1 **Identification and characterisation of the zebra finch (*Taeniopygia guttata*)**
2 **sperm proteome**

3
4 **Melissah Rowe^{1,2}, Sheri Skerget³, Matthew A. Rosenow⁴, and Timothy L. Karr⁵**

5
6 ¹ *Natural History Museum, University of Oslo, 0562 Oslo, Norway*

7 ² *Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo,*
8 *0316 Oslo, Norway*

9 ³ *Translational Genomics Research Institute, Phoenix, Arizona, USA*

10 ⁴ *Caris Life Sciences, Phoenix, Arizona, USA*

11 ⁵ *School of Life Sciences, Arizona State University, Arizona, USA*

12 Corresponding authors: melissah.rowe@nhm.uio.no; tkarr@asu.edu

13

14 Short title: The zebra finch sperm proteome (ZfSP)

15 Keywords: avian sperm; passerines; sexual reproduction; sperm proteins; positive selection

16

17

18

19

20

21 **Abstract**

22 Spermatozoa exhibit remarkable variability in size, shape, and performance. Our understanding of
23 the molecular basis of this variation, however, is limited, especially in avian taxa. The zebra finch
24 (*Taeniopygia guttata*) is a model organism in the study of avian sperm biology and sperm
25 competition. Using LC-MS based proteomics, we identify and describe 494 proteins of the zebra
26 finch sperm proteome (ZfSP). Gene ontology and associated bioinformatics analyses revealed a rich
27 repertoire of proteins essential to sperm structure and function, including proteins linked to
28 metabolism and energetics, as well as tubulin binding and microtubule related functions. The ZfSP
29 also contained a number of immunity and defence proteins and proteins linked to sperm motility
30 and sperm-egg interactions. Additionally, while most proteins in the ZfSP appear to be
31 evolutionarily constrained, a small subset of proteins are evolving rapidly. Finally, in a comparison
32 with the sperm proteome of the domestic chicken, we found an enrichment of proteins linked to
33 catalytic activity and cytoskeleton related processes. As the first described passerine sperm
34 proteome, and one of only two characterised avian sperm proteomes, the ZfSP provides a
35 significant step towards a platform for studies of the molecular basis of sperm function and
36 evolution in birds.

37 **Introduction**

38 Spermatozoa exhibit enormous morphological diversity across the animal kingdom [1]. Similarly,
39 sperm performance traits (*e.g.* swimming speed, longevity) show considerable variation among
40 species and populations, as well as among males within a population and within and among the
41 ejaculates of a single male [2-4]. Sperm function in fertilisation, however, is highly conserved
42 across all taxa, suggesting variation in sperm form and function is likely the result of a number of
43 other processes, such as post-copulatory sexual selection (*i.e.* sperm competition and cryptic female
44 choice; reviewed in [2-5]), chemotaxis, sperm-egg interactions [6,7], and paternal effects [8-10].
45 Indeed, recent studies have revealed both conserved and variable components of the sperm
46 proteome suggesting a form of “compartmentalized adaptation” [11,12]. These results have led to
47 the suggestion that, from an adaptive and evolutionary perspective, sperm can be considered a type
48 of “functional chimera” and the challenge now is to understand the genetic and molecular basis of
49 sperm variation from this perspective.

50 It is now widely appreciated that females may produce offspring sired by multiple males in a
51 broad range of internally and externally fertilising taxa [13]. In birds, for example, genetic
52 polyandry is a regular occurrence in approximately 86% of all surveyed passerine species [14].
53 When females mate multiply, ejaculates from rival males may spatially and temporally overlap and
54 generate both competition among males for fertilisation success (*i.e.* sperm competition, [15]) and
55 the potential for female control over paternity (*i.e.* cryptic female choice, [16]). These post-
56 copulatory processes constitute a powerful evolutionary force, and post-copulatory sexual selection
57 is credited with driving rapid diversification in a broad range of sexual traits (*e.g.* plumage,
58 genitalia, sperm morphology, [17-19]). Given the essential role of sperm in the fertilisation process,
59 sperm, and associated seminal fluid proteins of the ejaculate, are a major target for the impact of
60 sexual selection resulting from sperm competition and sperm-female interactions.

61 The proteome concept is simple: it is the full protein complement expressed by a genome
62 [20]. However, in practice the determination of a proteome, at any level of analysis (cellular, tissue

63 or organ), is a daunting task for a variety of biological and technical reasons; most notably, the
64 typically high complexity of proteomes and the large quantity of data generated in proteomic
65 studies. As sperm cells are accessible, easily purified, and, in contrast to other cells, extremely
66 differentiated with marked genetic, cellular, functional, and chromatin changes, they are particularly
67 well-suited for proteomic and evolutionary analyses as first shown by Dorus *et al.* [11] (see also
68 [21] for review). Indeed whole sperm cell proteome characterisation has now been conducted in a
69 range of vertebrate and invertebrate taxa (*e.g.* humans [22], mouse [23], rat [24], macaque [25],
70 horse [26], chicken [27,28], *Drosophila* [11,29], *C. elegans* [30], Lepidoptera [31-33], mussel [34],
71 spoon worm [35]). Importantly, through the use of both single-protein and whole cell proteomic
72 approaches, it has become apparent that sperm are equipped with numerous proteins that are critical
73 to sperm structure and function, including proteins involved in sperm motility, sperm capacitation,
74 sperm-egg fusion, and fertilisation (reviewed in [36]).

75 Sperm proteins have also been implicated in the process of reproductive isolation and
76 speciation. For example, in abalone (genus *Haliotis*), the sperm protein lysin, which is involved in
77 the dissolution of the egg vitelline envelope, binds to a receptor (VERL) on the egg vitelline
78 envelope in a species-specific manner [37,38]. Moreover, lysin appears to have evolved rapidly and
79 amino-acid sequences of lysins exhibiting remarkable levels of diversification across species [39],
80 likely driven by coevolution with the constantly changing VERL receptor [40]. Indeed, many
81 reproductive proteins appear to have evolved rapidly (reviewed in [41]), including sperm cell
82 surface proteins with roles in both immunity and reproductive systems [42,43].

83 To date, studies of reproductive proteins have been overwhelmingly focused on invertebrate
84 (*e.g.* marine broadcast spawners, *Drosophila*) and mammalian (*e.g.* mouse, human) taxa [44]. In
85 birds, studies of reproductive proteins are generally lacking [45], and the handful of available
86 studies are limited to species in the Galloanserae (fowl and ducks). In these taxa, there is support for
87 positive adaptive evolution in the avian egg coat protein, Zona Pellucida 3 (ZP3) [46], and a study
88 of the polyandrous Red Junglefowl (*Gallus gallus*) demonstrates that the seminal fluid proteome of

89 males is functionally complex [47]. Finally, two recent studies have characterised the sperm
90 proteome in domestic poultry with the aim of developing diagnostic tools to predict male fertility
91 [27,28]. In contrast, studies of Passeriformes are absent from the literature on reproductive proteins,
92 despite the large amount of data on sexual selection, mating system variation, and reproductive
93 biology in this group. The passerines are the largest and most diverse of the avian clades, and the
94 inclusion of species from this group in the study of reproductive proteins promises to shed further
95 light on the molecular basis of reproductive traits.

96 The zebra finch (*Taeniopygia guttata*) is a small passerine (Passeriformes: Estrildidae)
97 species and an important model organism in several research fields [48], including the neurobiology
98 of vocal learning and reproductive behaviour and biology, especially sperm biology. In zebra finch,
99 total sperm length, as well as length of the various sperm components (*i.e.* head, midpiece, and
100 flagellum length), is highly variable among males and highly heritable [49]. Moreover, the observed
101 variation in sperm morphology is linked to sperm swimming speed [50,51], differential sperm
102 storage by females [52], and fertilisation success [53]. Two recent studies have shown that the
103 variation in sperm length is linked to an inversion polymorphism on the Z chromosome, identifying
104 at least three segregating inversion haplotypes [54,55]. Consistent with the previous studies
105 examining the functional consequences of sperm variation [50-53], inversion haplotype is also
106 associated with sperm swimming speed [54,55] and fertilisation success [54]. Specifically, males
107 heterozygous for the inversion karyotype have faster swimming sperm [54,55] and higher
108 fertilisation success under both non-competitive and competitive mating conditions [54]. As such,
109 heterozygote advantage appears to maintain the inversion polymorphism and hence variation in
110 sperm size in this species. Importantly, while the Z chromosome inversion explains the majority of
111 the genetic variation in sperm morphology observed in zebra finches, knowledge of specific genes
112 that determine sperm variation remains limited.

113 Here, we use LC-MS discovery-based proteomics to describe the zebra finch sperm
114 proteome (ZfSP), with the aim of gaining a deeper understanding of the molecular basis of sperm

115 traits in this species. Finally, given that proteins on the sperm cell surface can be available for direct
116 molecular interactions (*i.e.* protein-protein), knowledge of the protein constituents of zebra finch
117 sperm may provide a first step towards understanding the molecular mechanisms underlying male
118 fertilisation success and sperm-female interactions, as well as post-mating, pre-zygotic reproductive
119 barriers in passerine birds.

120

121 **Materials and methods**

122 *Isolation of spermatozoa*

123 We acquired 50 sexually mature, male zebra finches from commercial breeders. Birds were
124 anesthetized using isoflourane inhalation and immediately euthanized via rapid decapitation. For
125 each male, the testes, ductus deferens, and seminal glomera were dissected from the body cavity
126 and placed in a Petri dish containing Phosphate Buffered Saline (PBS). The paired seminal glomera
127 were then isolated from the remainder of the reproductive tract and transferred to a new Petri dish
128 containing fresh PBS. Sperm were isolated from the seminal glomera by macerating tissue in a
129 microcentrifuge tube containing PBS using a plastic pestle. Purified sperm samples were obtained
130 by repeat cycles of gravity-based separation to settle large particulate matter followed by brief
131 centrifugation (5,000 rpm) using a swinging bucket rotor (Sorvall) to pellet sperm followed by
132 resuspension in PBS. This process was repeated 5 times and then sperm purity was assessed using a
133 DNA-based fluorescence assay. Specifically, to assess sperm purity, we mixed 5 μ l aliquots of our
134 sperm sample with an equal volume of PBS containing 1 μ g/ml of 4',6-diamidino-2-phenylindole
135 (DAPI; Invitrogen) and placed a small amount of this mixture on a microscope slide (10 slides in
136 total). We then examined 50 randomly chosen fields of view across the 10 slides using a Zeiss
137 Axioskop with fluorescence and DIC optics, and scored visible cells as either sperm or non-sperm.
138 Sperm cells were identified by the presence of DAPI-positive sperm nuclei (using epifluorescence)
139 and attached sperm tails (via DIC optics), while non-sperm cells were identified by the presence of
140 DAPI-positive non-sperm cell nuclei. In total, we scored 500 sperm cells and observed only 4 other

cellular structures (*e.g.*, round nuclei, cell bodies), thus confirming that the purification process yielded sperm samples of > 99% purity (Figure 1). The final sperm pellets were then stored at -80 °C until further processing. All procedures were approved by the Institutional Animal Care and Use Committee (IACUC) at Arizona State University (ASU protocol no. 10-1140R).

