Sperm phenotype of admixed and 'pure' subspecies males in the long-tailed finch

Callum Scott McDiarmid

Supervisor: Prof. Simon Griffith

Department of Biological Sciences,

Faculty of Science and Engineering

Macquarie University, Sydney NSW Australia

A thesis submitted for the degree of Masters of Research (MRes)

Submitted on 2 November 2018

Declaration

I certify that the material of this thesis has not been previously submitted as part of the requirements for a higher degree at any other university or institution.
This thesis contains no material previously published or written by any other person. I certify that all information sources and literature used are indicated in the thesis.

This project forms part of an ongoing project on long-tailed finches at Macquarie University. In 2015 Dr Simon Griffith and Dr Laura Hurley set up the hybrid and backcross populations that formed the basis of this project. I collected sperm samples in early 2018 from all males alive at the time, with assistance from Laura Hurley. For photographing and measuring sperm morphology and processing the large number of videos I received practical assistance from a number of people. I developed the parameter requirements for the ImageJ CASA system, processed a set of videos previously analysed using an alternative CASA system, and then processed all sperm motility videos with ImageJ CASA with assistance from Rebecca Violante. I photographed sperm cells with the assistance of Sophia Perez, Andrew Blunsden and Rebecca Violante. I also measured sperm morphology with assistance from James Robinson. I performed the data analyses, created the figures and prepared the manuscript with some guidance from Simon Griffith and Laura Hurley.

All other research described in this report is my own original work.

Callum Scott McDiarmid
2 November 2018

NOTE TO EXAMINERS
This thesis is written in the form of a journal article from Journal of Evolutionary Biology. The majority of the author guidelines have been followed, except for deviations required for the Macquarie University thesis formatting requirements. All figures and tables have been presented at the appropriate places in the text to enhance readability.

Acknowledgements

I would like to thank members of the Griffith Lab and post-graduate students from W19F for encouragement and support during the year. Particular thanks to Simon Griffith, for being an inspiring and helpful supervisor, and Lori Hurley, whose experience and expertise studying the sperm of these intriguing little birds was both valuable and exciting to an aspiring 'spermatologist'.

Abstract

Reproductive isolation is central to the speciation process. Recent research suggests that where heterospecific mating occurs, postmating prezygotic (PMPZ) barriers and hybrid infertility can act as powerful reproductive barriers. Two traits that may underlie PMPZ isolation and reduced hybrid fertility are sperm morphology and motility. The two subspecies of long-tailed finch (*Poephila acuticauda acuticauda* and *P. a. hecki*) offer an excellent opportunity to investigate mechanisms of reproductive isolation as their narrow hybrid zone in the wild suggests incomplete reproductive isolation, and they can be studied in captivity. In this study we examine sperm morphology and motility in experimentally created hybrid long-tailed finches and compare them to 'pure' males of each subspecies. We found that hybrid males had longer midpiece than *P. a. hecki* males and significantly longer flagellum length than both subspecies, which we discuss in the context of the expected genetic composition of these groups. However, we found no evidence that 'hybrid' males had low sperm velocity or proportion of motile sperm in vitro, or for a relationship between sperm morphology and velocity. We discuss the opportunity posed by this system to investigate the genomic basis of reproductive traits that could propel this research field forward.

Introduction

Speciation drives biodiversity and is a fundamental component of evolutionary biology. The widely adopted Biological Species Concept (Mayr, 1942) defines species as separate populations that are unable to interbreed, making the build-up of reproductive barriers between diverging populations a central element of the speciation process (Coyne & Orr, 2004). Mechanisms of reproductive isolation are typically grouped into one of three categories; those that limit heterospecific mating (premating barriers), those that limit heterospecific fertilisation (postmating prezygotic barriers), or those that limit hybrid fitness (postzygotic barriers). Premating barriers have received extensive attention, particularly in the wild, and include mechanisms such as spatial or temporal differences in breeding behaviour (Patten *et al.*, 2004; Seddon & Tobias, 2010; López-Rull *et al.,* 2016; Uy *et al.,* 2018). Postzygotic barriers have been investigated through comparative phylogenetic studies (Price & Bouvier, 2002; Fitzpatrick, 2004), and for a few 'classic' model taxa, through detailed laboratory studies (e.g. house mice White *et al.*, 2012; Bhattacharyya *et al.*, 2013; and *drosophila* Sun *et al.*, 2004; Lopez-Maestre *et al.*, 2017). Postmating prezygotic (PMPZ) barriers have received less attention, and result from incompatibilities between males and females post-mating, for example in female sperm storage (Tyler *et al.*, 2013) or sperm-egg interactions (Palumbi & Metz, 1991; Southern *et al.*, 2018). Recent evidence suggests that when premating isolation is incomplete, PMPZ incompatibility can be a powerful barrier to the creation of hybrid offspring (McDonough *et al.*, 2016; Devigili *et al.*, 2017), and thus deserves more research attention (Garlovsky & Snook, 2018).

Speciation is believed to most generally to occur via two or three key steps: 1. a population is separated in geographical space, 2. in allopatry the populations diverge due to selective or neutral processes, and 3. in some cases they then enter into secondary contact (Coyne & Orr, 2004; Price, 2008). Depending on whether reproductive barriers have evolved, secondary contact can result in populations that either readily interbreed once again, hybridise but with fitness consequences, or do not interbreed at all. Here we use the inclusive definition of 'hybrid' employed by (Price, 2008) that includes both inter-species and inter-subspecies crosses. Studying hybridisation between deeply diverged taxa can be valuable, for example for investigating the physiology and genetics of hybridisation (e.g. Ishishita *et al.*, 2016), however it may not reflect those reproductive barriers that first arose (Butlin *et al.*, 2012; Soudi *et al.*, 2016; Coyne, 2018). Areas of secondary contact between recently diverged taxa with incomplete reproductive isolation are called hybrid zones, and these provide an opportunity to examine the speciation process and the mechanisms and strength of reproductive barriers, and their potential to promote divergence (Abbott *et al.*, 2013). When studying the ecology of hybrid zones one can practically consider reproductive barriers as either limiting the breeding success of heterospecific pairs and thus the production of hybrids (i.e. pre- and

postmating prezygotic barriers) or limiting the reproductive success of hybrids themselves (i.e. post-zygotic barriers).

The numerous and varied impacts hybridisation has on individuals can shape how speciation proceeds (Abbot *et al.,* 2013). New genomic techniques are making apparent that hybridisation is common over evolutionary time, for example there is a the growing number of studies showing that adaptive genomic components have originated from historic hybridisation events, including between snowshoe hares and black-tailed jackrabbits (*Lepus americanus and L. californicus*; Jones *et al.*, 2018), among swordtail fish species (*Xiphophorus* spp; Schumer *et al.*, 2018) and between humans and Neanderthals (*Homo sapiens and H. neanderthalensis*; Ackermann *et al.*, 2016). Hybridisation has attracted significant research attention that has illuminated some general trends to hybridisation in the speciation process (Abbott *et al.*, 2013). Firstly, the genetic basis for hybrid sterility and inviability tends to be disproportionately influenced by the sex chromosomes (Z/X chromosome; Coyne, 2018). This is likely to be at least partially due to the tendency for the genes underlying sexually dimorphic traits, for example the reproductive characters involved in sterility, accumulating on the sex chromosomes (Connallon & Clark, 2010; Dean & Mank, 2014; Kim *et al.*, 2017). A second generalisation is that if one sex is more affected by hybridisation it is almost always the heterogametic sex (i.e. Haldane's Rule; Coyne & Orr, 1989; Coyne, 2018) but see (Moran *et al.*, 2017). In ZW systems (such as birds) females are the heterogametic sex, and in XY systems (such as mammals) it is males. Investigating hybrid female infertility and inviability in avian systems will likely be rewarding area of research (see Eroukhmanoff *et al.*, 2016; Mořkovský *et al.*, 2018). Thirdly, as divergence time between hybridising taxa increases so will the degree of hybrid offspring dysfunction (Price & Bouvier, 2002). The specific rate of hybrid dysfunction accumulation varies across taxa, for example deeper divergence is required to produce dysfunctional hybrids in birds and frogs compared to that in mammals (Fitzpatrick, 2004). In birds, largely based on data from crosses between captive individuals, full F1 infertility was suggested to arise after an average of 7 million years (Price & Bouvier, 2002). Effective hybrid infertility in the wild is likely to occur much earlier for two main reasons. First, these data are mostly derived from F1 hybrids, but later generations of hybrids (e.g. F2 and F3, backcrosses) are expected to have lower fitness (Presgraves, 2003; Oka *et al.*, 2004). F1 hybrids still have a full chromosome from each parent species, whereas later generations have admixed chromosomes due to recombination, potentially exposing recessive genotypes and novel genetic combinations (Turner & Schwahn, 2011; White *et al.*, 2012). Secondly, these were mostly non-competitive fertilisation scenarios, whereas in many species females will mate with multiple males within a single reproductive cycle (Birkhead & Møller, 1998). This creates competitive fertilisation scenarios where even a slightly decreased sperm fertilisation efficiency or sperm-female compatibility can greatly reduce a male's reproductive success (Dixon *et al.*, 2003; Geyer & Palumbi, 2005; Pryke *et al.*, 2010; Lymbery *et*

al., 2017). This suggests that slight detriments to key traits such as sperm velocity and motility, or slightly lower compatibility due to unusual sperm morphology, may result in unexpectedly effective reproductive isolation.

Despite PMPZ mechanisms receiving relatively limited attention in the broader field of speciation biology (Birkhead & Brillard, 2007), they are intimately linked to the popular research topic of postcopulatory sexual selection. For successful fertilisation, a complex set of interactions must occur between the ejaculate and female reproductive tract for internal fertilisers, or between the ejaculate and ovarian fluid plume in external fertilisers (Reinhardt *et al.*, 2015). In birds, for example, following insemination sperm must: 1. pass through the vagina, 2. enter and be stored in sperm storage tubules (SSTs), 3. be transported between the SSTs and infundibulum, 4. penetrate the perivitelline layer (PVL) of the ovum, and finally 5. locate and fuse with the female pronucleus within the ovum (Birkhead & Brillard, 2007). As sperm are one of the fastest evolving animal cells (Pitnick *et al.*, 2009a; Rowe *et al.*, 2015a), and reproductive genes and proteins also evolve rapidly (Metz & Palumbi, 1996; Turner & Hoekstra, 2008), populations in allopatry may diverge in a way that partially or completely interrupts one of these five necessary interactions. Across animals there is some evidence of this, including sperm death or abnormal behaviour in the reproductive tract of a heterospecific females (Steele & Wishart, 1992; Gregory & Howard, 1994; Pease *et al.*, 2016), unsuccessful female sperm storage (Price *et al.*, 2000; Sagga & Civetta, 2011), and incompatibilities between the sperm and egg proteins (Metz & Palumbi, 1996; Vacquier, 1998). Two key postcopulatory traits with a demonstrated influence on fertilisation success, that are therefore prime candidates for PMPZ reproductive isolation, are sperm morphology and sperm motility (Birkhead & Brillard, 2007; Pizzari & Parker, 2009).

Across sexually producing animals sperm have a similar overall structure (i.e. a head to carry DNA and either flagellum or cilia to provide propulsion), however in their size and morphology sperm are one of the most variable animal cells (Pitnick *et al.*, 2009a). Sperm cells of passerine birds have a fairly conserved structure and shape (Briskie & Montgomerie, 1992) but vary enormously in size (at least between 42.7 µm and 291 µm; Briskie *et al.*, 1997; Dixon & Birkhead, 1997). Across passerine species the total length of sperm covaries with the size of the female sperm storage tubules (SST; Briskie & Montgomerie, 1992), and in fact SST size may be driving the evolution of total sperm length (Briskie *et al.*, 1997). Coevolution between sperm size and sperm storage organ size also appears to occur in a number of invertebrate taxa (Dybas & Dybas, 1981; Presgraves *et al.*, 1999; Miller & Pitnick, 2002; Minder *et al.*, 2005) and mammals (Gomendio & Roldan, 1993; Anderson *et al.*, 2006). This tight, and potentially rapid, coevolution of sperm storage organs and sperm length may drive divergence among species, potentially resulting in incompatibility during hybridization (Birkhead & Brillard, 2007). This is emphasized by the fact that during intraspecific

crosses in passerine birds sperm movement through the vagina and entering female SSTs accounts for the biggest loss of sperm (~98% of the original ejaculate is lost; Birkhead & Hunter, 1990; Briskie & Montgomerie, 1993; Hemmings & Birkhead, 2017). Furthermore, zebra finch sperm morphology predicts which sperm successfully passes through the vagina and enters female sperm storage, both within and among ejaculates (Bennison *et al.*, 2015; Hemmings *et al.*, 2016). Crossing the vagina and entering SSTs has also been demonstrated to be a strong barrier for heterospecific sperm in experimental crosses between deeply diverged species of fowl and duck (Steele & Wishart, 1992; Sellier *et al.*, 2005). For sperm length to be involved in heterospecific incompatibility relies on divergence in sperm length (and/or SST size). This has largely been supported in the findings of a number of recent studies on passerine birds describing variation in sperm size among subspecies and populations (e.g. Lüpold *et al.*, 2010; Laskemoen *et al.*, 2012; Hogner *et al.*, 2013; Støstad *et al.*, 2016; Supriya *et al.*, 2016)). Notably, this has rarely been extended to hybrids found in natural hybrid zones (but see Cramer *et al.*, 2014) despite potentially being key to understanding reproductive isolation between diverging populations and thus speciation.

The genomic basis of key traits contributing to reproductive isolation is pertinent to understanding both the ramifications of hybridisation and how selection can act on those traits (Butlin *et al.*, 2012). Researchers investigating PMPZ isolation may therefore greatly benefit from the recent success by two independent research groups in identifying the genomic basis of sperm morphology in the zebra finch (Knief *et al.*, 2017; Kim *et al.*, 2017). Evolutionary theory predicts that genes coding for sexually dimorphic traits will accumulate on the avian Z chromosome (Connallon & Clark, 2010; Ellegren, 2011). This was the case for the two zebra finch studies; the Z chromosome explained 67-90% of variance in the sperm midpiece, tail, and total length (Kim *et al.*, 2017). Furthermore, all 108 genes that were differentially expressed between selection lines for long and short sperm were located on the Z-chromosome, suggesting that even those autosomal genes influencing sperm morphology have their expression controlled by Z-linked genes (Kim *et al.*, 2017). These differentially expressed Z-linked genes are mostly located within a large polymorphic inversion that covers over 80% of the zebra finch Z-chromosome (Kim *et al.*, 2017). Furthermore, a male's Z-chromosome inversion karyotype (i.e. what copy of the inversion he has) strongly predicts his sperm cell length (Knief *et al.*, 2017; Kim *et al.*, 2017). Hence, the inversion appears to be acting as a 'supergene' that keeps together a number of co-adapted genes by reducing the rate of recombination (Hoffmann & Rieseberg, 2008). The maintenance of stable polymorphisms is an evolutionary quandary, and in this system, the question is: what is preventing the inversion karyotype encoding the 'best sperm morphology from increasing in frequency to fixation, eliminating the alternative karyotype? Both studies (Knief *et al.*, 2017; Kim *et al.*, 2017) found that the sperm morphology encoded by heterokaryotypes tended to produce the fastest swimming sperm,

and thus suggested that a 'heterozygote advantage' for sperm velocity might be maintaining the polymorphic inversion. It is intriguing that Z-chromosome and sperm morphology are causally linked in the zebra finch as they are both known to evolve particularly rapidly, and have been proposed to drive species divergence and reproductive isolation (Saetre *et al.*, 2003; Presgraves, 2008; Rowe *et al.*, 2015a).

Sperm motility is commonly quantified by both the proportion of motile sperm in an ejaculate, as presumably only motile sperm are useful for fertilisation, and the velocity of those sperm that are motile (Pizzari & Parker, 2009; Fitzpatrick & Lüpold, 2014). Having sperm with a high velocity is a key determinant of fertilisation success, with evidence from non-competitive settings (Donnelly *et al.*, 1998; Malo *et al.*, 2006), as well as in competitive contexts (Gage *et al.*, 2004; Boschetto *et al.*, 2010) including in birds (Birkhead *et al.*, 1999; Froman *et al.*, 1999; Denk *et al.*, 2005; Bennison *et al.*, 2015). This may result from one or multiple mechanisms, including that highly motile sperm remain in female SSTs for longer periods of time (Froman *et al.*, 2002; Pizzari *et al.*, 2008), or high motility may be required to successfully pass through the challenging environment of the vagina (Bakst *et al.*, 1994; Hemmings & Birkhead, 2017). Despite this importance for fertilization success, sperm velocity appears a reasonably dynamic trait (Cramer *et al.*, 2014; Reinhardt *et al.*, 2015; Hurley *et al.*, 2018) that can vary with male age (Johnson & Gemmell, 2012), environmental conditions (e.g. diet; Helfenstein *et al.*, 2010; Tomášek *et al.*, 2017) and genetic issues (e.g. inbreeding Losdat *et al.*, 2014; Opatová *et al.*, 2015). This variability and susceptibility to perturbations suggests that sperm velocity may be impacted in the ejaculates of hybrid males with novel genetic combinations (Turner & Schwahn, 2011). So while it is important to determine whether a difference in sperm morphology might be driving a difference in sperm velocity in hybrid males, it seems unlikely to be a highly strong effect.

It has long been hypothesized that the size of a sperm's morphological components will relate to its swimming velocity. This is intuitively expected to occur via influencing thrust production (e.g. the flagellum), the available energy (e.g. the mitochondria-containing midpiece) or the drag (e.g. head; Humphries *et al.*, 2008; Pizzari & Parker, 2009; Simmons & Fitzpatrick, 2012). It has also been considered that sperm swimming velocity may be influenced by the ratios among these morphological components (Humphries *et al.*, 2008). Comparative studies have suggested that total sperm length correlates with increased velocity across a number of taxa, including fish (Fitzpatrick *et al.*, 2009), mammals (Gomendio & Roldan, 1993) and birds (Lüpold *et al.*, 2009b). There is also quantitative genetic evidence suggesting that in zebra finches sperm velocity and morphology are co-inherited (Mossman *et al.*, 2009). However, many studies have found conflicting results, for example a comparative study found that while sperm with longer midpieces did have higher levels of energy (adenosine triphosphate; ATP), this did not translate into faster swimming sperm (Rowe

et al., 2013). Similarly, a comparative study of North American and European passerines found that sperm morphology and velocity appeared to evolve independently (Kleven *et al.*, 2009). Recently it was even identified that in the zebra finch, longer sperm midpieces were proportionally 'thinner', so that midpiece volume remains consistent irrespective of midpiece length (Mendonca *et al.*, 2018). Evidence is even more inconsistent within species and the significant effects identified tend to be weak (e.g. Lüpold *et al.*, 2009a). Intraspecific studies up to this point have also faced the methodological limitation that morphology and velocity must be assessed per male rather than per sperm cell. This is at least in part due to the logistical challenge that using the microscope cameras currently available equipment the same sperm cell cannot be assayed for both morphology and velocity with high precision. Measuring morphology from a single frame from the sperm velocity video would allow this and has been attempted (Fitzpatrick *et al.*, 2010), however has low resolution meaning that precision is lost. Sperm velocity and morphology in passerine birds has received particular attention, and we have summarized the intraspecific published studies (n=13) in Table 1. Of the studies summarized in Table 1, it appears that having a larger sample size generally increase the chance of finding a relationship (as we might expect for a weak effect).

