial predictor variable (followed by Tukey post hoc tests) to compare sperm traits between the Norwegian population, egg-laying in the German population, and the nestling feeding period in the German population.

RESULTS
In the German population, sperm total length decreased with increasing time interval between the first brood laying date of a focal male’s social female and sperm sampling date (LME with sperm sample identity and male identity as random effects: \( \chi^2 = 14.9, df = 1, P < 0.001, \) slope ± SE: \(-0.09 \pm 0.02; \) see Fig. 1a). Slopes of the within- \((-0.11 \pm 0.02)\) and between-individual \((-0.07 \pm 0.03)\) effect were statistically indistinguishable \((\chi^2 = 0.82, df = 1, P = 0.37)\) and very similar to the population-level slope (see above), indicating that longitudinal changes within individuals (i.e. individual phenotypic plasticity) and not selective (dis-) appearance of individuals with particularly short or long spermatozoa produced the seasonal dynamics in sperm total length. This was confirmed by restricting the population-level analysis to ten males with paired samples during both the egg-laying and the nestling period \((\chi^2 = 10.9, df = 1, P = 0.001, \) slope ± SE: \(-0.10 \pm 0.02; \) data points linked by lines in Fig. 1a).

Based on our data set with multiply-sampled males and controlling for the fixed effect of time of the season and the random effect of sperm sample identity, we found a significant \((\chi^2 = 5.67, df = 1, P = 0.009)\) repeatability of 0.31 (95% CI: 0.04–0.54) of sperm total length within males. Furthermore, controlling for the effect of time of the season, we found a significant \((\chi^2 = 6.29, df = 1, P = 0.006)\) repeatability of 0.68 (95% CI: 0.18–0.92) of mean sperm total length per sperm sample within males.

Analyses of sperm components revealed that changes in flagellum length (with contributions of both midpiece but particularly tail), rather than changes in head length, explained the patterns observed for sperm total length (Table 1; see also Appendix Fig. A1). As a consequence, the flagellum/head length ratio decreased (Table 1, Fig. 1b) and the midpiece/total length ratio tended to increase over the breeding cycle (Table 1, Fig. 1c).
Table 1. Test statistics for seasonal variation in sperm morphology in a German population of great tits: (a) models using continuous predictor variables (sperm sampling date relative to the day the first egg was laid by a focal male’s social female) and (b) models with two-level factorial predictor variables reflecting the two distinct sampling periods (including paired t-tests based on means per sperm sample)

<table>
<thead>
<tr>
<th>Trait</th>
<th>Data set</th>
<th>(a) Models with continuous predictor variables</th>
<th>(b) Models with factorial predictor variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>LME</td>
<td>Within-/between-subject effects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\chi^2$</td>
<td>df</td>
</tr>
<tr>
<td>Total length</td>
<td>Full</td>
<td>14.90</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Paired</td>
<td>10.85</td>
<td>1</td>
</tr>
<tr>
<td>Head length</td>
<td>Full</td>
<td>0.03</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Paired</td>
<td>2.01</td>
<td>1</td>
</tr>
<tr>
<td>Flagellum length</td>
<td>Full</td>
<td>19.42</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Paired</td>
<td>12.91</td>
<td>1</td>
</tr>
<tr>
<td>Midpiece length</td>
<td>Full</td>
<td>5.05</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Paired</td>
<td>3.94</td>
<td>1</td>
</tr>
<tr>
<td>Tail length</td>
<td>Full</td>
<td>17.58</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Paired</td>
<td>14.91</td>
<td>1</td>
</tr>
<tr>
<td>Flagellum/head length</td>
<td>Full</td>
<td>4.71</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Paired</td>
<td>1.05</td>
<td>1</td>
</tr>
<tr>
<td>Midpiece/total length</td>
<td>Full</td>
<td>7.23</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Paired</td>
<td>7.54</td>
<td>1</td>
</tr>
</tbody>
</table>

**Note:** Figures are given for the full data set (Full) and for a subset of males sampled twice (Paired). Linear mixed effects (LME) models were used with sperm sample identity nested in male identity as random effects to account for the non-independence of measurements obtained from the same sperm sample and the same individual. Within-between-subject effects in (a) refers to tests for a difference between such effects in LME models which contained two separate predictor variables for both the effects following within-individual centring of the covariate according to van de Pol and Wright (2009). Flagellum length is the sum of sperm midpiece and tail length.
Figure (a) shows the mean sperm total length ± SE (µm) over different sampling dates (days after first egg date). Figure (b) presents the mean flagellum/head length ± SE over similar sampling dates.
Replacing relative sampling date by a two-level factorial predictor reflecting the two distinct sampling periods led to very similar results and identical conclusions (Table 1b; see also Table 2).