Protein solubilization and quantification

Sperm pellets were thawed, dissolved in standard SDS sample buffer (Invitrogen) with 1mM DTT included as a reducing agent, and protein concentration quantified using the EZQ® Protein Quantitation Kit (Invitrogen, Carlsbad, CA) following the standard kit protocol. Protein fluorescence was read using a laser-based scanner (Typhoon Trio+; Amersham Biosciences/GE Healthcare) equipped with a 488 nm laser and an emission filter (610 nm bandpass filter) and analyzed using ImageQuant™ TL image analysis software. Sample concentrations were then calculated by interpolation from a standard curve created from control samples of known protein concentration (0.02–5 mg/mL) and expressed in µg of protein.

1-Dimensional SDS- Polyacrylamide Gel Electrophoresis

We ran SDS-PAGE gels of triplicate samples (50 µg each technical replicate) on a 1 mm 10% NuPAGE® Novex® Bis-Tris Mini Gel using the XCell SureLock Mini-Cell system (Invitrogen) as per manufacturer's instructions for reduced samples. The gel was then stained using SimplyBlue™ SafeStain (Invitrogen), and destained following the manufacturer's instructions. The three lanes were then separated from the gel and cut into 16 sections of equal size. Next, half of each section was transferred to a 96-well plate, resulting in a total of 48 gel sections and stored at -80 °C until later analysis.

In-gel protein digestion and mass spectrometry

166 Each gel section was cut into smaller (~1 x 1 x 1 mm) pieces, transferred to a 0.6 ml
167 microcentrifuge tube (one tube per gel section), and subject to a standard in-gel proteolytic
168 digestion following Shevchenko *et al.*, [56]. Briefly, gel sections were first further destained using
169 50 mM ammonium bicarbonate / 50% acetonitrile and then dehydrated with 100% acetonitrile.
170 Proteins were then reduced and alkylated with dithiothreitol and iodoacetamide, and subsequently
171 digested with trypsin (15 ng/ μ l, Promega). Next, protein digests were extracted with 5% formic
172 acid/50% acetonitrile and dried down using a vacuum centrifuge. Finally, the resulting peptides
173 were eluted in 20 μ l of 0.1% formic acid.

174 Extracted peptides were analyzed by nanoflow reverse phase liquid chromatography using a
175 nanoAcquity LC system (Waters, Milford, MA) coupled in-line with the linear trap quadrupole
176 LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific). The nanoAcquity LC system
177 was equipped with a Symmetry C18 (Waters Corporation) trap column (5 μ m; 180 μ m x 20 mm)
178 and a BEH130 C18 (Water Corporation) analytical column (1.7 μ m; 100 μ m x 100 mm). Mobile
179 phases A and B consisted of 0.1% formic acid in water and 0.1% formic acid in acetonitrile,
180 respectively. Samples were loaded onto the trap column for desalting and preconcentration with
181 99%/1% mobile phase A:B at a flow rate of 5 μ l/min for 3 min. Peptides were separated on the
182 analytical column at a flow rate of 500 nl/min with a two-step linear gradient consisting of 7% B to
183 25% B in 72 min and 25% B to 45% B in 10 min. The electrospray ion source consisted of a
184 nanospray head (Thermo Fisher Scientific) coupled with a coated PicoTip fused silica spray tip with
185 OD 360 μ m, ID 20 μ m, and 10 μ m diameter emitter orifice (New Objective, Inc.). Samples were
186 analyzed using positive ion spray voltage and heated capillary temperature of 1.9 kV and 220 $^{\circ}$ C,
187 respectively.

188 Mass spectrometry data were collected with the instrument operating in data dependent
189 MS/MS mode. MS survey scans (m/z 300–2000) were acquired in the Orbitrap analyzer with a
190 resolution of 60 000 at m/z 400 and an accumulation target of 1×10^6 . This was followed by the
191 collection of MS/MS scans of the 15 most intense precursor ions with a charge state ≥ 2 and an

192 intensity threshold above 500 in the LTQ with an accumulation target of 10 000, an isolation
193 window of 2 Da, normalized collision energy of setting of 35%, and activation time of 30ms.
194 Dynamic exclusion was used with repeat counts, repeat duration, and exclusion duration of 1, 30 s,
195 and 60 s, respectively.

196

197 *Peptide and protein identification*

198 Protein mass spectra data files were analyzed using MaxQuant (Version 1.5.3.30, [57,58]), and the
199 resulting peptide spectra searched against the *T. guttata* protein database (uniprot.org). For the
200 MaxQuant analysis, carbamidomethyl (C) was specified as a fixed modification, minimum peptide
201 length was set at 7, and for both peptide and protein quantification the false discovery rate (FDR)
202 was set to 1%. We required a minimum of 1 unique peptide for protein identification, and set the
203 peptide mass tolerance to 20.0 ppm and fragment mass tolerance to 0.5 Da (see supplementary table
204 S1 for MaxQuant parameters).

205 Protein identifications from MaxQuant consisted of zebra finch UniProt entry IDs. Gene
206 names were assigned to 78% (384/494) of these identified proteins. We also obtained a list of cross-
207 referenced Ensembl Gene and Protein IDs using UniProt [59]; zebra finch Ensembl IDs were
208 assigned to 83% (408/494) of the identified proteins. Next, and because few analysis tools currently
209 support analysis of genome-scale data for the zebra finch, we retrieved chicken (*Gallus gallus*)
210 orthologs using the g:Orth module in g:profiler [60,61] and used these IDs in our analysis (unless
211 otherwise stated). Chicken Ensembl gene IDs were assigned to 94% (382/408) of proteins, and of
212 these 97% (370/382) were 1:1 orthologs with unique IDs.

213

214 *Gene ontology (GO) and enrichment analysis*

215 We first classified the components of the ZfSP into major protein classes using the PANTHER
216 resource [62]. Next, we conducted a network analysis on the genes found in the ZfSP using
217 Cytoscape (v3.3.0, [63]) and the ClueGO plugin (v2.2.5, [64]) with the *Gallus gallus* (9031) marker

218 set, and within this analysis tested for enrichment of GO terms using a right-sided hypergeometric
219 test with a Benjamini-Hochberg multiple test correction. For molecular function, GO tree levels
220 (min = 1; max = 4) and GO term restriction (min#genes = 5, min% = 3%) were set and terms were
221 grouped using a Kappa Score Threshold of 0.3. For cellular components, GO tree levels (min = 1;
222 max = 5) and GO term restriction (min#genes = 5, min% = 5%) were set and terms were grouped
223 using a Kappa Score Threshold of 0.4. Finally, for biological processes, GO tree levels (min = 1;
224 max = 4) and GO term restriction (min#genes = 5, min% = 5%) were set and terms were grouped
225 using a Kappa Score Threshold of 0.4. For all networks, resulting groups consisted of a minimum of
226 two terms and groups sharing $\geq 50\%$ of terms were merged. Finally, to investigate the 86 proteins
227 that could not be assigned a zebra finch Ensembl ID, we obtained GO functional annotations and
228 gene descriptions for these proteins using PANNZER [65] and manually explored the resultant gene
229 list.

230

231 *Evolutionary analysis: estimating rates of nonsynonymous change (dN/dS)*

232 To gain information on the impact of selection on the ZfSP, we performed pairwise comparisons of
233 one-to-one orthologous flycatcher (*Ficedula albicollis*) coding sequences. We chose to use the
234 flycatcher in these analyses as it is the closest relative of the zebra finch for which such information
235 is available and the use of longer phylogenetic branches in these analyses makes the inference of
236 substitutions less accurate. For this analysis, orthology was determined using the maximum
237 likelihood phylogenetic gene trees methods of Ensembl (version 90) [66], and dN and dS values
238 were retrieved from the BioMart interface of Ensembl (<http://ensembl.org/biomart/martview>) [67].
239 From these values, we calculated the non-synonymous/synonymous substitutions ratio (dN/dS) of
240 ZfSP genes by dividing the non-synonymous substitution rate (dN) by the synonymous substitution
241 rate (dS). The dN/dS ratio is an indicator of the selection pressure on a gene, with dN/dS equal to 1
242 indicating neutral evolution, dN/dS < 1 purifying selection (*i.e.* changes that are actively selected
243 against), and dN/dS > 1 diversifying positive selection (*i.e.* changes that are favoured by selection).

244 We then created two datasets for subsequent analysis. The first dataset (hereafter ‘high-confidence’)
245 included only those ortholog pairs tagged as high confidence orthologies following the criteria
246 established in Ensembl. The second dataset (hereafter ‘low confidence’) included all ortholog pairs
247 tagged as low confidence orthologies based on the same standard Ensembl criteria; we considered
248 high and low confidence orthologies separately in order to avoid the potential influence of
249 inaccurate orthologies on our analysis.

