This is a repository copy of *A sex-linked supergene controls sperm morphology and swimming speed in a songbird*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/119459/

Version: Accepted Version

Article:
Kim, K.W. orcid.org/0000-0003-1489-8264, Bennison, C., Hemmings, N. et al. (6 more authors) (2017) A sex-linked supergene controls sperm morphology and swimming speed in a songbird. Nature Ecology & Evolution, 1. pp. 1168-1176.

https://doi.org/10.1038/s41559-017-0235-2

Reuse
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Published in Kim et al. 2017. Nature Ecology & Evolution **1**, 1168-1176.

(DOI: 10.1038/s41559-017-0235-2)

Author Accepted Manuscript (there may be slight differences with published one)

A sex-linked supergene controls sperm morphology and swimming speed in a songbird

Kang-Wook Kim[1*], Clair Bennison[1], Nicola Hemmings[1], Lola Brookes[1], Laura L Hurley[2], Simon C Griffith[2], Terry Burke[1], Tim R Birkhead[1] and Jon Slate[1*]

1. Department of Animal & Plant Sciences, University of Sheffield, Sheffield, S10 2TN, UK
2. Department of Biological Sciences, Macquarie University, Sydney, New South Wales, Australia.

*Correspondence:

Kang-Wook Kim (k.kim@sheffield.ac.uk) and Jon Slate (j.slate@sheffield.ac.uk)

Department of Animal & Plant Sciences, University of Sheffield, Sheffield, S10 2TN, UK

Tel: +44(0114) 222 0113 (K-WK); +44 (0114) 222 0048 (JS)

Supplementary information: 4 tables, 6 figures

Supplementary materials: Two Excel files for GWAS, eQTL and eigenGWAS, and summary of gene expression analysis. One .mpg file that contains videos of motile sperm of alternative karyomorphs.

Sperm competition is an important selective force in many organisms. As a result, sperm have evolved to be among the most diverse cells in the animal kingdom. However, the relationship between sperm morphology, sperm motility and fertilisation success is only partially understood. The extent to which between-male variation is heritable is largely unknown, and remarkably few studies have investigated the genetic architecture of sperm traits, especially sperm morphology. Here we use high-density genotyping and gene expression profiling to explore the considerable sperm trait variation that exists in the zebra finch Taeniopygia guttata. We show that nearly all of the genetic variation in sperm morphology is caused by an inversion polymorphism on the Z chromosome acting as a 'supergene'. These results provide a striking example of two evolutionary genetic predictions. First, that in species where females are the heterogametic sex, genetic variation affecting sexually dimorphic traits will accumulate on the Z chromosome. Second, recombination suppression at the inversion allows beneficial dominant alleles to become fixed on whichever haplotype they first arise, without being exchanged onto other haplotypes. Finally, we show that the inversion polymorphism will be stably maintained by heterozygote advantage, because heterozygous males have the fastest and most successful sperm.

Sperm are perhaps the most diverse cells in the animal kingdom, with enormous morphological variation between taxa, between species, between males and within an ejaculate[1]. Considerable interest in sperm diversity has arisen following the realisation that sperm competition (post-copulatory sexual selection) is a powerful selective force in many organisms[2], and that sperm morphology has co-evolved with female reproductive tract morphology[3]. The zebra finch is a model species for studies of sperm biology. Sperm length is repeatable within an ejaculate, yet variable between different males; most morphological traits (head, midpiece, tail and total length) are highly heritable[4]. Furthermore, there is a documented phenotypic and genetic correlation between morphology and sperm swimming velocity ('motility')[5]. In artificially selected lines, pronounced differences in total sperm length are apparent after just three generations of divergent selection, and males with long sperm have the greatest probability of fertilisation success in sperm competition

trials[6]. Additionally, the zebra finch has its genome sequenced, assembled and annotated[7], and so the toolkit to explore the genetics of phenotypic variation is available.

In this study we set out to understand the genetic architecture of sperm morphology and motility in the zebra finch. Our aim was to combine genome wide association mapping with analyses of gene expression in order to identify the genetic loci that contribute to phenotypic variation in sperm traits, a goal that has remained elusive in most vertebrates. We used birds from selection lines for long and short sperm, as using individuals from opposite ends of the phenotypic distribution should increase the power to detect both segregating loci and genes that are highly differentially expressed in testes of males with different sperm morphology. During the course of the study it became clear that most of the genetic variation in sperm traits was located on the Z chromosome, consistent with evolutionary theory[8-10], and that it was contained within a chromosomal inversion. Therefore, we investigated whether the inversion: (i) had similar effects in wild zebra finches, (ii) caused differences in fertilisation success in competition between males, and (iii) affected other traits such as embryo and chick survival.

Results and discussion

To understand the genetic architecture of sperm morphology we first developed a high-density SNP chip (see Methods). Chromosome partitioning analyses[11] revealed that the Z chromosome explained 67-90% of the additive genetic variance in midpiece, tail and total length of sperm (Fig. 1a, Supplementary Table 1), despite the Z chromosome representing just ~7% of the zebra finch genome. Therefore sperm morphology does not fit a classical polygenic trait architecture where a chromosome's contribution to trait variance is expected to scale linearly with its size[11]. A genomewide association study (GWAS) of morphological traits in the pre-selection lines (676 individuals, 219,279 SNPs), revealed numerous SNPs that explained phenotypic variation above the genomewide statistical significance threshold (Fig. 1b, Supplementary Fig. 1, Supplementary Data Set 1), but these were almost exclusively distributed across most of the Z chromosome.

We next examined which regions of the genome explained most of the response to three generations of artificial selection for long and short sperm. The expectation was that the regions of the genome responsible for changes in total sperm length (the trait under divergent selection) would show the greatest population genetic differentiation. We tested this by performing an EigenGWAS analysis[12] (see Methods). As with the analyses of morphological traits, there were numerous significant SNPs, distributed across most of the length of the Z chromosome (Fig. 1c), indicating that Z-linked genes drove the between-line response to divergent selection.

To further understand the genetic basis of sperm trait variation we used microarrays to examine gene expression profiles in the testes of short- and long-line males. We identified 108 genes that were differentially expressed between the long and short lines (Bonferroni corrected, $\alpha = 0.05$, Fig. 1d, Supplementary Data Set 2). Among the differentially expressed genes, overrepresented Gene Ontology (GO) terms include GO:0008283 (cell proliferation), GO:0051301 (cell division) and GO:0090307 (mitotic spindle assembly). Differentially expressed genes associated with these terms include CCSAP (centriole, cilia and spindle-associated protein), TPX2 (microtubule-associated protein) and SPICE1 (spindle and centriole associated protein 1). The sperm's flagellum has a cytoskeleton, the axoneme, composed of centrioles formed from microtubules. The most differentially expressed gene (in terms of fold change and statistical significance), dimethylglycine dehydrogenase (DMGDH), is on the Z chromosome. DMGDH is an enzyme that catalyzes the demethylation of dimethylglycine to form sarcosine, and is not an obvious candidate gene to affect sperm traits. However, in a study where rats were exposed to the industrial solvent Ethylene glycol monomethyl ether (EGME), a phenotypic consequence of exposure was deformed and reduced numbers of sperm.[13]. The most dramatic metabolic changes were in elevated levels of dimethylglycine and sarcosine. Therefore, it seems likely that variation in DMGDH expression can influence catalysis of a compound that affects sperm traits. Thus, the GO term analysis and identities of differentially expressed genes include good functional candidates for influencing sperm morphological variation. There was a 4-5-fold excess of differentially expressed genes on the Z chromosome (21 of 738 genes) relative to on autosomes (87 of 13,486 genes; $\chi^2 = 53.6$, d.f. = 1, $P = 2.5 \times 10^{-13}$). Notably, among the

108 differentially expressed genes, expression was usually up-regulated in the long-line males relative to short-line males (up-regulated in 94 genes, down-regulated in 14 genes, binomial test $P = 1.0 \times 10^{-15}$; Supplementary Fig. 2). As with the analyses of genetic architecture and response to selection, the significant Z-linked loci were distributed across most of the chromosome (Fig. 1d). These data provide unequivocal support for the theory that masculinising genes should be overrepresented and up-regulated on the avian Z chromosome[8,14,15]; notably, not all previous empirical studies have found support for this theory[16]. For example, in a study that examined gene expression in the testes of wild turkeys (Meleagris gallopavo), Z-linked genes did not play a larger than expected role in the differences between dominant and subordinate males[16].

While gene expression analyses of testes reveal which genes are involved in the pathways responsible for sperm morphological variation, they do not identify the regions in the genome that are responsible for variation in gene expression. To bridge the gap between expression data and genetic variation, we first used principal component analysis (PCA) to reduce the expression data of the 108 differentially expressed genes to a single quantitative trait. We then performed an expression QTL (eQTL) mapping analysis[17,18] to identify regions that regulate the gene expression of sperm traits. All significant SNPs were on the Z chromosome (Fig. 1e), even though 87 of the 108 differentially expressed genes were on the autosomes, suggesting that the Z chromosome not only contains an excess of genes that are differentially expressed in long and short-line males, but that it also regulates expression of the autosomal genes that differ between the selection lines (i.e. the Z chromosome is responsible for trans-acting regulation of gene expression).

The Z chromosome clearly causes most of the transcriptomic, genetic and phenotypic variation in sperm morphology. Intriguingly, though, the Z chromosome effects appear to be spread across nearly the entire chromosome, rather than being caused by one or a few major genes. Linkage disequilibrium extends over large physical distances on the zebra finch Z chromosome[19,20], suggestive of a low recombination rate. What suppresses recombination in this Z chromosome? Chromosomal inversions are common in finches[21] and previous cytological studies have reported a polymorphic

chromosomal inversion on the zebra finch Z chromosome[22]. Recent population genetic analyses indicate that the polymorphism is segregating in both wild and domesticated birds[23]. Principal component analysis on the Z chromosome SNPs confirmed the presence of an inversion (Supplementary Fig. 3). Six discrete clusters (Fig. 2a) are consistent with three previously identified segregating inversion haplotypes (Z_A, Z_B, Z_C) that result in six possible male genotypes (Z_AZ_A, Z_AZ_B, Z_AZ_C, Z_BZ_B, Z_BZ_C, Z_CZ_C; hereafter we use the notation AA, AB etc for brevity)[23]. Haplotype frequencies were 0.378, 0.408 and 0.215 for A, B and C respectively.

BB and CC homozygous males have an extended block of very low heterozygosity relative to AA homozygotes (Fig. 2b). Individuals carrying one A haplotype and one alternative haplotype (termed here heterokaryomorphs) have high heterozygosity; i.e. alternate SNP alleles are fixed on different haplotypes. BC heterozygotes show a similar pattern. (Fig. 2b, Supplementary Fig. 4). Genetic diversity is higher among A haplotypes than for haplotypes B and C. These data are entirely consistent with an inversion polymorphism with breakpoints occurring at approximately 5.8 Mbp and 68.9 Mbp, where A is ancestral, and B and C are independently derived. We mapped putative break points for the genomic rearrangements to locations with a 3-68 kbp resolution, based on the pattern of heterozygosity and genotypes (Supplementary Table 2). Crucially, the region containing the inversion polymorphism is the same region in which all of the significant GWAS, eigenGWAS and expression GWAS results were detected. Thus, it seems likely that an inversion polymorphism, where recombination is suppressed between haplotypes, is acting as a supergene that controls nearly all the variation in sperm morphology.

Evolutionary theory predicts that dominant alleles that are beneficial to males will accumulate on the Z chromosome in species with ZW sex-determination[10]. A recombination-suppressing inversion could provide the perfect conditions for the Z chromosome to play an even greater role in sperm variation. Whenever a new male-beneficial mutation occurs, it will reach a high frequency on the haplotype in which it arose, but will be unable to recombine onto other haplotypes. Therefore, each haplotype will accumulate different male-beneficial dominant alleles. Males heterozygous for

two different haplotypes will express dominant alleles on both haplotypes and therefore should benefit from heterozygote advantage. Crucially, the heterozygote advantage will become stronger over time, as more and more male-beneficial alleles accumulate on different haplotypes.

