Animal Behaviour 111 (2016) 33–40

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Variation in male fertility in a polymorphic moth, *Parasemia plantaginis*

Rémi Chargé [a, *], Nina Wedell [b], Carita Lindstedt [a], Liisa Hämäläinen [a], Elsi Övermark [a], Johanna Mappes [a]

[a] Centre of Excellence in Biological Interactions, Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland
[b] Centre for Ecology and Conservation, University of Exeter, Penryn, U.K.

ARTICLE INFO

Article history:
Received 18 March 2015
Initial acceptance 24 April 2015
Final acceptance 22 September 2015
Available online
MS. number: 15-00222R

Keywords:
aposematism
Erebid moths
fertility
mating success
polymorphism
spermatophore

The maintenance of multiple morphs in warning signals is enigmatic because directional selection through predator avoidance should lead to the rapid loss of such variation. Opposing natural and sexual selection is a good candidate driving the maintenance of multiple male morphs but it also includes another enigma: when warning signal efficiency differs between male morphs, why would females choose a phenotype with lower survival? We tested the hypothesis that indirect responses to selection on correlated characters through sexual selection may substantially shape the evolution of male coloration. If male phenotypes differ in their fertilization ability, female choice against the best surviving phenotype can evolve. The wood tiger moth, *Parasemia plantaginis*, has two coexisting male morphs in Europe. Previous studies have shown that yellow males are better defended against predators, but that white males have a higher mating success. We examined differences in fertility between white and yellow males in terms of sperm production, number of sperm transferred and rate of sperm replenishment, and association between these fertility traits, female mate choice and reproductive output. If white morphs have greater fertility than yellow males, then this could explain why females prefer to mate with white males. However, we did not find any difference between male colour morphs either in mating probability, fertility (i.e. sperm availability and sperm transferred) or reproductive success (i.e. number of eggs laid and hatching success). We discuss our results in relation to context-dependent mating success and maintenance of colour polymorphism within populations.

Colour polymorphism is widespread in both vertebrates and invertebrates (Roulin, 2004; Svensson & Abbott, 2005). The maintenance or coexistence of several colour morphs is of particular interest because the mechanisms involved are still a matter of debate (Bond, 2007; Roulin, 2004). Several, nonmutually exclusive, hypotheses have been proposed as drivers of colour polymorphism such as frequency-dependent selection, via for example predation, or sexual selection (Endler, 1983; Kingston, Rosenthal, & Ryan, 2003; Olendorf et al., 2006). In aposematic species, individuals display discriminable, generally conspicuous, warning signals to inform predators about their unprofitability (e.g. toxicity, unpalatability; Mappes, Marples, & Endler, 2005; Ruxton, Speed, & Sherratt, 2004). Because of the benefits for both prey (predation avoidance) and predator (avoidance of toxic prey), aposematic species are expected to show uniform conspicuous colour patterns driven by positive frequency-dependent selection (Beatty, Beirinckx, & Sherratt, 2004; Kapan, 2001; Rowland, Ihalainen, Lindström, Mappes, & Speed, 2007). Nevertheless, many aposematic species have polymorphic warning coloration. How warning colours are maintained in aposematic species may thus be an important step to understand how the general diversity of signals originates and is maintained in the wild.

Several hypotheses have been proposed to explain colour polymorphism in aposematic species such as effect of variable background on signal perception (Ruxton, Sherratt, & Speed, 2004), variation in predator tolerance or naivety towards toxicity (Endler & Mappes, 2004), and the cost of producing warning signals and toxic defences (Lindstedt et al., 2011; Speed & Ruxton, 2007). Surprisingly, very few studies have examined the role of sexual selection through female choice in the diversity of intraspecific colour patterns in aposematic species (but see Maan & Cummings, 2008; Nokelainen, Hegna, Reudler, Lindstedt, & Mappes, 2011). It has been suggested that aposematic signals, after first arising from

* Correspondence: R. Chargé, Department of Biological and Environmental Science, P.O. Box 35, FI-40014, University of Jyväskylä, Finland.
E-mail address: remi.charge@yahoo.fr (R. Chargé).

http://dx.doi.org/10.1016/j.anbehav.2015.10.014
0003-3472/

natural selection (e.g. predation avoidance), could secondarily acquire a role in sexual selection. For instance, Maan and Cummings (2008) found in the polymorphic poison frog *Oophaga pumilio* that divergence in female mate choice has resulted in aposematic colour polymorphism in geographically distinct populations. Maintenance of colour polymorphism by female mate choice becomes puzzling when females prefer the less efficient aposematic morph in males (Nokelainen et al. 2011). Intuitively, females should prefer to mate with males with efficient warning coloration, as it would ensure better protection for their offspring if coloration is heritable and enhances survival against predators. However, female preference for weaker warning signals could be selected for if the male shows higher reproductive quality such as fertility (see for instance Kaitala, 1991; Simmons, 2011). In these conditions, females could prefer to mate with males displaying the weaker warning signal if it signals a higher lifetime reproductive success.

Here we investigated mechanisms that could explain female choice for males with a weaker warning signal. As a study species, we used the aposematic wood tiger moth, *Parasemia plantaginis* (family Erebidae, subfamily Arctiine). Males of this species exhibit discrete wing coloration on both a local and a broad geographical scale (Hegna, Galarza, & Mappes, 2015). European populations feature two distinct genetic male morphs, yellow and white (Galarza, Nokelainen, Ashrafi, Hegna, & Mappes, 2014). In previous experiments, females have been shown to have a preference for males with white hind wings compared to males with yellow hind wings when males were stressed before mating trials (Nokelainen et al. 2011), and when male colour frequency was white biased (Gordon, Kokko, Rojas, Nokelainen, & Mappes, 2015). Why females show such mating preference despite yellow males providing more efficient predation avoidance is not known. According to sexual selection models, females may gain benefits related to female fecundity ('direct' benefits) and/or offspring viability arising from the transmission of paternal genes ('indirect' genetic benefits; Johnstone, 1995; Kokko, Jennions, & Brooks, 2006). Among direct benefits, females may seek to maximize the fertilization success of her eggs, particularly in species producing hundreds of eggs, as in many insects. Because male sperm production can be costly, choosy females are expected to show preferences for males in good nutritional condition, free from parasites and nonsenescent, in order to target males with high fertility (Garratt & Brooks, 2012). In addition, females may also avoid mating with sperm-depleted males (Elzinga, Chevasco, Grapputo, & Mappes, 2011; Kaitala & Wiklund, 1994; Marcotte, Delisle, & McNeil, 2007; Velde, Damiens, & Van Dyck, 2011). As many insects are promiscuous, the ability of males to recover quickly from sperm depletion following mating might hence be important to ensure fertilization success in successive copulations (Wedell & Cook, 1999).