250 Using our ‘high-confidence’ dataset, we searched for evidence of elevated dN/dS values in
251 the ZfSP. We then investigated genes with clear molecular function identified by the GO analysis
252 and combined genes from multiple functional groups to form the following six general functional
253 categories: DNA/RNA binding, catalysis and proteolysis, oxidoreductases, protein binding,
254 cytoskeleton and axenome motility, and uncharacterised (genes were permitted to occur in multiple
255 functional categories as suggested by the results of the GO analysis). We then compared the mean
256 dN/dS in each of these six functional categories to the mean dN/dS for the entire ZfSP using one-
257 sample Wilcoxon tests (to account for non-normal distribution of data) and compared the
258 distribution of dN/dS values in each functional category to the distribution of values in the entire
259 ZfSP using two-sample Kolmogorov-Smirnov tests. Given observations that proteins within sperm
260 are generally either high conserved or rapidly evolving [11,12], we also tested for bimodality in the
261 distribution of dN/dS values in each of the functional categories and the entire ZfSP using the
262 bimodality coefficient in the R package *modes* [68]. We also obtained dN and dS values, and
263 calculated the dN/dS ratio, for all genes in the zebra finch genome (as before dN and dS values for
264 the zebra finch genome were obtained by downloading data on zebra finch-flycatcher one-to-one,
265 high-confidence orthologs from Ensembl (version 90) [67]). We then tested whether the mean
266 dN/dS value of genes in the ZfSP differed from the genome wide mean dN/dS value by comparing
267 the distribution of dN/dS values between ZfSP genes and the overall genomic background using
268 permutation tests using the R package *coin* [69] in the R statistical package v. 3.5.0 [70]. Next, we
269 also examined our low-confidence dataset for evidence of elevated dN/dS values in the ZfSP.

270 Finally, for comparison, we also calculated mean dN, dS, and dN/dS values for the pairwise
271 comparison with chicken, using identical methods as described above; though in this instance we
272 created a single dataset based on high-confidence orthologies (as defined by Ensembl).

273

274 *Comparison with the chicken sperm proteome*

275 To compare the ZfSP to the published chicken sperm proteome, we obtained the chicken sperm
276 proteome IDs from the published literature [27], and converted all unique UniProt Entry IDs to their
277 corresponding Ensembl IDs using bioDBnet [71]. This resulted in a total of 294 chicken Ensembl
278 IDs for comparison to the ZfSP chicken orthologs. We conducted a functional network comparison
279 between proteins identified in the zebra finch to those identified in the chicken using ClueGO [64],
280 with the following parameters: Molecular Function GO levels 1-4, initial group size = 2, groups
281 sharing $\geq 50\%$ of terms, Kappa Score Threshold = 0.3. For the zebra finch cluster, we set term
282 restrictions at min#genes = 5 and min% = 3%, while for the chicken cluster we lowered the
283 selection criteria slightly to account for the slightly lower number of identifiers in the chicken
284 sperm proteome and set the restrictions at min#genes = 3 and min% = 2%. We repeated this
285 analysis using a different published semen proteome [28] and our results were qualitatively similar.

286

287 **Results**

288 *Sperm protein identification*

289 Mass spectrometry identified a total of 4599 peptides and 1824 unique peptides across the three
290 technical replicates, leading to the identification of 494 proteins in purified zebra finch sperm
291 samples (supplementary table S2 and S3), which represents the minimum number of proteins in
292 sperm from this species. Sequence coverage of identified proteins ranged from 0.4% to 74.8%, with
293 an average sequence coverage of 16.8%. Although the cutoff for the minimum number of unique
294 peptides identified per protein was set to 1, the average number of unique peptide hits per protein

295 was nearly four times this value (mean: 3.7, range: 1-84), suggesting our MS data were of generally
296 good quality.

297

298 *Gene ontology (GO) and functional analysis*

299 The major protein classes characterizing the ZfSP included categories essential to sperm structure
300 and function (Figure 2); proteins linked to either metabolism/energetics (*e.g.* oxidoreductases,
301 hydrolases, transferases) or tubulin binding and microtubule related functions (*e.g.* cytoskeletal)
302 comprised 42.8% of the total gene hits against protein classes. The protein class oxidoreductase was
303 a particularly dominant feature of the ZfSP, forming the second largest category. Several additional
304 categories were also well represented in the ZfSP (*e.g.* transporters, nucleic acid binding,
305 transfer/carrier proteins, enzyme modulators), along with a number of more minor categories (*e.g.*
306 lyase, defense/immunity proteins, calcium binding protein, chaperones; Figure 2). As with sperm
307 proteomes in several other taxa, however, the largest category was proteins of unclassified function
308 (Figure 2).

309 Gene ontology annotations were assigned to 94.86% of gene products. ClueGO analysis
310 using GO molecular function resulted in a 94-node network distributed into 15 broad functional
311 groups. The dominant functional categories were those associated with RNA binding,
312 oxidoreductase activity, hydrolase activity (acting on acid anhydrides) and purine ribonucleotide
313 binding (Figure 3a, supplementary figure S1a). In addition, we identified numerous molecular
314 function enrichments (Table 1; see supplementary table S4 for details of GO analysis for all GO
315 terms), including enrichment in genes involved in catalytic activity (*e.g.* electron carrier activity, 10
316 genes, $p = 1.61\text{E-}5$; oxidoreductase activity, 56 genes, $p = 2.07\text{E-}15$; NADH dehydrogenase
317 activity, 12 genes, $p = 4.27\text{E-}10$; catalytic activity, 157 genes, $p = 5.68\text{E-}05$), cytoskeleton related
318 processes (*e.g.* structural molecular activity, 35 genes, $p = 2.89\text{E-}7$; structural constituent of the
319 ribosome, 20 genes, $p = 3.94\text{E-}7$), and binding (*e.g.* RNA binding, 61 genes, $p = 8.13\text{E-}7$; anion
320 binding, 87 genes, $p = 1.84\text{E-}5$).

Cellular component GO analysis resulted in a 121-node network placed into 18 broad functional groups. The dominant categories according to cellular component GO analysis were mitochondrion and cortical actin cytoskeleton, followed by several more minor but well represented groups (*e.g.* actin cytoskeleton, mitochondrial membrane part; Figure 3b, supplementary figure S1b). Cellular component analysis showed significant enrichment in gene products from several functional categories (Table 2; see supplementary table S5 for details of GO analysis for all GO terms), including genes associated with the mitochondrial membrane (*e.g.* mitochondrion, 126 genes, $p = 9.20\text{E-}45$; mitochondrial part, 78 genes, $p = 4.81\text{E-}41$; mitochondrial membrane, 64 genes, $p = 8.28\text{E-}38$; mitochondrial protein complex, 19 genes, $p = 1.01\text{E-}13$), the cytoskeleton (*e.g.* cortical actin cytoskeleton activity, 8 genes, $p = 2.05\text{E-}5$; actin cytoskeleton, 22 genes, $p = 1.51\text{E-}4$; actomyosin, 7 genes, $p = 7.68\text{E-}4$), and endoplasmic reticulum part (*e.g.* endoplasmic reticulum part, 34 genes, $p = 7.92\text{E-}9$; endoplasmic reticulum lumen, 10 genes, $p = 2.03\text{E-}7$). Similarly, we observed a significant enrichment in a number of GO terms associated with membrane proteins, including membrane raft (10 genes, $p = 6.07\text{E-}3$), membrane protein complex (45 genes, $p = 1.10\text{E-}7$), and membrane microdomain (10 genes, $p = 6.07\text{E-}3$).

Network analysis using GO biological processes resulted in a 139-node network allocated into 21 functional groups. The dominant categories according to biological process included generation of precursor metabolites and energy, symbiosis, and single-organisms catabolic process (Figure 3c, supplementary figure S1c), and several functional categories showed significant enrichment in gene products (Table 3; see supplementary table S6 for details of GO analysis for all GO terms), with the greatest enrichment found in proteins involved in particular cellular processes (*e.g.* generation of precursor metabolites and energy, 35 genes, $p = 1.37\text{E-}19$; cellular respiration, 24 genes, $p = 2.64\text{E-}16$; protein folding, 19 genes, $p = 3.36\text{E-}8$) and a number of metabolic processes, including oxidation-reduction process (62 genes, $p = 3.77\text{E-}16$) and small molecule metabolic process (80 genes, $p = 2.68\text{E-}13$).

346 Finally, using PANNZER, we obtained GO functional annotations and gene descriptions for
 347 85 of the 86 proteins that could not be assigned an Ensembl ID (supplementary table S7). Amongst
 348 these annotations, we observed a large number of genes linked to oxidative phosphorylation,
 349 including several subunits of the NADH dehydrogenase complex I (*e.g.* NDUFA11, NDUFA2,
 350 NDUFA8, NDUFB3, NDUF2C, NDUF5S), as well as several ribosomal proteins (*e.g.* RPL15,
 351 RPL26, RPL7a), suggesting that some of these proteins may not be well annotated within the zebra
 352 genome. In addition, we noted several proteins with likely roles in innate immunity or antioxidant
 353 defense, including β 2-microglobulin (B2M), Glutathione peroxidase 1 (GPX1), Superoxide
 354 dismutase-1 (SOD1), and Glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Other notable
 355 proteins included Histone H3.3, which has been shown to be essential for chromatin assembly in the
 356 male pronucleus [72], and Brain-type creatine kinase (CKB), which is known to effect sperm cell
 357 energy metabolism and has been reported to be more abundant in the sperm cells of chickens
 358 characterized by high sperm mobility than those with low sperm mobility [73].
 359

360 *Evolutionary analysis*

361 For sperm proteome genes, we obtained dN/dS values for 231 zebra finch-flycatcher high-
 362 confidence orthologs (supplementary table S8). Pairwise divergence estimates between zebra finch
 363 and flycatcher orthologs resulted in a mean dN of 0.033 (s.d. = 0.037), a mean dS of 0.204 (s.d. =
 364 0.266), and an overall mean dN/dS ratio of the ZfSP was 0.179. For the whole genome pairwise
 365 comparison with flycatcher, we obtained dN/dS values for 8923 orthologs (supplementary table S9)
 366 and found the genome wide nonsynonymous mean of dN= 0.039 and synonymous mean of dS=
 367 0.190, while the genome wide mean dN/dS ratio was 0.236. Thus, mean dN/dS for the ZfSP was
 368 significantly lower than that observed for the overall genomic background of the zebra finch (0.179
 369 vs. 0.236, one-tailed permutation test, $Z = 0.53$, $p = 0.007$). Similarly, the median dN/dS of the
 370 ZfSP was lower than the dN/dS values obtained for the zebra finch genomic background (0.131 vs.
 371 0.156, supplementary figure S2). For zebra finch-chicken orthologs (high-confidence orthologs

only), both the nonsynonymous average ($dN = 0.059$) and synonymous average ($dS = 0.697$) levels of divergence were slightly higher, and the average dN/dS ratio for the ZfSP was 0.113 (supplementary table S9).