To formally evaluate the effect of the Z chromosome inversion on sperm traits we fitted inversion karyotype as an explanatory term in models of sperm morphology and motility. There were small, but significant, differences in headpiece length between karyotypes ($F_{5,670} = 2.91$, $P = 0.013$), but profound differences in midpiece ($F_{5,670} = 157.03$, $P < 3 \times 10^{-110}$), tail ($F_{5,670} = 293.37$, $P < 1 \times 10^{-165}$) and total length ($F_{5,670} = 155.94$, $P < 1 \times 10^{-109}$; Fig. 3a-b). Sperm from AA males are characterised by short midpieces, long tails and the greatest overall length. BB, BC and CC males have sperm with short tails and short overall length. AB and AC males (heterokaryomorphs) have sperm with long midpieces, intermediate tail length and relatively long overall length (Fig. 3a-b). The phenotypic differences are so pronounced that the midpiece:tail ratio, a predictor of motility, differs ~4-fold between AA birds and BB/BC/CC birds. We have previously shown that the midpiece:tail ratio explains variation in sperm motility[24], with birds with intermediate midpiece:tail ratios having the fastest sperm. Exploring this relationship further reveals that heterokaryomorphic AB and AC birds have the fastest sperm motility (Fig. 3c, $P < 1 \times 10^{-5}$, Supplementary Data Set 3). The response to selection observed in the selection lines was driven by an increase in frequency of haplotype A in the long sperm line and haplotype B in the short lines (Fig. 4).

A male's Z chromosome karyotype was not only associated with inter-male variation in sperm traits, but also intra-male sperm morphological variation. We quantified how variable a male's sperm were by estimating the coefficient of variation (standard deviation / mean) of the 5 sperm measured in each male. Male karyotype explained significant differences in within-male variation in sperm morphology (Supplementary Figure 5). AA males tend to have sperm with more variable midpiece length and less variable tail length than males of other karyotypes. The total sperm length of males with karyotypes AA, AB and AC is less variable than males lacking haplotype A. Recent work has shown that in zebra finches the sperm that reach the ovum at the site of fertilisation tend to be a

distinct subset of those within the ejaculate[25]; the 'fertilising' subset tend to be less variable than the overall ejaculate.

Although zebra finches are native to Australia, they have been domesticated numerous times and the original location of birds used to found aviary populations such as ours is rarely known. Therefore, we wanted to eliminate the possibility that the Z chromosomal inversion is only segregating in captive birds. We typed a panel of wild, wild-derived and domesticated birds from New South Wales, Australia and found all three Z haplotypes segregating in wild birds (frequencies of A, B and C were 0.730, 0.199 and 0.122), with approximately the same effects on sperm morphology as observed in our population (Supplementary Fig. 6). Thus, the polymorphism segregates in the wild, and we can reject the scenario that it only exists in domesticated populations because they were founded from multiple populations in which different Z chromosomes have diverged in allopatry.

An obvious question is what maintains the Z inversion polymorphism? Heterokaryomorphs with one ancestral and one derived chromosome (AB and AC) have the fastest sperm, so heterozygote advantage for sperm velocity may be sufficient to maintain the variation. Previously, we performed sperm competition trials between pairs of selection line birds, showing that birds from the long line had greater fertilisation success than birds from the short line[6]. Retrospectively, examining the data in light of the karyotypic information shows that most short-line birds were BB and long-line males were AA, AB or AC (Fig. 4). AA and AB males outcompeted BB males, with heterozygous AB males being significantly more successful than the other long-line types (Supplementary Table 3). We also tested whether there was a trade-off between a male karyotype's effect on sperm traits and whether it had an effect on egg hatching success and chick survival; while there were weakly significant effects (Supplementary Table 4) there was no systematic evidence that heterokaryomorphic males had lower fitness than homozygous males and no evidence that the enhanced fertilisation success of AB was compromised by having less fit offspring. Therefore, the data are consistent with heterozygote advantage for increased male sperm velocity/fertilisation success maintaining much of the genetic variation in sperm traits. Of course, the inversion region is very large, and undoubtedly affects other

traits as well, although previous analyses of morphological traits have not found sex-specific effects, nor have they found that the Z chromosome explains such a disproportionately large amount of genetic variation as we find for sperm traits in this study[23].

An unresolved question is which genes determine variation in sperm traits? The Z chromosome inversion explains nearly all of the genetic variation in sperm morphology, contains the expression QTL associated with differences between the long and short lines and it has driven the response to the artificial selection. However, the inversion contains at least 648 genes. Furthermore, the extreme linkage disequilibrium between SNPs within the inversion (Supplementary Fig. 4b) means that they provide almost equal support for associations with sperm traits, making it difficult to pinpoint specific genes that cause these associations. Nonetheless, some SNPs within the inversion are marginally more statistically significant than the rest. We interpret these results with some caution, given the strong LD, but summarise them briefly below. SNPs strongly associated with sperm morphology are located at 7.01Mbp, approximately 90kbp upstream of the nearest gene, GADD45G (7.100-7.101Mbp), which is known to be expressed in testes and is essential for sex determination and testis development[26]. Similarly, a SNP at 8.943Mbp affects sperm morphology and is approximately 80kbp from LRRC2 (9.064-9.129Mbp), a gene known to be expressed in mammalian testes[27], and with reported effects on male fertility[28]. SNP AX-146728334, at 9.997Mbp is significantly associated with midpiece, tail and total length and is within an intron of C9orf3, another gene known to be expressed in mammalian testes[29]. SNP AX-146856158 (19.546 Mbp) is the third most strongly associated SNP with midpiece length and is within an intron of the F-box and leucine rich repeat protein 17 (FBXL17), a protein-ubiquitin ligase. FBXL17 was recently shown to cause variation in sex reversal in the half-smooth tongue sole (Cynoglossus semilaevis)[30]. SNP AX-146979388 (65.252Mbp) is strongly associated with midpiece length and is the SNP closest to doublesex and mab-3 related transcription factor 2 (DMRT2). DMRT2 is a sex-determining gene[31] – mutations can cause sex-reversal - and is expressed in the mammalian testis. In the expression QTL analysis three SNPs were marginally more significantly associated with variation in differential gene expression than the rest of the SNPs on the inversion. Two, AX-146978822 and AX-146728701, at 68.569Mbp

and 11.630Mbp, respectively, are located most closely to unannotated genes. The third, AX-146868012, at 29.767 Mbp, is approximately 200kbp upstream of leucine rich repeat and Ig domain containing 2 (LINGO2), a gene that is expressed in testes and has been tentatively associated with variation in male birth rate and family size in a GWAS in a human population that prohibits contraception[32]. In the eigenGWAS analysis the most significant SNPs were at 68.2-68.5Mbp, close to the peak of the most significant eQTL. There are 5 genes in this region, but only two are annotated. ZNF462 at 68.167-68.215Mbp, encodes a zinc finger protein, is expressed in sperm[33], and in humans is at a locus that explains variation in age at menarche[34]. RAD23B at 68.310-68.343Mbp, encodes a nucleotide excision repair protein and is highly expressed in human testes and sperm[35]. In mice, knocking out RAD23B causes a failure of spermatogenesis[36].

Our data and those from an independent study[37], together reveal several surprising features about the genetics of sperm traits. First, despite sperm morphology having a continuous phenotypic distribution typical of a polygenic trait, here a single supergene has resulted in an atypical genetic architecture. Second, while theory has long predicted that the sex chromosomes may play an important role in the architecture of sexually-selected, sexually dimorphic or sexually antagonistic traits[10,14], existing data have provided largely equivocal support[9]. Here, the Z chromosome, which represents just 7% of the zebra finch genome, plays a central role in determining variation and evolution of sperm − the ultimate sexually-dimorphic trait. Our findings can be explained by, and support, evolutionary theory. In female-heterogametic species, Z-linked dominant alleles with beneficial effects in males are predicted, even if detrimental to females, because (i) dominant alleles are always expressed in males and (ii) Z chromosomes are present in males twice as often as in females. The inversion polymorphism maintains these alleles through heterozygote advantage for fertilisation success. Notably, we found evidence for autosomal inversion polymorphisms in the zebra finch genome (e.g. on Chromosomes 8, 9, 11 and 13; Supplementary Fig. 3) but these do not explain phenotypic variation in sperm traits. One previous explanation for the extreme variation in sperm morphology in zebra finches has been that because they are relatively monogamous, directional selection through sperm competition is relaxed[38]. Sperm trait variation will be more rapidly eroded in

species with intense post-copulatory sexual selection, than in species with reduced sperm competition. However, our data indicate that without the Z chromosome inversion, additive genetic variation would be up to an order of magnitude lower (extrapolated from Supplementary Table 1); therefore, the inversion polymorphism offers an alternative (or perhaps complementary) explanation for the pronounced between-male variation in zebra finch sperm.

Our data add to the emerging evidence that inversion polymorphisms can play a profound role in the establishment and maintenance of genetic diversity within and between populations and species[39-43], although the other examples do not involve sex chromosomes. Our study is perhaps the most extreme and complete example of a sex chromosome affecting sexually-selected trait variation. We predict that similar patterns would be seen for sexually-dimorphic traits in other taxonomic groups with ZW sex-determination systems, especially when Z chromosome inversions are present. Similarly, in species with XY sex-determination, similar patterns on the X chromosome should be seen for feminized traits. Finally, the loci that we have identified as being involved in variation in sperm morphology and testis gene expression are likely to play significant roles in sperm morphological variation in other species.

References

1 Pitnick, S., Hosken, D. J. & Birkhead, T. R. Sperm morphological diversity in *Sperm Biology: An Evolutionary Perspective* (eds Tim R. Birkhead, David J. Hosken, & Scott Pitnick) 69-149 (Academic Press, 2009).

2 Parker, G. A. Sperm competition and its evolutionary consequences in insects. *Biol. Rev. Camb. Philos. Soc.* **45**, 525-567 (1970).

3 Miller, G. T. & Pitnick, S. Sperm-female coevolution in *Drosophila*. *Science* **298**, 1230-1233 (2002).

4 Birkhead, T. R., Pellatt, E. J., Brekke, P., Yeates, R. & Castillo-Juarez, H. Genetic effects on sperm design in the zebra finch. *Nature* **434**, 383-387 (2005).

5 Mossman, J., Slate, J., Humphries, S. & Birkhead, T. Sperm morphology and velocity are genetically co-determined in the zebra finch. *Evolution* **63**, 2730-2737 (2009).

6 Bennison, C., Hemmings, N., Slate, J. & Birkhead, T. Long sperm fertilize more eggs in a bird. *Proc. R. Soc. Lond. B* **282**, 20141897 (2015).

7 Warren, W. C. *et al.* The genome of a songbird. *Nature* **464**, 757-762 (2010).

8 Dean, R. & Mank, J. E. The role of sex chromosomes in sexual dimorphism: discordance between molecular and phenotypic data. *J. Evol. Biol.* **27**, 1443-1453 (2014).