Sperm total length differed between German samples from the egg-laying period, German samples from the nestling period, and Norwegian samples ($\chi^2 = 18.4$, df = 2, $P < 0.001$). Tukey post hoc analysis revealed that the Norwegian samples differed significantly only from the German samples obtained during egg-laying ($z = 3.52$, $P = 0.001$), but not from those obtained during the nestling period ($z = 0.57$, $P = 0.83$; Fig. 2, see also Table 2).

**DISCUSSION**

Across animal taxa, evidence for gamete plasticity is scarce in general (see compilation of relevant work in table 1 in Marshall, 2015). Examples include age-dependent effects on sperm size in insects (Green, 2003) and fishes (Gasparini et al., 2010; Mehlis and Bakker, 2013), and temperature-dependent effects on sperm size in insects (Blanckenhorn and Hellriegel, 2002), fishes (Adriaenssens et al., 2012), and molluscs (Minoretti et al., 2013). Only four studies, however – all using bird model systems – were able to demonstrate seasonal variation in sperm morphology (Lüpold et al., 2012; Cramer et al., 2013;
Table 2. Sperm morphometrics for spermatozoa sampled in a German population of the great tit during the egg-laying period of the males’ social females (Germany Laying, $N = 20$), the respective nestling feeding period (Germany Feeding, $N = 24$), and a Norwegian population (Norway, $N = 10$)

<table>
<thead>
<tr>
<th>Trait</th>
<th>Level of analysis</th>
<th>Germany Laying Mean ± SD</th>
<th>Range</th>
<th>Germany Feeding Mean ± SD</th>
<th>Range</th>
<th>Norway Mean ± SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total length (µm)</td>
<td>Population</td>
<td>99.69 ± 4.24</td>
<td>89.01–115.90</td>
<td>97.26 ± 3.57</td>
<td>83.90–107.60</td>
<td>96.80 ± 3.07</td>
<td>81.55–106.50</td>
</tr>
<tr>
<td></td>
<td>Sample means</td>
<td>99.71 ± 3.09</td>
<td>96.43–105.60</td>
<td>97.25 ± 1.98</td>
<td>93.21–100.40</td>
<td>96.81 ± 1.29</td>
<td>95.09–98.31</td>
</tr>
<tr>
<td>Head length (µm)</td>
<td>Population</td>
<td>14.38 ± 0.90</td>
<td>11.43–19.05</td>
<td>14.42 ± 0.89</td>
<td>12.14–17.36</td>
<td>14.44 ± 1.09</td>
<td>11.91–17.44</td>
</tr>
<tr>
<td></td>
<td>Sample means</td>
<td>14.38 ± 0.60</td>
<td>13.19–15.54</td>
<td>14.42 ± 0.63</td>
<td>13.37–15.68</td>
<td>14.44 ± 0.87</td>
<td>12.92–16.33</td>
</tr>
<tr>
<td>Flagellum length (µm)</td>
<td>Population</td>
<td>85.31 ± 3.91</td>
<td>73.03–96.96</td>
<td>82.83 ± 3.38</td>
<td>69.67–92.30</td>
<td>82.36 ± 2.95</td>
<td>69.14–90.69</td>
</tr>
<tr>
<td></td>
<td>Sample means</td>
<td>85.32 ± 2.83</td>
<td>82.52–91.23</td>
<td>82.83 ± 1.89</td>
<td>78.90–85.90</td>
<td>82.37 ± 1.20</td>
<td>80.94–84.23</td>
</tr>
<tr>
<td>Midpiece length (µm)</td>
<td>Population</td>
<td>59.54 ± 3.30</td>
<td>45.08–69.15</td>
<td>58.63 ± 3.34</td>
<td>46.70–66.87</td>
<td>56.85 ± 4.77</td>
<td>38.45–67.43</td>
</tr>
<tr>
<td></td>
<td>Sample means</td>
<td>59.57 ± 2.07</td>
<td>56.85–64.52</td>
<td>58.62 ± 1.89</td>
<td>54.60–61.45</td>
<td>56.81 ± 3.30</td>
<td>50.26–61.16</td>
</tr>
<tr>
<td></td>
<td>Sample means</td>
<td>25.75 ± 2.16</td>
<td>21.09–30.01</td>
<td>24.21 ± 1.87</td>
<td>21.53–28.27</td>
<td>25.54 ± 3.73</td>
<td>19.97–32.05</td>
</tr>
<tr>
<td>Flagellum/head length</td>
<td>Population</td>
<td>5.94 ± 0.40</td>
<td>4.54–7.19</td>
<td>5.76 ± 0.41</td>
<td>4.66–6.97</td>
<td>5.74 ± 0.49</td>
<td>4.56–7.14</td>
</tr>
<tr>
<td></td>
<td>Sample means</td>
<td>5.95 ± 0.26</td>
<td>5.57–6.47</td>
<td>5.75 ± 0.29</td>
<td>5.25–6.29</td>
<td>5.74 ± 0.37</td>
<td>4.98–6.40</td>
</tr>
<tr>
<td>Midpiece/total length</td>
<td>Population</td>
<td>0.60 ± 0.03</td>
<td>0.46–0.68</td>
<td>0.60 ± 0.03</td>
<td>0.47–0.70</td>
<td>0.59 ± 0.04</td>
<td>0.42–0.68</td>
</tr>
<tr>
<td></td>
<td>Sample means</td>
<td>0.60 ± 0.02</td>
<td>0.57–0.64</td>
<td>0.60 ± 0.02</td>
<td>0.57–0.63</td>
<td>0.59 ± 0.03</td>
<td>0.53–0.63</td>
</tr>
</tbody>
</table>