Despite the central role sperm obviously plays in male fertility and the importance of hybrid infertility for understanding speciation, to the best of our knowledge the only naturally hybridising passerine systems in which sperm traits of hybrids have been assessed is the collared (*Ficedula albicollis*) and pied flycatchers (*F. hypoleuca*; Ålund *et al.*, 2013), and a small (n=2 hybrid males) exploratory investigation with house sparrows and Spanish sparrows (*Passer domesticus* and *P hispaniolensis*; Cramer *et al.*, 2014). There was no evidence that the hybrid sparrows had decreased velocity or unusual sperm morphology, but this is highly tentative given the sample size (Cramer *et al.*, 2014). The collared and pied flycatchers are roughly 1-2 MY divergent and produce inviable hybrid females and hybrid males with reduced fertility (Qvarnström *et al.*, 2010; Ellegren *et al.*, 2012). In the study of hybrid ejaculates (total N=13), wild male hybrids either did not produce sperm (N=11) or produced entirely immotile deformed sperm (N=2), and had zero fitness in the wild hybrid zone (Ålund *et al.*, 2013). Given such strong hybrid infertility, this system does not provide much scope for examining patterns of hybrid sperm size, or investigating it's genomic basis. Hence it would be valuable to also examine more recently diverged taxa - preferably ones that are known to produce hybrids but with some indication of reproductive isolation, and where larger sample sizes are possible.

Table 1: Published intraspecific studies in passerine species testing for a relationship between sperm morphology and velocity

Species	Scientific Name	N	Media	Statistical approach	H	MP	F ^T	Tot	H/Fl	H/M	M/Fl	M/T	Int.	N(tests)	Corr Multiple Test?	Citation
Gouldian finch	*Erythrura gouldiae*		PBS	Separate models	NS	NS	Marg		Marg				no	8		Immler et al 2010
Great tit	*Parus major*	75	DMEM	Separate models	NS	.+	NS	NS	NS	NS	NS		no	14	Yes FDR	Losdat and Helfenstein, 2018
House and Spanish Sparrows	*Passer domesticus, P. hispaniolensis*	41	PBS & DMEM	Separate models	NS	.-	.-	.-	.+			NS	no	6	no	Cramer et al 2014
House Sparrow	*Passer domesticus*	48	DMEM	Separate models	NS	NS	NS	NS	.-	NS	NS		no	7	no	Helfenstein et al 2010
Long-tailed finch acuticauda	*Poephila a. acuticauda*	18	DMEM	Separate models	NS	.+	NS	NS	NS		.+		no	7	no	Rowe et al 2015
Long-tailed finch, hecki	*Poephila a. hecki*	23	DMEM	Separate models	NS	.+	.+	.+	.+		.+		no	7	no	Rowe et al 2015
New world blackbirds	*Q. quiscula*	20	DMEM	Separate models		NS	NS		NS		NS		unclear	4		Lupold et al 2009
New world blackbirds	*X. xanthocephalus*	24	DMEM	Separate models		NS	NS		NS		NS		unclear	4		Lupold et al 2009
New world blackbirds	*A phoeniceus*	120	DMEM	Separate models		Marg	Marg		NS		NS		unclear	12	Yes - then all NS	Lupold et al 2009
Tree swallow	*Tachycineta bicolor*	38	DMEM	Separate regressions, then a model	NS	NS	NS^		NS			.+/NS	no	1	Bonferoni	Laskemoen et al 2010
Zebra finch	*Taeniopygia guttata*	108	dmem	Separate models, checked int app	NS	NS	.+	.+	.+				2*&3*	5		Mossman et al 2009
Zebra finch	*Taeniopygia guttata*	153	DMEM	Single model		.+	.+^2						2*	1	no need	Knief et al 2016
Zebra finch	*Taeniopygia guttata*	183	HAMS	Individual linear models	.+	NS	.+^2	.+^2	.+^2		.+^2		no		Bonferoni	Bennison et al 2016
Zebra finch	*Taeniopygia guttata*	97	Other Studies	BASED ON PREV STUDies												Kim et al 2017

.+ indicates significant positive effect
.- indicates a significant negative effect
Marg: indicates marginally significant effect, or inconsistently significant effect across different groups or time points
^ indicates where a study used 'tail' instead of flagellum, but these are highly correlated so likely would give similar results
Int: whether the interaction between morphological components was tested
N: the total number of males included in the study
Media: the media the sperm were swimming in when velocity was assessed
Statistical approach: The statistical method used to test for a relationship
N(tests): the total number of statistical analyses performed

Corr Multiple Test?: Did the authors use some form of statistical correction for multiple tests (can result in drastically lowered power)
Kim et al. 2017 was excluded as it appeared that some of its data was shared from previous studies in that group, so was not independent data
Column names H, MP F, Tot, H/Fl, H/M, M/Fl, M/T: sperm components significantly related to velocity, head (H), midpiece (MP), flagellum (F), total sperm length (Tot), head:flagellum ratio (H/Fl), head:midpiece ratio (H/M), midpiece:flagellum (M/Fl), midpiece:tail (M/T).
PBS: phosphate buffered saline
DMEM: Dulbecco's Modified Eagle Medium
: significant interaction terms (2: two way, 3*: three way interaction)

The two subspecies of long-tailed finch (*Poephila acuticauda acuticauda* and *P. a. hecki,* family Estrildidae) offers such an opportunity, as they diverged roughly 0.3 MY ago (Jennings & Edwards, 2005), form a narrow hybrid zone in the wild (suggesting some degree of reproductive isolation; Griffith & Hooper 2015; Hooper *et al.*, 2018) but multiple generations of hybrids can be readily created in captivity (pers. obs.). The two subspecies have historically been distinguished by bill colour; western *P. a. hecki* has a red bill and the eastern *P. a. acuticauda* has a yellow bill (Griffith & Hooper, 2017). Where the bill colour distributions come into contact, there is a narrow hybrid zone containing orange billed individuals (150 km in width, or ~8% of the species total range; Griffith & Hooper, 2017). Recent genomic work has identified that the two subspecies have low differentiation on autosomes, but differ across over 75% of the Z-chromosome, possibly driven by segregating chromosomal inversions (Hooper *et al.*, 2018). While there is also a very narrow hybrid zone between the eastern and western Z-chromosome karyotypes (115km, ~6% of total species range), this genetic hybrid zone is displaced 250km to the West of the bill colour hybrid zone. The displacement appears to result from the introgression of three putative colour genes into the eastern subspecies (Hooper *et al.*, 2018). Overall, these narrow hybrid zones suggest that for both bill colour and the genetics more broadly there is some reproductive isolation between the subspecies. Heterosubspecific pairs can be bred in captivity to create hybrids that can themselves breed for at least several generations, although the fertilisation success and survival rate has not been quantified. A recent study (Hurley et al. *in press*) found that the number of sperm reaching the perivitelline layer (PVL) of the ovum (an indicator of successful sperm-female interaction; Bennison et al. 2015) was lower when hybrids were paired with 'pure' subspecies individuals compared to pairs of 'pures'. This indicates that the reproductive isolation suggested by the narrow wild hybrid zone may be at least partially due to PMPZ barriers. There is already some preliminary evidence that the two subspecies of long-tailed finch differ slightly in sperm traits (based on allopatric populations; (Rowe *et al.*, 2015b), although critically hybrid male sperm was not examined.

The long-tailed finch has large inversions on its Z-chromosome (Hooper *et al.*, 2018). The specific details of these inversions have so far been difficult to determine as the sequencing data had to be mapped onto the zebra finch genome, which differs from the long-tailed finch by several chromosomal rearrangements (Hooper *et al.*, 2018). In fact the two long-tailed finch subspecies differ for 75% of their Z chromosome, and it appears likely that the Z chromosome of each subspecies has a polymorphic inversion (Hooper *et al.*, 2018). Chromosomal inversions can act to reduce recombination with heterokaryotypes (chromosomes without the same copy of the inversion), and so recombination between subspecies may be limited across large portions of their Z-chromosome. If long-tailed finch sperm morphology is largely encoded on the deeply divergent Z

chromosome, as found for zebra finches (Knief *et al.*, 2017; Kim *et al.*, 2017), we might expect hybrids to have unusual or dysfunctional sperm phenotypes.

Experimentally creating hybrids in captivity, such as is possible in the long-tailed finch system, has three key advantages over using wild hybrids: 1. larger sample sizes can be obtained (e.g. Ålund *et al.* (2013) could only locate 13 hybrid males between the pied and collared flycatchers), 2. in the wild there is greater opportunity for selection to bias which individuals are available to sample, and 3. in the wild those individuals that resort to hetero-specific pairings or hybridising may be a non-random subset of individuals, that may feasibly have lower fitness or ejaculate quality. The long-tailed finch thus offers a valuable opportunity to study the potential for sperm to contribute to reproductive isolation. In this study we examine sperm morphology and motility in experimentally created hybrid long-tailed finches and compare them to 'pure' subspecies males, which we then consider in the context of their expected Z chromosome karyotype and the likely consequences for the wild hybrid zone (Figure 1). Specifically, we ask:

1. Do hybrid individuals have depressed sperm velocity or proportion of motile sperm compared to a parental subspecies? If so, it suggests that hybrid infertility may be contributing to incomplete reproductive isolation in this system.
2. Do the two males of the subspecies differ in sperm morphology, as identified in a previous study (Rowe *et al.*, 2015b)? This understanding will allow us to interpret our findings for the sperm morphology of hybrid males and will contribute to the literature describing sperm differences found within species.
3. Do hybrid males differ in sperm morphology from either or both parental subspecies, and if so how? We can consider any differences or similarities in the context of the likely underlying Z-chromosome karyotypes, and this will potentially provide the basis for future studies into the genomic basis of sperm morphology in this species. It is also possible that highly unusual total sperm length could influence the sperm-female SST interactions that occur between insemination and fertilisation.
4. Is there a relationship between sperm morphology and sperm velocity? This result will contribute to the growing number of studies examining this expected but inconsistently observed relationship, and any significant relationship may affect the interpretation of our other findings.

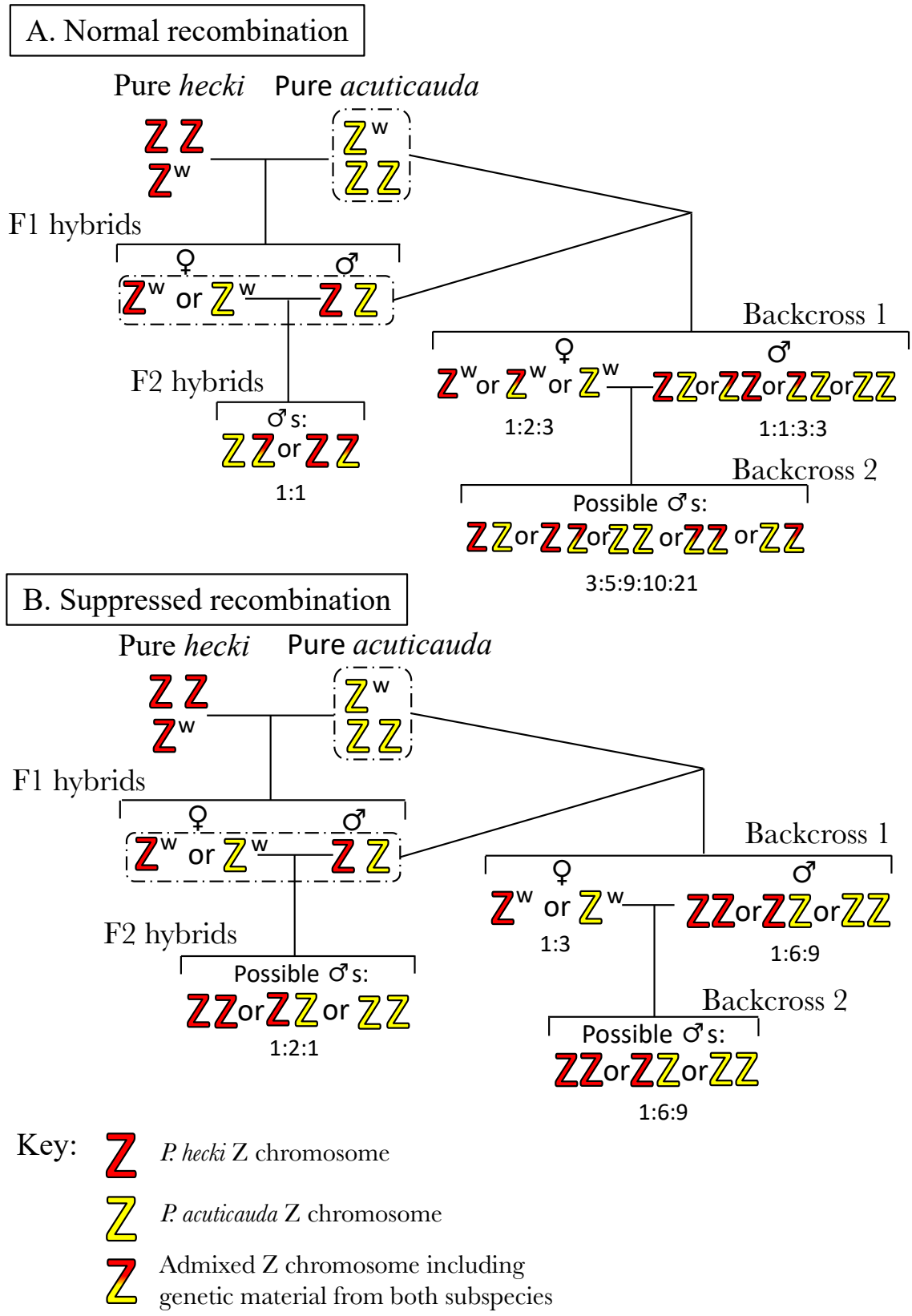


Figure 1: Schematic representation of the expected Z chromosomes in the breeding populations at Macquarie university, under the conditions of 'normal' recombination of one recombination event per chromosome per generation (A), and under the conditions of no recombination (B). This highlights the impact recombination rate can have on hybridisation and admixture.

Methods

Bird breeding details and genetic background

The male long-tailed finches (*Poephila acuticauda acuticauda* and *P. a. hecki)* used in this study were part of a large captive population maintained at Macquarie University (Sydney, Australia), derived from wild-caught individuals brought into captivity in 2009 and 2010. To improve writing flow and help readers follow this study we will refer to the two subspecies by their bill colour throughout (i.e. *P. a. acuticauda* as 'yellow', and *P. a. hecki* as 'red').

Breeding was set up in 2013 as per the schematic in Figure 1; 'pure' individuals from each subspecies were crossed to produce F1 hybrids (one red and one yellow parent). A number of these F1 hybrids were then crossed with 'pure' yellows, creating what we call Backcross 1s (B1s; one yellow and one F1 hybrid parent). At the beginning of 2016 these four groups (red, yellow, F1 hybrids and Backcross 1s) were then established to free-breed in large outdoor aviaries, each with between 20 and 30 pairs. We will refer to the offspring of F1 hybrids as F2 hybrids, and the offspring of Backcross 1s as Backcross 2s (B2s). Sampling for the current study was performed in early 2018, at which point there were fewer than this still alive and available to sample (F1 hybrids n=9, F2 hybrid males n=18, Backcross 1 males n=8, and backcross 2 males n=7).

While on the autosomes there is almost no differentiation between the long-tailed finch subspecies, 75% of the Z chromosome differs between the subspecies (Hooper et al 2017), making the Z chromosome the genetically defining feature of each subspecies. Another reason that we will focus on the Z chromosome is that the genetic basis of sperm phenotypic components is likely to reside on the Z chromosome (Knief *et al.*, 2017; Kim *et al.*, 2017). W chromosomes are also likely different between the subspecies and would be an intriguing line of future investigation but will not influence sperm production so are not considered further here.

We will also simplify our discussion by referring to all Z-chromosomes that have recombined in a way other than with a homologous 'pure' chromosome (i.e. has had the chance to mix with a different type of Z chromosome) as an as 'admixed' Z chromosome. This is an oversimplification, because recombination occurring at different chromosomal locations will result in various chromosomal compositions, however for simplicity and because we are completely blind to these differences in our males anyway, we consider all Z chromosomes as either yellow, red or admixed. It is important to note that although the hybrid and backcross groups will include males with a variety of Z-chromosome combinations (Figure 1), in almost all cases their bill colour will remain

some shade of orange. Z-chromosome types are therefore effectively undistinguishable based on bill colour (the main phenotypic indicator of subspecies or hybrids in this system).

The chromosomal combinations found in each group (F1 and F2 hybrids, Backcross 1 and 2) will differ depending on the rate of recombination. There is evidence in a passerine (collared flycatcher) suggesting that recombination is obligate per chromosome per set of meiotic divisions (Kawakami *et al.*, 2014). However, there is also evidence suggesting that rate of recombination on the long-tailed finch Z chromosome may be suppressed compared to autosomes (Singhal *et al.*, 2015), likely due to the large inversion on it. We hence describe two the expected Z chromosome karyotypes among our groups in two scenarios, one with the normal rate of recombination (1 per chromosome per generation; Figure 1A.) and one assuming no recombination (Figure 1B). We note that the real recombination rate is going to be much more similar to the 'normal' rate of recombination, and the 'no recombination' example is to demonstrate the influence of recombination in the system.

In both scenarios, each population of pure sub-species will just have pure Z-chromosomes of their respective colour. F1 hybrid males will have a whole copy of both a red and a yellow Z chromosome, that they have inherited from their parents without any opportunity for heterosubspecific recombination. The steps following this depend on whether or not recombination is occurring.
If recombination is occurring 'normally', i.e. once per chromosome per generation (Kawakami *et al.*, 2014; Figure 1. A), F1 hybrid males will pass on admixed Z chromosomes, whereas females are hemizygous and their single Z chromosome cannot recombine. All F2 hybrid males, (i.e. those with two F1 hybrid parents), thus inherit a 'pure' Z chromosome from their mother and an admixed Z chromosome from their father. During the backcross mating, similarly male F1 hybrids pass on admixed Z chromosome and female F1 hybrids pass on a 'pure' Z chromosome, and male offspring will in Backcross 1 also receive a pure yellow Z chromosome from the pure yellow group. Hence in Backcross 1s (B1s) we expect a disproportionately high number of yellow Z chromosomes overall, but half the males will have one admixed chromosome. For backcrosses 2s (B2s; with two B1 parents), recombination in Backcross 1 will further admix most of the chromosomes, meaning that only 1 in 5 B2 males (9/48) will have two pure yellow Z chromosomes, and only 1 in 16 B2 males will have one of each pure Z chromosome. Hence F1 hybrid males are distinct from any other group as all males will have one red and one yellow Z chromosome. Backcross 1 males are also distinct as almost half have pure Z chromosomes with the others containing some admixture. There is a less distinct difference between F2 hybrid males and backcross 2 males; F2 hybrid males all contain an admixed chromosome, as do roughly 75% of the backcross 2 males.

Alternatively, if the recombination is suppressed then there will be no admixture between the yellow and red Z chromosomes in any of the groups, so that the only possible Z chromosome combinations a male can have is two yellow, two red, or one of each. This means that F2 hybrids, Backcross 1s and backcross 2s all share the same possible combinations of Z chromosome (although occurring in different proportions), and all three groups contain individuals with two matching 'pure' Z chromosomes.