Among the 231 high-confidence orthologs identified, none of the ZfSP genes exhibited dN/dS values > 1 (supplemental table S8). However, 13 genes had a dN/dS value greater than 0.5 (Table 4), and these same 13 genes were identified as having elevated dN/dS values by an outlier analysis. Previous studies have demonstrated that genes with dN/dS values > 0.5 are likely to have undergone adaptive evolution and are thus candidates for more detailed analysis of selection [74,75]. Among the genes exhibiting elevated dN/dS values in the ZfSP, a number have been reported in the sperm proteome of a range of species (*e.g.* mouse: USMG5, UQCRC2, [76]; macaque: FABP4, FUNDC2, UQCRC2, [25]) and have been linked to reproduction in a range of species. For example, UQCRC2 has been linked to fertility in bulls [77], while thioredoxin (TXN) and thioredoxin-related genes have been linked to oxidative stress, DNA damage, and impaired sperm motility [78]. Moreover, in our dataset, many of the genes with $dN/dS > 0.5$ are genes that code for proteins involved in energetics and metabolism or are involved in protein-protein interactions linked to catalysis, energetics, and oxidoreductase activity (Figure 4).

Among the ZfSP functional categories, and using only high-confidence ortholog pairs, the highest levels of evolutionary constraint (*i.e.* the lowest dN/dS values) were genes involved in catalysis and proteolysis (mean $dN/dS = 0.129$). Average evolutionary rates for each of the functional categories, with the exception of genes of unknown function, were significantly lower than the average evolutionary rate observed across the entire sperm proteome (all $p < 0.05$; Figure 5). For genes of unknown function, the average dN/dS value tended to be higher than the mean dN/dS value for the entire sperm proteome, but not significantly so (Wilcoxon one-sample t-test: $V = 889$, $p = 0.32$; Figure 5). Amongst the six functional categories, we observed the highest dN/dS values (> 0.8) in the categories DNA/RNA binding and oxidoreductases. The distribution of dN/dS values in the functional categories DNA/RNA binding, catalysis, cytoskeleton, protein binding, and

398 oxidoreductase did not differ significantly from the distribution of values in the sperm proteome as
399 a whole (all $D > 0.11$, all $p > 0.09$). In contrast, the distribution of dN/dS values in the category
400 genes of unknown function differed significantly from the dN/dS distribution of the entire ZfSP (D
401 $= 0.21$, $p = 0.04$). Finally, we observed significant bimodality in the distribution of dN/dS values in
402 each of the six functional categories (all bimodality coefficients > 0.605 , range $0.605 - 0.673$), as
403 well as significant bimodality in the distribution of dN/dS values for the entire sperm proteome
404 (bimodality coefficient for the ZfSP: 0.606).

405 We identified 57 low-confidence ortholog pairs (supplementary table S8). Amongst these, a
406 single gene showed a dN/dS value > 1 . This gene, SPATA18, has been reported in the sperm
407 proteome of mice [76], macaques [25], humans [22], and chickens [28], is a key regulator of
408 mitochondrial quality, and down-regulation of SPATA18 is linked to low sperm motility in humans
409 [79]. Amongst the low-confidence orthologs, outlier analysis identified two genes with elevated
410 dN/dS values, SPATA18 and Phosphoinositide phospholipase C (PLCZ1), while a further three
411 genes exhibited dN/dS values exceeding 0.5, including LRPPRC, IGLL1, and an uncharacterised
412 protein. Most notably, PLCZ1 encodes a protein that appears essential for spermatogenesis and
413 oocyte activation and is associated with Spermatogenic failure 17, which is an autosomal recessive
414 infertility disorder resulting in male infertility due to oocyte activation failure.

415

416 *Comparison to chicken sperm proteome*

417 Overlap between the zebra finch and chicken sperm proteome was small. Specifically, out of the
418 370 zebra finch gene products that were identified as chicken orthologs, only 61 were shared
419 between the two species, while 309 were unique to the ZfSP and a further 225 were unique to the
420 chicken sperm proteome (Figure 6a). To explore the relationship between proteins identified in the
421 zebra finch to proteins identified in the chicken, we conducted a functional network comparison
422 based on GO molecular function terms. Gene ontology annotations were assigned to 94.86% of
423 gene products in the zebra finch and 90.21% of gene products in the chicken. The resulting network

analysis resulted in a 125-node network organized into 17 functional groups (Figure 6b). Seventeen nodes (14%) showed no compositional bias towards proteins identified in either the zebra finch or chicken sperm proteome (Figure 6b). These unspecified nodes fell primarily into the functional groups ubiquitin-protein transferase activity and electron carrier activity, with further unspecified nodes in the groups small molecule binding, ion channel binding, hydrolase activity, and oxidoreductase activity. Eighty nodes (64%) were enriched for proteins identified in the zebra finch. Enrichment towards zebra finch was particularly strong for proteins known to be highly abundant in sperm, including proteins involved in catalytic activity (*e.g.* oxidoreductases) and cytoskeleton related processes (*e.g.* structural molecular activity, actin binding), which includes members of the actin and tubulin family of proteins (Figure 6b). In contrast, only 28 nodes (22%) were enriched for proteins identified in the chicken, with the strongest enrichment observed in the functional group ubiquitin-protein transferase activity and to a lesser extent ion channel binding. Two functional groups appeared unique to the chicken sperm proteome: oxygen binding and hydro-lase activity (Figure 6b).

438

439 **Discussion**

Sperm are the most diverse cell type known [1] and essential to the process of fertilisation, arguably one of the least understood of fundamental biological processes [80]. To date, the study of the molecular biology of sperm has largely focused on understanding the basis of fertility/infertility, particularly in humans [21,81]. More recently, sperm cell proteomics has contributed significantly to our understanding of the molecular basis of sperm form and function in a range of vertebrate and invertebrate taxa [11,29,32,82]. Moreover, recent research has highlighted the role of reproductive proteins, such as those that constitute the sperm proteome, in ejaculate-female and sperm-egg interactions, as well as the potential implications these molecular interactions have for post-copulatory processes such as cryptic female choice and conspecific sperm precedence [7,41,83].

Birds have long been at the centre of research on sexual selection [84] and speciation [85].

450 Considerable attention has also been paid to post-copulatory sexual selection in birds [86], and
451 numerous studies have investigated the role of sperm competition in driving evolutionary
452 diversification in sperm morphology and function in both galliformes [87] and passerines [17,88].
453 In contrast, studies of avian reproductive proteins are generally limited (see [46,47] for exceptions),
454 and sperm proteome characterisation is currently limited to domestic poultry [27,28]. This is despite
455 the suggestion that the investigation of avian reproductive proteins is a promising avenue of
456 research likely to reveal significant insight into the rapid evolution of reproductive proteins and
457 speciation [45].

458 We identified 494 proteins in the ZfSP. While the size of the ZfSP is smaller than other
459 recently published sperm proteomes, which report between *c.* 800-1350 proteins [23-29,43,82,89],
460 it is consistent with findings in the marine mussel *Mytilus galloprovincialis* [34] and higher than the
461 number of proteins identified in the sperm proteome of the spoon worm *Urechis unicinctus* [35].
462 However, because this study represents, to the best of our knowledge, the first description of a
463 passerine sperm proteome, it remains to be determined how representative this proteome will be
464 when compared to sperm proteomes from other passerine species. Depth of proteome coverage can
465 be influenced by sampling bias errors due to a number of factors, such as a difficulty in identifying
466 proteins of low abundance or very basic/acidic proteins. Indeed, the first attempt to characterise the
467 *D. melanogaster* sperm proteome (DmSPI) identified only 342 unique proteins [11], a number that
468 was significantly expanded to 1108 proteins in a follow up study (DmSPII) [29]. Importantly, that
469 study demonstrated that the original DmSPI was a reliable representative of the *D. melanogaster*
470 sperm proteome and suggested that such small or incomplete proteomes could still provide useful
471 and broad based perspectives of cellular proteomes [29]. Given the rapid advances in proteomic
472 technologies, the use of multiple search engines to expand coverage [90], and enhanced
473 computational tools, we expect the ZfSP reported here will increase in size in the future.

474 Gene ontology analysis revealed that the ZfSP is statistically enriched for genes involved in
475 a number of key functions. In particular, proteins linked to central metabolism and energetics (*e.g.*

476 oxidoreductases, hydrolases, transferases) and structural proteins (*e.g.* tubulin binding, cytoskeleton
477 related processes) were a highly enriched, dominant feature of the ZfSP. These findings are highly
478 consistent with those reported in a range of mammalian and insect sperm proteomes [11,23-
479 25,29,32,82], as well as the recent chicken sperm proteome [28]. Taken together, these findings
480 highlight the importance of these conserved proteins to sperm function, particularly the high energy
481 and motility needs of sperm.

482 A recurrent observation in evolutionary biology research is that traits pertaining to
483 reproduction have evolved more rapidly compared to other types of characters. For example, rapid
484 evolution has been reported for genital morphology in insects [91] and colouration in birds [92] and
485 fish [93]. Furthermore, at the molecular level, within- and between-species comparisons of gene
486 and protein sequences has shown that proteins linked to sexual reproduction (the so-called
487 ‘reproductive proteins’) are more divergent than those expressed in non-reproductive tissues, and in
488 many cases this rapid divergence appears to be driven by adaptive evolution [41,94]. Overall, the
489 ZfSP showed no evidence of positive selection (as indicated by $dN/dS > 1$) in pairwise comparisons
490 between zebra finch and flycatcher (high-confidence orthologs only). Thus, the ZfSP appears to be
491 evolving relatively conservatively with a high degree of selective constraint. Although at first this
492 may seem counterintuitive, our finding is consistent with results for the *Drosophila* sperm proteome
493 [11,95]. Our finding is also in line with results for the mouse sperm proteome, for which a single
494 gene was reported to have a dN/dS value greater than one [43]. Moreover, the high level of
495 selective constraint on sperm proteome genes likely reflects the role of numerous sperm proteins in
496 essential functions, such as primary metabolism and energetics, which are expected to be universal
497 to sperm cells from across the animal kingdom.