To study possible differences in fertilization benefits between white and yellow male morphs, we first conducted an experiment (the spermatophore experiment) in which we tested whether white males transfer larger spermatophores to females than yellow males. We then conducted an experiment (the fertility experiment) to investigate in detail the fertilization benefits that may arise from mating with white males. Specifically, we investigated differences between morphs in the amount of sperm produced and transferred to the female, the time needed by males to recover from sperm transfer after mating and the reproductive output. We assumed that if more efficient warning coloration is costly for yellow males to produce, they may have a lower reproductive quality (because of conflicting resource allocation) associated with lower reproductive success than white males. In this case, the reproductive costs of warning signal production could induce conflicting selection pressures from sexual and natural selection, potentially maintaining polymorphism in aposematic warning signals in this moth species. We thus predicted that white males would have higher fertility than yellow males and that females showing mating preference for white males would secure fertility benefits.

METHODS

Study Species

Colour polymorphism in *P. plantaginis* varies throughout the world. In Finland, males present two distinct morphs on their hind wings: yellow and white (Galarza et al. 2014). Both sexes produce defensive fluids to protect themselves against predators (birds, spiders, ants etc., Lindstedt et al. 2011) and yellow adult males have been shown to be more repellent to predators than white males (Nokelainen et al. 2011). Male mating attempts do not always lead to copulation suggesting a potential for female mate choice. Copulations usually take several hours between dusk and dawn. Males provide females with both fertile eupyrene and nonfertile apyrene spermatozoa contained within a spermatophore. Females mate multiply, providing an opportunity for sperm competition (Jennions & Petrie, 2000). Females lay on average 250 eggs within 3–4 days after copulation that hatch about 7 days later (at room temperature in the laboratory: minimum 17 °C, maximum 34 °C, mean 24 °C). About 50 days later (in laboratory conditions) larvae pupate and the pupal stage lasts ca. 10 days. As adults do not feed, all the resources needed for sperm production, wing colour pigment production and other traits are gathered during the larval stage and distributed to the different traits during metamorphosis.

Stock Conditions

Individuals used in the experiments originated from the laboratory stock population held in captivity at the University of Jyväskylä, Finland. Founders originated from 50 wild-caught already mated females sampled in Finland during summer 2010 (see Lindstedt et al. 2011). Genetic diversity was maintained by supplementing laboratory stock annually with wild individuals caught in Finland during the summers of 2011 and 2012. Three generations were produced per year during spring, mid-summer and late summer, with the latter one overwintering. Generations do not overlap. The breeding was conducted as follows (at room temperature, as described above, and natural photoperiod: length of day 10–19 h). Once adults emerged, one female and one male (either yellow or white) were placed together into a transparent plastic container (size 10 × 7 cm and 7 cm high) for mating. Adults died soon after egg laying and were removed from the container. Hatched larvae were reared in separate sibling groups (up to 30 individuals per box) and fed with dandelion and lettuce leaves. Food was available ad libitum and was changed daily. At the pupal stage, individuals were weighed and individually moved into plastic tubes (4 × 7 cm) until adult emergence. Subsequently, freshly emerged adults were placed in low-temperature-controlled cabinets (ca. 10 °C) with constant darkness until used in the experiments or in matings to maintain the captive stock. The species is a capital feeder (i.e. adults do not feed). During mating trials, all pairs were checked under dim red light. Conditions of experimental animals were kept as good as possible as we aimed to test their behaviour in optimal conditions. At the end of the experiments, individuals were killed by freezing and stored at −80 °C until dissection. The following experiments comply with current laws of Finland.

Experiment 1: The Spermatophore Experiment

During the summer of 2011, 171 pairs consisting of one virgin female and one virgin male were formed from captive-born adults.

Two mating trials were conducted, on different days, with individuals from the first generation (38 and 40 pairs) and two other trials with individuals from the second generation (44 and 49 pairs). Each trial involved 50% of white and 50% of yellow males. Female age ranged from 0 to 8 days, and male age from 0 to 7 days. Pairs were transferred into individual transparent containers in the afternoon (around 1700 hours) since mating usually starts around this time. All pairs were checked regularly every 30 min to record whether a pair had mated and the mating latency (i.e. time elapsed until the copulation starts) as a proxy for female preference (Sharma, Griffin, Hollis, Tregenza, & Hosken, 2012; Taylor, Wedell, & Hosken, 2010); mating duration was also noted. Observations were recorded until the last pair had separated (in the early morning of the following day).

Immediately after the copulation, all mated females were killed by freezing and stored at −80 °C until dissection (which was carried out within 17 days). The frozen female was opened from the side to avoid any potential damage to the bursae copulatrix. All dissections were conducted under a light microscope (and in modified Barth's saline solution in the sperm experiment, Gurdon, 1968). The spermatophore was carefully removed from the bursae copulatrix and the excess of saline solution around the spermatophore was quickly dried using a paper tissue. The spermatophore was subsequently weighed to the nearest 0.01 mg (wet weight).