While recombination may be reduced on the Z chromosome (Singhal *et al.*, 2015), it is likely to be drastically more similar to the 'normal' recombination example. In the present study, due to our relatively low sample sizes for F1 hybrid (N=9) and Backcross 1 males (N=8), we will perform a preliminary investigation comparing all four of our hybrid and backcross groups, but subsequently remove F1 hybrid and Backcross 1 males from further analysis. While F1 hybrids (with one of each 'pure' Z chromosome) and backcrosses 1s (some admixed individuals, but almost half of males have two pure yellow Z chromosomes), are quite distinct, the F2 hybrid and Backcross 2 groups will contain highly similar array of males, where most of the males contain at least one admixed Z chromosome. For our analyses, we will thus combine F2 hybrids and backcross 2s into a single group due to their similar chromosomal composition, which we will use for our main investigation and formal analyses, which we will refer to as 'Mixed Hybrids'.

For examining the costs and consequences for hybridisation, later generation hybrids can experience worse dysfunction than F1 hybrids, as recombination can create novel genetic combinations and expose recessive-recessive incompatibilities. There is already some evidence that F2 incompatibilities contribute to reproductive isolation via hybrid dysfunction and may be more common than other incompatibilities arising in hybrids (Presgraves, 2003; Oka *et al.*, 2004; Turner & Schwahn, 2011; White *et al.*, 2012). So, by studying later generation individuals we will be unlikely to miss powerful hybrid dysfunction that only appears in the subsequent generation, which could not be said if we just examined F1 individuals. So, while some studies use F1 generation hybrids for logistical reasons (e.g. Price & Bouvier, 2002) this can underestimate the degree of hybrid dysfunction and thus reproductive isolation that would be found in wild hybrid zones (Turelli & Orr, 2000; Wiley *et al.*, 2009).

Sperm sample collection

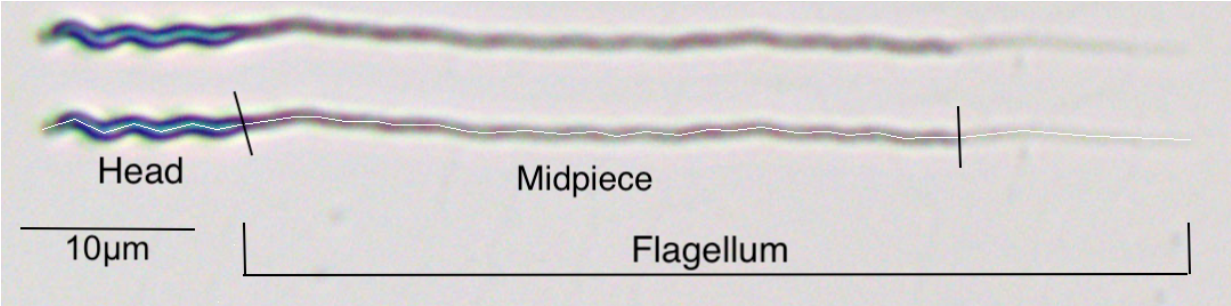
All males in this study were sexually mature adults and sampled during the breeding season while having access to mates. Sperm samples were collected by cloacal massage (Wolfson 1952), and subsequently processed for velocity and morphology measurements following standard procedures (e.g. Rowe *et al.*, 2015b). Following sample collection into a capillary, based on the visual estimate

of ejaculate quantity the sample was immediately diluted into either 25, 50, 100, 150 or 200 µl of preheated (40°C) phosphate buffer solution (PBS) to attain a concentration roughly optimal for a video recording. For samples where motility was going to be assessed, 6 ul of the sample was then immediately loaded into a pre-heated slide chamber (depth 20 µm; Leja, Netherlands) on a heated stage plate set to 40°C (TP-S, Tokai Hit, Shizuoka, Japan). Under 400x magnification using a phase contrast microscope (CX41, Olympus, Japan) and connected digital camera (Legria HF G25, Canon, Japan) we recorded 5 s of video in six unique fields of view (30 s total recording per sample). A small aliquot of the remaining diluted sample was then fixed in 5% buffered formaldehyde for later use measuring sperm morphology. Given that we do not expect a difference in motility between the two subspecies (Rowe *et al.*, 2015b) and that our main interest with sperm motility is whether our Mixed Hybrid males have lower motility compared to 'pure' individuals, due to time constraints video recordings for red males was note taken, and only yellow and Mixed Hybrid males were analysed with computer assisted sperm analysis (CASA; see below).

Sperm morphology

To measure sperm morphology, we first smeared 15 µl of a fixed sperm sample onto a microscope slide, and left for 24 h to air dry. This was then rinsed under running distilled water and left it to air dry for another 24 h. Digital images were then captured using a phase contrast microscope (Olympus BX50, Olympus Japan) with a connected camera (14MP Aptima COMS, RisingCam). Digital images were captured at very high resolution (4096 x 3288 pixels), meaning that sperm morphology could subsequently be measured very precisely by magnifying the images (1 µm was greater than 10 pixels long). Following Laskemoen et al. (2012) we used digital image analysis (ImageJ 1.50i; Schneider *et al.*, 2012) on morphologically normal individual sperm cells to measure the length of the head, mid-piece and 'tail' (naked flagellum), and calculated flagellum length (midpiece + tail) and total sperm length (head + flagellum; Figure 2). Finally, we calculated the among male coefficient of variation (CV_{am} = 100*standard deviation/mean; Lifjeld *et al.*, 2010) for the different groups (red, yellow, Mixed Hybrids), for the two 'pure' subspecies pooled and for all individuals pooled.

Figure 2: Representative long-tailed finch sperm, indicating the morphological measures we used.



To calculate how many sperm per male are required to capture average sperm length per male we first photographed and measured 30 sperm each from 14 males. We then used a resampling procedure to examine how the 95% confidence intervals (CIs) around the mean decreased with the increased number of sperm measured per male. Specifically, for a single trait on a single male (e.g. head length, Male 1) we would randomly re-sample the 30 measured sperm cells two at a time (1,000 times), then three at a time (1,000 times), then four at a time (1,000 times), up to twenty sperm cells at a time (1,000 times). Figure 3 shows an example with two individuals, whereby the variation around the mean drops off with the increasing number of sperm cells measured. We performed this for all 14 males for all four sperm traits of interest (head, midpiece, flagellum and total sperm length). The results of this analysis suggested that measuring 10 sperm per male would be adequate to capture the intra-individual average length for each of these traits (Figure 3, Supplementary Figures 1-4), which is consistent with recommendations in the literature (Laskemoen *et al.*, 2007; Bennison *et al.*, 2014). Many studies rely on published information regarding how many sperm cells they should measure per individual, however given the inherent variability of sperm phenotypes, this approach to re-assess the required number for each species independently is preferable. Finally, we performed this resampling analysis across individuals to determine how the variation around the mean dropped with increasing number of males sampled, which should prove useful when comparing to the previous study on sperm in this species and when planning future work.

Sperm velocity and proportion motile

We used computer-assisted sperm analysis software (CASA ImageJ plugin; Wilson-Leedy & Ingermann, 2007) to calculate sperm swimming velocity and the proportion of motile sperm per ejaculate from the recorded videos. We had recorded six unique fields of view for each sample, and for each field of view we analysed 1 s (at 25 frames s^{-1}). Each 1 s clip was first converted to its 25 individual frames as an 'image sequence' (VirtualDub, Version 1.10.4). To ensure that sperm cells were correctly identified we set the following detection parameters in the CASA software: sperm size between 50 and 350 pixels, maximum velocity between frames 30 pixels, and minimum track length 10 frames. To attain parameter values, we first used trial and error to get approximately appropriate values, followed by extensive stepwise changing of each parameter independently for a set of 15 videos, in each case comparing the original video to the tracks made by the CASA software. The final values were those that prevented sperm cells being tracked twice and tracks jumping from one sperm to another and limited how many sperm were excluded due to not meeting these criteria (extreme parameter values eliminated large numbers of sperm that could provide useful data). We visually inspected each video before and after analysis to ensure that no non-sperm

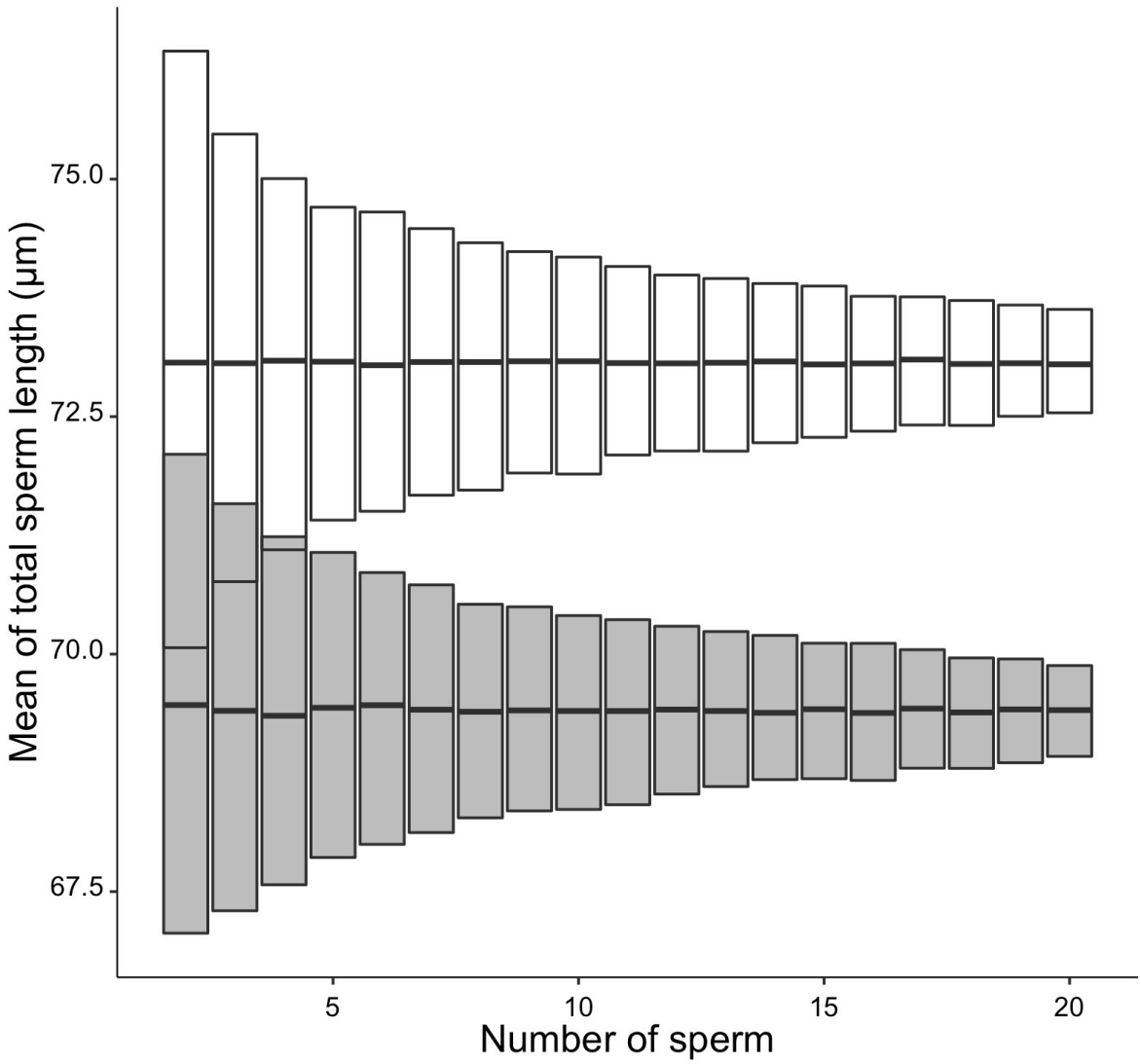


Figure 3: Two representative individuals (one in grey, one in white) in resampling analysis, here for total sperm length within males. It demonstrates how the spread of calculated average sperm length decreased with the increasing number of sperm sampled. On the basis of this we measured 10 sperm per male. Boxes represent 95% confidence intervals (CI).

particles (e.g. the occasional red blood cell or faecal particle) were being assessed as sperm, and to ensure all sperm were only being tracked once. To control for the effects of drift sperm cells with a straight line velocity (VSL) of less than 25 µm/s, and a curvilinear velocity (VCL) of less than 30 µm/s were considered immotile. This was determined by both running the software on 18 videos with entirely immotile sperm, and by comparing the tracks and videos when using this values. These cut offs are consistent with published studies (Cramer *et al.*, 2014; Hurley *et al.*, 2018).

This CASA software (an ImageJ plugin) is freely available, which offers a great opportunity for researchers, as commercially available medical CASA programs (e.g. CEROS vs.12, Hamilton Thorn Research; Sperm Class Analyzer 5.4.0.0, Microptic) can be expensive. When this free software was compared to commercial programs for fish (Boryshpolets *et al.*, 2013) and stallions (Giaretta *et al.*, 2017), it gave correlated but consistently different outputs for velocity values between the programs. As this comparison has not previously been made in a passerine bird, we analysed a set of 13 videos of zebra finch sperm that had previously been analysed with the Microptic SCA software (zebra finch and long-tailed finch sperm appear highly similar morphologically; (Rowe *et al.*, 2015b). The VCL output was highly similar (R^2=0.88), and so while we are confident this software will work well (as for Losdat & Helfenstein, 2018) we caution against direct comparisons of absolute sperm velocity assessed using different software. As for sperm morphology, we used a resampling approach to determine the minimum number of sperm cells required to capture variation within an ejaculate, which we identified as 20 (in line with a published study (Hurley *et al.*, 2018; Supplementary Figure 5). Hence, samples with fewer than 20 motile sperm were excluded from further analysis.

CASA software gives three velocity readings per sperm: Curvilinear velocity (VCL, the actual speed of the tracked sperm), straight-line velocity (VSL, the overall speed from starting to ending point across all frames) and average path velocity (VAP, a 'smoothed' path that is calculated by the ratio of VCL and VSL). As passerine sperm swims almost as a straight line (e.g. as opposed to some marine invertebrate sperm that swims in circles, or mammal sperm that appear to wobble from side to side) we followed publish studies (Rowe *et al.*, 2015b; Hurley *et al.*, 2018) and used VCL for analyses as it is likely to represent actual sperm velocity better than the other calculated approximations. These three velocity parameters are highly intercorrelated (R^2=0.82, 0.89, 0.94), and so would produce similar results. We also calculated a measure of the fastest 10% of sperm cells for each male (following Mossman *et al.*, 2009; Bennison *et al.*, 2016), as this may represent the most viable component of any particular male's sperm sample. Following published studies (Hurley *et al.*, 2018) we calculated the proportion of motile sperm as the total number of motile sperm divided by the total number of sperm identified by CASA per sample.

Statistical Methods

Data was imported into R and organised using R package 'dplyr' (Wickham *et al*., 2017).
Data exploration was performed as described in (Zuur *et al.*, 2010). Briefly, this includes checking for outliers, homogeneity of variance, normality and collinearity. We fitted linear models with Gaussian errors with one morphometric trait as the response and the other as the predictor to look at the correlation between them. There was a high correlation between sperm flagellum and total sperm length (R^2=0.95), and low correlation between midpiece and flagellum (R^2=0.13), midpiece and total (R^2=0.13) and very low between head and all other components (R^2<0.1). In most of our models these component lengths are the response variables in separate models so the correlation between flagellum and total sperm length does not pose a direct statistical issue (although it will affect our interpretation). The one model in which they are used together is to investigate the relationship between sperm velocity and morphology, where sperm components are the explanatory variables. In this case we exclude total length and just use head, midpiece and flagellum length (see below). To determine how much of the total variation our models explained (i.e. model fit), we calculated marginal R^2 (for fixed effects) and conditional R^2 (for full fixed effects + random effects, i.e. full model), and the random effect intra-class correlation coefficients (ICC) for each random effect (Nakagawa & Schielzeth, 2010; 2012).

Sperm morphology

For analyses of sperm morphology, we excluded F1 hybrids and Backcross 1s due to low sample size, as they could not readily be pooled due to different chromosomal combinations. We did pool F2 hybrids and Backcross 2s (due to similarly high proportion of males containing an admixed Z chromosome; Figure 1) to form the 'Mixed Hybrids' group. We then tested to see whether sperm morphology varied between the two subspecies, and between the Mixed Hybrid males and 'pure' type males (reds and yellows), using contrasts when fitting the GLMM. Using the measures of all sperm cells we ran generalised linear mixed effect models (GLMM; lmer package in R; Bates et al 2015) with Gaussian errors, with sperm component (head length, midpiece length, flagellum length and total length) as the response variable, with group (red, yellow or Mixed Hybrid) as a fixed factor and Male ID as a random effect. While ANOVAs could also have been appropriate, we used linear mixed models so that analyses were consistent throughout.

While we also calculated the among male variation (CV_{am}) for each group (CV_{am} = 100*standard deviation/mean; Lifjeld *et al.*, 2010), and for all individuals in the study pooled together, CV_{am} could not be statistically compared across groups as N=1 for each group.

Sperm motility

To assess sperm motility and velocity between groups we used a similar approach, running VCL per male and VCL for the top 10% fastest sperm per male as separate models. As mentioned above, due to time constraints videos were not collected or processed for red males and so the available sperm motility data is from yellow males and the Mixed Hybrid group. This is acceptable for our purposes, because our main point of interest is whether hybrids have decreased sperm motility, and we do not expect a difference between red and yellow males (Rowe *et al.*, 2015b). We ran a generalised linear mixed effect model with Goussian errors, with per-sperm VCL or the subset of the top 10% fastest sperm as the response variable, group (Mixed Hybrid or yellow) as a fixed effect, and male ID as a random effect. To test whether the proportion of motile sperm differed with group (Mixed Hybrid or yellow), we ran a linear model with binomial family, with the proportion of motile per male as the response variable and group as a fixed effect.

Testing for a relationship between sperm morphology and velocity

This is a statistically challenging problem, as there are such a number of morphological components, or ratios between them, that have been hypothesised to influence sperm swimming speed. Researches in this field have typically performed multiple linear models with velocity as the dependent variable (one for each morphological measure), however they then face the decision of applying a correction for multiple tests, hence loosing statistical power for what is already clearly a week effect (Nakagawa, 2004), or they do not and run the risk of type 1 error. We prefer the method similar to that employed by (Knief *et al.*, 2017), whereby the key morphological variables are included with their interaction terms in a single model, that then has nonsignificant (based on p value <0.05) interaction terms removed via a stepwise process based on statistical significance (Zuur *et al.*, 2009). These interaction terms should account for any important ratios between morphological components. Using this approach, (Knief *et al.*, 2017) identified the main relationship between morphology and velocity identified in an independent study (Kim *et al.*, 2017).

Due to methodological constraints, morphology and velocity cannot be calculated for the same sperm cell, so each needs to be averaged per male before comparison. As there is much higher variation in sperm velocity than in sperm morphology, sperm morphological components (the independent variables in our subsequent models) were mean-centred (using 'rescale' function in R). While there is suggestion that the ratio of sperm components (e.g. head:flagellum length) may influence sperm velocity, to minimise collinearity and over-fitting the model, we did not include these ratios as separate fixed effects but expect that the interaction between them should provide this information.