498 Although the ZfSP generally appears to be evolutionarily constrained (as indicated by low
499 dN/dS values), and thus potentially under purifying selection, variation in evolutionary rates did
500 occur among subsets of individual functional categories, as has been observed in a range of
501 vertebrate and invertebrate sperm proteomes [11,34,43]. Considering genes with dN/dS values

greater than 0.5 (suggestive of site-specific positive selection; see [74,75,96,97]), we identified a set of proteins involved in energetics and metabolism or proteins involved in protein-protein interactions linked to catalysis, energetics, and oxidoreductase activity. Similar results have been observed for genes involved in metabolic and energetic functions in both the *Drosophila* and mouse sperm proteomes [11,43] and more generally in the sperm proteomes of a range of mammal species [12]. In mouse, sperm membrane proteins also appear to evolve more rapidly relative to other sperm genes [43]. Although we do not explicitly classify sperm membrane proteins for our evolutionary analysis, we did find the ZfSP was significantly enriched in a number of membrane proteins (*e.g.* PRKAR1A, VDAC2, HSPD1) and further identified a number of surface membrane proteins, including nine solute carriers (SLC25A11, SLC2A3, SLC25A22, SLC25A1, SLC25A3, SLC25A12, SLC25A4, SC25A10, and SLC25A13). Nonetheless, it is likely that membrane proteins may constitute a relatively minor contribution to the ZfSP, as such proteins are generally underrepresented in shotgun proteomic studies [98], see also [29]. More generally, if the sperm proteome we identified here represents a subset of the ‘true’ ZfSP, then the overall low evolutionary rate we observed may be explained by sampling bias. A further consideration is that dN/dS calculations have relatively low statistical power when only pairwise comparisons are made [74]. Regardless, the consistency of our findings with those reported in *Drosophila* and mammals [11,12,43], and our observation of significant bimodality in the distribution of dN/dS values in the ZfSP, suggests that genes in the ZfSP are either highly conserved or evolving rapidly, thus supporting the idea of compartmentalized adaptation in the ZfSP.

In the zebra finch, genetic variation in sperm length has been linked to a Z chromosome inversion polymorphism [54,55]. At this time, however, knowledge of the specific genes underlying particular sperm phenotypes remain limited [99]. Nonetheless, Kim *et al.* [55] identify a number of possible candidate genes within these inversions. While none of these candidate genes were identified in our sperm proteome, it is likely that many of them are responsible for *trans*-acting regulation of autosomal genes [55] and thus their absence from the ZfSP is unsurprising. In

contrast, we did identify three of the 108 genes that were found to be differentially expressed in the testes of males with long versus short sperm: CCT5, FASTKD1, and IPO5. Most notably, CCT5 is a molecular chaperonin encoding the Chaperonin containing TCP1 subunit 5 protein, appears to play a role in the folding of actin and tubulin, and has previous been identified in the sperm proteome of the chicken [28], mouse [71], and macaque [25]. Finally, we also identified the protein cAMP-dependent protein kinase type I-alpha regulatory subunit (Prkar1a), which has been shown to be associated with sperm midpiece length in deer mice (*Peromyscus*, [100]).

We also identified a number of genes more generally linked to sperm form and function. These included proteins that have an important role sperm development, such as the RAB proteins, which are involved in acrosome formation [101] and appear essential for sperm tail function and male fertility [102]. Similarly, we identified a number of proteins linked to sperm motility and sperm-egg interactions (*e.g.* CANX, ATP2B4, TEK3, DNAH17). Notably, we identified several heat shock proteins (HSPs), including members of the heat shock protein families – HSP70 (HSPA9, HSPA5, HSPA8, HYOU1), HSP60 (HSPD1), HSP10 (HSPD1), and HSP90 (HSP90AA1, HSP90B1). In humans, three of these proteins (HYOU1, HSPA5, HSPD1) have been identified as sperm surface proteins, and, most notably, HSPA5 has a putative role in mediating sperm-egg interactions [103]. Furthermore, in Japanese quail, heat shock protein 70 (HSP70) has been shown to activate sperm flagella movement through an interaction with a sperm surface protein, voltage-dependent anion channel protein 2 (VDAC2) which we also identify here, and this interaction appears to play an important role in the release of sperm from the sperm storage tubules of females [104].

We also identified a member of the A Disintegrin and Metalloproteinase (ADAM) family of integral membrane proteins, ADAM2 (also known as fertilin β), which has been reported in the sperm proteomes of rat [24] and mouse [43]. In mammals, ADAM2 appears to plays a role in sperm zona pellucida binding [105] and possibly also egg membrane recognition and fusion [106], and shows evidence of rapid evolution [43,107,108]. In zebra finch, ADAM2 is involved in strong

554 protein-protein interactions with several genes, including three Zona Pellucida (ZP) glycoproteins
555 (ZP1, ZP2, and ZP3, see supplementary figure S3). In birds, the egg membrane (*i.e.* the perivitelline
556 layer) is composed of several ZP glycoproteins (*e.g.* ZP3, ZP1, and ZPD), all of which appear to be
557 involved in sperm-zona interactions, including sperm activation and entry into the oocyte [109,110].
558 ZP3 has also been shown to be diverging by positive adaptive evolution [46]. Taken together,
559 further investigations of ADAM proteins, as well as other sperm proteins identified in zebra finch
560 known to mediate sperm-egg interactions (*e.g.* acrosin (ACR)), are likely to provide insight into
561 fertilisation and ejaculate-female interactions, such as conspecific sperm precedence.

562 Zebra finch sperm also contains albumin (ALB), which has been observed in mammalian
563 sperm proteomes [23-25,43] and is a major constituent of the seminal fluid proteome of the Red
564 Jungle Fowl [47] and domestic chicken [111,112]. In mammals, serum albumin appears to play a
565 critical role in sperm capacitation [113]. Serum albumin is also classified as an immune regulatory
566 or defense protein [111], and in domestic turkeys addition of serum albumin under *in vitro*
567 conditions appears to increase the proportion of motile sperm and sperm velocity [112,114].
568 Finally, we identified several additional genes in the ZfSP linked to antioxidant function (*e.g.*
569 thioredoxin [TXN, TXN2], superoxide dismutase [SOD]), innate immunity (*e.g.* complement C1q
570 binding [C1QBP], macrophage migration inhibitory factor (MIF) variant), antimicrobial function
571 (*e.g.* ovotransferrin [LTF], Histone 2A[H2AFJ], Pyruvate kinase [PKM], Glyceraldehyde-3-
572 phosphate dehydrogenase [GAPDH]), and adaptive immune function (*e.g.* β 2-microglobulin),
573 which is again consistent with results of several sperm and ejaculate proteome studies [42,47].

574 Finally, we observed significant differences in the sperm proteome of the zebra finch and
575 the chicken. The reasons underlying these differences are unclear, though there are notable
576 differences in sperm morphology between the two species that may contribute to some of proteomic
577 differences observed here (*e.g.* enrichment of structural proteins). For example, zebra finch sperm is
578 shorter than chicken sperm (*c.* 70 vs. 90 μ m, zebra finch and chicken respectively), and in zebra
579 finch, like other passerines, sperm are spiral- or corkscrew-shaped with a characteristic helical

580 membrane that wraps around the acrosome, while in chickens (and other galliformes) sperm are
581 sauropsid-shaped and lack a helical membrane [115]. We note, however, that a number of other
582 factors may contribute to the proteomic differences between these taxa, including sperm
583 competition [82], changes due to domestication of the chicken [47], or even simply differences
584 between studies in sample preparation.

585

586 **Conclusions**

587 In conclusion, this study has generated an inventory of 494 gene products that are present in zebra
588 finch sperm. As is the case in a range of mammalian and insect sperm proteomes, zebra finch sperm
589 includes a number of proteins essential to sperm structure and function, such as proteins linked to
590 metabolism and energetics. Similarly, a number of proteins linked to critical sperm functions, such
591 as sperm motility and sperm-egg interactions, as well as a diverse array of immunity proteins are
592 present in zebra finch sperm. Although this list likely represents a partial list of proteins in zebra
593 finch sperm, it provides considerable insight into avian sperm biology and function and sets the
594 stage for future studies examining the molecular basis of sperm form and function in passerine
595 birds.

596

597

598 **Funding**

599 This work was supported by the Research Council of Norway (grant number 230434 to MR); and a
600 Kristine Bonnevie Scholarship from the University of Oslo (to MR).

601

602 **Acknowledgements**

603 We wish to thank Steve Dorus for his excellent advice, input and expertise during the course of this
604 project.

605

607 **References**

- 608 [1] Pitnick S, Hosken DJ, Birkhead TR. Sperm Morphological Diversity. In: Birkhead TR,
609 Hosken DJ, Pitnick S (Eds.), Sperm Biology: an Evolutionary Perspective. Academic
610 Press, Oxford, 2009: pp. 247–394.
- 611 [2] Snook RR. Sperm in competition: not playing by the numbers. Trends in Ecology and
612 Evolution. 2005; 20:46–53.
- 613 [3] Pizzari T, Parker GA. Sperm competition and sperm phenotype. In: Birkhead TR, Hosken
614 DJ, Pitnick S (Eds.), Sperm Biology: an Evolutionary Perspective. Academic Press,
615 Oxford, 2009: pp. 205–244.
- 616 [4] Fitzpatrick JL, Lüpold S. Sexual selection and the evolution of sperm quality. MHR:
617 Basic science of reproductive medicine. 2014; 20: 1180–1189.
- 618 [5] Simmons LW, Fitzpatrick JL. Sperm wars and the evolution of male fertility.
619 Reproduction. 2012; 144: 519–534.
- 620 [6] Vacquier VD, Swanson WJ, Lee Y-H. Positive darwinian selection on two homologous
621 fertilization proteins: what is the selective pressure driving their divergence? J Mol Evol.
622 1997; 44: S15–S22.
- 623 [7] Vacquier VD, Swanson WJ. Selection in the rapid evolution of gamete recognition
624 proteins in marine invertebrates. Cold Spring Harbor Perspectives in Biology. 2011; 3:
625 a002931
- 626 [8] Loppin B, Lepetit D, Dorus S, Couble P, Karr TL. Origin and Neofunctionalization of a
627 Drosophila Paternal Effect Gene Essential for Zygote Viability. Current Biology. 2005;
628 15: 87–93.
- 629 [9] Crean AJ, Bonduriansky R. What is a paternal effect? Trends in Ecology and Evolution.
630 2014; 29: 554–559.
- 631 [10] Fitch KR, Yasuda GK, Owens KN, Wakimoto BT. Paternal effects in Drosophila:
632 implications for mechanisms of early development. Curr Top Dev Biol. 1998; 38: 1–34.