Note: Descriptive statistics are given for population-wide estimates (Population, i.e. not accounting for sperm sample identity) as well as based on mean values per sperm sample (Sample means). Flagellum length is the sum of sperm midpiece and tail length.
Laskemoen et al., 2013a; Edme et al., 2018). In contrast to our results in great tits, Lüpold et al. (2012) found absolute and relative flagellum length in red-winged blackbirds Agelaius phoeniceus to increase within individual males across the breeding season, and Cramer et al. (2013) likewise reported a seasonal increase in the flagellum/head length ratio in house wrens Troglodytes aedon (note, however, that the latter study could not distinguish effects within individuals from selective (dis-) appearance). Laskemoen et al. (2013a) found no evidence for seasonal variation in sperm morphology in barn swallows Hirundo rustica. Finally, Edme et al. (2018) demonstrated an increase in sperm total length over the breeding season in the collared flycatcher Ficedula albicollis. Thus our study is one of the first to demonstrate that individual phenotypic plasticity underlies seasonal variation in sperm morphology in any taxon. However, more studies are required to address why responses appear to differ by species.

One possible non-adaptive explanation for the observed seasonal decrease in sperm length in our study could be that males have trimmed back investment in reproductive organs during the first brood nestling period, possibly mediated by corresponding seasonal trajectories of hormonal profiles (e.g. Wingfield and Farner, 1978; Puxten et al., 2007; Kempenaers et al., 2008). In this case, shrinking reproductive tissues may constrain sperm design and result in the production of shorter spermatozoa (see Lüpold et al., 2009). The seasonal decrease in sperm total length was mainly due to a decrease in tail length and it is conceivable that this morphologically least complex part of the sperm cell, which is produced last during spermatogenesis in the elongation phase, is affected most by shrinkage of the testes and their seminiferous tubules. However, when we were collecting sperm samples of great tits as well as coal tits Periparus ater and blue tits Cyanistes caeruleus during the nestling period of the first brood in our study area, males demonstrated a well-developed cloacal protuberance, and provided large experimental ejaculates with highly motile sperm (see the Appendix for representative sample videos of three great tit males included in our study; note that video equipment was available only during the nestling period of the German population, which is why sperm motility could not be included in analyses). In the year of our study, 64% of

Fig. 2. Tukey’s Honestly Significant Differences (±95% family-wise confidence intervals, CI) between mean sperm total length of spermatozoa sampled in a German great tit population during the egg-laying period of the males’ social females (Germany Laying, N = 20), the respective nestling feeding period (Germany Feeding, N = 24), and a Norwegian population (Norway, N = 10). Results from Tukey post hoc tests for a linear mixed effects model including sperm sample identity nested in male identity as random effects are shown.
22 identified females whose males were sperm-sampled during the first brood nesting period were found initiating second clutches. Laying dates for these clutches (calculated back from either incomplete clutches or a combination of clutch size and hatching date while assuming a 12-day incubation period) were separated on average (± SD) 10.9 ± 4.1 days (range: 3–17) from the date of sperm sampling. Given the potential of extended sperm storage in female passerines (e.g. Birkhead and Møller, 1992) and probable male uncertainty over when exactly social and potential extra-pair mating partners initiate second clutches and become fertile again, we expect the sperm phenotypes observed during the first brood nesting period to be fully functional and well comparable to those sampled earlier (see also below). 