9 Mank, J. E. Sex chromosomes and the evolution of sexual dimorphism: lessons from the genome. *Am. Nat.* **173**, 141-150 (2009).
10 Rice, W. R. Sex-chromosomes and the evolution of sexual dimorphism. *Evolution* **38**, 735-742 (1984).
11 Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genet.* **43**, 519-525 (2011).
12 Chen, G. B., Lee, S. H., Zhu, Z. X., Benyamin, B. & Robinson, M. R. EigenGWAS: finding loci under selection through genome-wide association studies of eigenvectors in structured populations. *Heredity* **117**, 51-61 (2016).
13 Takei, M. *et al.* Ethylene glycol monomethyl ether-induced toxicity is mediated through the inhibition of flavoprotein dehydrogenase enzyme family. *Toxicol. Sci.* **118**, 643-652 (2010).
14 Connallon, T. & Clark, A. G. Sex linkage, sex-specific selection, and the role of recombination in the evolution of sexually dimorphic gene expression. *Evolution* **64**, 3417-3442 (2010).
15 Ellegren, H. Emergence of male-biased genes on the chicken Z-chromosome: Sex-chromosome contrasts between male and female heterogametic systems. *Genome Res.* **21**, 2082-2086 (2011).
16 Pointer, M. A., Harrison, P. W., Wright, A. E. & Mank, J. E. Masculinization of gene expression is associated with exaggeration of male sexual dimorphism. *PLoS Genet.* **9**, e1003697 (2013).
17 Rockman, M. V. & Kruglyak, L. Genetics of global gene expression. *Nature Rev. Genet.* **7**, 862-872 (2006).
18 Schadt, E. E. *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297-302 (2003).
19 Backstrom, N. *et al.* The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Res.* **20**, 485-495 (2010).
20 Stapley, J., Birkhead, T. R., Burke, T. & Slate, J. Pronounced inter- and intrachromosomal variation in linkage disequilibrium across the zebra finch genome. *Genome Res.* **20**, 496-502 (2010).
21 Hooper, D. M. & Price, T. D. Rates of karyotypic evolution in Estrildid finches differ between island and continental clades. *Evolution* **69**, 890-903 (2015).
22 Itoh, Y., Kampf, K., Balakrishnan, C. N. & Arnold, A. P. Karyotypic polymorphism of the zebra finch Z chromosome. *Chromosoma* **120**, 255-264 (2011).
23 Knief, U. *et al.* Fitness consequences of polymorphic inversions in the zebra finch genome. *Genome Biol.* **17**, 199 (2016).
24 Bennison, C., Hemmings, N., Brookes, L., Slate, J. & Birkhead, T. Sperm morphology, adenosine triphosphate (ATP) concentration and swimming velocity: unexpected relationships in a passerine bird. *Proc. R. Soc. Lond. B* **283**, 20161558 (2016).
25 Hemmings, N., Bennison, C. & Birkhead, T. R. Intra-ejaculate sperm selection in female zebra finches. *Biol. Lett.* **12**, 20160220 (2016).
26 Johnen, H. *et al.* Gadd45g is essential for primary sex determination, male fertility and testis development. *PLoS One* **8**, e58751 (2013).
27 Song, R. *et al.* Male germ cells express abundant endogenous siRNAs. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 13159-13164 (2011).
28 Singh, A. P., Harada, S. & Mishina, Y. Downstream genes of Sox8 that would affect adult male fertility. *Sex. Dev.* **3**, 16-25 (2009).
29 Diaz-Perales, A. *et al.* Identification of human aminopeptidase O, a novel metalloprotease with structural similarity to aminopeptidase B and leukotriene A_4 hydrolase. *J. Biol. Chem.* **280**, 14310-14317 (2005).
30 Jiang, L. & Li, H. Single locus maintains large variation of sex reversal in half-smooth tongue sole (*Cynoglossus semilaevis*). *G3 (Bethesda)* **7**, 583-589 (2017).
31 Calvari, V. *et al.* A new submicroscopic deletion that refines the 9p region for sex reversal. *Genomics* **65**, 203-212 (2000).

32 Kosova, G., Scott, N. M., Niederberger, C., Prins, G. S. & Ober, C. Genome-wide association study identifies candidate genes for male fertility traits in humans. *Am. J. Hum. Genet.* **90**, 950-961 (2012).
33 Ostermeier, G. C., Goodrich, R. J., Moldenhauer, J. S., Diamond, M. R. & Krawetz, S. A. A suite of novel human spermatozoal RNAs. *J. Androl.* **26**, 70-74 (2005).
34 Perry, J. R. B. *et al.* Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. *Nature Genet.* **41**, 648-650 (2009).
35 Huang, X. Y. *et al.* Expression of a novel RAD23B mRNA splice variant in the human testis. *J. Androl.* **25**, 363-368 (2004).
36 Ng, J. M. Y. *et al.* Developmental defects and male sterility in mice lacking the ubiquitin-like DNA repair gene mHR23B. *Mol. Cell. Biol.* **22**, 1233-1245 (2002).
37 Knief, U. *et al.* Strong heterotic effects of a sex-chromosome inversion on sperm characteristics and siring success. *Nat. Ecol. Evol.* **In Press** (2017).
38 Calhim, S., Immler, S. & Birkhead, T. R. Postcopulatory sexual selection is associated with reduced variation in sperm morphology. *PLoS One* **2**, e413 (2007).
39 Joron, M. *et al.* Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**, 203-206 (2011).
40 Küpper, C. *et al.* A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genet.* **48**, 79-83 (2016).
41 Lamichhaney, S. *et al.* Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nature Genet.* **48**, 84-88 (2016).
42 Tuttle, E. M. *et al.* Divergence and functional degradation of a sex chromosome-like supergene. *Curr. Biol.* **26**, 344-350 (2016).
43 Wang, J. *et al.* A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**, 664-668 (2013).
44 Henderson, C. R. Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423-447 (1975).
45 Kruuk, L. E. B. Estimating genetic parameters in natural populations using the 'animal model'. *Phil. Trans. R. Soc. Lond. B* **359**, 873-890 (2004).
46 ASReml User Guide. Release 3.0. VSN International (VSN International, Hemel Hempstead, 2009).
47 Immler, S., Griffth, S. C., Zann, R. & Birkhead, T. R. Intra-specific variance in sperm morphometry: a comparison between wild and domesticated Zebra Finches *Taeniopygia guttata*. *Ibis* **154**, 480-487 (2012).
48 Abramoff, M. D., Magalhaes, P. J. & Ram, S. J. Image processing with ImageJ. *Biophotonics International* **11**, 36-42 (2004).
49 Bruford, M. W., Hanotte, O., Brookfield, J. F. Y. & Burke, T. Multilocus and single-locus DNA fingerprinting in *Molecular genetic analysis of populations: A practical approach* (ed A. R. Hoelzel) 225-269 (IRL Press, 1998).
50 Baird, N. A. *et al.* Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376 (2008).
51 Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124-3140 (2013).
52 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
53 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
54 Danecek, P. *et al.* The Variant Call Format and VCFtools. *Bioinformatics* **27**, 2156-2158 (2011).
55 Moser, G. *et al.* Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model. *PLoS Genet.* **11**, e1004969 (2015).

56 Aulchenko, Y. S., de Koning, D.-J. & Haley, C. Genomewide rapid association using mixed model and regression: A fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics* **177**, 577-585 (2007).
57 Ma, J. Z. & Amos, C. I. Investigation of inversion polymorphisms in the human genome using principal components analysis. *PLoS One* **7**, e40224 (2012).
58 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.* **88**, 76-82 (2011).
59 Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. cluster: Cluster Analysis Basics and Extensions. R package version 2.0.4. (2016).
60 Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575 (2007).
61 Bates, D., Maechler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Software* **67**, 1-48 (2015).
62 Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest: Tests in linear mixed effects models. R package version 2.0-32. (2016).
63 Warnes, G. R. *et al.* gplots: Various R programming tools for plotting data. R package version 3.0.1. (2016).
64 Ronnegard, L. *et al.* Increasing the power of genome wide association studies in natural populations using repeated measures - evaluation and implementation. *Methods Ecol. Evol.* **7**, 792-799 (2016).
65 Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
66 Birkhead, T. R. & Fletcher, F. Male phenotype and ejaculate quality in the zebra finch *Taeniopygia guttata*. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **262**, 329-334 (1995).
67 Hadfield, J. D. MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *J. Stat. Software* **33**, 1-22 (2010).

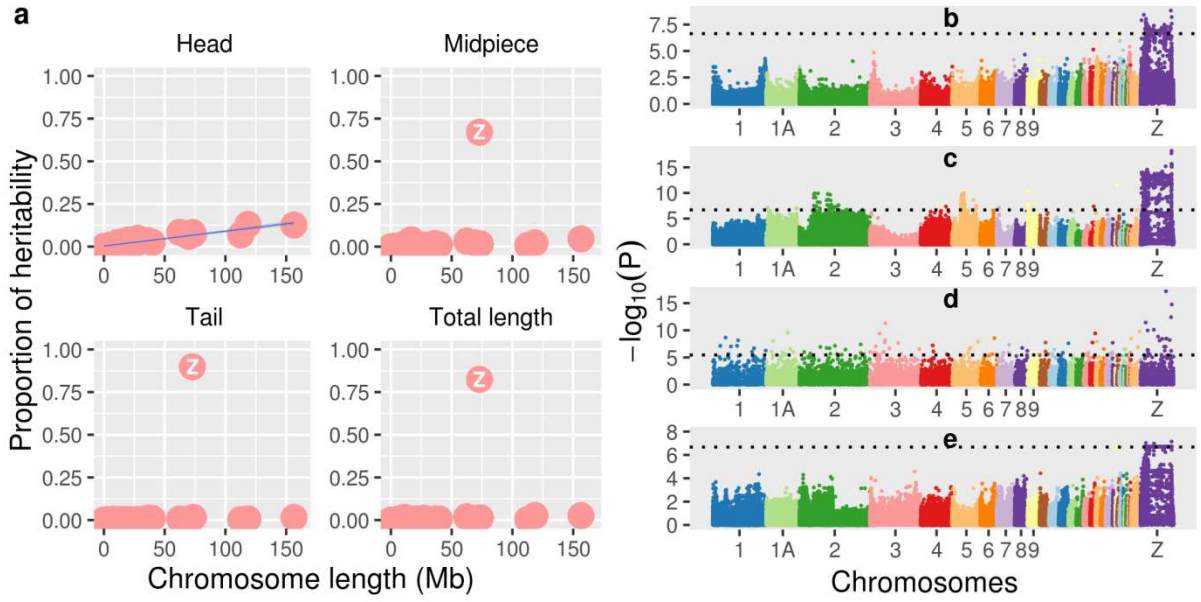


Figure 1 | Mapping variation in sperm morphology and gene expression. a, The relationship between chromosome size and proportion of additive genetic variance explained in pre-selection line birds (n = 676). **b**, GWAS to show the association between SNPs and total sperm length in pre-selection line birds (n = 676). **c**, EigenGWAS showing the association between SNPs and the primary eigenvector of autosomal genetic differentiation after three generations of divergent artificial selection (n = 149). **d**, Test for the differential gene expression between testes of short (n = 20) and long line (n = 22) males. Raw P values are plotted at the starting position of the transcripts. **e**, eQTL analysis on the primary eigenvector of PCA for the 108 genes that are differentially expressed between lines (n = 42). For **b**, **c** and **e**, P values were corrected for a genomic inflation factor, λ. In b-e, a genomewide Bonferroni adjusted significance threshold of P = 0.05 is indicated by dotted lines.

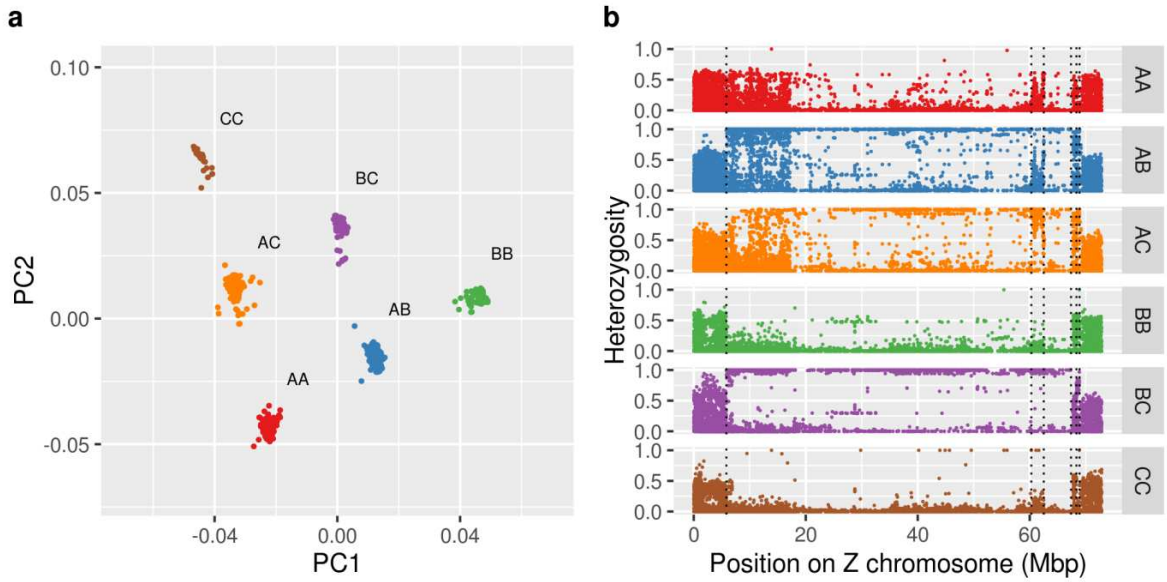


Figure 2 | Characterisation of a Z inversion polymorphism. a, Principal component analysis and clustering for 676 pre-selection line males (n_{AA} = 91, n_{AB} = 205, n_{AC} = 132, n_{BB} = 85, n_{BC} = 122, n_{CC} = 41) using all SNPs on the Z chromosome. Each cluster is indicated by colour and the names of each cluster were chosen to show the homo- (AA, BB and CC) or heterokaryotypes (AB, AC and BC) based on the pattern of heterozygosity and location on the PCA plot. **b**, The observed heterozygosity of Z chromosome SNPs for each Z chromosome karyotype. The six dashed vertical lines indicate putative breakpoints (BP_1-BP_6 from left), inferred from the pattern of heterozygosity and genotype (see Supplementary Table 2, Supplementary Fig. 4 for details on the breakpoints and the pattern of LD).