- 633 [11] Dorus S, Busby SA, Gerike U, Shabanowitz J, Hunt DF, Karr TL. Genomic and
634 functional evolution of the *Drosophila melanogaster* sperm proteome. Nature Genetics.
635 2006; 38: 1440–1445.
- 636 [12] Bayram HL, Claydon AJ, Brownridge PJ, Hurst JL, Mileham A, Stockley P. et al., Cross-
637 species proteomics in analysis of mammalian sperm proteins, Journal of Proteomics.
638 2016; 135: 38–50.
- 639 [13] Birkhead TR, Møller AP. Sperm Competition and Sexual Selection, Academic Press, San
640 Diego; 1998.
- 641 [14] Griffith SC, Owens IPF, Thuman KA. Extra pair paternity in birds: a review of
642 interspecific variation and adaptive function. Molecular Ecology. 2002; 11: 2195–2212.
- 643 [15] Parker GA. Sperm competition and its evolutionary consequences in the insects.
644 Biological Reviews. 1970; 45: 525–567.
- 645 [16] Thornhill R. Cryptic female choice and its implications in the scorpionfly *Harpobittacus*
646 *nigriceps*. The American Naturalist. 1983; 122: 765–788.
- 647 [17] Rowe M, Albrecht T, Cramer ERA, Johnsen A, Laskemoen T, Weir JT, et al.,
648 Postcopulatory sexual selection is associated with accelerated evolution of sperm
649 morphology. Evolution. 2015; 69: 1044–1052..
- 650 [18] Seddon N, Botero CA, Tobias JA, Dunn PO, MacGregor HEA, Rubenstein DR, et al.,
651 Sexual selection accelerates signal evolution during speciation in birds. Proc. Biol. Sci.
652 2013; 280: 20131065.
- 653 [19] Fitzpatrick JL, Almbro M, Gonzalez-Voyer A, Kolm N, Simmons LW. Male contest
654 competition and the coevolution of weaponry and testes in pinnipeds. Evolution. 2012;
655 66: 3595–3604.
- 656 [20] Wilkins MR, Sanchez J-C, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF,
657 et al., Progress with Proteome Projects: Why all Proteins Expressed by a Genome Should

658 be Identified and How To Do It. Biotechnology and Genetic Engineering Reviews. 1996;
659 13: 19–50.

660 [21] Oliva R, de Mateo S, Estanyol JM. Sperm cell proteomics. Proteomics. 2009; 9: 1004–
661 1017.

662 [22] Amaral A, Castillo J, Ramalho-Santos J, Oliva R. The combined human sperm proteome:
663 cellular pathways and implications for basic and clinical science. Human Reproduction
664 Update. 2013; 20: 40–62.

665 [23] Baker MA, Hetherington L, Reeves GM, Aitken RJ. The mouse sperm proteome
666 characterized via IPG strip prefractionation and LC-MS/MS identification. Proteomics.
667 2008; 8: 1720–1730.

668 [24] Baker MA, Hetherington L, Reeves G, Müller J, Aitken RJ. The rat sperm proteome
669 characterized via IPG strip prefractionation and LC-MS/MS identification. Proteomics.
670 2008; 8: 2312–2321.

671 [25] Skerget S, Rosenow M, Polpitiya A, Petritis K, Dorus S, Karr TL. The Rhesus Macaque
672 (*Macaca mulatta*) Sperm Proteome. Molecular and Cellular Proteomics. 2013; 12: 3052–
673 3067.

674 [26] Swegen A, Curry BJ, Gibb Z, Lambourne SR, Smith ND, Aitken RJ. Investigation of the
675 stallion sperm proteome by mass spectrometry. Reproduction. 2015; 149: 235–244.

676 [27] Soler L, Labas V, Thélie A, Grasseau I, Teixeira-Gomes A-P, Blesbois E. Intact cell
677 MALDI-TOF MS on sperm: A new molecular test for male fertility diagnosis. Animal
678 Reproduction Science. 2016; 169: 115–51.

679 [28] Labas V, Grasseau I, Cahier K, Gargaros A, Harichaux G, Teixeira-Gomes A-P., et al.,
680 Qualitative and quantitative peptidomic and proteomic approaches to phenotyping
681 chicken semen. Journal of Proteomics. 2015; 112: 313–335.

- 682 [29] Wasbrough ER, Dorus S, Hester S, Howard-Murkin J, Lilley K, Wilkin E, et al., The
683 *Drosophila melanogaster* sperm proteome-II (DmSP-II). Journal of Proteomics. 2010; 73:
684 2171–2185.
- 685 [30] Ma X, Zhu Y, Li C, Xue P, Zhao Y, Chen S, et al., Characterisation of *Caenorhabditis*
686 *elegans* sperm transcriptome and proteome. BMC Genomics. 2014; 15: 168.
- 687 [31] Karr TL, Walters JR, Panning for sperm gold: Isolation and purification of apyrene and
688 eupyrene sperm from lepidopterans. Insect Biochemistry and Molecular Biology. 2015;
689 63: 152–158.
- 690 [32] Whittington E, Zhao Q, Borziak K, Walters JR, Dorus S. Characterisation of the
691 *Manduca sexta* sperm proteome: Genetic novelty underlying sperm composition in
692 Lepidoptera. Insect Biochemistry and Molecular Biology. 2015; 62: 183–193.
- 693 [33] Whittington E, Forsythe D, Borziak K, Karr TL, Walters JR, Dorus S. Contrasting
694 patterns of evolutionary constraint and novelty revealed by comparative sperm proteomic
695 analysis in Lepidoptera. 2017; 18: 931.
- 696 [34] Zhang Y, Mu H, Lau SCK, Zhang Z, Qiu J-W. Sperm proteome of *Mytilus*
697 *galloprovincialis*: Insights into the evolution of fertilization proteins in marine mussels,
698 Proteomics. 2015; 15: 4175–4179.
- 699 [35] Qin Z, Zhang Y, Mu H, Zhang Z, Qiu J-W. The sperm proteome of the echiuran *Urechis*
700 *unicinctus* (Annelida, Echiura). Proteomics. 2018; Jul 23:e1800107.
- 701 [36] Dorus S, Karr TL, Sperm proteomics and genomics. In: Birkhead TR, Hosken DJ, Pitnick
702 S (Eds.), Sperm Biology: an Evolutionary Perspective. Academic Press, Oxford, 2009:
703 pp. 435-469.
- 704 [37] Swanson WJ, Vacquier VD. The abalone egg vitelline envelope receptor for sperm lysin
705 is a giant multivalent molecule. Proceedings of the National Academy of Sciences. 1997;
706 94: 6724–6729.

707 [38] Kresge N, Vacquier VD, Stout CD. Abalone lysin: the dissolving and evolving sperm
708 protein. *BioEssays*. 2001; 23: 95–103.

709 [39] Metz EC, Robles-Sikisaka R, Vacquier VD. Nonsynonymous substitution in abalone
710 sperm fertilization genes exceeds substitution in introns and mitochondrial DNA.
711 *Proceedings of the National Academy of Science*. 1998; 95: 10676–10681.

712 [40] Swanson WJ, Aquadro CF, Vacquier VD. Polymorphism in Abalone Fertilization
713 Proteins Is Consistent with the Neutral Evolution of the Egg's Receptor for Lysin (VERL)
714 and Positive Darwinian Selection of Sperm Lysin. *Molecular Biology and Evolution*.
715 2001; 18: 376–383.

716 [41] Swanson WJ, Vacquier VD. The rapid evolution of reproductive proteins. *Nature*
717 *Reviews Genetics*. 2002; 3: 137–144.

718 [42] Dorus S, Skerget S, Karr TL. Proteomic discovery of diverse immunity molecules in
719 mammalian spermatozoa. *Syst Biol Reprod Med*. 2012; 58: 218–228.

720 [43] Dorus S, Wasbrough ER, Busby J, Wilkins EC, Karr TL. Sperm proteomics reveals
721 intensified selection on mouse sperm membrane and acrosome genes. *Molecular Biology*
722 *Evolution*. 2010; 27: 1–12.

723 [44] Wilburn DB, Swanson WJ. From molecules to mating: Rapid evolution and biochemical
724 studies of reproductive proteins. *Journal of Proteomics*. 2016; 135: 12–25.

725 [45] Edwards SV, Kingan SB, Calkins JD, Balakrishnan CN, Jennings WB, Swanson WJ. et
726 al., Speciation in birds: Genes, geography, and sexual selection. *Proceedings of the*
727 *National Academy of Sciences*. 2005; 102: 6550–6557.

728 [46] Calkins JD, El-Hinn D, Swanson WJ. Adaptive Evolution in an Avian Reproductive
729 Protein: ZP3. *J Mol Evol*. 2007; 65: 555–563.

730 [47] Borziak K, Álvarez-Fernández A, Karr TL, Pizzari T, Dorus S. The Seminal fluid
731 proteome of the polyandrous Red junglefowl offers insights into the molecular basis of

732 fertility, reproductive ageing and domestication. Nature Scientific Reports. 2016; 6:
733 35864

734 [48] Zann RA, The Zebra Finch: a synthesis of field and laboratory studies, Oxford University
735 Press, 1996.

736 [49] Birkhead TR, Pellatt EJ, Brekke P, Yeates R, Castillo-Juarez H. Genetic effects on sperm
737 design in the zebra finch. Nature. 2005; 434: 383–387.

738 [50] Mossman J, Slate J, Humphries S, Birkhead TR. Sperm morphology and velocity are
739 genetically codetermined in the zebra finch. Evolution. 2009; 63: 2730–2737.