The observed seasonal dynamics in sperm size could potentially also represent an adaptive response to changes in the level of sperm competition (Crean and Marshall, 2008; Immler et al., 2010; Marshall, 2015) assuming the latter varies with season. This appears to be the case in our study population, where the probability that a nestling in a second brood was sired by an extra-pair male was four times that of first broods [binomial generalized linear mixed model, \( P < 0.001 \), our own unpublished data from the year 2012 (see also Lubjuhn et al., 2001)]. There is, however, no clear picture as to which sperm morphological traits, or trait values, promote competitive fertilization success in general (Pitnick et al., 2009) and in passerine birds in particular (cf. table 6 in Saetre et al., 2018). In passerine birds, for example, results from tree swallows *Tachycineta bicolor* showing a positive correlation between the ratio midpiece/total sperm length and competitive fertilization success (Laskemoen et al., 2010) would support an adaptive explanation, while evidence from the zebra finch *Taeniopygia guttata* demonstrating that longer – and thus faster (Mossman et al., 2009) – sperm are more successful (Bennison et al., 2015) would not. In order to conclusively test the hypothesis that individual phenotypic plasticity in avian sperm morphology represents an adaptive response to variation in levels of sperm competition, an experimental approach is needed that not only creates experimental social environments predictive of differential levels of sperm competition (Immler et al., 2010), but also includes fitness assays to probe the competitive fertilization success of different sperm phenotypes under a match/mismatch paradigm (Groothuis and Taborsky, 2015).

Conditional on the detected seasonal effects, we have documented significant between-ejaculate repeatability in both sperm total length (0.31) and mean sperm length per sample (0.68) within males (see also Birkhead and Fletcher, 1995; Morrow and Gage, 2001; Lüpold et al., 2012; Laskemoen et al., 2013). This demonstrates a high degree of within-individual consistency in sperm length despite the observed phenotypically plastic changes across the breeding season and likely reflects the substantial heritability of sperm morphological traits (Birkhead et al., 2005; Edme et al., 2018). In fact, our repeatabilities, reflecting estimates of upper bounds of expected realized heritabilities in ecologically relevant settings, align well with the only available heritability estimates from a natural bird population: 0.21 ± 0.07 (± SE) for sperm total length and 0.44 ± 0.14 (± SE) for mean sperm total length per sample in the collard flycatcher (Edme et al., 2018).

Another important consequence of the observed seasonal variation in sperm dimensions is that whether or not populations appeared to differ in sperm morphology in our study depended on which of the German sub-samples was used. Our analysis thus tellingly illustrates that seasonal variation in sperm dimensions within populations has the potential to confound between-population comparisons unless sampling date in relation to the reproductive cycle is controlled for. As the Norwegian samples were collected during the pre-laying and/or egg-laying period, we propose that both populations might indeed differ
in sperm total length and thus our study adds to the few studies that have suggested population differentiation in avian sperm traits (Lüpfeld et al., 2011; Schmoll and Kleven, 2011; Hogner et al., 2012; but see Lifjeld et al., 2012; Laskemoen et al., 2013a; Gohli et al., 2015). More generally – and applicable well beyond avian study systems and seasonality – accounting for potentially widespread environmental effects on sperm phenotype will advance our understanding of sources of sperm morphological variation in a within-species context.

DATA ACCESSIBILITY

Additional data and video streams illustrating typical patterns of sperm motility of three exemplary great tit Parus major males sampled during the first brood nestling period are available at evolutionary-ecology.com/data/3131Appendix.pdf.

AUTHOR CONTRIBUTIONS

T.S. conceived the study, participated in the design of the study, supervised and participated in fieldwork, performed the statistical analysis, and wrote the manuscript. O.K. participated in the design of the study and in fieldwork. M.R. participated in fieldwork and developed and executed the microscopy protocol. All authors proofread earlier versions of the manuscript and approved its final version.

ACKNOWLEDGEMENTS

This paper is dedicated to Theo C.M. Bakker on the occasion of his retirement as a full professor of behavioural ecology. T.S. would like to thank Theo for acting as a reviewer of his PhD thesis and habilitation thesis and for inspiring discussions during his PhD and early postdoc time in Bonn. We thank Nikolas Vellnow for field assistance and Steven Ramm, Peter Korsten, Michael L. Rosenzweig, and two anonymous reviewers for comments on earlier versions of the manuscript. This work conforms to the legal requirements of Germany and Norway and was supported by a Forschungs- und Innovationsförderung (FIF) grant of Bielefeld University’s Faculty of Biology to T.S.

REFERENCES