We ran two linear models (LMs) with Gaussian errors, either with mean VCL per male or mean VCL of the fastest 10% of sperm per male as the fixed effect, and began with a three way interaction between head length, midpiece length and flagellum length as our explanatory factors. Non-significant explanatory variables were then removed in a stepwise fashion to create a minimal model (Zuur *et al.*, 2009), that was then compared to the null model (fixed effects replaced with a '1') using an ANOVA to assess the significance of the fixed effects to explain sperm velocity. We also ran independent linear models with Gaussian errors with VCL against sperm morphological traits to facilitate a comparison with the previous results found for yellow males (Row *et al.*, 2015b), the traits being head length, midpiece length, flagellum length, total sperm length, the flagellum:head ratio and the midpiece:flagellum ratio.

Results

Sperm morphology

Using the Marginal R^2 values and ICC values can inform us how much of the total variation was explained by our fixed effect (Marginal R^2) and random effect (ICC). While not analysed due to low sample sizes, for illustrative purposes we have included a figure of the hybrid male sperm parsed into Hybrid F1s, Hybrid F2s, Backcross 1s and Backcross 2s (Figure 4). As outlined in the methods, Hybrid F2s and Backcross 2s were pooled to give the group 'Mixed Hybrids' for analysis.

For our models comparing the groups red (N males =38), yellow (N males =35) and Mixed Hybrid (N males =25), considering all sperm cells measured (N=970), Male ID explained 45% of variance in head length, 53% of mid-piece variation, 66% of flagellum length, and 68% of total sperm length (values based on ICC value of male ID as a random effect; Table 2). Red and yellow males did not significantly differ in head length (t_{94}=0.029, p=0.78), flagellum length (t_{94}=-0.37, p=0.71), or total sperm length (t_{94}=-0.30, p=0.76), but did significantly differ in midpiece length (t_{94}=-4.01, p<0.001; Table 2, Figure 5). Hybrid males did not differ from 'pure' (yellow and red) males in head length (t_{94}=1.57, p=0.12), but did significantly differ in midpiece (t_{94}=-2.16, p=0.033), flagellum (t_{94}=-3.65, p<0.001) and total sperm length (t_{94}=-3.21, p=0.002; Table 2, Figure 5). It is worth noting that total sperm length and flagellum length are highly correlated (R^2=0.95).

The among male CV (CV_{am}) for total sperm length could not be statistically compared (as N=1 per group), however variation was greatest among red males (CV_{am} =6.22), followed by Mixed Hybrid males (CV_{am} =5.26) and yellow males with the smallest (CV_{am} =3.23). When grouped together, the two subspecies (red and yellow males) had an intermediate CV (CV_{am} =5.03), as did all males in the study pooled together (CV_{am} =5.42).

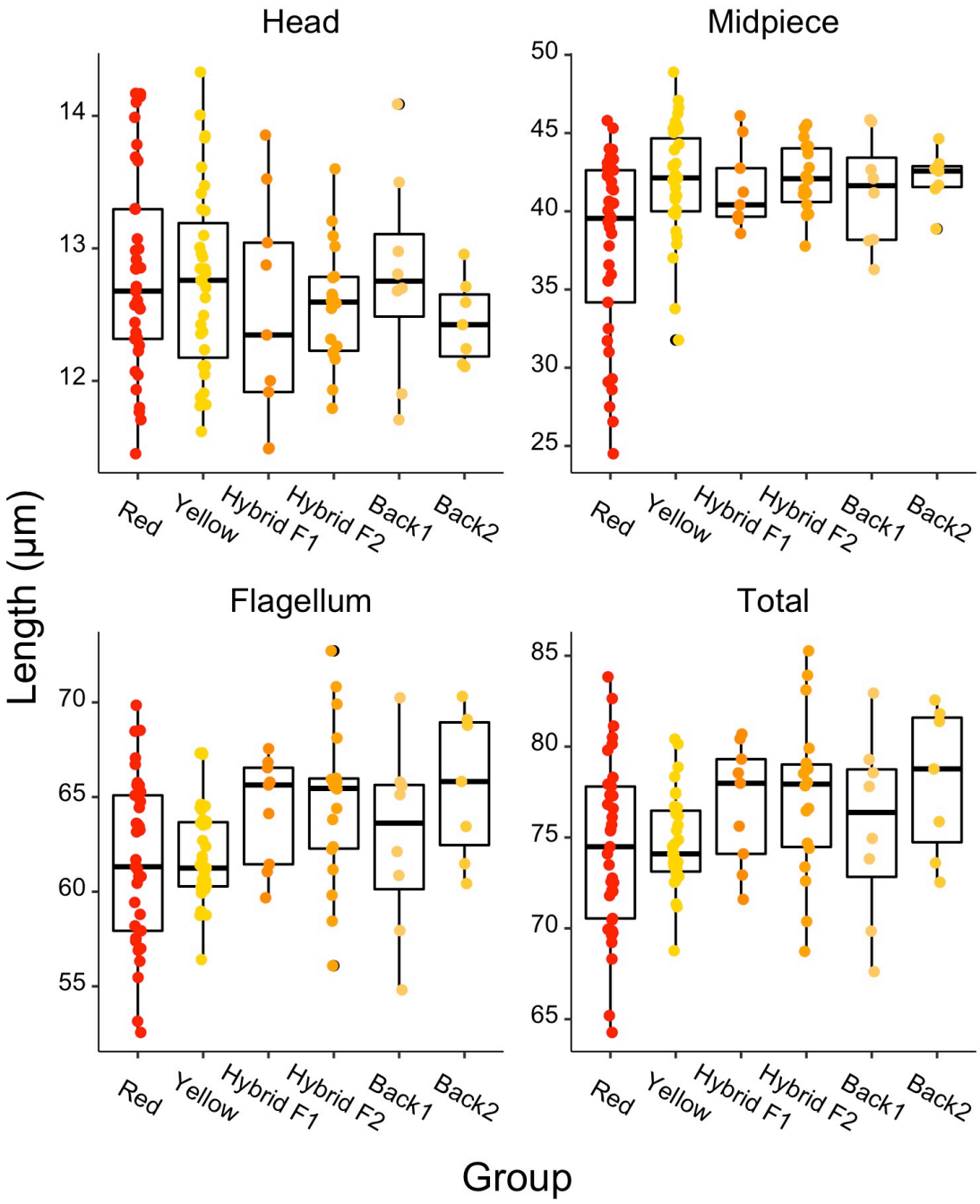


Figure 4: Sperm morphology measurements for all six possible groups (red males, yellow males, Hybrid F1 males, Hybrid F2 males, Backcross 1 males and Backcross 2 males) for illustrative purposes, as the hybrid groups contained too few samples to statistically compare. See Figure 5 for morphological data as analysed. Dots are individual males, with boxplots reflecting the median, the 25th percentile (Q1) and 75th percentile (Q3), with whiskers 1.5 times the interquartile range above Q1 and 1.5 times the interquartile range below Q3.

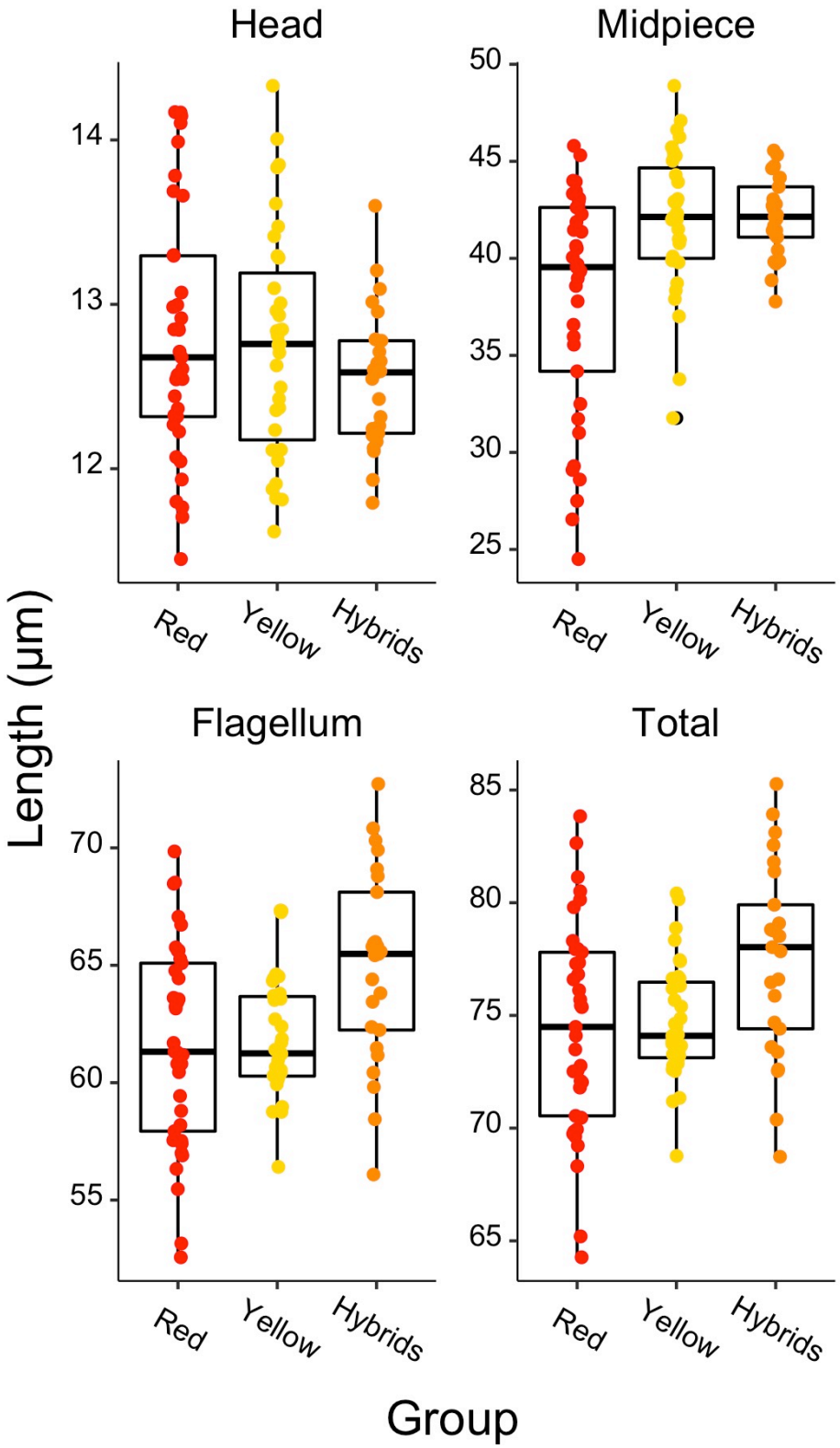


Figure 5: Sperm morphology measurements for all the three groups used in our main analysis (red males, yellow males and Mixed Hybrid males). Dots are individual males, with boxplots reflecting the median, the 25th percentile (Q1) and 75th percentile (Q3), with whiskers 1.5 times the interquartile range above Q1 and 1.5 times the interquartile range below Q3.

Table 2: Generalised linear mixed models to examine whether sperm morphology differed bewteen either red and yellow males (R-Y), or between hybrid and 'pure' (red and yellow) males (Pure-Hyb) using planned contrasts, with Male ID included as a random effect.

Head	Estimate	Std. Error	Df	t value	Pr
Intercept	12.70	0.068	94	187.12	
R-Y	0.022	0.078	94	0.029	0.78
Pure-Hyb	0.080	0.051	94	1.57	0.12

	Var	St Dev	n	ICC
Marg R^2	0.011			0.01
Male ID	0.39	0.62	97	0.45
Residual	0.47	0.69		
Cond R^2			970	0.46

Mid-piece	Estimate	Std. Error	Df	t value	Pr
Intercept	40.67	0.45	94	90.80	
R-Y	-2.06	0.51	94	-4.01	<0.001
Pure-Hyb	-0.73	0.34	94	-2.16	0.033

	Var	St Dev	N	ICC
Marg R^2	4.10			0.12
Male ID	17.71	4.21	97	0.53
Resid	11.73	3.43		
Cond R^2			970	0.65

Flag	Estimate	Std. Error	Df	t value	Pr
Intercept	62.84	0.38	94	163.48	
R-Y	-0.16	0.44	94	-0.37	0.71
Pure-Hyb	-1.05	0.29	94	-3.65	<0.001

	Var	St Dev	N	ICC
Marg R^2	1.93			0.10
Male ID	13.42	3.67	97	0.66
Resid	4.88	2.21		
Cond R^2			970	0.76

Total	Estimate	Std. Error	Df	t value	Pr
Intercept	75.54	0.40	94	187.31	
R-Y	-0.14	0.46	94	-0.30	0.76
Pure-Hyb	-0.97	0.30	94	-3.21	0.002

	Var	St Dev	N	ICC
Marg R^2	1.65			0.08
Male ID	14.78	3.84	97	0.68
Resid	5.34	2.31		
Cond R^2			970	0.75

Sperm motility

Male ID accounted for 35% of the total variance in the sperm velocity (VCL) data (N=13,814), but accounted for 83% when we considered just the subset of the fastest 10% of sperm per male (N=1,395; Table 3A). The fixed effect of group, i.e. yellow (N males =33) or Mixed Hybrids (N males =19), did not significantly predict either total VCL (Marg. R^2<0.001, $t_{49.81}$=0.18, p=0.86) or the VCL of the top 10% of sperm (Marg. R^2<0.001, $t_{49.15}$=0.203, p=0.84; Table 3A; Figure 6). Group also did not significantly predict the proportion of motile sperm per male (t_{56}=0.305, p=0.76; Table 3B; Figure 7). This means there was no significant difference between 'pure' yellow males and the Mixed Hybrids in terms of sperm swimming speed, the swimming speed of the fastest 10% of each ejaculate, or the proportion of motile sperm per ejaculate.

A relationship between sperm morphology and velocity

The average VCL per male was not significantly predicted by head length ($t_{3,42}$=1.25, p=0.22), midpiece length ($t_{3,42}$=0.89, p=0.38) or flagellum length ($t_{3,42}$=-0.4, p=0.69; Table 4A), and all interactions were removed via stepwise removal of nonsignificant interaction terms. Similarly, in our final reduced model the VCL for the top 10% fastest sperm per male was also not significantly predicted by head length ($t_{3,42}$=1.26, p=0.22), midpiece length ($t_{3,42}$=1.08, p=0.28) and flagellum length ($t_{3,42}$=0.073, p=0.94; Table 4A). Our linear models to compare to the previous results found for yellow males (Rowe *et al.,* 2015b), where each model had a different morphological trait predicting average VCL per male, found that VCL was not significantly predicted by any of the sperm morphological traits we tested, including midpiece and midpiece:flagellum ratio that were significantly positively associated with VCL in the previous study of this species (Table 4B).

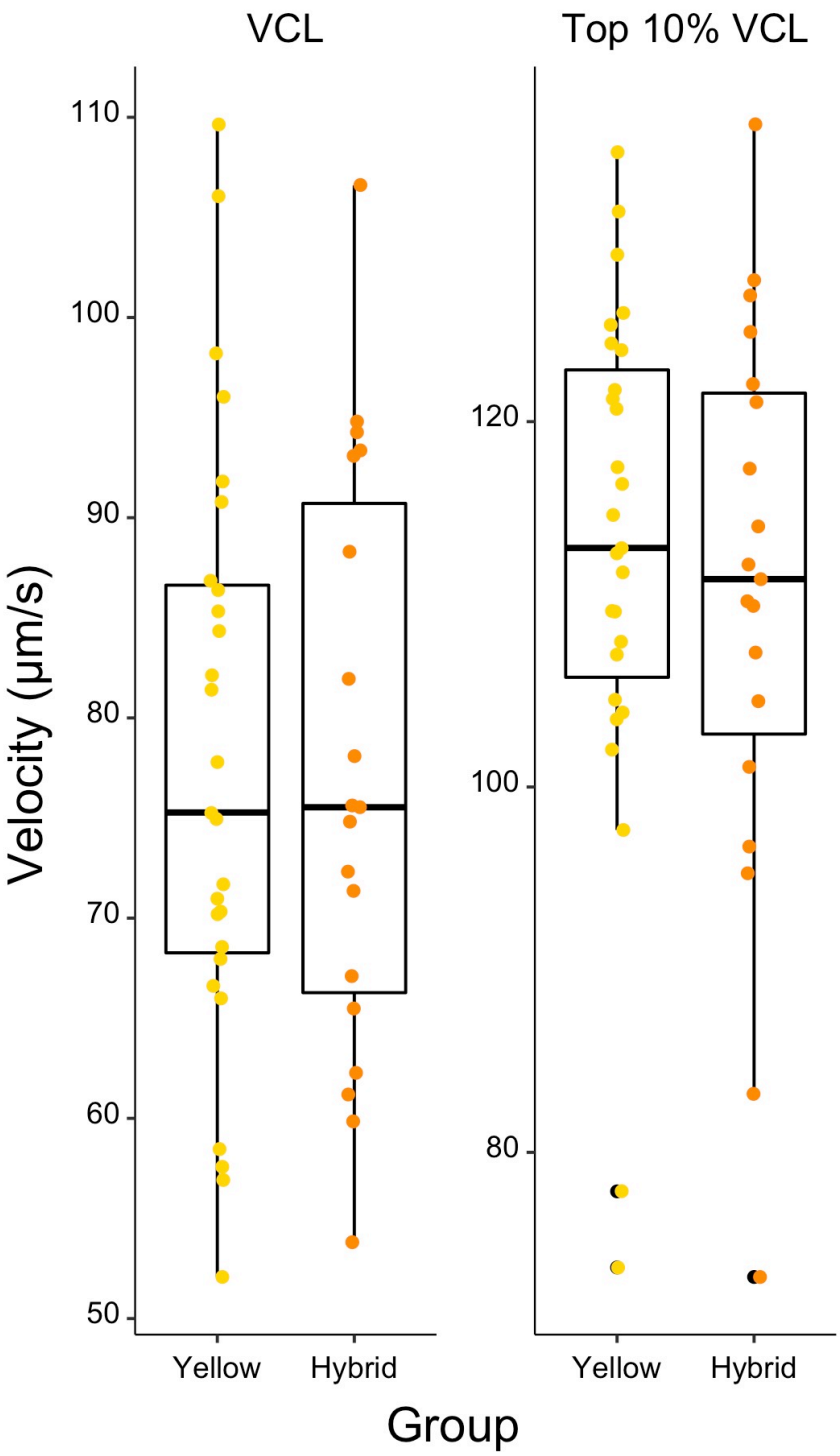


Figure 6: Sperm velocity values for Mixed Hybrid males and yellow males. Dots are individual males, with boxplots reflecting the median, the 25th percentile (Q1) and 75th percentile (Q3), with whiskers 1.5 times the interquartile range above Q1 and 1.5 times the interquartile range below Q3.