Experiment 2: The Fertility Experiment

During the second generation in 2012 three mating trials were conducted with 61, 65 and 75 pairs, respectively. In contrast to the first experiment, we restricted variation in individual age: with 93% of individuals 0–1 days old and the rest ranging from 2 to 3 days old. The fertility experiment was done in the same way as the spermatophore experiment, except all the pairs were checked every 15 min.

Mated females were either frozen immediately after copulation to check the content of the spermatophore or left in the plastic container to allow egg laying. Frozen females were dissected as in experiment 1, but in addition the developing eggs were counted. The spermatophore was opened and the eupyrene sperm bundles were counted under (x40) magnification following the method developed by Cook & Wedell, 1996. We assumed that each bundle in the spermatophore comprised 256 fertile eupyrene sperm cells as observed in other Lepidopteran species (Gage & Cook, 1994 in *Plodia interpunctella* and in *Plodia rapae*; Da Cruz-Landim & Fernandez-Winckler, 2008 in *Achroia grisella*). The sample was then diluted in approximately 30 ml of distilled water, weighed to calculate the exact dilution, and gently agitated to disperse sperm. Three samples of 10 µl each were plated out on slides and dried under dust covers for 24 h. The dry slides were dipped for about 5 s in distilled water to dissolve salt crystals and dried again for 24 h. The nonfertile apyrene sperm were counted under dark-field phase contrast at (×100) magnification.

We counted the sperm released into the male reproductive tract after mating as follows. At the end of each mating, males were randomly assigned to one of six treatments to examine the replenishment pattern of sperm: five to eight mated males per trial (with 50% white and 50% yellow males) were either immediately killed by placing them in a freezer or kept at room temperature until frozen at either ca. 5, 10, 30, 40 or 50 h after copulation. In addition, we froze 12 unmated males (six white and six yellow) at the end of each trial to assess the number of sperm stored in the reproductive tract of virgin males to have a baseline of the amount of sperm prior to a mating. The entire reproductive tract was dissected into a cavity slide filled with Barth's modified solution. Eupyrene and apyrene sperm counts were made following the method described above. For one male (trial 1, virgin, yellow wings), the number of apyrene sperm was very large in the three samples and the sperm count was stopped at 800 cells for convenience. This male was excluded when assessing the repeatability of the apyrene sperm count (see below).

To measure the fertility, eggs from singly mated females allowed to oviposit were counted after 4 days from the first laid egg and the number of larvae was recorded from the 7th and 14th day after the first laid egg to verify hatching. The maximum number of larvae between the two counts (i.e. after the 7th and 14th day) was used as the total number of larvae produced.

Repeatability of sperm count

Because sperm numbers are estimated, repeatability was evaluated for 31 males by counting the eupyrene bundles twice. Similarly, the repeatability of the apyrene sperm count was assessed using the same sample (each of three samples counted twice in five males) and between samples within a male ($N = 134$). Repeatability (i.e. intraclass correlation coefficient, ICC) of sperm counts was high when the same sample was counted twice ($N = 27$, ICC [confidence interval, CI] = 0.95 [0.90–0.98] for eupyrene sperm and $N = 15$, ICC [CI] = 0.98 [0.94–0.99] for apyrene sperm). The ICC for the apyrene count between samples within an individual was also relatively high ($N = 134$, ICC [CI] = 0.80 [0.75–0.85]).

Statistics

General procedure

All statistics were conducted with R software (v3.01, The R Foundation for Statistical Computing, Vienna, Austria http://www.r-project.org). Linear models were run using lm or glm functions. We visually inspected model assumptions and the appropriate error distribution was chosen accordingly, as well as by comparing the Akaike information criteria (AIC). Outliers were inspected based on the leverage plot (i.e. Cook's distance). We performed a full model, all possible submodels and an intercept model to select the models describing the data most accurately. We removed missing values from the data sets to enable accurate comparison between nested models. All submodels with an AIC delta lower than 4 compared to the best model were used to estimate coefficients averaging based on the Akaike weights (package MuMIn). The covariates included in each model are given in the tables. We tested interactions between all covariates (except trial) and colour morph. When the interaction term was kept in the 'best' submodels, we interpreted the significance level of the lower-order coefficient estimates at the mean of the covariates.

Mating latency showed a binomial distribution with no overlapping data points in the fertility experiment (more or less than 2 h in mating latency). To analyse this dependent variable, we recoded it into a discrete response variable with two levels: early mating individuals (mean ± SD: 0.51 ± 0.46 h, $N = 12$) and late mating individuals (mean ± SD: 6.88 ± 1.55 h, $N = 96$) following a binomial distribution. We further examined the difference in mating latency between white and yellow males within the late mated pairs ($N = 96$), while the model in the early mated pairs was not tested because of the low sample size ($N = 12$).

RESULTS

Mating Success

In the spermatophore experiment, 114 of 171 pairs mated, while 107 of 201 pairs mated in the fertility experiment. In both experiments, there was no difference in mating probability between white and yellow males: 67% and 66% in the spermatophore

Table 1
Difference in mating success (i.e. mated versus unmated) between male colour morphs in the spermatophore and the fertility experiments

Experiment	Trait	Estimate±SE	*z*	*P*
Spermatophore				
	Intercept	0.67±0.21	0.22	0.001
	Male mass[a]	0.01±0.01	0.97	0.332
	Male age	0.01±0.07	0.11	0.910
	Male colour (yellow)	0.01±0.33	0.03	0.975
	Male mass[a]*male colour	−0.02±0.01	1.58	0.113
	Trial[b]			
Fertility				
	Intercept	1.34±0.27	4.96	<0.001
	Male mass[a]	0.01±0.01	1.18	0.239
	Male colour (yellow)	−0.13±0.44	0.30	0.763
	Male mass[a]*male colour	−0.01±0.01	0.76	0.449
	Trial 2	0.36±0.52	0.69	0.489
	Trial 3	0.34±0.52	0.64	0.523

[a] Trait centred to the mean.
[b] Trait not present in the 'best' models (ΔAIC < 4).

experiment and 81% and 77% in the fertility experiment, respectively (Table 1). There was no difference in mating latency and mating duration between white and yellow males (Table A1, A2). Older males and larger females both showed shorter mating latency (Table A1).