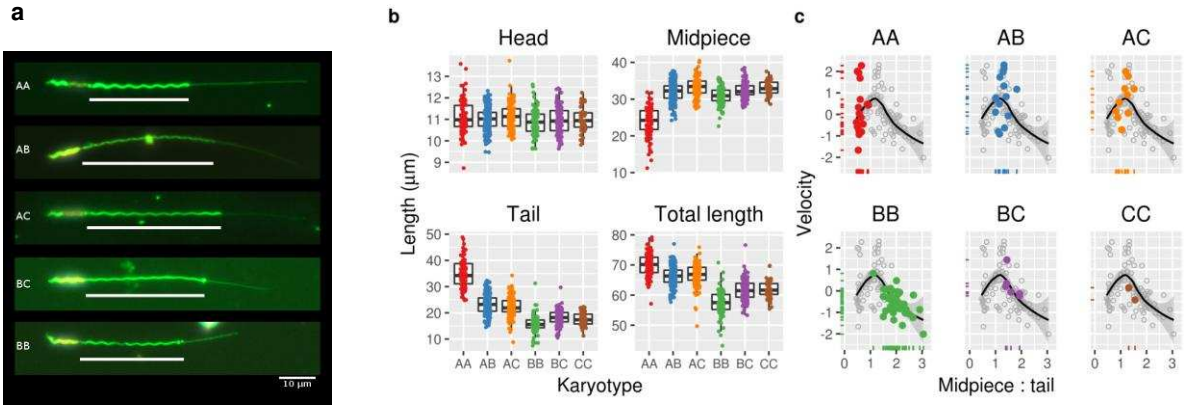


Figure 3 | The effects of karyotypes on sperm morphology and motility. a, Fluorescent-microscopy image of sperm from AA, AB, AC, BC and BB males, stained with MitoTracker Green FM (Molecular Probes). Midpieces are indicated by horizontal lines. **b**, Effects of karyotypes on four sperm morphological traits in 676 pre-selection line males (n_{AA} = 91, n_{AB} = 205, n_{AC} = 132, n_{BB} = 85, n_{BC} = 122, n_{CC} = 41) **c**, Effects of karyotypes on sperm velocity for 97 males (n_{AA} = 24, n_{AB} = 16, n_{AC} = 11, n_{BB} = 38, n_{BC} = 6, n_{CC} = 2), which is measured by the mean primary principal component of three velocity measures (see Methods). The names of the karyotypes are shown at the top of each panel. Grey dots show all data points and each panel shows a different karyotype, colour coded as in Fig. 2. A loess smooth line fitted to all data points is shown, with shaded standard error. See Supplementary material for videos of motile sperm of males of each karyotype.

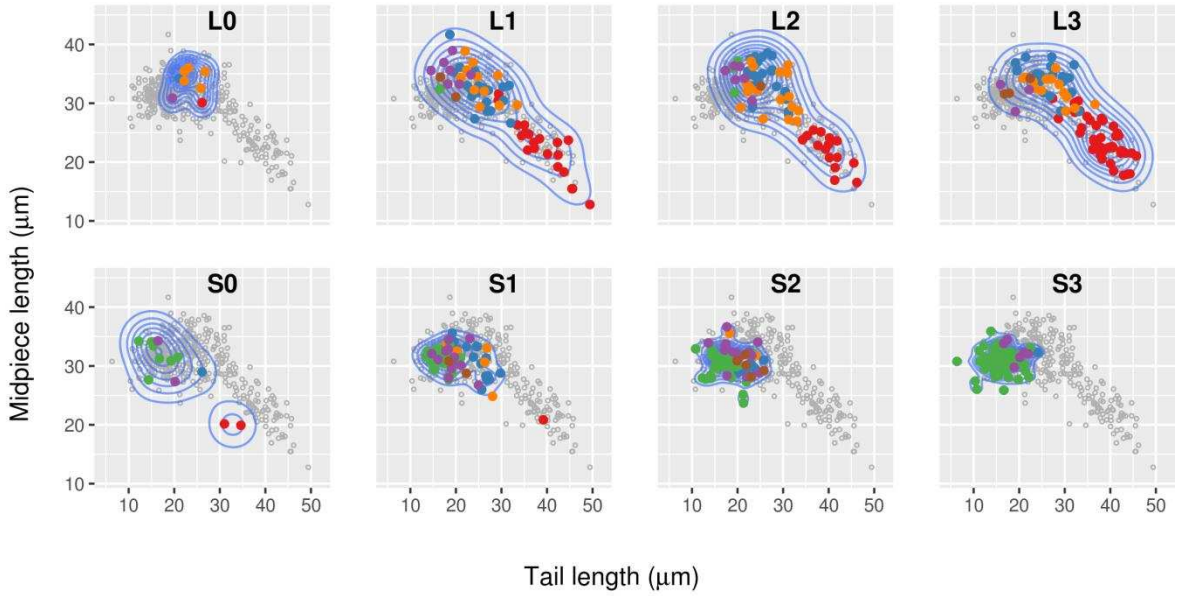


Figure 4 | Distribution of Z karyotypes in morphometric space during artificial selection for total sperm length. Panels illustrate how karyotype frequencies and sperm morphology changed in the long- (n = 226) and short-line (n = 201) birds in three generations of artificial selection (long sperm line: L0-L3; short sperm line: S0-S3). Grey dots represent all individual birds and the coloured dots are coded according to their karyotypes as in Fig. 2. Blue contour lines show the density of samples in the morphometric space in each generation. Note that the founder generations (L0 and S0) were selected according to the breeding values of pre-selection line birds (see Methods).

Acknowledgements

We thank Lynsey Gregory, Gemma Newsome and Phil Young for help with animal care. Gerhard van der Horst and Jim Mossman provided CASA training and advice. Rachel Tucker and Lewis Ottaway assisted with DNA extractions. Sebastian Manley and Eleanor McLaren assisted with sperm measurements. Claire Bloor, Alessandro Davassi, and Geoff Scopes, all of Affymetrix, provided help with the SNP chip design and quality control. Alison Downing, Karim Gharbi, Helen Gunter, Judith Risse, Richard Talbot and Urmi Trivedi of Edinburgh Genomics assisted with SNP genotyping and gene expression microarray scanning. The study was funded by grants BB/I02185X/1 from the Biotechnology and Biological Sciences Research Council (to JS) and ERC-2010-AdG from the European Research Council (to TRB), and by a PhD studentship from the Natural Environment Research Council (to CB).

Author Contributions

CB, NH and LB collected and measured sperm data. CB, NH and TRB designed and implemented the selection line experiments. K-WK designed the SNP chip and performed molecular work. K-WK and JS analysed the data. TRB managed the long-term study of zebra finches in Sheffield. LLH and SCG collected the samples from the Australian population. K-WK and JS wrote the paper with contributions from all other authors. TB, TRB and JS conceived the study.

Author Information

Correspondence and requests for materials should be addressed to K-WK (k.kim@sheffield.ac.uk) and JS (j.slate@sheffield.ac.uk).

Data availability: Sperm morphological and velocity data, and genotype PLINK files can be downloaded from Dryad (doi:10.5061/dryad.p4238). Microarray data are available on the Gene Expression Omnibus (GEO) under accession number GSE96970. GWAS and gene expression summary data are available as supplementary materials.

Competing interests: The authors declare no financial competing interests.

Methods

Animals: We studied outbred birds born in 1999-2002 (n = 810; hereafter pre-selection birds) and artificially selected birds born in 2006-2011 (n = 550; Long sperm line: L_0-L_3; Short sperm line: S_0-S_3) that were part of a domesticated zebra finch population maintained at the University of Sheffield since 1985. No power analysis was performed prior to the study. Throughout the study, traits were measured blind to treatment and to values of other traits. All male birds from each line were genotyped and phenotyped i.e. there was no potential for non-random selection of birds.

Previously, a bidirectional artificial selection experiment, for total sperm length was performed for three generations to produce long and short sperm male lines as described elsewhere[6]. Briefly, the founders of each line (cohort 0) were selected based on the estimated breeding values (EBVs) using pedigree information and the 'animal model'[44,45], implemented in AsREML v3.0[46]. In each subsequent generation (cohort 1-3), thirty pairs with extreme breeding values were selected from each line to breed for the next generation. Despite the short period of selection, the divergence of the sperm length between selection lines has significantly increased[6].

In addition, we studied zebra finches from Australia to confirm our main findings. Domesticated and wild birds have previously been shown to have similar sperm morphology[47]. Wild birds (n = 50) were a mixture of adult birds and their F1 and F2 offspring bred in captivity at Macquarie University. They were taken into captivity from two locations in far-west New South Wales in 2007 and 2010 (under Scientific Licence S11374 from the NSW National Parks and Wildlife Service). All males were held in a large aviary with a similar number of females (10m x 36m x 2.5m) and were breeding at the time that the samples were taken (in December 2014). The domesticated birds (n = 49) originally came from Australian aviculturists and represent the domesticated population of zebra finches that have been held and bred in captivity over the past hundred years. These males were also held in aviaries with females and were actively breeding at the time of sampling (between October 2014 and May 2015.

Sperm morphology: For the morphology analysis, sperm samples were collected from the ejaculatory duct of one seminal glomerus (SG) by dissection or faeces of each male and fixed in 5% formalin[5,6]. Five undamaged sperm, without developmental abnormalities, were photographed using the Spot Diagnostics image analysis system version 3 (Diagnostic Instruments. Inc), or light microscopy (Leitz Laborlux S) at x400 and an infinity 3 camera (Luminera Corporation). The length of head, midpiece and tail were measured to the nearest 0.01μm using ImageJ[48]. The flagellum length is the sum of midpiece and tail. The total length is the sum of the three components.

Sperm motility: Live sperm were collected from the left SG of euthanized selection line male birds and suspended in phosphate buffered saline solution (PBS). Computer assisted sperm analysis (CASA) was used to measure three parameters of swimming velocity with a Sperm Class Analyzer® (Microptic, Barcelona, Spain) as described in elsewhere[6]: curvilinear velocity (VCL), average path velocity (VAP) and straight line velocity (VSL). A principal component analysis (PCA) was used to combine these measures as they are highly correlated[5] and the first principal component was used as the measure of velocity.

DNA and RNA extraction: The carcasses of pre-selection birds had been kept frozen at -80°C and the brains or breast muscles were removed for DNA extraction. Blood samples of selection line birds and the Australian birds were collected and stored in 100% ethanol at RT. DNA was extracted using a standard ammonium acetate method[49].