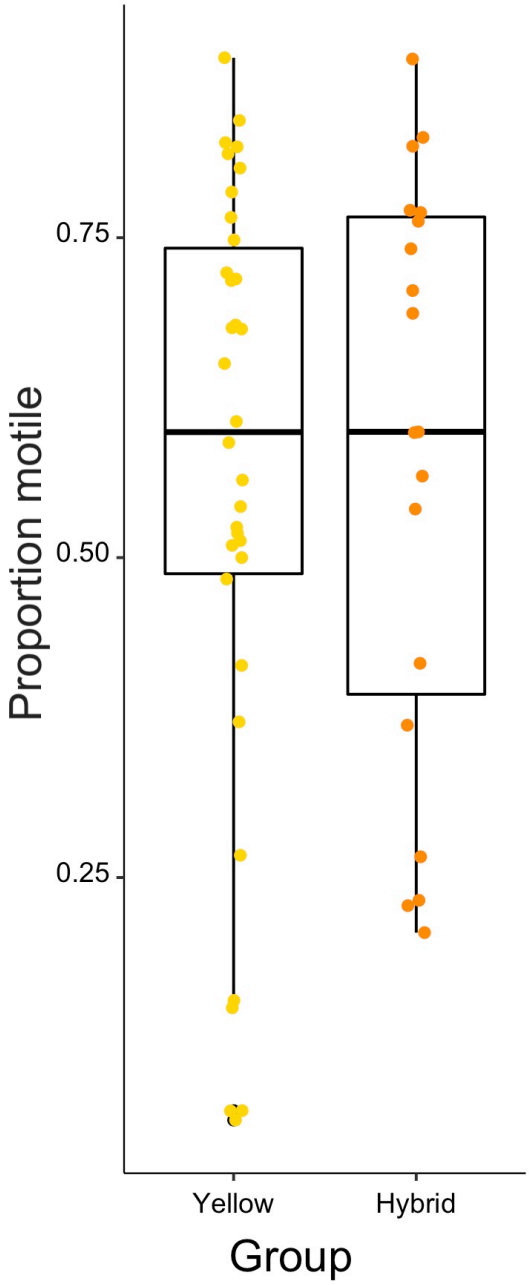


Figure 7: The proportion of motile sperm for Mixed Hybrid males and yellow males. Dots are individual males, with boxplots reflecting the median, the 25th percentile (Q1) and 75th percentile (Q3), with whiskers 1.5 times the interquartile range above Q1 and 1.5 times the interquartile range below Q3.

Table 3: Model outputs for assessing sperm motility values between Mixed Hybrid males and yellow males. A. Generalised linear mixed models with Goussian errors to examine whether group (Mixed Hybrid or yellow) significantly predicted sperm swimming velocity (VCL) or velocity of the top 10% fastest sperm from an ejaculate, with Male ID as a random effect, and B. a linear model with Goussian errors testing whether group significantdi predictedi the proportion of motile sperm.

A:VCL	Estimate	Std. Error	df	t-value	Pr
Intercept	76.89	2.61	49.81	29.46	
Group	0.77	4.21	49.72	0.18	0.86

	Var	St Dev	N	ICC
Marg R^2	0.14			<0.001
Male ID	214.6	14.65	52	0.35
Resid	405.9	20.15		
Cond R^2			13814	0.35

Top 10% VCL	Estimate	Std. Error	Df	t-value	Pr
Intercept	110.33	2.73	49.27	40.52	
Group	0.89	4.39	49.15	0.203	0.84

	Var	St Dev	N	ICC
Marg R^2	0.19			<0.001
Male ID	233.4	15.28	52	0.83
Resid	46.3	6.82		
Cond R^2			1395	0.83

B: Prop	Estimate	Std. Error	t-value	P-value
Intercept	0.502	0.046	10.92	
Group	0.022	0.073	0.305	0.761

Table 4: Model outputs from testing whether there is a relationship bewteen sperm morphology and velocity in this study. A) Output from generalise linear mixed models to examine whether sperm morphological components (head, midpiece and tail length) significantly predict the average sperm swimming velocity per male, or the average velocity of the top 10% fastest sperm per male. These are the reduced models, resulting from a stepwise reduction based on p-value >0.05, where the original models started with full three way interactions between head, midpiece and tail components. B) To facilitate comparison with a previous study we also tested for a relationship in yellow males between sperm morphological components and sperm swimming velocity (VCL) using separate linear models to test each predicted sperm length or ratio.

A:

VCL~morph	Estimate	Std. Error	t value	P-value
Intercept	77.70	2.17	35.86	
Head	5.72	4.59	1.25	0.22
Midpiece	4.11	4.61	0.89	0.38
Flagellum	-1.79	4.45	-0.40	0.69

Top 10% VCL~morph	Estimate	Std. Error	t value	P-value
Intercept	110.82	2.26	48.99	
Head	6.02	4.79	1.26	0.22
Midpiece	5.19	4.81	1.08	0.28
Flagellum	0.34	4.65	0.073	0.94

B:

Sperm trait	Coefficient (± SE)	t	P-value
Head length	5.40 ± 4.42	1.22	0.23
Midpiece length	-0.41 ± 1.06	-0.38	0.71
Flagellum length	-0.44 ± 1.18	-0.38	0.71
Total sperm length	-0.09 ± 1.23	-0.08	0.94
Flagellum:head	-7.37 ± 8.12	-0.91	0.37
Midpiece:flagellum	-13.16 ± 67.72	-0.19	0.84

Discussion

Recent studies have demonstrated that PMPZ isolation can play an important role early in the speciation progress (Cramer *et al.*, 2016a; McDonough *et al.*, 2016; Turissini *et al.*, 2017; Garlovsky & Snook, 2018). The two subspecies of long-tailed finch that we have examined form a narrow hybrid zone in the wild and hybrids (Hooper *et al.*, 2018) appear to suffer decreased sperm-female compatibility as suggested by a detailed captive study that found that significantly lower numbers of sperm reached the ovum in crosses between hybrids and 'pure' males than between 'pure' males (Hurley *et al.*, *in press*). Here, we have followed up on this finding by addressing whether there are significant differences in the sperm of males from pure and admixed backgrounds. This is one of the first studies to examine sperm morphology and velocity in Passerine hybrids that are found in a naturally occurring wild hybrid zone (but see also Ålund *et al.*, 2013; Cramer *et al.*, 2014). We found evidence for two independent differences in sperm morphology between the two 'pure' subspecies and our Mixed Hybrid group. Mixed Hybrid males had an average midpiece length roughly equal to yellow males but significantly longer than red male sperm, and Mixed Hybrid males had significantly longer average flagella (and total sperm length) than both 'pure' subspecies. Below we discuss these differences in the context of the expected Z-chromosome composition of these groups. Despite these differences in morphology, we found no evidence that Mixed Hybrid males had a lower sperm performance, based on our measurements of sperm velocity and the proportion of motile sperm in vitro. Finally, we found no evidence for a relationship between sperm morphology and sperm velocity, unlike a previous study in this species (Rowe *et al.*, 2015b), adding further weight to the idea that the empirical study of the relationship between sperm form and function across studies of Passerine birds is littered with inconsistent results, and in need of further consideration (reviewed below).

Our results for sperm morphology contrast to what was found by Rowe *et al*. (2015b), who also investigated sperm phenotype of the two long tailed finch subspecies (but without samples from hybrids). Rowe *et al*., (2015b) did not find the significant difference in sperm midpiece length identified in this study but did find that yellow males had significantly longer flagellum length and total sperm length (these are highly correlated), and ratio of midpiece:flagellum (and head width, which was not assessed here). As the two studies followed the same methodology to assess sperm morphology, the most obvious potential explanation for this discrepancy is the increased sample size of this study (N=37 reds, N=35 yellows; Rowe *et al*., 2015: N=23 reds, N=19 yellows). To more quantitatively determine the sample size required to attain representative measures of sperm morphology per group, we performed a resampling analysis of the data that we collected on head, midpiece, flagellum and total sperm length for yellow and red males separately. The resulting figures (Figure 8) suggest that while sampling 20 males is well into the asymptote of the curve for

all components in both subspecies, the 95% confidence interval is still notably larger (~twice the size) than when the sample size is 30 males (Figure 8). Specifically, the 95% confidence intervals (CIs) for midpiece length have a spread of about 4 µm in reds, and 2 µm in yellows, and for flagellum length (and total length) about 2 µm in reds and 1 µm in yellows. Rowe *et al.,* (2015b) found that flagellum (and total) length was about 3.5 µm different between the subspecies and hence is right on the borderline of what we expect could be detected with that sample size. In comparison, however, the larger sample size of the present study is expected to have much greater power to detect differences. Visually inspecting the spread of morphological data from this study (Figure 5), however, suggests that group averages could be disproportionately influenced by the presence or absence of just 5 or 6 individuals with extreme morphology, which may be what is driving the different findings of this study and that of Rowe *et al*., (2015b). That is, in a smaller sample, results can be disproportionately affected by the selection or omission of a few outliers.

We found a number of interesting differences in the sperm morphology of Mixed Hybrid and each 'pure' subspecies group, particularly when considering the corresponding arrangement of Z-chromosomes (Figure 1). Genetic differentiation between the two subspecies on the autosomes appears very low ($F_{ST} \approx 0$) whereas approximately 75% of length of the Z-chromosome is different between the subspecies ($F_{ST} \approx 0.4$; Hooper *et al.,* 2018), and because of this high differentiation we will refer to them as 'red' and 'yellow' Z-chromosomes, respectively. We expect this high differentiation to potentially impact sperm morphology in the long-tailed finch hybrid zone, as it was recently identified that in the closely related zebra finch, the Z chromosome accounts for 67-90% of variance in sperm head, midpiece and flagellum length (Kim *et al.*, 2017). In the present study the only difference in sperm morphology between the 'pure' groups was that red males have significantly shorter sperm midpieces than do yellow males (Figure 5). Surprisingly, we did not find that Mixed Hybrid males had a midpiece length that was intermediate between the two, but instead was on average equal to that of 'pure' yellow males. This high value for Mixed Hybrid does not appear to be simply driven by the Backcross 2 males (that are likely to contain a disproportionately high number of yellow Z-chromosomes), as F2 hybrid males that contain roughly equal numbers of red and yellow Z-chromosomes also appear to have yellow-like midpiece lengths (Figure 4). While they were not statistically compared to 'pure' males due to low sample size, hybrid F1s do appear to have intermediate midpiece length, (Figure 4). Hybrid F1s are the only 'hybrid' group in which we can be sure that no recombination has occurred between the yellow and red Z-chromosomes, tentatively suggesting that the long midpiece of Mixed Hybrid males results from recombination mixing genetic material between red and yellow Z-chromosomes. The other difference we found

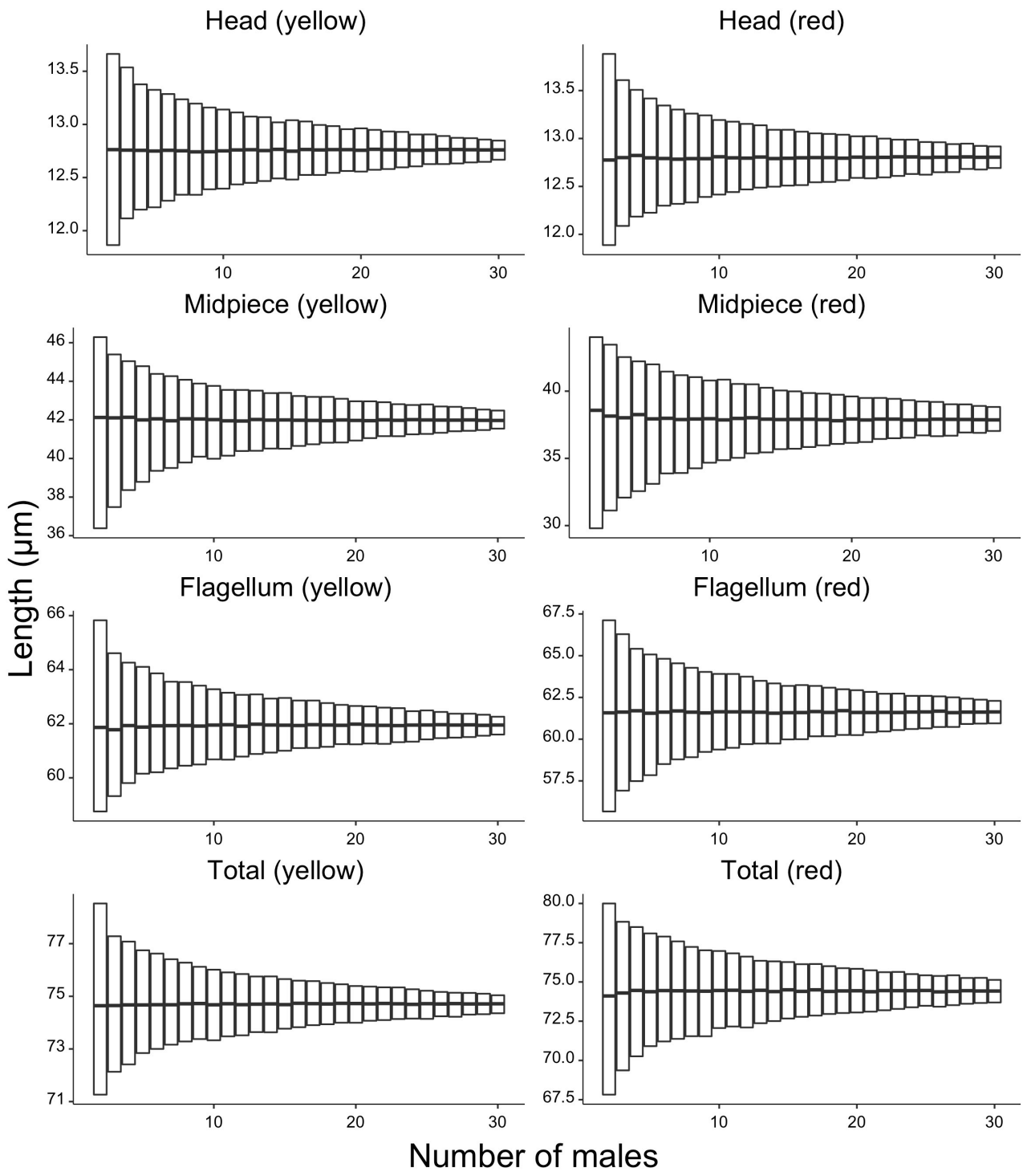


Figure 8: The output of among-male resampling analysis for sperm velocity (VCL). Boxes represent 95% confidence intervals (CI).

was that Mixed Hybrid males had significantly longer flagellum and total sperm length than both 'pure' subspecies. As total sperm length is comprised of flagellum length and head length (that did not differ between the groups), this difference in total sperm length is likely a direct result of the difference in flagellum length. Mixed Hybrid flagellum length thus clearly does not appear to be an intermediate phenotype between red and yellow males and may be an example of a transgressive phenotype. In this case visual examination of the data when all four hybrid and backcross groups were parsed out (Figure 4) suggests that even unrecombined F1 hybrid males had unusually long flagellum length compared to 'pure' males. The notable overlap in flagellum length of some Mixed Hybrid male and the 'pure' males likely results from the array of different Z-chromosome combinations likely present in the Mixed Hybrid group, including some effectively 'pure' individuals (based on their Z-chromosomes). In contrast, these same Mixed Hybrid males had midpieces that had relatively little variation among males. Overall it appears that midpiece and flagellum length are responding differently in the array of Z-chromosome types held by males in the Mixed Hybrid group.

A possible alternative explanation for the differences in sperm morphology found in this study is that they are driven by founder effects, i.e. that the F1 hybrid population was founded by individuals that by chance had long flagella and yellow-like midpieces. We find this unlikely, as there is no indication from this data of this (although many of the original F1s were not still alive at time of sampling), and the most extreme flagellum lengths found in Mixed Hybrid individuals are longer than the longest red sperm, and few yellow males (who made half the genetic contribution to F1s) have very long sperm. It would be have been valuable to have sperm samples from all the original founding individuals and F1 hybrid males, but unfortunately at that time we were not working on sperm, and did not have the techniques to collect it. To more precisely investigate these differences in sperm morphology and confirm whether they are in fact related to the male's Z-chromosomal make-up, it will be necessary to genetically identify the Z-chromosome type carried by each male. This would allow us to parse out the Mixed Hybrid group by chromosomal karyotype, as this group currently likely contains males ranging from two pure yellow Z-chromosomes to two admixed Z-chromosomes that have been through two generations of recombination (see Figure 1). By identifying male karyotype, we could then compare across actual groups to investigate how karyotype influences sperm morphology. One approach to this would be to designing a SNPchip containing ancestrally informative single nucleotide polymorphism (SNP) markers along the Z-chromosome (as done by Knief *et al.*, 2017). Each marker could then indicate if a male were 'red' or 'yellow' at that location on the Z-chromosome, also giving an indication of admixture if a high density of markers were used. This method could be valuably supplemented by employing an RNAseq approach to compare gene expression in the testes of males with different

phenotypes or chromosomal combinations (as per Dean & Mank, 2014; Kim *et al.*, 2017), which could then be mapped onto the zebra finch reference genome (as per Hooper *et al*., 2018). The long-tailed finch system is well placed for gene expression work, as captive individuals across groups can be kept in controlled conditions, and sampled at specific sampling points, eliminating noise created by the many factors that can influence gene expression. This could be used for a number of valuable research areas: A) By comparing males with long and short sperm we can investigate whether these are located on the Z chromosome, and identify candidate genes that are influential in sperm length (and compare to findings in the zebra finch); 2) by comparing different sperm lengths within and among groups (i.e. 'pure' males and hybrid males) we can identify whether the same genetic mechanisms are causing an increase in sperm flagellum length within subpecies and in hybrids, and more generally how hybrid gene expression relates to the 'pure' subspecies; 3) this could differentiate between novel genetic combinations being expressed in hybrids and whether the same gene combinations were just upregulated in expression – same when comparing between long and short sperm within a subspecies. This may generate novel hypotheses about the causes of lower fertility success of these subspecies.

It is unclear whether these differences in morphology are likely to have a direct impact on male fertilisation success, and thus influence reproductive isolation in the long-tailed finch hybrid zone (Hooper *et al*., 2018). Both sperm midpiece and flagellum have previously been found to correlate with in vitro swimming speed for passerine species (Table 1) including in this species (Rowe *et al*., 2015b), however in this study we found no relationship between sperm morphology and velocity, as we will discuss below. Sperm length may also influence fertilisation success through compatibility with female ST size: total sperm length coevolves with SST size across taxa, which has been suggested to potentially drive incompatibility among diverged taxa (Briskie *et al.*, 1997; Miller & Pitnick, 2002; Birkhead & Brillard, 2007). It is possible that the highly elongated sperm of some Mixed Hybrid males would have resulted in decreased success in entering or remaining in female SSTs, provided that there is a tight relationship between sperm length and SST in this species (which is currently unknown but would be a good target for future work). For many Mixed Hybrid males, however, their total sperm length falls within the 'normal' range that is stored by female SSTs. Whether sperm length influences fertilisation success could be tested explicitly via fertilisation trials in which male 'competitors' are paired based on differing in sperm length, (see (Bennison *et al.*, 2014) but matching Z-chromosome type. Matching genetic background would act to control for other traits, such as reproductive proteins (Swanson & Vacquier, 2002), that may differ between Z-chromosome types. This should be feasible, as there were multiple males in our study whose sperm do not overlap, although selection lines to produce many males with extreme sperm lengths (as done in the zebra finch; (Bennison *et al.*, 2014) would increase our power. The

successful interaction of 'long' and 'short' sperm with the female reproductive tract (i.e. successfully passing through the vagina, entering and exiting storage, and being transported to the ovum) could be determined by counting the number of 'long' and 'short' sperm respectively that are caught in the perivitelline layer (PVL) of the ovum (Bennison *et al*., 2016).