Sperm Storage

The number of eupyrene and apyrene sperm stored did not differ between unmated (i.e. virgin) white and yellow males (Table 2).

Spermatophore Size and Number of Sperm Transferred

There was no difference in the ability to provide a spermatophore between white and yellow males (spermatophore experiment: 50 spermatophores from white males and 50 from yellow; fertility experiment: 21 spermatophores from white males and 18 from yellow). There was no difference between white and yellow males either in the spermatophore size or in the amount of sperm transferred (Table 3, Table A3, Fig. 1). Older and larger males provided females with larger spermatophore (Table A3), and larger females received more eupyrene sperm than smaller females (Table 3).

Sperm Replenishment Rate

When we compared how fast males replenished their sperm supply following a copulation, we found that mated males recovered sperm numbers at a similar rate to that of virgin males (24 921 ± 2 544 eupyrene sperm) after about 10 h. Recently mated males had about half the number of eupyrene sperm (12 342 ± 1 200 cells) in their reproductive tract. White males did not recover their sperm supply faster than yellow males, since the interaction recovery time*colour morph was not significant and not retained in the 'best' models (Table A4, Fig. 2).

Reproductive Output

Male colour morph at the mean male body mass did not explain the number of eggs laid by the female (Table A5) or the variance in hatching success (number of larvae out of number of eggs laid, Table A5). However, there was a significant interaction between a male's mass and colour morph suggesting that small white males have a slightly better hatching success than larger white males and an opposite relation in yellow males (Table A5). However, the result should be taken with caution, because there was no effect of colour morph either at the minimum or at the maximum male body mass (results not shown).

As experiments were not conducted in controlled chambers, there were some variation in average temperature, humidity and light among treatments, which caused the significant trial effect.

DISCUSSION

We tested the hypothesis that indirect responses to selection on correlated characters through varying fertilization success may provide an explanation for the coexistence of multiple colour morphs in the aposematic wood tiger moth. More specifically, we tested whether the reproductive costs, expressed as reduced male fertility, producing a more efficient warning colour pattern could explain why females sometimes prefer to mate with males with weak warning signals (Gordon et al., 2015; Nokelainen et al. 2011). Our results did not support this hypothesis and contradict previous findings because reproductive quality of white and yellow males did not differ in terms of sperm quality (e.g. ejaculate size, sperm replenishment rate, or number of sperm transferred) or associated reproductive success (number of eggs laid and hatching success). Although there was no clear difference in a male's fertility traits between yellow and white colour morphs, there was a trend ($P = 0.07$) for yellow males to store more sperm than white males in unmated adults. This result should be explored more deeply in future investigations in order to confirm the trend.

It is worth noting that the experimental conditions were different in the present and previous studies. In Nokelainen et al.'s (2011) study, males were forced to produce defensive fluids in the groups in which there was a bias in mating success towards white males. In Gordon et al.'s 2015 study, the two colour morphs were

Table 2
Differences in eupyrene and apyrene sperm storage between male colour morphs in the fertility experiment

Dependent variable	Trait	Estimate±SE	*z*	*P*
Eupyrene sperm storage				
	Intercept	21594.03±4077.80	5.11	<0.001
	Male mass[a]	−42.68±100.79	0.41	0.684
	Male colour (yellow)	9586.94±5057.82	1.80	0.0719
	Male mass[a]*male colour	−198.88±142.54	1.32	0.187
	Trial[b]			
Apyrene sperm storage				
	Intercept	465081±101670	4.35	<0.001
	Male mass	−296±1130	0.25	0.804
	Male colour (yellow)	55281± 72524	0.72	0.470
	Trial[b]			

[a] Trait centred to the mean.
[b] Trait not present in the 'best' models (ΔAIC < 4).

Table 3
Difference in number of eupyrene sperm transferred to the female between male colour morph

Trait	Estimate±SE	*z*	*P*
Intercept	10965.3±5637.2	1.88	0.060
Male mass	33.0±77.8	0.41	0.683
Male colour (yellow)	5052.2±3477.3	1.40	0.161
Female mass[a]	132.4±50.9	2.52	0.012
Mating duration[a]	−1403.8±827.3	1.64	0.101
Female mass[a]∗male colour	−94.2±94.6	0.96	0.337
Mating duration[a]∗male colour	1352.4±1555.7	0.84	0.403
Trial 2	−3777.6±4347.8	0.84	0.402
Trial 3	6981.1±4199.7	1.60	0.109

[a] Trait centred to the mean.

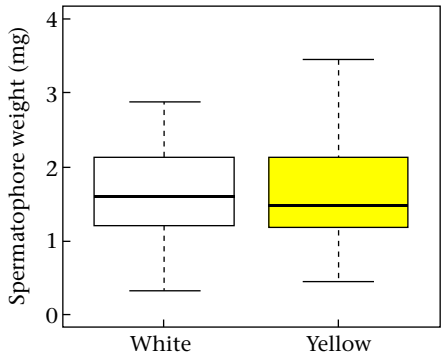

Figure 1. Spermatophore weight produced by white and yellow male (from the female preferences experiment). The line across the box indicates the median. Boxes denote the interquartile range with the 75th and 25th percentiles. The vertical whiskers represent the range of outlying data points up to 1.5 times the interquartile range.

maintained at different frequencies and a higher mating probability in white males was observed only when white males were much more numerous than yellow males. In both these experiments, the authors did not find a mating advantage for white males in the control group (i.e. either when males did not have to defend