Males of each third generation selection line that had extreme sperm length (mean total length ± s.d. = 74.7 ± 2.4 (n = 22) and 54 ± 3.9 μm (n = 22) for long and short sperm line respectively) were selected for RNA expression analyses. Testes were collected from both sides of males and stored in -80°C after treating with RNALater (Qiagen, Hilden, Germany). The testes were homogenized using TissueLyser (Qiagen) and stainless steel beads. RNA was extracted using TRIzol® (Thermo Fisher

Scientific, Waltham, Massachusetts, United States) and RNeasy Mini kits (Qiagen). The quality of RNA was checked using a Bioanalyzer (Agilent, Santa Clara, California, United States) and high quality samples with RNA integrity value > 7 were used for further analysis.

RADSeq and marker development: In a preliminary experiment, we used restriction-site-associated DNA (RAD) tag sequencing[50] to genotype 550 selection line birds and to obtain SNP information specific to this population. RAD libraries were constructed using SbfI restriction enzyme and sequenced on HiSeq2000 platform (Illumina, San Diego, California, United States) at Edinburgh Genomics to produce 100bp paired end reads. The short sequence reads were cleaned, mapped and SNPs were called using a custom pipeline. First, the quality of the short sequence reads (2 x 100bp) were checked using FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) then the raw reads were cleaned and de-duplicated using process_radtags and clone_filter utility programs of Stacks[51]. Cleaned sequence reads were then mapped on to the zebra finch reference genome assembly (taeGut3.2.4) using BWA ver. 0.7.6a[52]. Single nucleotide variants were called using SAMTOOLS ver.0.1.18[53] and the SNP filtering criteria were determined based on the observed distribution of SNP depth per site (minimum depth = 5, maximum mean depth = 25), minimum SNP quality > 250, SNP calling rate > 80% and minor allele frequency = 0.05 using custom R scripts and VCFTOOLS ver.0.1.12[54].

SNP chip development and typing: We tested different sources of SNPs, including 54,958 SNPs filtered from RADSeq, 2,385 SNPs of EST sequences and 684,001 SNPs of non-EST from dbSNP (http://www.ncbi.nlm.nih.gov/snp) for compatibility with the chip design in accordance with the Affymetrix Axiom array design criteria. A total of 610,626 SNPs were plated onto the 600K SNP Axiom Genotyping Array (Affymetrix, Santa Clara, California, USA). 1,344 male zebra finches including 810 pre-selection line and 435 selection line birds, in addition to 99 Australian birds (50 birds from wild population and 49 birds from captive populations) were genotyped on an Affymetrix GeneTitan system at Edinburgh Genomics. The quality of genotype data was checked using the best practice workflow of the Axiom Analysis Suite program (AxAS, Affymetrix) with following

parameters: DQC >= 0.82, sample call rate >= 97% and SNP call rate >= 95%. Genotypes were determined and filtered by analyzing clusters based on contrast and strength of signal using a Bayesian genotyping algorithm (BRLMM-P) implemented in the AxAS. Any SNPs that did not form well-separated clusters as measured by Fisher's linear discriminant analysis (FLD < 3.6) were excluded. SNPs were classified into six categories, based on the features of clusters and only polymorphic SNPs with high resolution (n = 247,811) and the SNPs with no minor homozygotes (n = 74,909) were selected (total 322,720 SNPs, 52.9% of the genotyped SNPs) and total 1,202 individuals were retained for further analysis.

Conversion rates were different among the source of SNPs. RAD sequencing-derived SNPs showed the highest conversion rate (72.2%) and the lowest proportion of the monomorphic homozygote SNPs. This is probably because the SNPs were selected from a previously tested data set and the flanking sequences were obtained from birds within our study population. In contrast, the reference sequence-derived SNPs showed the lowest conversion rate (51.7%). There are two factors that may cause the low conversion rate for this latter source of SNP. First, the source individual (the reference sequence bird) was from another population and SNPs were called from just this one male[7]. Therefore real SNPs may be absent in our population, and there may also be a high false positive rate in the SNPs reported in the database. Second, the proportion of effectively unconverted low resolution SNPs is higher than SNPs sourced from the RADSeq experiment. This could be because the sequences flanking the SNP in our study population contain polymorphic sites or fixed differences from the reference sequence, causing a reduction in the genotyping quality. Among the RAD-derived SNPs, loci with identified polymorphisms in the flanking sequences were effectively excluded in the filtering process of the SNP chip design, while the EST-derived sequences are likely to be relatively highly conserved as they contain coding regions. Therefore, flanking site polymorphism should be less frequent in these categories of SNP.

Data analyses: All statistical analyses and visualization were performed using R version 3.2.2 (R Core Team 2015) and relevant packages.

Quantitative genetic analysis: To determine the quantitative genetic parameters of sperm morphological traits, we used a Bayesian mixture model which allows all SNPs to be fitted simultaneously. Models were implemented in the BayesR[55] software, with the Markov chain Monte Carlo (MCMC) chain being run for 60,000 iterations, a burn-in period of 20,000 iterations, and then 1,000 estimates being obtained by sampling the remaining iterations every 40th iteration. The -permute flag was specified, so that the order that SNP effect sizes were estimated was randomised each iteration. The effect size of each SNP, expressed as an amount of variation explained, was estimated as:

$$V_{SNP} = 2 \times MAF \times (1\text{-}MAF) \times \beta^2$$

where MAF is the minor allele frequency and β is the effect size of an allelic substitution reported by BayesR. SNP effects were summed for each chromosome and for the entire genome. Thus, it was possible to estimate the amount of variation explained by each SNP individually, by all of the SNPs on each chromosome and by all SNPs combined. Summing the effects of all SNPs provides an estimate of the additive genetic variance (V_A) and therefore the heritability (defined as $h^2 = V_A/V_A + V_R$) where V_R is the residual variance. Summing the effects of SNPs on each chromosome enables chromosome partitioning, where the relative contribution of each chromosome to the overall heritability can be determined[11]. The proportion of additive genetic variance explained by the Z chromosome is defined as V_Z/V_A where V_Z is the summed effect of SNPs on the Z chromosome. Analyses were performed on the same set of pre-selection line males and SNPs as in the GWAS analyses.

Genome Wide Association Studies (GWAS): Additional quality control filtering was performed to remove SNPs with Minor Allele Frequency (MAF) < 0.02, SNPs out of Hardy Weinberg Equilibrium (HWE) at $P < 0.001$ and individuals that were > 0.95 identical by state (duplicated samples). A total of 676 pre-selection males typed at 219,279 SNPs with an annotated position in the zebra finch genome were retained, and used in the genome wide association analysis. The association between SNPs and four morphological traits (head, midpiece, tail and total sperm length) was assessed using

genome-wide rapid association using mixed model and regression (GRAMMAR) implemented in the R package GenABEL ver. 1.8[56]. In the first step residuals from a polygenic model were extracted fitting a genetic relationship matrix (GRM). The GRM accounts for population structure and relatedness between individuals, in order to reduce test statistic inflation and the risk of spurious associations. The residuals were then treated as phenotypes in a genome-wide association analysis using a linear model. We treated the Z chromosome as an autosomal chromosome because only males (two Z chromosomes) were included in the analysis.

For the analyses of the midpiece, tail and total length the inversion polymorphism in the Z chromosome caused an excess of significant test statistics across a large central region of the Z chromosome (Supplementary Fig. 1). This arises because most SNPs over a 60Mbp region are in high LD with unknown causal variants affecting a large proportion of the variance in sperm morphology. After fitting the karyotype of the Z chromosome as a covariate, the distribution of P values was closer to the uniform distribution, demonstrating that the Z-linked SNPs tag the inversion polymorphism and that their significant effects disappear in models where the inversion is already fitted, providing additional evidence that the inversion polymorphism determines sperm morphology (Supplementary Fig. 1).

Principal Component Analysis and karyotype clustering: We used a principal component analysis-based approach for detecting inversions[57]. PCAs for individual chromosomes were performed using GCTA[58], and distinct clusters were identified at several chromosomes (Supplementary Fig. 3). In particular, a clustering analysis using the pam function in the R package cluster[59] identified six clear clusters with the Z chromosome SNPs (Fig. 2). Based on the Z chromosome clustering, the samples were subsetted, and for each subset the heterozygosity per SNP and the linkage disequilibrium (LD) between SNPs were estimated using the --hardy and --r2 functions in PLINK ver. 1.9[60]. Cluster analysis of the Z chromosome reported in the main text was performed by using all samples (n = 1,202), including Australian birds to ensure consistent cluster labelling between the different datasets.

The effects of Z-inversion karyotypes on sperm morphology were tested with a linear model using the lm function in R and by fitting karyotype as a fixed effect.

EigenGWAS: We tested which genomic regions caused the response to divergent selection in the selection lines by performing an EigenGWAS analysis[12]. EigenGWAS first uses PCA on the SNP data to measure population genetic structure, before treating the individual Eigen vectors as a phenotype that can be analysed in a GWAS[12]. Regions that exceed statistical significance (after lambda correction for inflated test statistics) cause more of the structure in the data than is expected by genetic drift, and are therefore underlying the response to selection. Analyses were performed on Cohort 3 males (n = 149) from the long and short lines. We used the first principal component, calculated in GCTA from autosomal SNPs, as the phenotype. EigenGWAS was performed with the GEAR software[12]. Removing the Z chromosome from the PCA and retaining it for the GWAS is conservative, as any signatures of selection at Z chromosome SNPs cannot be attributed to the Z chromosome inversion having a large influence on the PCA.

Transcriptome analysis: cRNA prepared from testis mRNA was hybridized to an Affymetrix Zebra finch Gene 1.1 ST 96 array plate (Affymetrix, Santa Clara, California, United States) and scanned at ARK-Genomics (now Edinburgh Genomics) using the GeneTitan Instrument (Affymetrix). The array contains probe sets from 18,595 genes. When genes were represented by multiple probes, probe expression levels were combined to determine gene expression and normalized using a robust multi-array average (RMA) algorithm implemented in the Affymetrix Expression Console ver 1.3. Overall expression levels per sample were compared and any outlier samples, as judged by Pearson's correlation, were excluded. Expression levels of 69 testes from 42 males were retained and used for further analyses.

The gene expression differences between lines were examined in a linear mixed-effects model (LMM) framework, using the R package lme4[61]. For each locus, normalized log expression values were fitted as the response variable, line and testis side (left or right) as fixed effects and the male

identity as a random effect. The Kenward-Roger approximation, implemented in lmerTest[62], was used to evaluate the significance of the fixed effect. Bonferroni corrected significance levels ($\alpha = 0.05$) were used to correct for the multiple tests. To visualize the expression differences between lines, a heat map with hierarchical clustering for the RNA expression was generated using the heatmap.2 function of gplots[63] and the hclust library in R (Supplementary Fig. 2).

The genomic regions associated with the overall expression differences between lines were identified by using a GWAS approach with the focal phenotype being the first principal component of expression. First, a PCA was performed for the genes that showed significant differences between lines ($P < 0.05$) using prcomp function in R. Then in a linear mixed-effect GWAS model implemented in the R package RepeatABEL (http://www.genabel.org/packages/RepeatABEL)[64], the primary eigenvector of the PCA was treated as phenotype, the selection line and the testis side (left or right) were fitted as fixed effects and the male identity was fitted as a random effect to account for the repeated measures of male testes.

Gene ontology (GO) term enrichment analysis: Human orthologues of the genes that were differentially expressed between selection lines were obtained by using the Ensembl transcript ID and the Uniprot databases (www.uniprot.org). GO term enrichment analysis was performed using GOrilla[65] by submitting the 108 differentially expressed genes as a target set and the 14,224 transcripts that have an annotated position in the zebra finch genome as the background set (see Supplementary Material for the detailed results of the transcriptome and gene ontology analysis).

Effect of karyotype on sperm velocity: To examine the effect of karyotypes on the sperm velocity, the observations of 11,254 sperm (mean observations per male = 116, s.d. = 69.7) from 97 males ($n_{AA} = 24$, $n_{AB} = 16$, $n_{AC} = 11$, $n_{BB} = 38$, $n_{BC} = 6$, $n_{CC} = 2$) were analysed in a linear mixed-effect model to account for the multiple measurements of sperm per male, using the R package lmerTest[62]. The primary eigenvector of PCA performed on three measures of velocity (VCL, VAP and VSL) was used as a response variable. The karyotype was fitted as a fixed effect and the male identity was fitted as a

random effect. The significance level of the fixed effect was estimated using a Kenward-Roger approximation.