The among male coefficient of variation for total sperm length (CV_{am}) that we have found in the long-tailed finch, is similar to the highest found across passerine species (Lifjeld *et al*., 2010). CV_{am} has been demonstrated to associate with degree of extra pair paternity across populations, where high sperm competition is predicted to eliminate the variability of sperm. For example, zebra finches have high CV_{am} (Calhim *et al.*, 2007) and one of the lowest rates of EPY in any wild passerine (1.5%; Griffith *et al.*, 2010). Despite their similarity in CV_{am} predicting a similar level of EPY in zebra finches and long-tailed finches, the only study of EPY rate in wild long-tailed finches found that extra-pair males sired 12.8% of 391 offspring (van Rooij *et al.*, 2016). Unfortunately, neither the sperm CV_{am} or the rate of EPY has been assessed for any other Australian grassfinch, so it is currently unknown how well this correlation between CV_{am} and rate of EPY works in this family. It is interesting to note, however, that the site of this field study was recently identified as being within the phenotypically cryptic (i.e. for bill colour) genetic hybrid zone of the long-tailed finch (Hooper *et al.*, 2018). This raises the possibility that extra-pair parentage in the studied population may have been heightened in response to the risk of heterosubspecies crosses (Griffith & Immler, 2009; Griffith, 2010), as also found in the hybrid zone of the pied and collared flycatchers (Veen *et al.*, 2001).

Sperm swimming velocity appears to be both important for fertilisation success (Pizzari & Parker, 2009; Simmons & Fitzpatrick, 2012) and susceptible to environmental and genetic perturbations (reviewed in Reinhardt *et al.*, 2015). Despite this, we found no evidence that Mixed Hybrid males suffered from decreased sperm performance in vitro. Low sperm performance of hybrid males therefore does not appear to be responsible for the findings of a previous study in this system that found that pairings between F1 hybrid and 'pure' individuals had significantly fewer sperm reaching the perivitelline layer (PVL) of the ovum than did within-subspecies pairs (Hurley et al. *in press*).

While our results indicate that Mixed Hybrid males have fully functioning sperm in vitro, this represents a highly simplistic and artificial environment compared to the physically and chemically complex environment of the female reproductive tract (Pitnick *et al.*, 2009b). A valuable approach that gets closer to assessing sperm performance in vivo, is to assess sperm swimming speed within female reproductive tract fluid, a technique that has recently been pioneered in passerines (Cramer

et al., 2014; 2016a; b) and other taxa (guppies Gasparini & Pilastro, 2011; salmon and trout Yeates *et al.*, 2013; a mussel Lymbery *et al.*, 2017). The sperm of hybrid males may react differently in this environment, and furthermore 'pure' sperm performance could also be tested in homospecific and heterospecific female reproductive tract fluid (as per Cramer *et al.*, 2016a). There are three naturally hybridising passerine species in which hybrid sperm performance has now been assessed. In both the hybrids between the long-tailed finch sub-species (here) and hybrids between the house and Spanish sparrow (preliminary data from only two males; Cramer *et al.,* 2014) appear to have full sperm velocity and motility values in vitro, whereas the hybrid males produced by heterospecific crosses between the pied and collared flycatcher either do not produce sperm, or have extremely disfigured and immotile sperm cells (Ålund *et al.*, 2013). If hybrid dysfunction accumulates at a similar rate across species of passerines, this would suggest that hybrid sperm immobility arises at some period between a divergence of 0.3 MYA (this species and the house-Spanish sparrows: Jennings & Edwards, 2005; Cramer *et al.*, 2014) and roughly 1-2 MYA (pied and collared flycatchers; Qvarnström *et al.*, 2010; Ellegren *et al.*, 2012).

Despite the degree of variation in the sperm morphology that we measured, we found no significant relationship between head, midpiece or flagellum lengths (or their interaction) and either the average sperm swimming speed (VCL) for a male or the average VCL of the fastest 10% of sperm from a male. To facilitate a direct comparison with the study by (Rowe *et al.*, 2015b) we also performed their statistical approach, which is also more commonly adopted when investigating this research question: i.e. using separate linear models between sperm velocity and each morphological component or ratio of interest (Table 1). All these models also found no significant relationship. This is in direct contrast to the study by Rowe *et al.,* (2015b), who found a significant relationship between VCL and both midpiece length and the midpiece:flagellum ratio in yellow males, and between VCL and all of midpiece length, flagellum length, total sperm length, head:flagellum ratio and midpiece:flagellum ratio in red males. While this is a striking inconsistency, it does not appear as surprising when placed in the context of similar studies in passerines (summarised in Table 1). Despite receiving more attention than in any other taxonomic group, the results of published studies remain highly inconsistent (Table 1). One possible reason this study found no significant relationship when a previous study did could be due to the media used. While a number of recent studies have adopted assessing sperm velocity in phosphate buffer solution (PBS) rather than Dulbecco's Modified Eagle Medium (DMEM; e.g. Cramer *et al.*, 2016a; Sætre *et al.*, 2018), this has only been extended into studies investigating the relationship between sperm morphology and velocity in passerines in two cases (Immler *et al.*, 2010; Cramer *et al.*, 2014; see Table 1). One of these studies (Immler *et al.*, 2010) found inconsistent results within their study, finding a relationship between morphology and velocity at one time point and not another. The other (Cramer

et al., 2014) diluted some samples in PBS and some in DMEM and is the only study to have found a significantly negative relationship between morphological component lengths and swimming velocity (Table 1). Sperm will be less stimulated in PBS compared to DMEM that is specifically designed to keep sperm alive and motile. While purely speculative, it is possible that when sperm are performing below their full capacity (i.e. in the more passive media of PBS) it might create noise that masks any real relationship between sperm morphology and velocity. It would be valuable to test whether ejaculates diluted in PBS swim *consistently* slower than if they are diluted into DMEM. As mentioned above, it is also possible that in vitro conditions are just too dissimilar to the female reproductive tract, so that any real relationship between sperm morphology and sperm velocity in vivo (as we might expect there to be) are not apparent.

By far the clearest relationship between morphology and motility has been found in the zebra finch (Table 1), where sperm midpiece and flagellum significantly predict sperm velocity, although head length, head:flagellum ratio, and midpiece:flagellum ratio have been found to correlate with sperm velocity (Bennison *et al.,* 2016; Mossman *et al.,* 2009; Knief *et al.,* 2016). These studies have had very large sample sizes (N>100), and the high CV_{am} of zebra finches (Calhim *et al.,* 2007) provides a spread of morphological values that results in greater ability to identify any existing relationship. One research group even developed selection lines for long and short sperm, giving even more power to their analyses. As suggested by others (Lüpold *et al.,* 2009; Cramer *et al.,* 2014) this overall trend suggests that the importance of sperm morphology on velocity may be reasonably weak and that other factors, such as ejaculate quality or interactions with the female reproductive tract, primarily determine swimming velocity.

In conclusion, this study is one of very first to investigate hybrid sperm phenotype in a naturally hybridising system of passerine birds. Sperm midpiece length was significantly different between the two subspecies, but intriguingly our Mixed Hybrid group was not intermediate between them but was aligned with yellow males. While the two 'pure' subspecies did not significantly differ in flagellum length, Mixed Hybrid males had significantly longer flagellum than either, possibly a transgressive phenotype. We found no evidence that Mixed Hybrid males had low sperm velocity or proportion of motile sperm in vitro, or for a relationship between sperm morphology and velocity. There are number of valuable approaches that could be employed in future research, particularly testing sperm behaviour in contexts that are closer to that they experience *in vivo*, and a number of valuable genomic approaches that could provide insight into how postcopulatory processes are influenced by hybridisation impact the speciation process.

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N., *et al.* 2013. Hybridization and speciation. *J. Evol. Biol.* **26**: 229–246.

Ackermann, R.R., Mackay, A. & Arnold, M.L. 2016. The Hybrid Origin of "'Modern'" Humans. *Evolutionary Biology* **43**: 1–11. Springer US.

Anderson, M.J., Dixson, A.S. & Dixson, A.F. 2006. Mammalian sperm and oviducts are sexually selected: evidence for co-evolution. *J. Zoology* **270**: 682–686.

Ålund, M., Immler, S., Rice, A.M. & Qvarnström, A. 2013. Low fertility of wild hybrid male flycatchers despite recent divergence. *Biol Letters* **9**: 20130169–20130169. The Royal Society.

Bakst, M.R., Wishart, G. & Brillard, J.-P. 1994. Oviducal sperm selection, transport, and storage in poultry. *Poultry Science Review* **5**: 117–143.

Bennison, C., Hemmings, N., Brookes, L., Slate, J. & Birkhead, T.R. 2016. Sperm morphology, adenosine triphosphate (ATP) concentration and swimming velocity: unexpected relationships in a passerine bird. *Proc. R. Soc. B* **283**: 20161558–6.

Bennison, C., Hemmings, N., SLATE, J. & Birkhead, T.R. 2015. Long sperm fertilize more eggs in a bird. *Proc. R. Soc. B* **282**: 20141897–20141897. The Royal Society.

Bennison, C., Hemmings, N., SLATE, J. & Birkhead, T.R. 2014. Long sperm fertilize more eggs in a bird. *Proc. R. Soc. B* **282**: 20141897–20141897. The Royal Society.

Bhattacharyya, T., Gregorova, S., Mihola, O., Anger, M., Sebestova, J., Denny, P., *et al.* 2013. Mechanistic basis of infertility of mouse intersubspecific hybrids. *PNAS* E468–E477.

Birkhead, T.R. & Brillard, J.-P. 2007. Reproductive isolation in birds: postcopulatory prezygotic barriers. *Trends Ecol Evol* **22**: 266–272.

Birkhead, T.R. & Hunter, F.M. 1990. Mechanisms of sperm competition. *Trends Ecol Evol* **5**: 48–52.

Birkhead, T.R. & Møller, A.P. 1998. *Sperm competition and sexual selection*.

Birkhead, T.R., Martinez, J.G., Burke, T. & Froman, D.P. 1999. Sperm mobility determines the outcome of sperm competition in the domestic fowl. *Proc. R. Soc. B* **266**: 1759–1764.

Boryshpolets, S., Kowalski, R.K., Dietrich, G.J., Dzyuba, B. & Ciereszko, A. 2013. Different computer-assisted sperm analysis (CASA) systems highly influence sperm motility parameters. *Theriogenology* **80**: 758–765. Elsevier Inc.

Boschetto, C., Gasparini, C. & Pilastro, A. 2010. Sperm number and velocity affect sperm competition success in the guppy (Poecilia reticulata). *Behav Ecol Sociobiol* **65**: 813–821.

Briskie, J.V. & Montgomerie, R. 1993. Patterns of sperm storage in relation to sperm competition in passerine birds. *The Condor* **95**: 442–454.

Briskie, J.V. & Montgomerie, R. 1992. Sperm size and sperm competition in birds. *Proc. R. Soc. B* **247**: 89–95.

Briskie, J.V., Montgomerie, R. & Birkhead, T.R. 1997. The evolution of sperm size in birds. *Evolution* 937–945.

Butlin, R.K., Debelle, A., Kerth, C., Snook, R.R. & The Marie Curie SPECIATION Network. 2012. What do we need to know about speciation? *Trends Ecol Evol* **27**: 27–39.

Calhim, S., Immler, S. & Birkhead, T.R. 2007. Postcopulatory Sexual Selection Is Associated with Reduced Variation in Sperm Morphology. *PLoS ONE* **2**: e413–6.

Connallon, T. & Clark, A.G. 2010a. SEX LINKAGE, SEX-SPECIFIC SELECTION, AND THE ROLE OF RECOMBINATION IN THE EVOLUTION OF SEXUALLY DIMORPHIC GENE EXPRESSION. *Evolution* **64**: 3417–3442.

Connallon, T. & Clark, A.G. 2010b. Sex linkage, sex-specific selectionn, and the role of recombination in the evolution of sexually dimorphic gene expression. *Evolution* **64**: 3417–3442. Wiley/Blackwell (10.1111).

Coyne, J.A. 2018. "Two Rules of Speciation" revisited. *Mol Ecol* **130**: 113–4. Wiley/Blackwell (10.1111).

Coyne, J.A. & Orr, H.A. 1989. Patterns of speciation in *Drosophila*. *Evolution* **42**: 362–381.

Coyne, J.A. & Orr, H.A. 2004. *Speciation*.

Cramer, E.R.A., Ålund, M., McFarlane, S.E., Johnsen, A. & Qvarnström, A. 2016a. Females discriminate against heterospecific sperm in a natural hybrid zone. *Evolution* **70**: 1844–1855.

Cramer, E.R.A., Laskemoen, T., Stensrud, E., Rowe, M., Haas, F., Lifjeld, J.T., *et al.* 2014. Morphology-function relationships and repeatability in the sperm of Passersparrows. *Journal of Morphology* **276**: 370–377.

Cramer, E.R.A., Stensrud, E., Marthinsen, G., Hogner, S., Johannessen, L.E., Laskemoen, T., *et al.* 2016b. Sperm performance in conspecific and heterospecific female fluid. *Ecol Evol* **6**: 1363–1377.

Dean, R. & Mank, J.E. 2014. The role of sex chromosomes in sexual dimorphism: discordance between molecular and phenotypic data. *J. Evol. Biol.* **27**: 1–11.

Denk, A.G., Holzmann, A., Peters, A., Vermeirssen, E.L.M. & Kempenaers, B. 2005. Paternity in mallards: effects of sperm quality and female sperm selection for inbreeding avoidance. *BEHECO* **16**: 825–833.

Devigili, A., Fitzpatrick, J.L., GASPARINI, C., Ramnarine, I.W., PILASTRO, A. & Evans, J.P. 2017. Possible glimpses into early speciation: the effect of ovarian fluid on sperm velocity accords with post-copulatory isolation between two guppy populations. *J. Evol. Biol.* **31**: 66–74.

Dixon, A. & Birkhead, T.R. 1997. Reproductive anatomy of the reed bunting: a species which exhibits a high degree of sperm competition through extra-pair copulations. *The Condor* **4**: 966–969.

Dixon, S.M., Coyne, J.A. & Noor, M.A.F. 2003. The evolution of conspecific sperm precedence in Drosophila. *Mol Ecol* **12**: 1179–1184.

Donnelly, E.T., Lewis, S.E.M., McNally, J.A. & Thompson, W. 1998. In vitro fertilization and pregnancy rates: the influence of sperm motility and morphology on IVF outcome. *Fertility and*

Sterility **70**: 305–314.

Dybas, L.K. & Dybas, H.S. 1981. Coadaptation and taxonomic differentiation of sperm and spermathecae in featherwind beetles. *Evolution* **35**: 168–174.

Ellegren, H. 2011. Emergence of male-biased genes on the chicken Z-chromosome: Sex-chromosome contrasts between male and female heterogametic systems. *Genome Res.* **21**: 2082–2086.

Ellegren, H., Smeds, L., Burri, R., Olason, P.I., Backström, N., Kawakami, T., *et al.* 2012. The genomic landscape of species divergence in Ficedula flycatchers. *Nature* **491**: 756–760. Nature Publishing Group.

Eroukhmanoff, F., Rowe, M., Cramer, E.R.A., Haas, F., Hermansen, J.S., Runemark, A., *et al.* 2016. Experimental evidence for ovarian hypofunction in sparrow hybrids. *Avian Research* 1–5. BioMed Central.

Fitzpatrick, B.M. 2004. Rates of evolution of hybrid inviability in birds and mammals. *Evolution* **58**: 1865–1870.

Fitzpatrick, J.L. & Lüpold, S. 2014. Sexual selection and the evolution of sperm quality. *MHR: Basic science of reproductive medicine* **20**: 1180–1189.

Fitzpatrick, J.L., Garcia-Gonzalez, F. & Evans, J.P. 2010. Linking sperm length and velocity: the importance of intramale variation. *Biol Letters* **6**: 797–799. The Royal Society.

Fitzpatrick, J.L., Montgomerie, R., Desjardins, J.K., Stiver, K.A., Kolm, N. & Balshine, S. 2009. Female promiscuity promotes the evolution of faster sperm in cichlid fishes. *PNAS* **106**: 1128–1132.

Froman, D.P., Feltmann, A.J., Rhoads, M.L. & Kirby, J.D. 1999. Sperm Mobility: A Primary Determinant of Fertility in the Domestic Fowl (Gallus domesticus). *Biology of Reproduction* **61**: 400–405.

Froman, D.P., Pizzari, T., Feltmann, A.J., Castillo-Juarez, H. & Birkhead, T.R. 2002. Sperm mobility: mechanisms of fertilizing efficiency, genetic variation and phenotypic relationship with male status in the domestic fowl, Gallus gallus domesticus. *Proc. R. Soc. B* **269**: 607–612.

Gage, M.J.G., Macfarlane, C.P., Yeates, S., Ward, R.G., Searle, J.B. & Parker, G.A. 2004. Spermatozoal Traits and Sperm Competition in Atlantic Salmon. *Curr Biol* **14**: 44–47.

Garlovsky, M.D. & Snook, R.R. 2018. Persistent postmating, prezygotic reproductive isolation between populations. *Ecol Evol* **8**: 9062–9073.

Gasparini, C. & Pilastro, A. 2011. Cryptic female preference for genetically unrelated males is mediated by ovarian fluid in the guppy. *Proc Biol Sci* **278**: 2495–2501. The Royal Society.

Geyer, L.B. & Palumbi, S.R. 2005. Conspecific sperm precendence in two species of tropical sea urchins. *Evolution* **59**: 97–105.

Giaretta, E., Munerato, M., Yeste, M., Galeati, G., Spinaci, M., Tamanini, C., *et al.* 2017. Implementing an open-access CASA software for the assessment of stallion sperm motility: Relationship with other sperm quality parameters. *Animal Reproduction Science* **176**: 11–19. Elsevier B.V.

Gomendio, M. & Roldan, E.R.S. 1993. Coevolution between male ejaculates and female

reproductive biology in eutherian mammals. *Proc. R. Soc. B* **252**: 7–12.

Gregory, P.G. & Howard, D.J. 1994. A postinsemination barrier to fertilisation isolates two closely related ground crickets. *Evolution* **48**: 705–710.

Griffith, S.C. 2010. The role of multiple mating and extra-pair paternity in creating and reinforcing boundaries between species in birds. *EMU* **110**: 1–9.

Griffith, S.C. & Hooper, D.M. 2017. Geographical variation in bill colour in the Long- tailed Finch: evidence for a narrow zone of admixture between sub-species. *EMU* **00**: 1–10. Taylor & Francis.

Griffith, S.C. & Immler, S. 2009. Female infidelity and genetic compatibility in birds: the role of the genetically loaded raffle in understanding the function of extrapair paternity. *J Avian Biol* **40**: 97–101.

Griffith, S.C., Holleley, C.E., Mariette, M.M., Pryke, S.R. & Svedin, N. 2010. Low level of extrapair parentage in wild zebra finches. *Anim Behav* **79**: 261–264. Elsevier Ltd.

Helfenstein, F., Podevin, M. & Richner, H. 2010. Sperm morphology, swimming velocity, and longevity in the house sparrow Passer domesticus. *Behav Ecol Sociobiol* **64**: 557–565. Springer-Verlag.