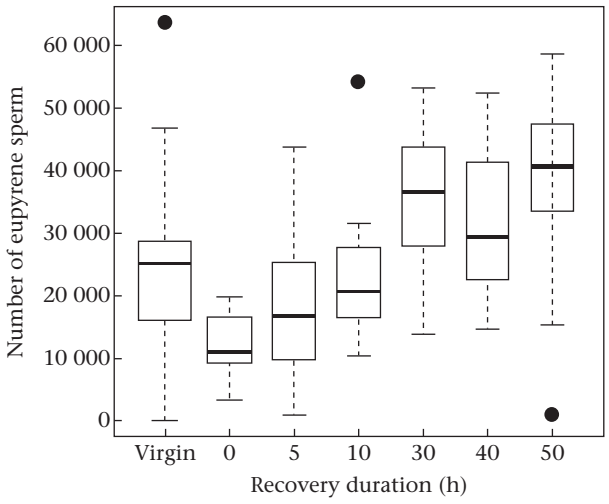

Figure 2. Postmating recovery pattern of fertile sperm: number of eupyrene sperm stored in the male reproductive tract in virgin and mated males at the end of the copulation (i.e. '0 h') and from several postcopulation recovery times. The line across the box indicates the median. Boxes denote the interquartile range with the 75th and 25th percentiles. The vertical whiskers represent the range of outlying data points up to 1.5 times the interquartile range and the black dots represent outliers.

themselves in Nokelainen et al. 2011 or when colour morphs were in the same ratio in Gordon et al., 2015). A recent study also shows that white males are more active fliers (Rojas, Gordon, & Mappes, 2015) and do not invest as much in their chemical defence as yellow males (Suisto, Lindstedt, Pakkanen, & Mappes, n.d.). Thus, the main difference between the three experimental designs is environmentally based conditions. Although speculative, it is possible that the benign conditions in our rearing regime and those in the control groups in Nokelainen et al. (2011) may reduce variation in male condition between the morphs. Variation in mating success may only become apparent when males have to face an environmental stress (see for instance Chargé, Sorci, Hingrat, Lacroix, & Saint Jalme, 2011; Mappes, Alatalo, Kotiaho, & Parri, 1996), perhaps reflecting more natural situations. Under our initial assumption, conflicting resource allocation would lead white males to invest more in reproduction whereas yellow males allocated more resources to warning coloration. In future, it would be interesting to conduct mating experiments under more natural (i.e. harsh) conditions to fully understand the complexity of female choice in this aposematic species.

We found that male body mass predicted spermatophore size, as found in other studies (LaMunyon & Eisner, 1994; Lauwers & Dyck, 2006). We also found that older males provided females with bigger spermatophores on their first mating. A positive relationship between male age and spermatophore size is expected in the Lepidoptera, assuming spermatophore size reflects sperm supply. Sperm is produced during the larval stages and released daily into the reproductive tract of adult males (Friedländer, Seth, & Reynolds, 2005; Giebultowicz, Weyda, Erbe, & Wergin, 1997). Interestingly, male age also explained variation in mating latency. Taken together, these results may indicate either active female choice towards males with a larger sperm supply in order to secure fertilization or that males with a full sperm supply are more ready to mate. Unfortunately, we cannot disentangle these two mechanisms. Similarly, larger females mated faster, or males were more willing to mate with larger females, maybe because of the higher future reproductive potential, as larger females had more developing eggs (Pearson correlation: $r = 0.84$ [IC: 0.71, 0.91], $t_{40} = 9.627$, $P < 0.001$). Female body mass was also related to sperm transfer, possibly in response to a higher risk of sperm competition in larger/more fecund females as they mate more frequently, or because of a greater reproductive potential (Cook & Wedell, 1999; Gage, 1998; Parker, Simmons, Stockley, McChristie, & Charnov, 1999; Reinhold, 2002; Svensson, Raberg, Koch, & Hasselquist, 1998).

To conclude, we did not find any physiological evidence for fertilization benefits that could explain the female preference for white males observed in previous studies (Gordon et al., 2015; Nokelainen et al. 2011). In our experiments, direct benefits received from mating with white and yellow males were equal. This was also shown in female behaviour, as females were equally willing to mate with both male morphs. Our results suggest that direct benefits gained from mating do not differ between morphs and thus cannot counterbalance the selective benefit derived from increased predator protection by yellow males, and therefore does not provide an explanation for how both morphs are maintained (see also Nokelainen, Lindstedt, & Mappes, 2013; Nokelainen, Valkonen, Lindstedt, & Mappes, 2014). However, since there is additional evidence suggesting that yellow males may pay a higher price of having a better warning signal when conditions are harsh (i.e. more natural; Gordon et al., 2015, Nokelainen et al. 2011, 2013; Suisto et al. n.d.), the next step is to test how the reproductive traits measured here differ when the quality of the environment for males is manipulated.

Acknowledgments

We thank Kaisa Suisto and Jimi Kirvesoja for the maintenance of laboratory stock and Catherine Soler for help in preparing mating containers in the sperm experiment. We are very grateful to Lorraine Bourget for valuable help in apyrene sperm count and to Andrés López-Sepulcre for his statistical expertise. We thank Sandra Varga and Swanne Gordon who gave valuable comments on the manuscript. This study was financed by the Academy of Finland (Finnish Centres of Excellence in Biological interactions research) project number SA-252411 (to J.M.) and the projects 136387 (2010-2011) and 257581 (to C.L.).

References

Beatty, C. D., Beirinckx, K., & Sherratt, T. N. (2004). The evolution of müllerian mimicry in multispecies communities. *Nature, 431*(7004), 63–66.

Bond, A. B. (2007). The evolution of color polymorphism: crypticity, searching images, and apostatic selection. *Annual Review of Ecology, Evolution, and Systematics, 38*(1), 489–514.

Chargé, R., Sorci, G., Hingrat, Y., Lacroix, F., & Saint Jalme, M. (2011). Immune-mediated change in the expression of a sexual trait predicts offspring survival in the wild. *PLoS One, 6*(10), e25305.