Effect of karyotype on within-male sperm morphological variation: Previous work has shown that sperm morphology is highly repeatable within males, both within and between ejaculates[4,66]. However, there remains some within-male variation, which was quantifiable because 5 sperm were measured per male. The coefficient of variance (s.d. / mean) was measured for 5 sperm morphological traits (head length, midpiece length, tail length total length and midpiece:tail ratio) for each male and used as an indicator of within-male variation in sperm morphology. The effect of karyotype on within-male morphological variation was assessed by ANOVA.

Effect of karyotype on fertilisation Success: In a previous experiment we performed fertilisation success competitions between pairs of males from the long and the short selection lines, using DNA profiling to determine the father of embryos[6]. Long-line males outcompeted short-line males and were more likely to fertilise eggs, regardless of mating order. We re-examined the data in light of the Z-chromosome karyotype information. Nearly all of the short-line males had karyotype BB while long-line males were mostly AA, AB or AC. Among long-line males, AB males appear to be the most successful against short-line males (Supplementary Table 3). However, simple estimates of karyotype success rate are potentially compromised because there were fewer male-male competing pairs than embryos and so the data are non-independent.

To overcome this problem, we used generalised linear mixed models (GLMMs), implemented in MCMCglmm[67]. For each embryo, paternity success of the long-line male was treated as a binomial outcome ('success', 'failure') with mating pair fitted as random effect and long-line male karyotype as a fixed effect. Models were run for 600,000 iterations with a burn-in period of 100,000 iterations and an estimate obtained from every 100th iteration thereafter. The residual variance was set to a fixed

prior of 1.0 and the 'competing pair' random effect was given a non-informative inverse-Gamma prior with V = 1 and nu = 0.002.

Effect of karyotype on fitness: Aviary data were available for 2,884 eggs from 575 clutches, sired by a total of 168 males and using a total of 187 different pairings. Three fitness variables were available: hatching success rate, chick survival rate from hatchling to independence ('fledging'), and overall survival rate from egg to independence. There were no CC males among this dataset.

The raw suggest that BB and AC have lower fitness (Supplementary Table 4), but individual eggs and chicks are non-independent. Due to repeated measures at the level of male and breeding pair we used generalised linear mixed models (GLMMs) to test whether Z chromosome karyotype caused variation in male fitness. In all models the response was fitted as multinomial distribution (using count data of survivors and non-survivors). Male identity and breeding pair identity were fitted as random effects, and male karyotype was fitted as a fixed effect. Models were run for 600,000 iterations with a burn-in period of 100,000 iterations and an estimate obtained from every 100^{th} iteration thereafter. No prior was specified and no model showed evidence of autocorrelation between samples.

Supplementary Table 1 | Genetic architecture of sperm traits estimated by fitting all SNP simultaneously in BayesR. V_A, additive genetic variance. V_e, environmental variance. h^2, narrow sense heritability. V_Z/V_A, proportion of additive genetic variance explained by the Z chromosome. N_SNPs, number of SNPs contributing to V_A. Estimates are mean and 95% credible intervals, estimated from 1,000 samples of the MCMC chain, 60,000 iterations, burn-in = 20,000 iterations, thinning interval = 40 iterations. Analyses were performed on 676 pre-selection birds.

Phenotype	V_A	V_e	h^2	V_Z/V_A	N_SNPs
head	0.16 (0.04-0.31)	0.24 (0.18-0.30)	0.41 (0.13-0.63)	0.083 (0.001-0.432)	3,204 (434-7,125)
midpiece	5.65 (2.87-9.22)	8.89 (7.27-10.65)	0.39 (0.22-0.55)	0.671 (0.306-0.910)	1,366 (322-3,019)
tail	15.60 (10.55-22.09)	11.03 (8.78-13.30)	0.59 (0.46-0.71)	0.899 (0.722-0.983)	1,280 (224-3,086)
total length	11.04 (7.75-15.44)	7.71 (6.15-9.27)	0.59 (0.47-0.70)	0.824 (0.609-0.964)	1,178 (248-3,141)

Supplementary Table 2 | Heterogeneous pattern of mean (s.d.) observed heterozygosity (H) among putative Z chromosome karyotypes, across different sections of the Z chromosome. The gaps between sections are the predicted locations of break points causing genomic rearrangements. Positions are those on the current zebra finch reference genome assembly (Taeniopygia_guttata-3.2.4). Sections 1 and 7 show similar patterns in all karyotypes. Sections 2-6 distinguish Haplotype A, B and C. See Fig. 2 and Supplementary Fig. 4. Analyses were performed on 676 pre-selection birds.

						Mean (s.d.) heterozygosity for karyotype of Z chromosome					
Section	Start (bp)	End (bp)	Size (bp)	Gap * (bp)	No. SNP	H_{AA}	H_{AB}	H_{AC}	H_{BB}	H_{BC}	H_{CC}
1	1	5,798,614	5,798,613		2,463	0.17 (0.19)	0.17 (0.19)	0.18 (0.19)	0.14 (0.19)	0.19 (0.22)	0.13 (0.17)
2	5,822,233	60,264,951	54,442,718	23,619	4,625	0.08 (0.15)	0.53 (0.45)	0.31 (0.4)	0.02 (0.07)	0.64 (0.47)	0.02 (0.07)
3	60,283,286	62,504,794	2,221,508	18,335	503	0.18 (0.18)	0.17 (0.18)	0.82 (0.21)	0.02 (0.05)	0.97 (0.17)	0.03 (0.09)
4	62,507,513	67,311,640	4,804,127	2,719	96	0.07 (0.15)	0.62 (0.44)	0.33 (0.41)	0.02 (0.05)	0.62 (0.47)	0.01 (0.02)
5	67,380,137	68,339,423	959,286	68,497	260	0.04 (0.1)	0.27 (0.3)	0.34 (0.28)	0.22 (0.16)	0.3 (0.21)	0.3 (0.22)
6	68,342,761	68,887,401	544,640	3,338	235	0.03 (0.07)	0.3 (0.31)	0.54 (0.44)	0.27 (0.18)	0.46 (0.35)	0.04 (0.07)
7	68,898,688	72,852,975	3,954,287	11,287	805	0.19 (0.18)	0.17 (0.17)	0.19 (0.18)	0.17 (0.18)	0.17 (0.17)	0.18 (0.18)

* the size of gap between the end position of previous section and the beginning of the section.

Supplementary Table 3 | Summary data on fertilisation success of long-line versus short-line males. All short line males were karyotype BB.

(a) Summary Data			
Long line karyotype	$N_{embryos}/N_{pairs}$	N successes/failures	Success Rate
AA	89/9	54/35	0.61 (0.50-0.71)
AB	43/4	36/7	0.84 (0.69-0.93)
AC	29/2	14/15	0.48 (0.29-0.67)
Total	161/15	104/57	0.65 (0.57-0.72)

(b) Model Summary – all karyotypes			
Random Effects	Estimate	95% credible interval	
Competing Pair	7.032	0.835-17.26	
Residual	1.00	1.00-1.00	
Fixed Effects	Posterior Mean*	95% credible interval*	P_{MCMC}
AA	1.043 (0.74)	-0.930-3.009 (0.31-0.96)	0.255
AB	3.425 (0.97)	0.072-6.875 (0.61-1.00)	**0.027**
AC	-0.156 (0.46)	-4.094-3.441 (0.02-0.97)	0.922

*Values in parentheses are model estimates converted back to the original scale using an inverse logit transformation.

Supplementary Table 4 | Summary data and models on survivorship from egg to hatchling, from hatchling to fledgling and from egg to fledgling. There was no significant difference in egg hatching success between offspring of different male karyotypes. Chicks sired by AC males had marginally significantly lower chick survival rates. BB and AC males had clutches with the lowest survival rates from egg to fledging. We urge caution interpreting these findings because we have not studied fitness in the wild, the karyotype of females was unknown in our sample, and female fitness was not considered. However all of the karyotypes are capable of producing surviving chicks with approximately equal survival to fledgling.

Summary Data

Karyotype	$N_{eggs}/N_{hatchlings}$	Hatch success	$N_{hatchlings}/N_{fledglings}$	Hatchling survival	$N_{eggs}/N_{fledglings}$	Egg to Fledgling survival
AA	554/283	0.51	283/220	0.78	554/220	0.40 (0.36-0.44)
AB	569/274	0.48	274/201	0.73	569/201	0.35 (0.31-0.39)
AC	376/162	0.43	160*/101	0.63	372/101	0.27 (0.23-0.32)
BB	1,102/499	0.45	499/320	0.64	1,102/320	0.29 (0.26-0.32)
BC	283/158	0.56	154*/109	0.71	279/109	0.39 (0.33-0.45)

Model (a) Hatching Success

Random Effects	Estimate	95% credible interval	
Male ID	2.17	<0.001-3.833	
Breeding Pair	1.16	<0.001-3.36	
Breeding Attempt	1.65	0.072-5.076	
Residual	2.46	1.666-3.233	
Fixed Effects	**Posterior Mean**	**95% credible interval**	**P_{MCMC}**
Intercept	1.080	-0.083-2.460	0.066
AB	-0.355	-1.374-0.651	0.500
AC	-1.088	-2.288-0.174	0.085
BB	-0.755	-1.625-0.242	0.104
BC	0.087	-1.362-1.223	0.889

Supplementary Table 4 (continues)

Model (b) Hatchling Survival

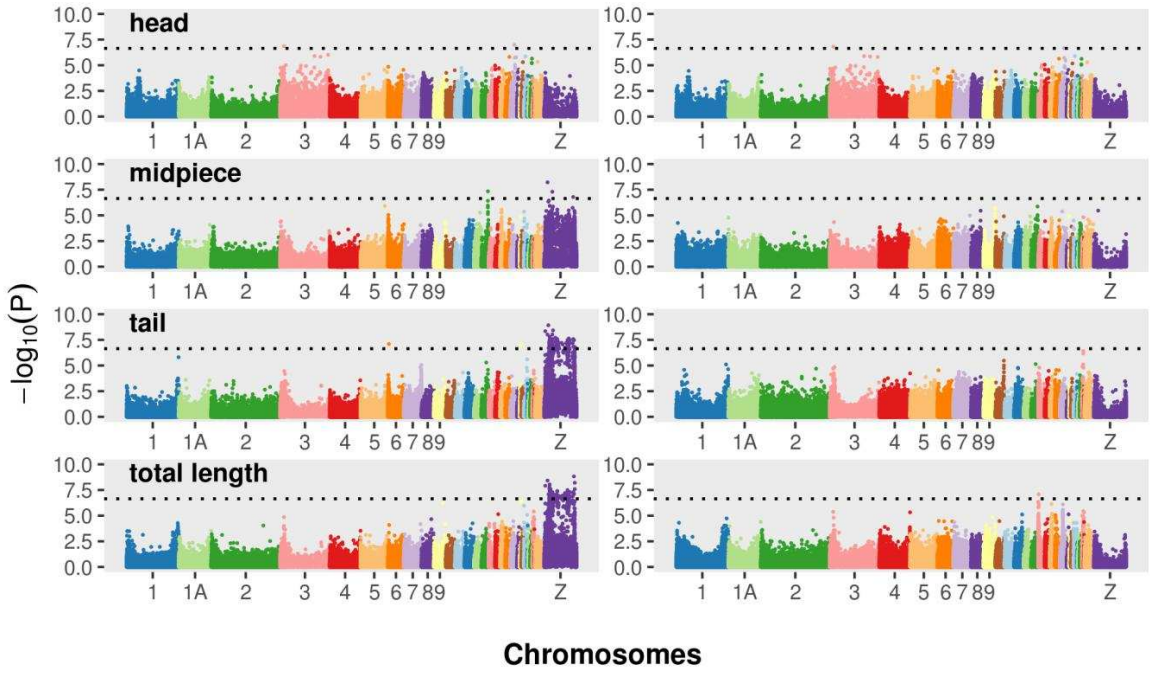
Random Effects	Estimate	95% credible interval	
Male ID	0.528	<0.001-1.389	
Breeding Pair	0.691	<0.001-1.548	
Breeding Attempt	0.019	<0.001-0.073	
Residual	0.752	0.161-1.394	
Fixed Effects	**Posterior Mean**	**95% credible interval**	**P_{MCMC}**
Intercept	1.707	1.120-2.257	**$<2.0 \times 10^{-4}$**
AB	-0.282	-1.067-0.456	0.475
AC	-1.010	-1.860- -0.049	**0.025**
BB	-0.652	-1.344-0.026	0.062
BC	-0.395	-1.350-0.492	0.408