740 [51] Bennison C, Hemmings N, Brookes L, Slate J, Birkhead TR. Sperm morphology,
741 adenosine triphosphate (ATP) concentration and swimming velocity: unexpected
742 relationships in a passerine bird. Proc. Biol. Sci. 2016; 283: 20161558

743 [52] Hemmings N, Birkhead TR. Differential sperm storage by female zebra finches
744 *Taeniopygia guttata*. Proc. R. Soc. B. 2017; 284: 20171032–6.

745 [53] Bennison C, Hemmings N, Slate J, Birkhead TR. Long sperm fertilize more eggs in a
746 bird. Proc. Biol. Sci. 2015; 282: 20141897.

747 [54] Knief U, Forstmeier W, Pei Y, Ihle M, Wang D, Martin K, et al., A sex-chromosome
748 inversion causes strong overdominance for sperm traits that affect siring success. Nat.
749 Ecol. Evol. 2017; 1: 1177-1184.

750 [55] Kim K-W, Bennison C, Hemmings N, Brookes L, Hurley LL, Griffith SC, et al., A sex-
751 linked supergene controls sperm morphology and swimming speed in a songbird. Nat.
752 Ecol. Evol. 2017; 1: 1168–1176.

753 [56] Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass
754 spectrometric characterization of proteins and proteomes. Nat Protoc. 2007; 1: 2856–
755 2860.

756 [57] Cox J, Matic I, Hilger M, Nagaraj N, Selbach M, Olsen JV, et al., A practical guide to the
757 MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat Protoc.*
758 2009; 4: 698–705.

759 [58] Cox J, Mann M, MaxQuant enables high peptide identification rates, individualized
760 p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol.*
761 2008; 26: 1367–1372.

762 [59] UniProt Consortium. UniProt: a hub for protein information, *Nucleic Acids Research.*
763 2015; 43: D204–12.

764 [60] Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler--a web-based toolset for
765 functional profiling of gene lists from large-scale experiments. *Nucleic Acids Research.*
766 2007; 35: W193–W200.

767 [61] Reimand J, Arak T, J. Vilo J. g:Profiler--a web server for functional interpretation of gene
768 lists (2011 update). *Nucleic Acids Research.* 2011; 39: W307–15.

769 [62] Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10:
770 expanded protein families and functions, and analysis tools. *Nucleic Acids Research.*
771 2016; 44: D336–42.

772 [63] Shannon P. Cytoscape: A Software Environment for Integrated Models of Biomolecular
773 Interaction Networks, *Genome Research.* 2003; 13: 2498–2504.

774 [64] Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al.,
775 ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and
776 pathway annotation networks. *Bioinformatics.* 2009; 25: 1091–1093.

777 [65] Koskinen P, Törönen P, Nokso-Koivisto J, Holm L. PANNZER: high-throughput
778 functional annotation of uncharacterized proteins in an error-prone environment.
779 *Bioinformatics.* 2015; 31: 1544–1552.

780 [66] Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. EnsemblCompara
781 GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome*
782 *Research*. 2008; 19: 327–335.

783 [67] Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, et al., Ensembl
784 2016. *Nucleic Acids Research*. 2016; 44: D710–D716.

785 [68] Deevi S, 4D Strategies. modes: Find the Modes and Assess the Modality of Complex and
786 Mixture Distributions, Especially with Big Datasets. R package version 0.7.0. 2016;
787 <https://CRAN.R-project.org/package=modes>

788 [69] Hothorn T, Hornik K, van de Wiel MA, Zeileis A. Implementing a Class of Permutation
789 Tests: The coin Package. *Journal of Statistical Software* 2008; 28: 1-23.

790 [70] R Core Team. R: Core Team (2018). R: A language and environment for statistical
791 computing. R Foundation for Statistical Computing, Vienna, Austria. URL
792 <https://www.R-project.org/>.

793 [71] Mudunuri U, Che A, Yi M, Stephens RM. bioDBnet: the biological database network,
794 *Bioinformatics*. 2009; 25: 555–556.

795 [72] Loppin B, Bonnefoy E, Anselme C, Laurençon A, Karr TL, Couble P. The histone H3.3
796 chaperone HIRA is essential for chromatin assembly in the male pronucleus. *Nature*.
797 2005; 437: 1386–1390.

798 [73] Froman DP, Feltmann AJ, Pendarvis K, Cooksey AM, Burgess SC, Rhoads DD. A
799 proteome-based model for sperm mobility phenotype. *Journal of Animal Science* 2011;
800 89: 1330–1337.

801 [74] Clark NL, Swanson WJ. Pervasive Adaptive Evolution in Primate Seminal Proteins.
802 *PLoS Genet*. 2005; 1: e35–8.

803 [75] Swanson WJ, Wong A, Wolfner MF, Aquadro CF. Evolutionary expressed sequence tag
804 analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive
805 selection. *Genetics*. 2004; 168: 1457–1465.

806 [76] Skerget S, Rosenow MA, Petritis K, Karr TL. Sperm Proteome Maturation in the Mouse
807 Epididymis. PLoS ONE. 2015; 10: e0140650–23.

808 [77] Park YJ, Kwon W-S, Oh S-A, Pang M-G. Fertility-Related Proteomic Profiling Bull
809 Spermatozoa Separated by Percoll, J. Proteome Res. 2012; 11: 4162–4168.

810 [78] Conrad M, Ingold I, Buday K, Kobayashi S, Angeli JPF. ROS, thiols and thiol-regulating
811 systems in male gametogenesis. BBA - General Subjects. 2014; 1850: 1566-74

812 [79] Bansal SK, Gupta N, Sankhwar SN, Rajender S. Differential Genes Expression between
813 Fertile and Infertile Spermatozoa Revealed by Transcriptome Analysis. PloS One. 2015.
814 10: e0127007–21.

815 [80] Vacquier VD. Evolution of Gamete Recognition Proteins. Science. 1998; 281: 1995–
816 1998.

817 [81] Chu DS, Liu H, Nix P, Wu TF, Ralston EJ, Yates III JR. et al., Sperm chromatin
818 proteomics identifies evolutionarily conserved fertility factors. Nature. 2006; 443: 101–
819 105.

820 [82] Vicens A, Borziak K, Karr TL, Roldan ERS, Dorus S. Comparative Sperm Proteomics in
821 Mouse Species with Divergent Mating Systems. Molecular Biology and Evolution. 2017;
822 34:1403-1416.

823 [83] McDonough CE, Whittington E, Pitnick S, Dorus S. Proteomics of reproductive systems:
824 Towards a molecular understanding of postmating, prezygotic reproductive barriers,
825 Journal of Proteomics. 2016; 135: 26–37.

826 [84] Andersson M, Andersson M. Female choice selects for extreme tail length in a
827 widowbird. Nature. 1982; 299: 818–820.

828 [85] Price T. Speciation in Birds. Roberts and Company, 2008.

829 [86] Birkhead TR, Møller AP. Sperm competition in birds: Evolutionary causes and
830 consequences, 1st ed., Academic Press, London, 1992.

831 [87] Immler S, Saint-Jalme M, Lesobre L, Sorci G, Roman Y, Birkhead TR. The evolution of
832 sperm morphometry in pheasants. *Journal of Evolutionary Biology*. 2007; 20: 1008–1014.

833 [88] Lüpold S, Calhim S, Immler S, Birkhead TR. Sperm morphology and sperm velocity in
834 passerine birds. *Proceedings of the Royal Society B-Biological Sciences*. 2009; 276:
835 1175–1181.

836 [89] Palmer MR, McDowall MH, Stewart L, Ouaddi A, MacCoss MJ, Swanson WJ. Mass
837 spectrometry and next-generation sequencing reveal an abundant and rapidly evolving
838 abalone sperm protein. *Molecular Reproduction and Development*. 2013; 80: 460-465.

839 [90] Shteynberg D, Nesvizhskii AI, Moritz RL, Deutsch EW. Combining results of multiple
840 search engines in proteomics. *Mol. Cell Proteomics*. 2013; 12: 2383–2393.

841 [91] Arnqvist G. Comparative evidence for the evolution of genitalia by sexual selection.
842 *Nature*. 1998; 393: 784–786.

843 [92] Price JJ, Whalen LM. Plumage evolution in the Oropendolas and Caciques: different
844 divergence rates in polygynous and monogamous taxa. *Evolution*. 2009; 63: 2985–2998.

845 [93] Seehausen O, Mayhew PJ, Van Alphen JJM. Evolution of colour patterns in East African
846 cichlid fish. *Journal of Evolutionary Biology*. 1999; 12: 514–534.

847 [94] Wyckoff GJ, Wang W, Wu C-I. Rapid evolution of male reproductive genes in the
848 descent of man. *Nature*. 2000; 403: 304–309.

849 [95] Karr TL. Fruit flies and the sperm proteome. *Human Molecular Genetics*. 2007; 16:
850 R124-133.

851 [96] Yang Z, Swanson WJ, Vacquier VD. Maximum-likelihood analysis of molecular
852 adaptation in abalone sperm lysin reveals variable selective pressures among lineages and
853 sites. *Molecular Biology and Evolution*. 2000; 17: 1446–1455.

854 [97] Findlay GD, Yi X, MacCoss MJ, Swanson WJ. Proteomics Reveals Novel *Drosophila*
855 Seminal Fluid Proteins Transferred at Mating. *PLoS Biol*. 2008; 6.