Cook, P. A., & Wedell, N. (1996). Ejaculate dynamics in butterflies: a strategy for maximizing fertilization success? *Proceedings of the Royal Society B: Biological Sciences, 263*(1373), 1047–1051.

Cook, P. A., & Wedell, N. (1999). Non-fertile sperm delay female remating. *Nature, 397*(6719), 486–486.

Da Cruz-Landim, C., & Fernandez-Winckler, F. (2008). Spermatogenesis chronology in *Achroia grisella* Fabricius (Lepidoptera, Pyralidae) comparing eupyrene and apyrene spermiogenesis. *Animal Biology, 58*(1), 1–12.

Elzinga, J. A., Chevasco, V., Grapputo, A., & Mappes, J. (2011). Influence of male mating history on female reproductive success among monandrous Naryciinae (Lepidoptera: Psychidae). *Ecological Entomology, 36*(2), 170–180.

Endler, J. A. (1983). Testing causal hypotheses in the study of geographical variation. In J. Felsenstein (Ed.), *Numerical taxonomy* (pp. 424–443). Berlin, Germany: Springer.

Endler, J. A., & Mappes, J. (2004). Predator mixes and the conspicuousness of aposematic signals. *American Naturalist, 163*(4), 532–547.

Friedländer, M., Seth, R. K., & Reynolds, S. E. (2005). Eupyrene and apyrene sperm: dichotomous spermatogenesis in Lepidoptera. *Advances in Insect Physiology, 32*, 206–308.

Gage, M. (1998). Influences of sex, size, and symmetry on ejaculate expenditure in a moth. *Behavioral Ecology, 9*(6), 592–597.

Gage, M. J. G., & Cook, P. A. (1994). Sperm size or numbers? Effects of nutritional stress upon eupyrene and apyrene sperm production strategies in the moth *Plodia interpunctella* (Lepidoptera: Pyralidea). *Functional Ecology, 8*(5), 594.

Galarza, J. A., Nokelainen, O., Ashrafi, R., Hegna, R. H., & Mappes, J. (2014). Temporal relationship between genetic and warning signal variation in the aposematic wood tiger moth (*Parasemia plantaginis*). *Molecular Ecology, 23*(20), 4939–4957.

Garratt, M., & Brooks, R. C. (2012). Oxidative stress and condition-dependent sexual signals: more than just seeing red. *Proceedings of the Royal Society B: Biological Sciences, 279*(1741), 3121–3130.

Giebultowicz, J. M., Weyda, F., Erbe, E. F., & Wergin, W. P. (1997). Circadian rhythm of sperm release in the gypsy moth, *Lymantria dispar*: ultrastructural study of transepithelial penetration of sperm bundles. *Journal of Insect Physiology, 43*(12), 1133–1147.

Gordon, S. P., Kokko, H., Rojas, B., Nokelainen, O., & Mappes, J. (2015). Colour polymorphism torn apart by opposing positive frequency-dependent selection, yet maintained in space. *Journal of Animal Ecology, 84*(6), 1555–1564.

Gurdon, J. B. (1968). Changes in somatic cell nuclei inserted into growing and maturing amphibian oocytes. *Journal of Embryology and Experimental Morphology, 20*(3), 401–414.

Hegna, R. H., Galarza, J. A., & Mappes, J. (2015). Global phylogeography and geographical variation in warning coloration of the wood tiger moth (*Parasemia plantaginis*). *Journal of Biogeography, 42*, 1469–1481.

Jennions, M. D., & Petrie, M. (2000). Why do females mate multiply? A review of the genetic benefits. *Biological Review, 75*, 21–64.

Johnstone, R. A. (1995). Sexual selection, honest advertisement and the handicap principle: reviewing the evidence. *Biological Reviews, 70*, 1–65.

Kaitala, A. (1991). Phenotypic plasticity in reproductive behaviour of waterstriders: trade-offs between reproduction and longevity during food stress. *Functional Ecology, 5*(1), 12.

Kaitala, A., & Wiklund, C. (1994). Polyandrous female butterflies forage for matings. *Behavioral Ecology and Sociobiology, 35*(6), 385–388.

Kapan, D. D. (2001). Three-butterfly system provides a field test of mullerian mimicry. *Nature, 409*(6818), 338–340.

Kingston, J. J., Rosenthal, G. G., & Ryan, M. J. (2003). The role of sexual selection in maintaining a colour polymorphism in the pygmy swordtail, *Xiphophorus pygmaeus*. *Animal Behaviour, 65*(4), 735–743.

Kokko, H., Jennions, M. D., & Brooks, R. (2006). Unifying and testing models of sexual selection. *Annual Review of Ecology, Evolution, and Systematics, 37*(1), 43–66.

LaMunyon, C. W., & Eisner, T. (1994). Spermatophore size as determinant of paternity in an arctiid moth (*Utetheisa ornatrix*). *Proceedings of the National Academy of Sciences of the United States of America, 91*(15), 7081–7084.

Lauwers, K., & Dyck, H. (2006). The cost of mating with a non-virgin male in a monandrous butterfly: experimental evidence from the speckled wood, *Pararge aegeria*. *Behavioral Ecology and Sociobiology, 60*(1), 69–76.

Lindstedt, C., Eager, H., Ihalainen, E., Kahilainen, A., Stevens, M., & Mappes, J. (2011). Direction and strength of selection by predators for the color of the aposematic wood tiger moth. *Behavioral Ecology, 22*(3), 580–587.

Maan, M. E., & Cummings, M. E. (2008). Female preferences for aposematic signal components in a polymorphic poison frog. *Evolution, 62*(9), 2334–2345.

Mappes, J., Alatalo, R. V., Kotiaho, J., & Parri, S. (1996). Viability costs of condition-dependent sexual male display in a drumming wolf spider. *Proceedings of the Royal Society B: Biological Sciences, 263*(1371), 785–789.