Model (c) Survival from egg to hatchling

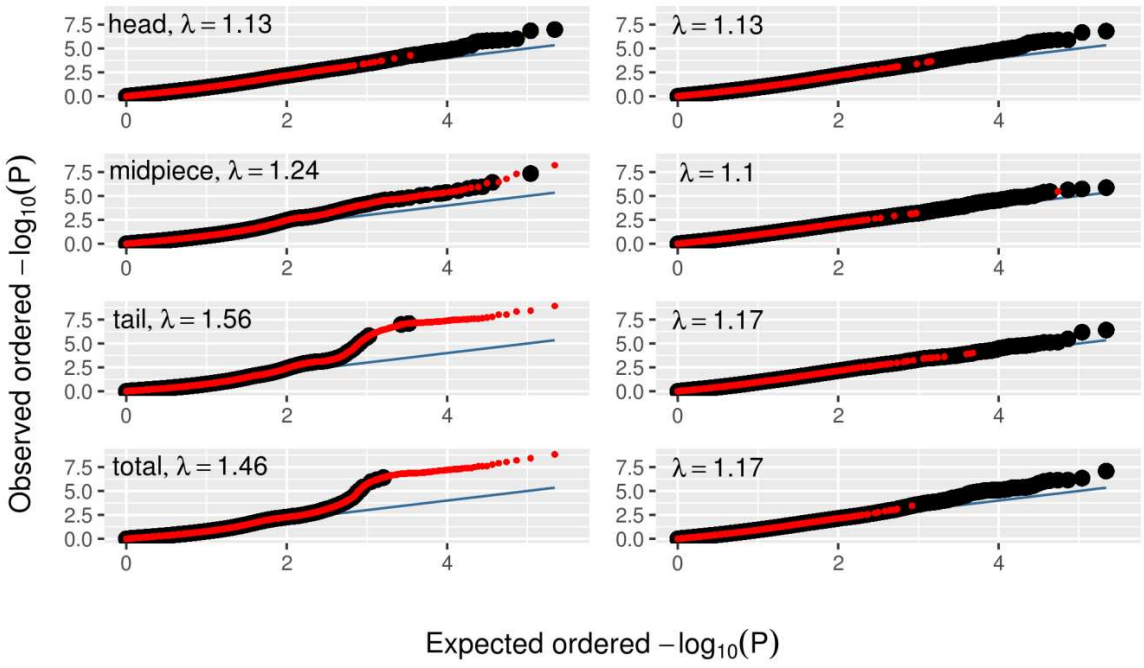
Random Effects	Estimate	95% credible interval	
Male ID	1.408	<0.001-2.789	
Breeding Pair	1.186	<0.001-2.965	
Breeding Attempt	1.346	0.019-4.309	
Residual	1.868	1.215-2.586	
Fixed Effects	**Posterior Mean**	**95% credible interval**	**P_{MCMC}**
Intercept	0.225	-0.925-1.417	0.725
AB	-0.334	-1.232-0.566	0.479
AC	-1.342	-2.471- -0.219	**0.021**
BB	-0.900	-1.720- -0.125	**0.027**
BC	-0.325	-1.490-0.759	0.580

* Numbers of hatchlings in the AB and BC rows vary slightly between columns as a small number of hatchlings were cross-fostered. These birds were removed from analyses of survival from hatchling to fledgling.

a



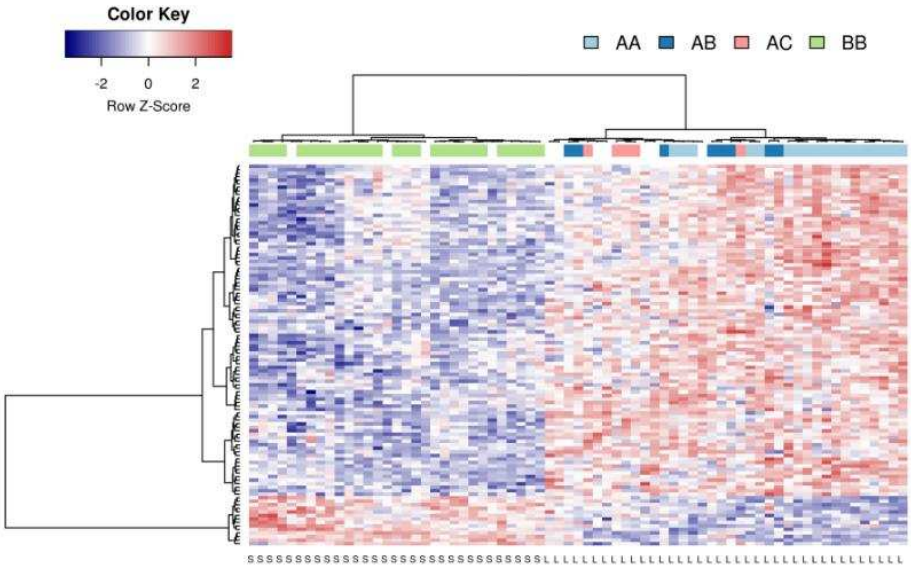
b



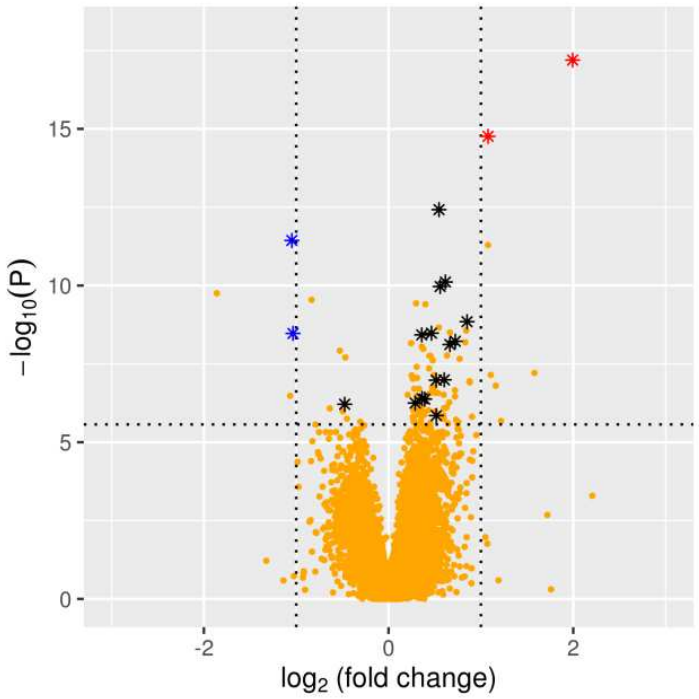
Supplementary Figure 1 | GWAS of four sperm morphological traits in pre-selection males (n = 676). a, The $-\log_{10}$(P value) of the association between the trait and the SNP is plotted for each SNP. P values were obtained using the GRAMMAR method and corrected for the genomic inflation factor (λ). Right panels show the results after fitting the karyotype of Z chromosome (see Fig. 2) as a covariate. Bonferroni adjusted genomewide significance thresholds (genomewide P = 0.05) are shown

with dashed lines. **b**, QQ plots for the GWAS of four sperm morphological traits in pre-selection males. In each panel, expected P values are shown with blue lines. Black dots represent the observed values of autosomal SNPs and the smaller red dots show SNPs on the Z chromosome. Right panels show the same models, but after fitting the Z chromosome karyotype as a covariate. Note that the genomic inflation (λ) is mostly caused by SNPs tagging the Z chromosome karyotype.

a

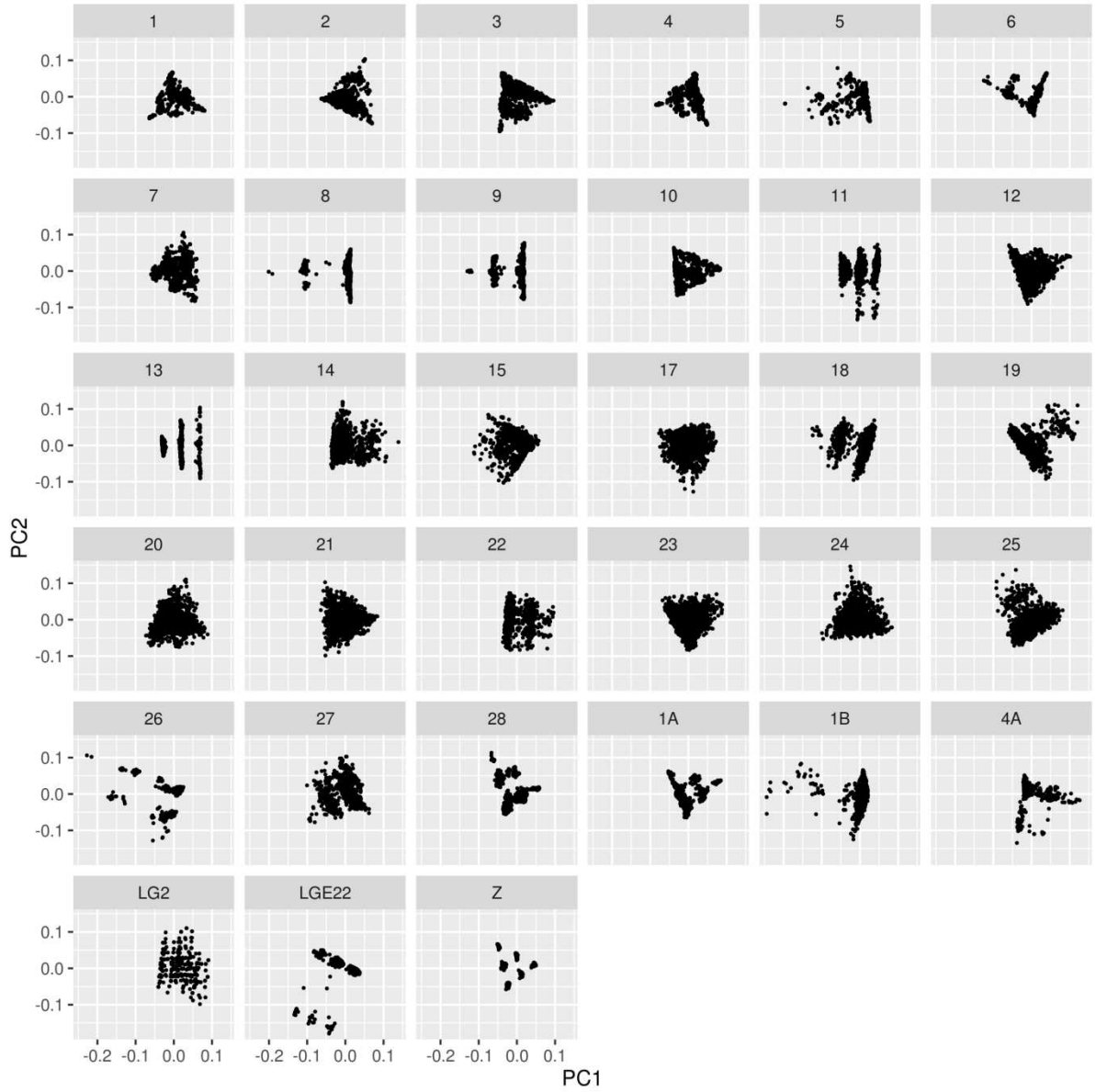


b



Supplementary Figure 2 | Gene expression differences between lines. a, Heatmap where each row indicates 108 differentially expressed transcripts at the Bonferroni adjusted significance threshold of P = 0.05, and each column indicates 69 testes of selection line males (n = 42). Hierarchical clustering at the top of the heatmap shows the two selection lines in which the individual karyotype of Z