Hemmings, N. & Birkhead, T.R. 2017. Differential sperm storage by female zebra finches Taeniopygia guttata. *Proc Biol Sci* **284**: 20171032–6. The Royal Society.

Hemmings, N., Bennison, C. & Birkhead, T.R. 2016. Intra-ejaculate sperm selection in female zebra finches. *Biol Letters* **12**: 20160220–6.

Hoffmann, A.A. & Rieseberg, L.H. 2008. Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annu. Rev. Ecol. Evol. Syst.* **39**: 21–42.

Hogner, S., Laskemoen, T., Lifjeld, J.T., Pavel, V., Chutný, B., García, J., *et al.* 2013. Rapid sperm evolution in the bluethroat (Luscinia svecica) subspecies complex. *Behav Ecol Sociobiol* **67**: 1205–1217.

Hooper, D.M., Griffith, S.C. & Price, T.D. 2018. Sex chromosome inversions enforce reproductive isolation across an avian hybrid zone. *Mol Ecol* 1–47. Wiley/Blackwell (10.1111).

Humphries, S., Evans, J.P. & Simmons, L.W. 2008. Sperm competition: linking form to function. *BMC Evol Biol* **8**: 319–11. BioMed Central.

Hurley, L.L., McDiarmid, C.S., Friesen, C.R., Griffith, S.C. & Rowe, M. 2018. Experimental heatwaves negatively impact sperm quality in the zebra finch. *Proc Biol Sci* **285**: 20172547–9. The Royal Society.

Immler, S., Pryke, S.R., Birkhead, T.R. & Griffith, S.C. 2010. Pronounced within-individual plasticity in sperm morphometry across social environments. *Evolution* **64**: 1634–1643.

Ishishita, S., Kinoshita, K., Nakano, M. & Matsuda, Y. 2016. Embryonic development and inviability phenotype of chicken-Japanese quail F1 hybrids. *Nature Publishing Group* 1–12. Nature Publishing Group.

Jennings, B. & Edwards, S.V. 2005. Speciation history of Australian grass finches (Poephila) inferred from thirty gene trees. *Evolution* **59**: 2033–2047.

Johnson, S.L. & Gemmell, N.J. 2012. Are old males still good males and can females tell the difference? *Bioessays* **34**: 609–619.

Jones, M.R., Mills, L.S., Alves, P.C., Callahan, C.M., Alves, J.M., Lafferty, D.J.R., *et al.* 2018. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* **360**: 1355–1358.

Kawakami, T., Smeds, L., Backström, N., Husby, A., Qvarnström, A., Mugal, C.F., *et al.* 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Mol Ecol* **23**: 4035–4058. Wiley/Blackwell (10.1111).

Kim, K.-W., Bennison, C., Hemmings, N., Brookes, L., Hurley, L.L., Griffith, S.C., *et al.* 2017a. A sex-linked supergene controls sperm morphology and swimming speed in a songbird. *Nature Ecology & Evolution* **1**: 1168–1176.

Kim, K.-W., Bennison, C., Hemmings, N., Brookes, L., Hurley, L.L., Griffith, S.C., *et al.* 2017b. A sex-linked supergene controls sperm morphology and swimming speed in a songbird. *Nature Ecology & Evolution* **1**: 1168–1176. Springer US.

Kleven, O., Fossøy, F., Laskemoen, T., Robertson, R.J., Rudolfsen, G. & Lifjeld, J.T. 2009. Comparative evidence for the evolution of sperm swimming speed by sperm competition and female sperm storage duration in passerine birds. *Evolution* **63**: 2466–2473. Blackwell Publishing Inc.

Knief, U., Forstmeier, W., Pei, Y., Ihle, M., Wang, D., Martin, K., *et al.* 2017. A sex-chromosome inversion causes strong overdominance for sperm traits that affect siring success. *Nature Ecology & Evolution* **1**: 1177–1184. Nature Publishing Group.

Laskemoen, T., Albrecht, T., Bonisoli-Alquati, A., Cepak, J., de Lope, F., Hermosell, I.G., *et al.* 2012. Variation in sperm morphometry and sperm competition among barn swallow (Hirundo rustica) populations. *Behav Ecol Sociobiol* **67**: 301–309.

Laskemoen, T., Kleven, O., Fossøy, F. & Lifjeld, J.T. 2007. Intraspecific variation in sperm length in two passerine species, the Bluethroat Luscinia svecica and the Willow Warbler *Phylloscopus trochilus*. *Ornis Fennica* **84**: 131–139.

Lifjeld, J.T., Laskemoen, T., Kleven, O., Albrecht, T. & Robertson, R.J. 2010. Sperm Length Variation as a Predictor of Extrapair Paternity in Passerine Birds. *PLoS ONE* **5**: e13456–8.

Lopez-Maestre, H., Carnelossi, E.A.G., Lacroix, V., Burlet, N., Mugat, B., Chambeyron, S., *et al.* 2017. Identification of misexpressed genetic elements in hybrids between Drosophila-related species. *Nature Publishing Group* 1–13. Nature Publishing Group.

Losdat, S. & Helfenstein, F. 2018. Relationships between sperm morphological traits and sperm swimming performance in wild Great Tits (Parus major). *J Ornithol* **159**: 805–814. Springer Berlin Heidelberg.

Losdat, S., Chang, S.M. & Reid, J.M. 2014. Inbreeding depression in male gametic performance. *J. Evol. Biol.* **27**: 992–1011.

Lüpold, S., Calhim, S., Immler, S. & Birkhead, T.R. 2009a. Sperm morphology and sperm velocity in passerine birds. *Proc. R. Soc. B* **276**: 1175–1181. The Royal Society.

Lüpold, S., Linz, G.M., Rivers, J.W., Westneat, D.F. & Birkhead, T.R. 2009b. Sperm competition selects beyond relative testes size in birds. *Evolution* **63**: 391–402. Wiley/Blackwell (10.1111).

Lüpold, S., Westneat, D.F. & Birkhead, T.R. 2010. Geographical variation in sperm morphology in the red-winged blackbird (Agelaius phoeniceus). *Evol Ecol* **25**: 373–390.

Lymbery, R.A., Kennington, W.J. & Evans, J.P. 2017. Egg chemoattractants moderate intraspecific sperm competition. *Evolution Letters* **1**: 317–327.

Malo, A.F., Gomendio, M., Garde, J., Lang-Lenton, B., Soler, A.J. & Roldan, E.R.S. 2006. Sperm design and sperm function. *Biol Letters* **2**: 246–249. The Royal Society.

Mayr, E. 1942. *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*.

McDonough, C.E., Whittington, E., Pitnick, S. & Dorus, S. 2016. Proteomics of reproductive systems: Towards a molecular understanding of postmating, prezygotic reproductive barriers. *Journal of Proteomics* **135**: 26–37. Elsevier B.V.

Mendonca, T., Birkhead, T.R., Cadby, A.J., Forstmeier, W. & Hemmings, N. 2018. A trade-off between thickness and length in the zebra finch sperm mid-piece. *Proc Biol Sci* **285**: 20180865. The Royal Society.

Metz, E.C. & Palumbi, S.R. 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Molecular Biology and Evolution* **13**: 397–406.

Miller, G.T. & Pitnick, S. 2002. Sperm-Female Coevolution in Drosophila. *Science* **298**: 1230–1233.

Minder, A.M., Hosken, D.J. & Ward, P.I. 2005. Co-evolution of male and female reproductive characters across the Scathophagidae (Diptera). *J. Evol. Biol.* **18**: 60–69.

Moran, P.A., Ritchie, M.G. & Bailey, N.W. 2017. A rare exception to Haldane's rule: Are X chromosomes key to hybrid incompatibilities? *Heredity* **118**: 554–562. Nature Publishing Group.

Mořkovský, L., Janoušek, V., Reif, J., Rídl, J., Pačes, J., Choleva, L., *et al.* 2018. Genomic islands of differentiation in two songbird species reveal candidate genes for hybrid female sterility. *Mol Ecol* **27**: 949–958. Wiley/Blackwell (10.1111).

Mossman, J., Slate, J., Humphries, S. & Birkhead, T.R. 2009. Sperm morphology and velocity are genetically codetermined in the zebra finch. *Evolution* **63**: 2730–2737. Blackwell Publishing Inc.

Nakagawa, S. 2004. A farewell to Bonferroni: the problems of low statistical power and publication bias. *BEHECO* **15**: 1044–1045.

Nakagawa, S. & Schielzeth, H. 2012. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol Evol* **4**: 133–142.

Nakagawa, S. & Schielzeth, H. 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biol Rev* **136**: 935–956. Blackwell Publishing Ltd.

Oka, A., Mita, A., Sakurai-Yamatani, N., Yamamoto, H., Takagi, N., Takano-Shimizu, T., *et al.* 2004. Hybrid Breakdown Caused by Substitution of the X Chromosome Between Two Mouse Subspecies. *Genetics* **166**: 913–924.

Opatová, P., Ihle, M., Albrechtová, J., Tomášek, O., Kempenaers, B., Forstmeier, W., *et al.* 2015. Inbreeding depression of sperm traits in the zebra finch Taeniopygia guttata. *Ecol Evol* **6**: 295–

304.

Palumbi, S.R. & Metz, E.C. 1991. Strong reproductive isolation between closely related tropical sea urchins (genus *Echinometra*). *Molecular Biology and Evolution* **8**: 227–239.

Pease, J.B., Guerrero, R.F., Sherman, N.A., Hahn, M.W. & Moyle, L.C. 2016. Molecular mechanisms of postmating prezygotic reproductive isolation uncovered by transcriptome analysis. *Mol Ecol* **25**: 2592–2608.

Pitnick, S., Hosken, D.J. & Birkhead, T.R. 2009a. Sperm morphological diversity. In: *Sperm biology: an evolutionary perspective* (T. R. Birkhead, D. J. Hosken, & S. Pitnick, eds), pp. 69–149.

Pitnick, S., Wolfner, M.F. & Suarez, S.S. 2009b. Ejaculate-female and sperm-female interactions. In: *Sperm biology: an evolutionary perspective* (T. R. Birkhead, D. J. Hosken, & S. Pitnick, eds), pp. 247–304.

Pizzari, T. & Parker, G.A. 2009. Sperm competition and sperm phenotype. In: *Sperm biology: an evolutionary perspective* (T. R. Birkhead, D. J. Hosken, & S. Pitnick, eds), pp. 207–245.

Pizzari, T., Worley, K., Burke, T. & Froman, D.P. 2008. Sperm competition dynamics: ejaculate fertilising efficiency changes differentially with time. *BMC Evol Biol* **8**: 332–7.

Presgraves, D.C. 2003. A Fine-Scale Genetic Analysis of Hybrid Incompatibilities in Drosophila. 1–18.

Presgraves, D.C. 2008. Sex chromosomes and speciation in Drosophila. *Trends in Genetics* **24**: 336–343.

Presgraves, D.C., Baker, R.H. & Wilkinson, G.S. 1999. Coevolution of sperm and female reproductive tract morphology in stalk-eyed Fies. *Proc. R. Soc. B* **266**: 1041–1047.

Price, C.S.C., Kim, C.H., Posluszny, J. & Coyne, J.A. 2000. Mechanisms of conspecific sperm precedence in *Drosophila*. *Evolution* **54**: 2028–2037.

Price, T.D. 2008. *Speciation in birds*.

Price, T.D. & Bouvier, M.M. 2002. The evolution of F1 postzygotic incompatibilities in birds. *Evolution* **56**: 2083–2089.

Pryke, S.R., Rollins, L.A. & Griffith, S.C. 2010. Females use multiple mating and genetically loaded sperm competition to target compatible genes. *Science* **329**: 964–267. American Association for the Advancement of Science.

Qvarnström, A., Rice, A.M. & Ellegren, H. 2010. Speciation in Ficedula flycatchers. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **365**: 1841–1852. The Royal Society.

Reinhardt, K., Dobler, R. & Abbott, J. 2015. An Ecology of Sperm: Sperm Diversification by Natural Selection. *Annu. Rev. Ecol. Evol. Syst.* **46**: 435–459.

Rowe, M., Albrecht, T., Cramer, E.R.A., Johnsen, A., Laskemoen, T., Weir, J.T., *et al.* 2015a. Postcopulatory sexual selection is associated with accelerated evolution of sperm morphology. *Evolution* **69**: 1044–1052.

Rowe, M., Griffith, S.C., Hofgaard, A. & Lifjeld, J.T. 2015b. Subspecific variation in sperm morphology and performance in the Long-tailed Finch (Poephila acuticauda). *Avian Research*

1–10. BioMed Central.

Rowe, M., Laskemoen, T., Johnsen, A. & Lifjeld, J.T. 2013. Evolution of sperm structure and energetics in passerine birds. *Proc Biol Sci* **280**: 20122616–20122616. The Royal Society.

Saetre, G.P., Borge, T., Lindroos, K., Haavie, J., SHELDON, B.C., Primmer, C., *et al.* 2003. Sex chromosome evolution and speciation in Ficedula flycatchers. *Proc. R. Soc. B* **270**: 53–59.

Sagga, N. & Civetta, A. 2011. Male-Female Interactions and the Evolution of Postmating Prezygotic Reproductive Isolation among Species of the Virilis Subgroup. *International Journal of Evolutionary Biology* **2011**: 485460–11. Hindawi.

Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**: 671–675. Nature Publishing Group.

Schumer, M., Xu, C., Powell, D.L., Durvasula, A., Skov, L., Holland, C., *et al.* 2018. Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* **360**: 656–660.

Sellier, N., Brun, J.-M., Richard, M.-M., Batellier, F., Dupuy, V. & Brillard, J.-P. 2005. Comparison of fertility and embryo mortality following artificial insemination of common duck females (Anas Platyrhynchos) with semen from common or Muscovy (Cairina Moschata) drakes. *Theriogenology* **64**: 429–439.

Simmons, L.W. & Fitzpatrick, J.L. 2012. Sperm wars and the evolution of male fertility. *Reproduction* **144**: 519–534. Society for Reproduction and Fertility.

Singhal, S., Leffer, E.M., Sannareddy, K., Turner, I., Venn, O., Hooper, D.M., *et al.* 2015. Stable recombination hotspots in birds. *Science* **350**: 928–932.

Soudi, S., Reinhold, K. & Engqvist, L. 2016. Strong cryptic prezygotic isolation despite lack of behavioral isolation between sympatric host races of the leaf beetle Lochmaea capreae. *Evolution* **70**: 2889–2898.

Southern, H.M., Berger, M.A., Young, P.G. & Snook, R.R. 2018. Sperm morphology and the evolution of intracellular sperm-egg interactions. *Ecol Evol* **8**: 5047–5058.

Steele, M.G. & Wishart, G. 1992. Evidence for a species-specific barrier to sperm transport within the vagina of the chicken hen. *Theriogenology* **38**: 1107–1114.

Støstad, H.N., Rekdal, S.L., Kleven, O., Laskemoen, T., Marthinsen, G., Johnsen, A., *et al.* 2016. Weak geographical structure in sperm morphology across the range of two willow warbler Phylloscopus trochilussubspecies in Scandinavia. *J Avian Biol* **47**: 731–741.

Sun, S., Ting, C.-T. & Wu, C.-I. 2004. The Normal Function of a Speciation Gene, Odysseus, and Its Hybrid Sterility Effect. *Science* 1–4.

Supriya, K., Rowe, M., Laskemoen, T., Mohan, D., Price, T.D. & Lifjeld, J.T. 2016. Early diversification of sperm size in the evolutionary history of the old world leaf warblers (Phylloscopidae). *J. Evol. Biol.* **29**: 777–789.

Swanson, W.J. & Vacquier, V.D. 2002. The rapid evolution of reproductive proteins. *Nat Rev Genet* **3**: 137–144.

Sætre, C.L.C., Johnsen, A., Stensrud, E. & Cramer, E.R.A. 2018. Sperm morphology, sperm motility and paternity success in the bluethroat (Luscinia svecica). *PLoS ONE* **13**: e0192644–

18.

Tomášek, O., Albrechtová, J., Němcová, M., Opatová, P. & Albrecht, T. 2017. Trade-off between carotenoid-based sexual ornamentation and sperm resistance to oxidative challenge. *Proc Biol Sci* **284**: 20162444–9. The Royal Society.

Turelli, M. & Orr, H.A. 2000. Dominance, Epistasis and the Genetics of Postzygotic Isolation. 1–17.

Turissini, D.A., McGirr, J.A., Patel, S.S., David, J.R. & Matute, D.R. 2017. The Rate of Evolution of Postmating-Prezygotic Reproductive Isolation in Drosophila. *Molecular Biology and Evolution* **35**: 312–334.

Turner, L.M. & Hoekstra, H.E. 2008. Causes and consequences of the evolution of reproductive proteins. *Int. J. Dev. Biol.* **52**: 769–780.

Turner, L.M. & Schwahn, D.J. 2011. Reduced male fertility is common but highly variable in form and severity in a natural house mouse hybrid zone. *Evolution* **66**: 443–458.

Tyler, F., Harrison, X.A., Bretman, A., Veen, T., Rodríguez-Muñoz, R. & Tregenza, T. 2013. Multiple post-mating barriers to hybridization in field crickets. *Mol Ecol* **22**: 1640–1649.

Vacquier, V.D. 1998. Evolution of gamete recognition proteins. *Science* **281**: 1995–1998.

van Rooij, E.P., Rollins, L.A., Holleley, C.E. & Griffith, S.C. 2016. Extra-pair paternity in the long-tailed finch Poephila acuticauda. *PeerJ* **4**: e1550–10. PeerJ Inc.

Veen, T., BORGE, T., Griffith, S.C., Sntre, G.-P., Bureš, S., Gustafsson, L., *et al.* 2001. Hybridization and adaptive mate choice in flycatchers. *Nature* **411**: 45–50.

White, M.A., Stubbings, M., Dumont, B.L. & Payseur, B.A. 2012. Genetics and evolution of hybrid male sterility in house mice. *Genetics* **191**: 917–934. Genetics.