Mappes, J., Marples, N., & Endler, J. (2005). The complex business of survival by aposematism. *Trends in Ecology & Evolution, 20*(11), 598–603.

Marcotte, M., Delisle, J., & McNeil, J. N. (2007). Effects of different male remating intervals on the reproductive success of *Choristoneura rosaceana* males and females. *Journal of Insect Physiology, 53*(2), 139–145.

Nokelainen, O., Hegna, R. H., Reudler, J. H., Lindstedt, C., & Mappes, J. (2011). Trade-off between warning signal efficacy and mating success in the wood tiger moth. *Proceedings of the Royal Society B: Biological Sciences, 279*(1727), 257–265.

Nokelainen, O., Lindstedt, C., & Mappes, J. (2013). Environment-mediated morph-linked immune and life-history responses in the aposematic wood tiger moth. *Journal of Animal Ecology, 82*, 653–662.

Nokelainen, O., Valkonen, J., Lindstedt, C., & Mappes, J. (2014). Changes in predator community structure shifts the efficacy of two warning signals in Arctiid moths. *Journal of Animal Ecology, 83*(3), 598–605.

Olendorf, R., Rodd, F. H., Punzalan, D., Houde, A. E., Hurt, C., Reznick, D. N., et al. (2006). Frequency-dependent survival in natural guppy populations. *Nature, 441*(7093), 633–636.

Parker, G. A., Simmons, L. W., Stockley, P., McChristie, D. M., & Charnov, E. L. (1999). Optimal copula duration in yellow dung flies: effects of female size and egg content. *Animal Behaviour, 57*(4), 795–805.

Reinhold, K. (2002). Modelling the evolution of female choice strategies under inbreeding conditions. In W. J. Etges, & M. A. F. Noor (Eds.), *Genetics of mate choice: From sexual selection to sexual isolation* (Vol. 9, pp. 189–195). Dordrecht, The Netherlands: Springer.

Rojas, B., Gordon, S., & Mappes, J. (2015). Frequency-dependent flight activity in the polymorphic wood tiger moth, *Parasemia plantaginis*. *Current Zoology, 61*(4), 765–772.

Roulin, A. (2004). The evolution, maintenance and adaptive function of genetic colour polymorphism in birds. *Biological Reviews, 79*(4), 815–848.

Rowland, H. M., Ihalainen, E., Lindström, L., Mappes, J., & Speed, M. P. (2007). Co-mimics have a mutualistic relationship despite unequal defences. *Nature, 448*(7149), 64–67.

Ruxton, G. D., Sherratt, T. N., & Speed, M. P. (2004). *Avoiding attack: The evolutionary ecology of crypsis, warning signals and mimicry*. Oxford, U.K.: Oxford University Press.

Ruxton, G. D., Speed, M., & Sherratt, T. N. (2004). Evasive mimicry: when (if ever) could mimicry based on difficulty of capture evolve? *Proceedings of the Royal Society B: Biological Sciences, 271*(1553), 2135–2142.

Sharma, M. D., Griffin, R. M., Hollis, J., Tregenza, T., & Hosken, D. J. (2012). Reinvestigating good genes benefits of mate choice in *Drosophila simulans*: good genes benefits of mate choice. *Biological Journal of the Linnean Society, 106*(2), 295–306.

Simmons, L. W. (2011). Resource allocation trade-off between sperm quality and immunity in the field cricket, *Teleogryllus oceanicus*. *Behavioral Ecology, 23*(1), 168–173.

Speed, M. P., & Ruxton, G. D. (2007). How bright and how nasty: explaining diversity in warning signal strength. *Evolution, 61*(3), 623–635.

Suisto, K., Lindstedt, C., Pakkanen, H., & Mappes, J. (n.d.). [Chemical defense of aposematic and polymorphic wood tiger moth *Parasemia plantaginis*; evidence for honest signalling]. Unpublished raw data.

Svensson, E. I., & Abbott, J. (2005). Evolutionary dynamics and population biology of a polymorphic insect: polymorphism and evolutionary dynamics. *Journal of Evolutionary Biology, 18*(6), 1503–1514.

Svensson, E. I., Raberg, L., Koch, C., & Hasselquist, D. (1998). Energetic stress, immunosuppression and the costs of an antibody response. *Functional Ecology, 12*, 912–919.

Taylor, M. L., Wedell, N., & Hosken, D. J. (2010). Attractive males do not sire superior daughters. *Evolutionary Ecology, 24*, 195–205.

Velde, L. V., Damiens, D., & Van Dyck, H. (2011). Spermatophore and sperm allocation in males of the monandrous butterfly *Pararge aegeria*: the female's perspective: male butterfly spermatophore and female mate choice. *Ethology, 117*(7), 645–654.

Wedell, N., & Cook, P. A. (1999). Butterflies tailor their ejaculate in response to sperm competition risk and intensity. *Proceedings of the Royal Society B: Biological Sciences, 266*(1423), 1033–1039.

Appendix

Table A1
Difference in mating latency between male colour morphs in the spermatophore and the fertility experiments