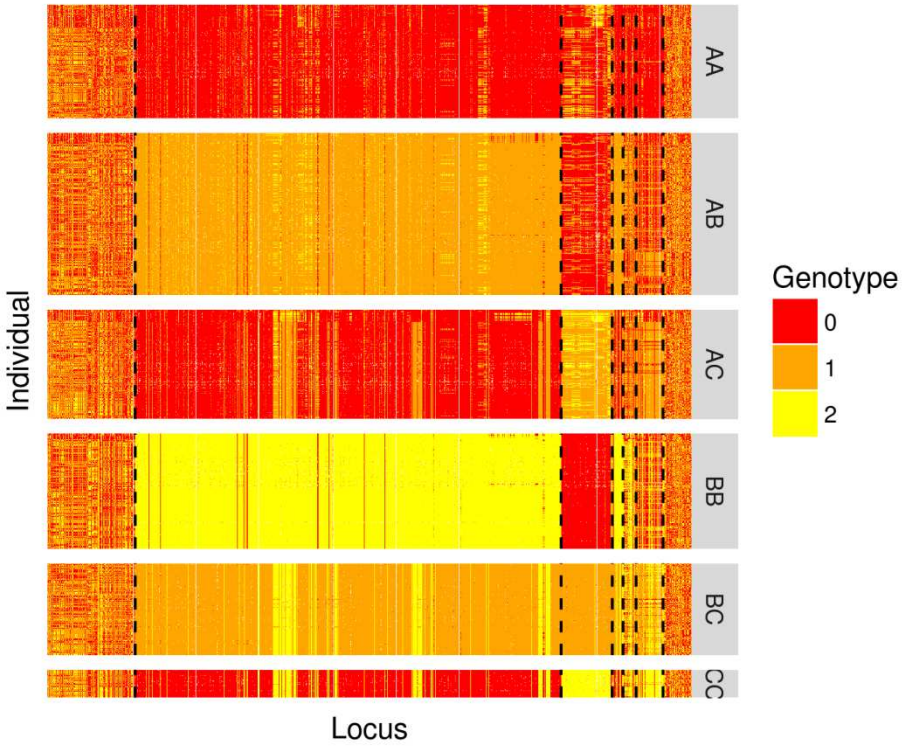
chromosome is indicated by colour (individuals with unknown genotype are unshaded) and the selection line (L for long and S for short) is indicated at bottom. Note that most differentially expressed genes are upregulated in the long line. **b.** Volcano plot of between-line gene expression differences, indicating the statistical significance and the $\log_2$ fold expression difference of each gene. Horizontal dotted line indicate a Bonferroni-adjusted significance threshold of P = 0.05. Vertical dotted lines show the two fold differences between lines. Star shaped dots represent genes on the Z chromosome.



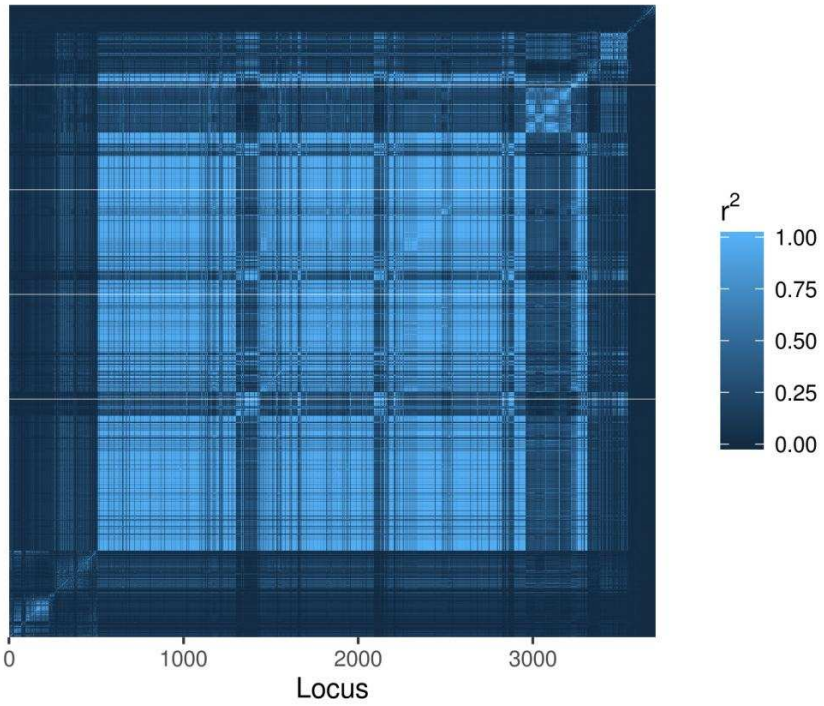
Supplementary Figure 3 | PCA for the captive population using all SNPs on each chromosome.

Pre-selection (n = 676) and selection line (n = 427) birds in Sheffield captive population were included in the analysis.

a

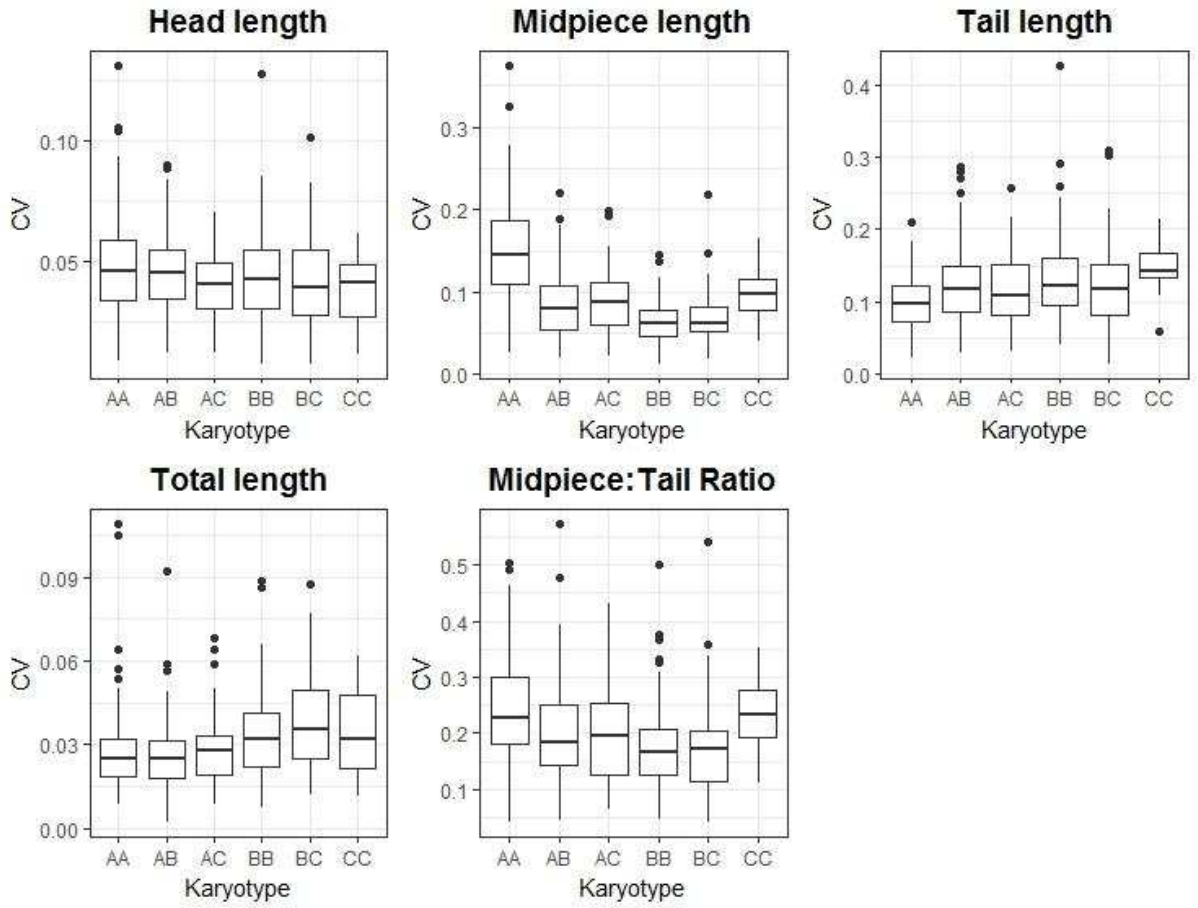


b

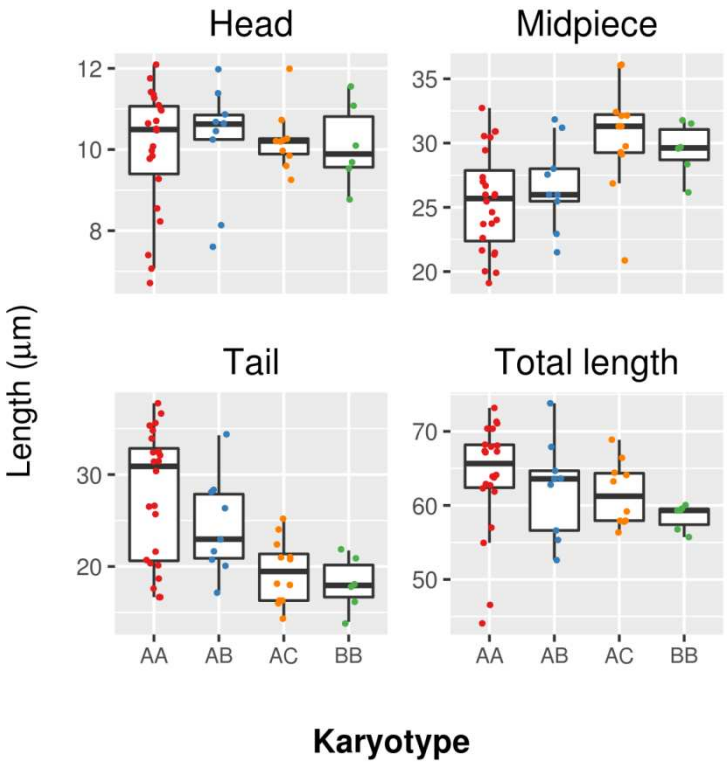


Supplementary Figure 4 | Genotypes and the pattern of LD of Z chromosome SNPs for 1,202 birds from the Sheffield and Australian populations. Of 8,987 Z-linked SNPs, 3,704 with MAF > 0.25 are plotted for clarity. **a**, Reference alleles were defined as the major alleles in AA individuals

and genotypes were coded as ‘0’ representing homozygotes for reference allele, ‘2’ representing homozygotes for the alternative allele and ‘1’ being a heterozygote. Six dashed lines indicate the putative breakpoints derived from the abrupt changes in patterns of genotypes and heterozygosity (see Fig. 2). See Supplementary Table 2 for the position of breakpoints. **b**, Pairwise linkage disequilibrium (r^2) between SNPs.



Supplementary Figure 5 | The effects of karyotype on within-male variation in sperm morphology. For the pre-selection line birds of known karyotype and for whom 5 sperm were measured (n = 418), the coefficient of variance was estimated for each morphological trait. The effect of karyotype on within-male variation was assessed by ANOVA. Head length ($F_{5,412}$ = 2.096, P = 0.065); Midpiece length ($F_{5,412}$ = 51.38, P < 2.2 x 10^{-16}); Tail length ($F_{5,412}$ = 5.617, P < 5.1 x 10^{-5}); Total length ($F_{5,412}$ = 5.997, P < 2.3 x 10^{-5}); Midpiece:Tail ratio ($F_{5,412}$ = 7.42, P < 1.2 x 10^{-6}).



Supplementary Figure 6 | The effects of karyotype on sperm morphology in Australian birds. Samples include 50 wild birds and 49 captive birds that showed little genetic differentiation in a PCA analysis.