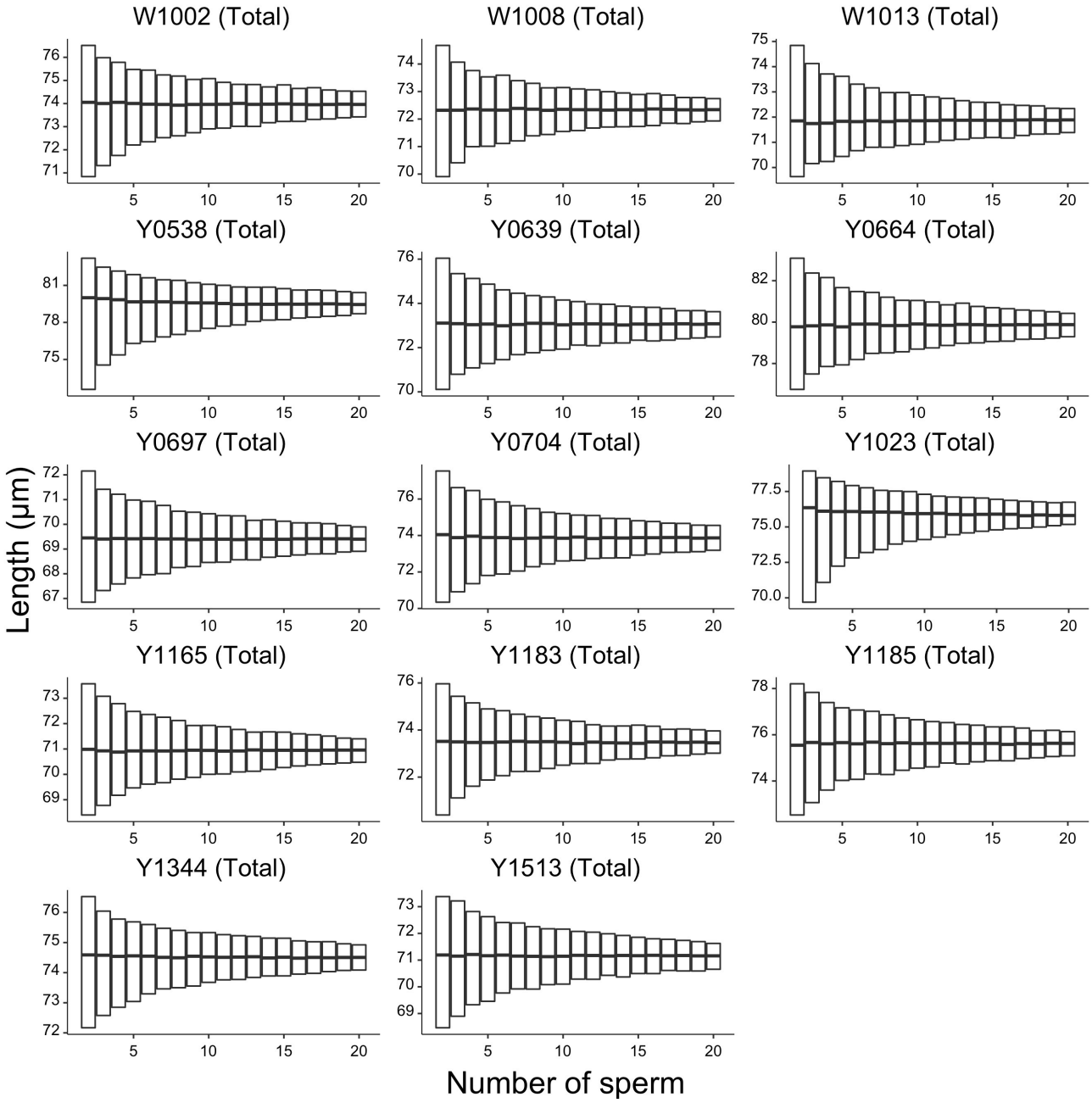
Wiley, C., Qvarnström, A., Andersson, G., BORGE, T. & Sætre, G.-P. 2009. Postzygotic isolation over multiple generations of hybrid descendents in a natural hybrid zone: how well do single-generation estimates reflect reproductive isolation. *Evolution* **63**: 1731–1739.

Wilson-Leedy, J.G. & Ingermann, R.L. 2007. Development of a novel CASA system based on open source software for characterization of zebrafish sperm motility parameters. *Theriogenology* **67**: 661–672.

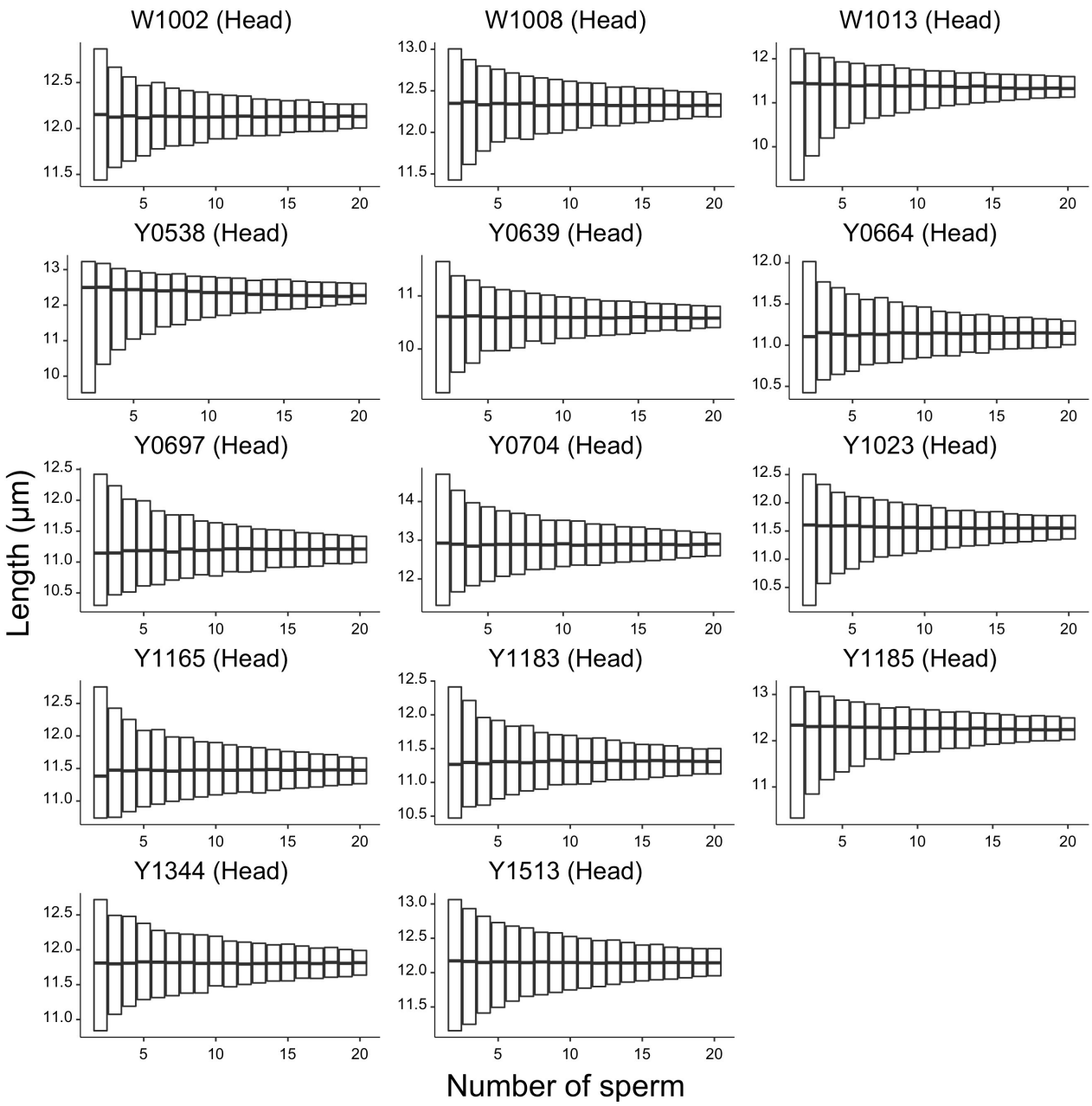
Yeates, S., Diamond, S.E., Einum, S., Emerson, B.C., Holt, W.V. & Gage, M.J.G. 2013. Cryptic choice of conspecific sperm controlled by the impact of ovarian fluid on sperm swimming behavior. *Evolution* **67**: 3523–3536.

Zuur, A.F., Ieno, E.N. & Elphick, C.S. 2010. A protocol for data exploration to avoid common statistical problems. *Methods Ecol Evol* **1**: 3–14.

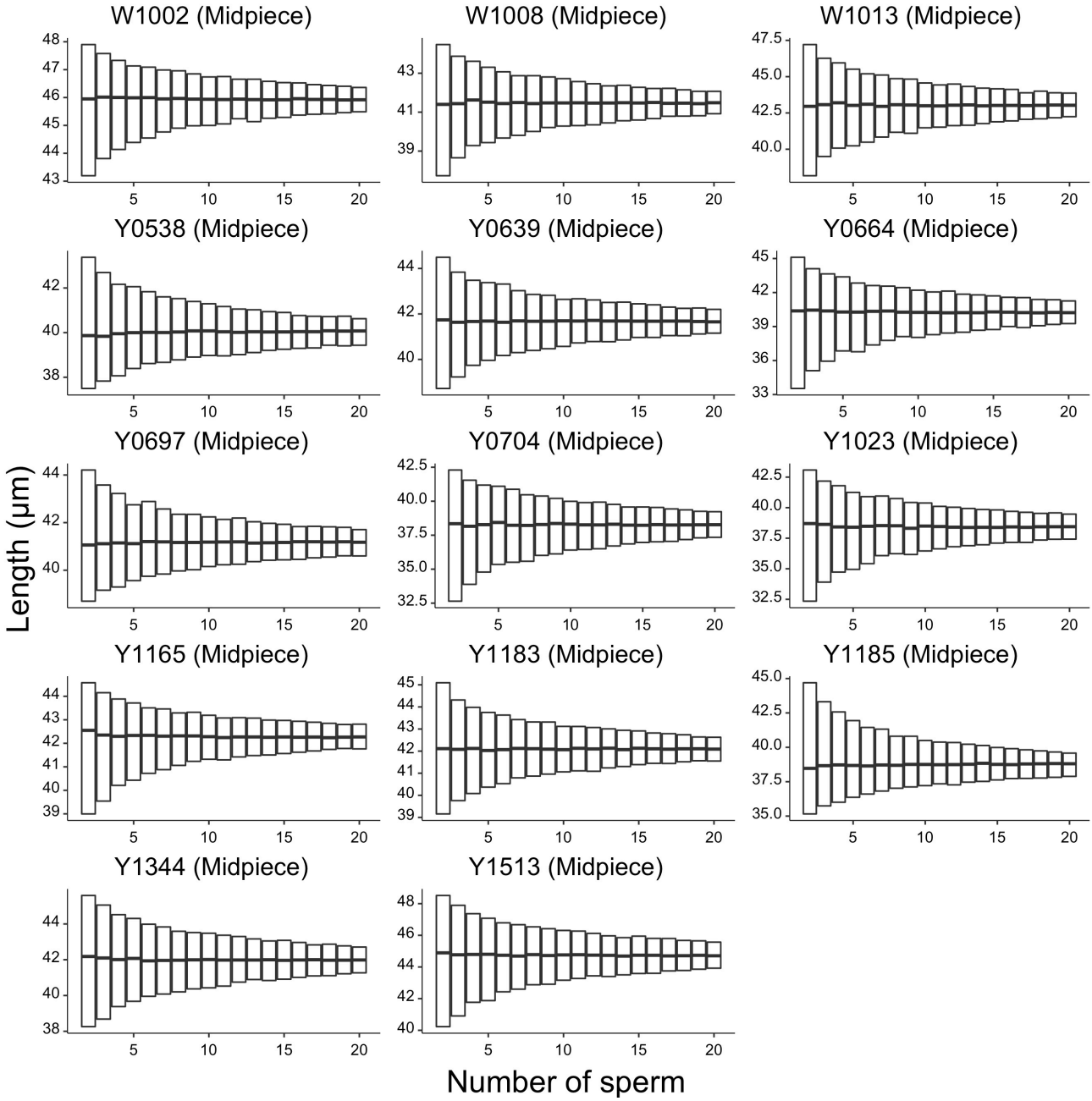
Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. & Smith, G.M. 2009. *Mixed effects models and extensions in ecology with R*. New York: Springer.



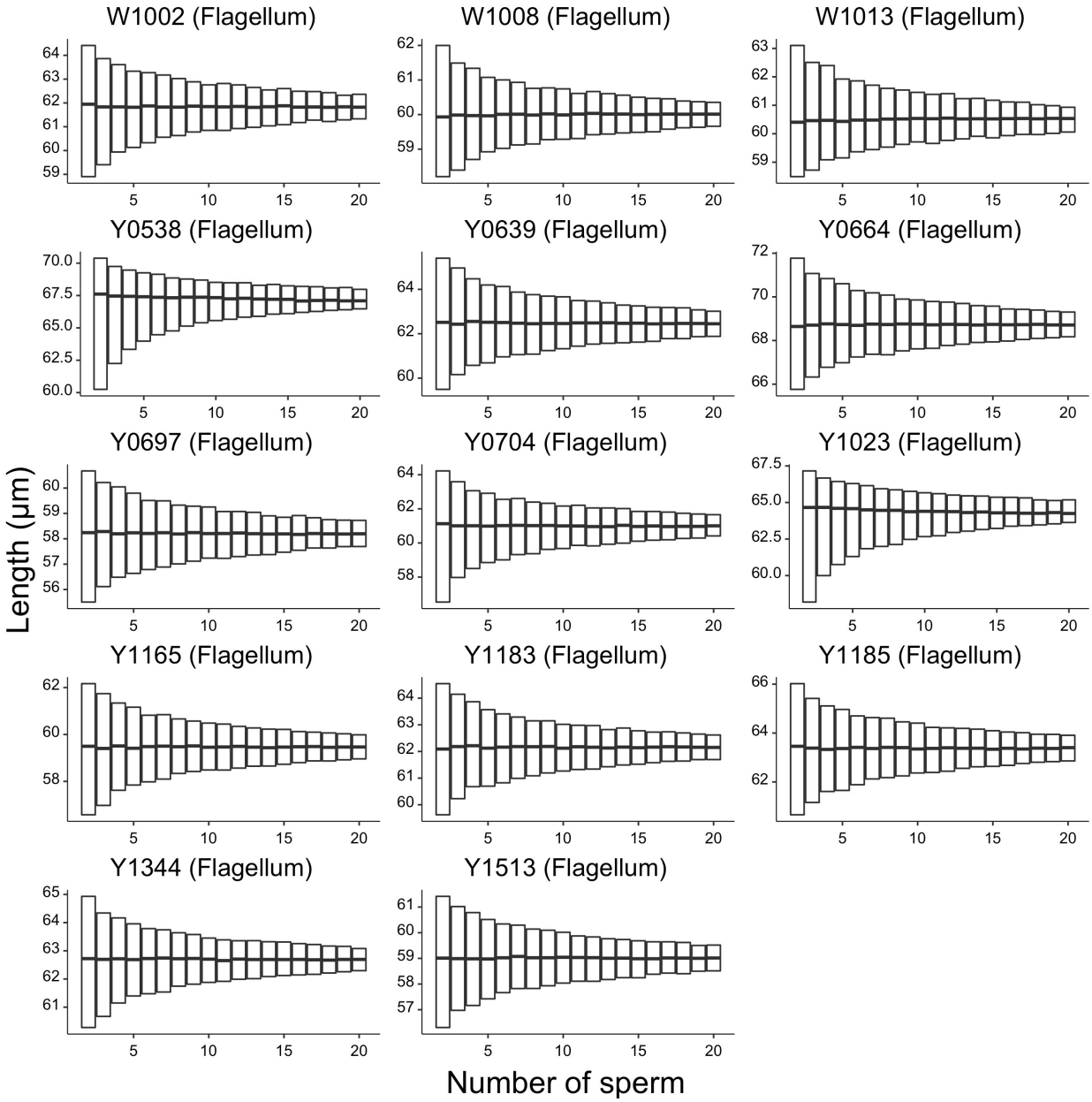
Supplementary Figure 1: The output of our within-male resampling analysis for total sperm length for all 14 males. Boxes represent 95% confidence intervals (CI).



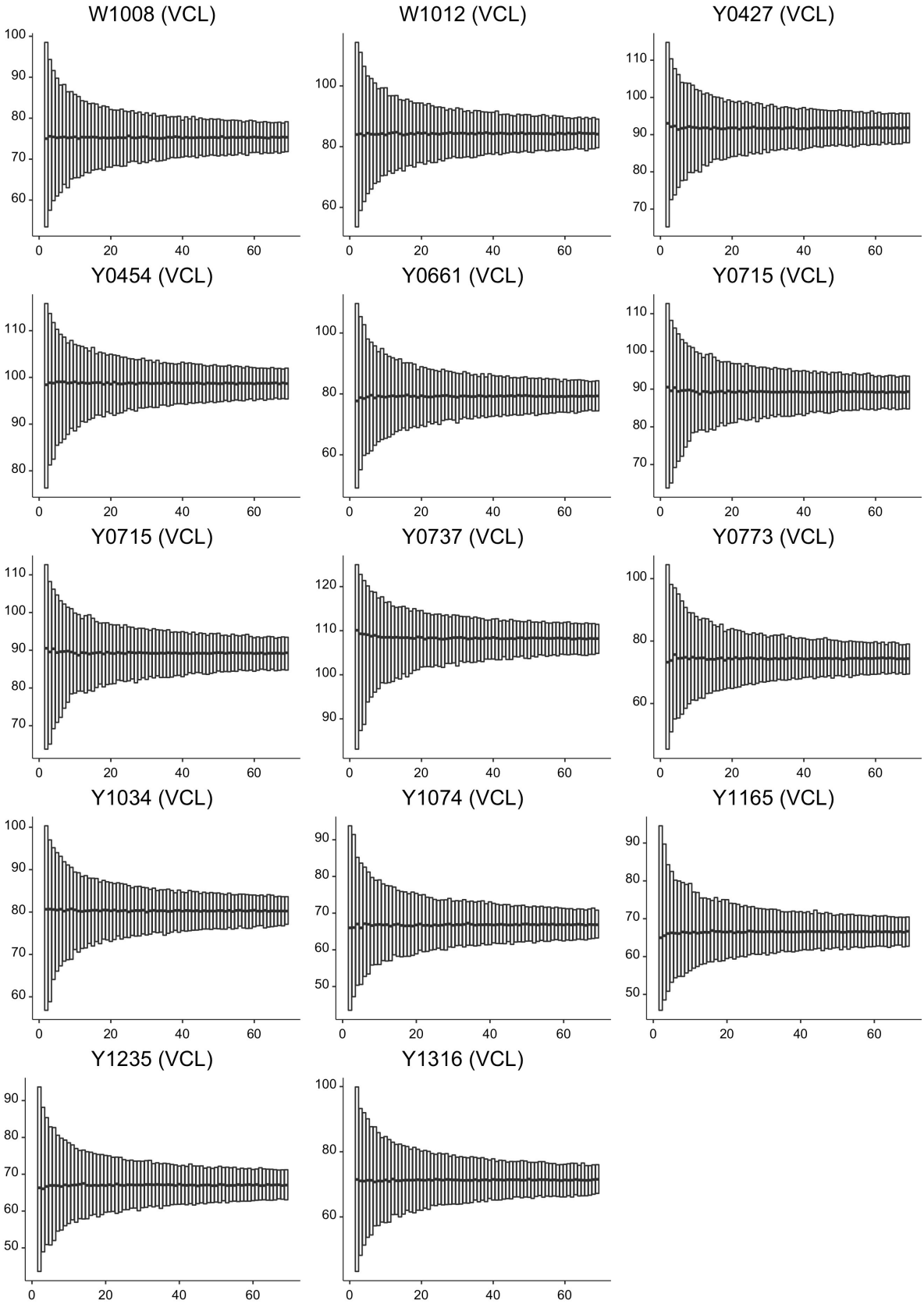
Supplementary Figure 2: The output of our within-male resampling analysis for sperm head length for all 14 males. Boxes represent 95% confidence intervals (CI).



Supplementary Figure 3: The output of our within-male resampling analysis for sperm midpiece length for all 14 males. Boxes represent 95% confidence intervals (CI).



Supplementary Figure 4: The output of our within-male resampling analysis for sperm flagellum length for all 14 males. Boxes represent 95% confidence intervals (CI).



Supplementary Figure 5: The output of our within-male resampling analysis for sperm velocity (VCL) for all 14 males. Boxes represent 95% confidence intervals (CI).