Experiment	Trait	Estimate±SE	*z*	*P*
Spermatophore				
	Intercept	13.46±1.35	9.87	<0.001
	Male mass[a]	0.03±0.03	0.98	0.325
	Male age	−1.01±0.33	3.00	0.003
	Male colour (yellow)	−1.63±1.65	0.98	0.330
	Female mass[a]	−0.07±0.03	2.44	0.015
	Female age[a]	$1.44.10^{-4}$±0.29	0.000	0.100
	Male mass[a]*male colour	−0.02±0.05	0.39	0.697
	Male age*male colour	0.37±0.60	0.61	0.540
	Female mass[a]*male colour	0.07±0.03	2.19	0.028
	Trial[b]			
Fertility (categorical)				
	Intercept	5.61±3.53	1.58	0.115
	Male mass[a]	−0.02±0.02	1.26	0.209
	Male colour (yellow)	−1.24±4.54	0.27	0.785
	Female mass	−0.02±0.01	1.58	0.113
	Male mass[a]*male colour	0.08±0.03	1.26	0.009
	Female mass*male colour	0.03±0.02	1.44	0.149
	Trial 2	−0.99±0.97	1.01	0.312
	Trial 3	−1.36±1.00	1.35	0.178
Fertility (continuous)				
	Intercept	7.51±0.25	29.91	<0.001
	Male mass[a]	$2.01.10^{-3}$±$5.77.10^{-3}$	0.34	0.732
	Male colour (yellow)	0.03±0.29	0.10	0.924
	Female mass[a]	$7.9.10^{-3}$±$4.45.10^{-3}$	1.75	0.080
	Female mass[a]*male colour	0.01±0.01	1.11	0.268
	Trial 2	−1.61±0.35	4.49	<0.001
	Trial 3	−0.41±0.35	1.16	0.246

Because of the nature of the response variable in the fertility experiment, mating latency was first analysed as early versus late mating and as a continuous variable within late mating.
[a] Trait centred to the mean.
[b] Trait not present in the 'best' models (ΔAIC < 4).

Table A2
Differences in mating duration between male colour morphs in the spermatophore and the fertility experiments

Experiment	Trait	Estimate±SE	*z*	*P*
Spermatophore				
	Intercept	7.64±1.06	7.16	<0.001
	Male mass[a]	−0.04±0.02	1.48	0.138
	Male age	0.73±0.23	3.12	0.002
	Male colour (yellow)	−1.46±1.05	1.37	0.170
	Female mass[a]	0.01±0.01	0.74	0.458
	Female age	−0.08±0.22	0.36	0.716
	Mating latency	0.14±0.07	2.12	0.034
	Male mass[a]*male colour	0.04±0.03	1.29	0.199
	Male age*male colour	−0.21±0.42	0.52	0.603
	Female mass[a]*male colour	0.01±0.02	0.61	0.539
	Female age*male colour	−0.40±0.39	1.02	0.308
	Mating latency*male colour	−0.03±0.13	0.27	0.829
	Trial[b]			
Fertility				
	Intercept	3.99±1.86	2.13	0.033
	Male mass	0.01±0.01	0.79	0.428
	Male colour (yellow)	-0.14±0.44	0.31	0.756
	Female mass	0.01±0.01	1.82	0.068
	Mating latency	−0.12±0.09	1.30	0.195
	Trial 2	−0.10±0.56	0.18	0.855
	Trial 3	0.47±0.54	0.85	0.393

[a] Trait centred to the mean.
[b] Trait not present in the 'best' models (ΔAIC < 4).

Table A3
Differences in spermatophore size (mg) provided to female between male colour morphs in the spermatophore and the fertility experiments

Experiment	Trait	Estimate±SE	*z*	*P*
Spermatophore				
	Intercept	-0.31±0.50	0.61	0.544
	Male mass	$0.01 \pm 2.22.10^{-3}$	4.07	<0.001
	Male age	0.13±0.02	5.28	<0.001
	Male colour (yellow)	−0.01±0.11	0.07	0.946
	Female mass	$6.20.10^{-4} \pm 1.26.10^{-3}$	0.49	0.626
	Female age	−0.04±0.02	1.88	0.060
	Mating duration	−0.01±0.01	0.60	0.551
	Trial 2	−0.08±0.16	0.48	0.633
	Trial 3	−0.36±0.16	2.19	0.028
	Trial 4	−0.22±0.16	1.39	0.164
Fertility				
	Intercept	0.61±0.45	1.33	0.183
	Male mass	$3.39.10^{-3} \pm 2.15.10^{-3}$	1.58	0.115
	Male colour (yellow)	−0.01±0.01	0.09	0.925
	Female mass	$1.89.10^{-3} \pm 1.25.10^{-3}$	1.46	0.144
	Mating duration	0.02±0.02	0.81	0.420
	Trial 2	−0.13±0.12	1.06	0.291
	Trial 3	0.13±0.11	1.14	0.253

Table A4
Difference in speed of sperm recovery between male colour morph

Trait	Estimate±SE	*z*	*P*
Intercept	15017.75±2034.35	7.30	<0.001
Male mass[a]	5.39±57.08	0.09	0.925
Male colour (yellow)	3096.29±2590.68	1.18	0.238
Recovery time	452.49±62.82	7.12	<0.001
Male mass[a] * male colour	120.74±81.50	1.46	0.143
Recovery time * male colour	47.52±116.02	0.41	0.686
Trial[b]			

[a] Trait centred to the mean.
[b] Trait not present in the 'best' models (ΔAIC < 4).

Table A5
Difference in number of eggs laid and hatching success between sire colour morphs

Dependent variable	Trait	Estimate±SE	*z*	*P*
Number of eggs laid				
	Intercept	5.32±0.58	9.06	<0.001
	Male mass	$-3.14.10^{-3} \pm 2.13.10^{-3}$	1.00	0.319
	Male colour (yellow)	0.11±0.15	0.69	0.492
	Female mass	$2.51.10^{-3} \pm 2.13.10^{-3}$	1.16	0.247
	Trial[a]			
Hatching success				
	Intercept	−0.31±1.38	0.22	0.827
	Male mass[b]	−0.023±0.01	1.91	0.056
	Male colour (yellow)	0.04±1.56	0.02	0.982
	Female mass	$9.24.10^{-3} \pm 5.41.10^{-3}$	1.68	0.094
	Male mass[b] * male colour	0.04±0.02	2.18	0.029
	Female mass[b] * male colour	0.01±0.01	1.01	0.312
	Trial 2	−1.29±0.39	3.22	0.001
	Trial 3	−1.49±0.42	3.45	0.001

[a] Trait not present in the 'best' models (ΔAIC < 4).
[b] Trait centred to the mean.