- 856 [98] Lu B, McClatchy DB, Kim JY, Yates III JR. Strategies for shotgun identification of
857 integral membrane proteins by tandem mass spectrometry. *Proteomics*. 2008; 8: 3947–
858 3955.
- 859 [99] Fisher HS. Supergene yields super sperm. *Nat. Ecol. Evol.* 2017; 1: 1064–1065.
- 860 [100] Fisher HS, Jacobs-Palmer E, Lassance J-M, Hoekstra HE. The genetic basis and fitness
861 consequences of sperm midpiece size in deer mice. *Nature Communications*. 2016;
862 7:13652.
- 863 [101] Kierszenbaum AL, Tres LL, Rivkin E, Kang-Decker N, van Deursen JMA. The
864 acroplaxome is the docking site of Golgi-derived myosin Va/Rab27a/b- containing
865 proacrosomal vesicles in wild-type and Hrb mutant mouse spermatids. *Biology of*
866 *Reproduction*. 2004; 70: 1400–1410.
- 867 [102] Lo JCY, Jamsai D, O'Connor AE, Borg C, Clark BJ, Whisstock JC, et al., RAB-Like 2
868 Has an Essential Role in Male Fertility, Sperm Intra-Flagellar Transport, and Tail
869 Assembly. *PLoS Genet*. 2012; 8: e1002969–11.
- 870 [103] S. Naaby-Hansen, J.C. Herr, Heat shock proteins on the human sperm surface, *Journal of*
871 *Reproductive Immunology*. 84 (2010) 32–40.
- 872 [104] Hiyama G, Matsuzaki M, Mizushima S, Dohra H, Ikegami K, Yoshimura T, et al., Sperm
873 activation by heat shock protein 70 supports the migration of sperm released from sperm
874 storage tubules in Japanese quail (*Coturnix japonica*). *Reproduction*. 2014; 147: 167–178.
- 875 [105] Rubinstein E, Ziyat A, Wolf J, Lenaour F, Boucheix C. The molecular players of sperm–
876 egg fusion in mammals. *Seminars in Cell & Developmental Biology*. 2006; 17: 254–263.
- 877 [106] Evans JP. The molecular basis of sperm-oocyte membrane interactions during
878 mammalian fertilization. *Human Reproduction Update*. 2002; 8: 297–311.
- 879 [107] Vicens A, Lüke L, Roldan ERS. Proteins Involved in Motility and Sperm-Egg Interaction
880 Evolve More Rapidly in Mouse Spermatozoa. *PLoS ONE*. 2014; 9: e91302.

- 881 [108] Finn S, Civetta A. Sexual Selection and the Molecular Evolution of ADAM Proteins, J
882 Mol Evol. 2010; 71: 231–240.
- 883 [109] Okumura H, Kohno Y, Iwata Y, Mori H, Aoki N, Sato C. et al., A newly identified zona
884 pellucida glycoprotein, ZPD, and dimeric ZP1 of chicken egg envelope are involved in
885 sperm activation on sperm-egg interaction. Biochem. J. 2004; 384: 191–199.
- 886 [110] Okumura H. Avian Egg and Egg Coat. In: Sasanami T (ed). Avian Reproduction: from
887 behaviour to molecules. Springer Singapore, Singapore, 2017: pp. 75–90
- 888 [111] Marzoni M, Castillo A, Sagona S, Citti L, Rocchiccioli S, Romboli I, et al., A proteomic
889 approach to identify seminal plasma proteins in roosters (*Gallus gallus domesticus*),
890 Animal Reproduction Science. 2013; 140: 216–223.
- 891 [112] Atikuzzaman M, Sanz L, Pla D, Alvarez-Rodriguez M, Rubér M, Wright D, et al.,
892 Selection for higher fertility reflects in the seminal fluid proteome of modern domestic
893 chicken. Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics.
894 2017; 21: 27–40.
- 895 [113] Travis AJ, Kopf GS. The role of cholesterol efflux in regulating the fertilization potential
896 of mammalian spermatozoa. J. Clin. Invest. 2002; 110: 731–736.
- 897 [114] Bakst MR, Cecil HC. Effect of bovine serum albumin on motility and fecundity of turkey
898 spermatozoa before and after storage. J. Reprod. Fertil. 1992; 94: 287–293.
- 899 [115] Jamieson BGM. Avian spermatozoa: structure and phylogeny. In Jamieson BGM (Ed.),
900 Reproductive biology and phylogeny of birds. Part A. Phylogeny, Morphology and
901 Fertilization. Enfield: Science Publishers, Inc, 2007: pp. 349–511.
- 902

903 **Table 1. Gene ontology analysis: molecular function.** Zebra finch sperm proteome, top 10
 904 significantly enriched terms. See supplementary table S3 for full results.

GO ID	GO term description	p-value*
GO:0016491	Oxidoreductase activity	2.33 E-17
GO:0003954	NADH dehydrogenase activity	9.59 E-12
GO:0016651	Oxidoreductase activity, acting on NAD(P)H	9.79 E-10
GO:0016903	Oxidoreductase activity, acting on aldehyde or oxo group of donors	4.59 E-09
GO:0005198	Structural molecule activity	1.62 E-08
GO:0003735	Structural constituent of ribosome	2.66 E-08
GO:0016655	Oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor	5.58 E-08
GO:0003723	RNA binding	7.31 E-08
GO:0016624	Oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulfide as acceptor	1.45 E-07
GO:0051536	Iron-sulfur cluster binding	8.78 E-07

905 ***(corrected with Benjamini-Hochberg)**

906
 907
 908 **Table 2. Gene ontology analysis: cellular component.** Zebra finch sperm proteome, top 10
 909 significantly enriched terms. See supplementary table S4 for full results.

GO ID	GO term description	p-value*
GO:0005739	Mitochondrion	9.20 E-45
GO:0044444	Cytoplasmic part	1.53 E-44
GO:0044429	Mitochondrial part	4.81 E-41
GO:0043209	Myelin sheath	1.97 E-40
GO:0065010	Extracellular membrane-bounded organelle	3.76 E-40
GO:0070062	Extracellular exosome	3.87 E-40
GO:1903561	Extracellular vesicle	4.39 E-40
GO:0043230	Extracellular organelle	4.48 E-40
GO:0005743	Mitochondrial inner membrane	6.15 E-38
GO:0031966	Mitochondrial membrane	8.28 E-38

910 ***(corrected with Benjamini-Hochberg)**

911
 912
 913
 914 **Table 3. Gene ontology analysis: biological process.** Zebra finch sperm proteome, top 10
 915 significantly enriched terms. See supplementary table S5 for full results.

GO ID	GO term description	p-value*
GO:0006091	Generation of precursor metabolites and energy	1.37 E-17
GO:0045333	Cellular respiration	2.64 E-16
GO:0055114	Oxidation-reduction process	3.77 E-16
GO:0009116	Nucleoside metabolic process	1.29 E-13
GO:0044281	Small molecule metabolic process	2.68 E-13
GO:1901657	Glycosyl compound metabolic process	2.87 E-13
GO:0015980	Energy derivation by oxidation of organic compounds	5.76 E-12
GO:1901564	Organonitrogen compound metabolic process	8.97 E-12
GO:0022900	Electron transport chain	2.37 E-11
GO:0006099	Tricarboxylic acid cycle	1.11 E-10

916 ***(corrected with Benjamini-Hochberg)**

917

918

Table 4. Evolutionary analysis of the zebra finch sperm proteome. Zebra finch sperm proteome genes with dN/dS > 0.5

ZF Ensembl gene ID	Gene name	Gene description	dN	dS	dN/dS
ENSTGUG000000003431	COX7A2L	Cytochrome c oxidase subunit 7A2 like	0.091	0.105	0.861
ENSTGUG000000002755	SSNA1	SS nuclear autoantigen 1	0.056	0.067	0.831
ENSTGUG000000010298	USMG5	Upregulated during skeletal muscle growth 5-like	0.050	0.066	0.763
ENSTGUG000000011667	FABP4	Fatty acid binding protein , adipocyte	0.030	0.046	0.649
ENSTGUG000000010822	NUSAP1	Nucleolar and spindle associated protein 1	0.257	0.422	0.611
ENSTGUG000000012628	CCDC77	Coiled-coil domain containing 77	0.116	0.202	0.572
ENSTGUG000000012019	N/A	N/A	0.057	0.105	0.543
ENSTGUG000000007376	N/A	N/A	0.117	0.218	0.537
ENSTGUG000000001294	TXN	Thioredoxin	0.099	0.186	0.536
ENSTGUG000000009075	UQCRC2	Ubiquinol-cytochrome c reductase core protein II	0.070	0.131	0.534
ENSTGUG000000006922	FUNDC2	FUN14 domain containing 2	0.020	0.036	0.522
ENSTGUG000000011286	MTHFD1L	Methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1-like	0.054	0.104	0.514
ENSTGUG000000008469	N/A	N/A	0.095	0.185	0.512

919

920 **Figure 1. Sperm isolation and purification.** (A) Seminal glomerulae used as starting material. (B)
921 Low magnification field of view of purified sperm stained with DAPI to reveal cell nuclei in
922 purified fractions. (C) Higher magnification view of purified sperm visualized using DIC to
923 visualize sperm tails (bracket) and epifluorescence microscopy to reveal nuclei (arrowhead).

924
925
926
927

928 **Figure 2. Major protein classes in the ZfSP.** Bar graph representation by number of genes in each
929 protein class recognized by PATHER (pantherdb.org).

931
932
933
934
935
936
937
938
939
940

941 **Figure 3. Gene ontology (GO) classification of gene products identified in ZfSP.** GO
 942 classification according to analysis with ClueGO for (A) molecular function, (B) cellular
 943 compartment, and (C) biological process. (See Supplementary Figure S1 for network views).
 944

945
 946

947 **Figure 4. STRING protein-protein interaction network for the 13 genes in the ZfSP with**
948 **dN/dS values > 0.5.** Figure was produced using STRING evidence view, interaction score (0.400),
949 and maximum number of interactions in the 1st shell set to no more than 10.

954 **Figure 5. Evolutionary analysis of the ZfSP.** Average dN/dS values between the zebra finch and
955 the flycatcher for the entire sperm proteome (ZfSP) and the individual functional categories of ZfSP
956 proteins. Genes without clear functions were omitted. We compared the overall dN/dS ratio of the
957 ZfSP to each functional group using a Wilcoxon one-sample t test (see main text for statistical
958 results).

963 **Figure 6. Comparison of the chicken and zebra finch sperm proteome.** (A) Overlap in genes
 964 between the zebra finch (red) and the chicken (blue) sperm proteome (B) Enriched cellular
 965 component GO categories for proteins in the chicken and zebra finch sperm proteomes. Network
 966 diagram indicates functional enrichment of GO categories (based on ClueGO analysis) for proteins
 967 in the zebra finch sperm proteome (red) versus the chicken sperm proteome (blue). The varying
 968 shades of colour represent nodes and clusters of functional GO categories biased towards the zebra
 969 finch (red) or the chicken (blue).

970 **A**

971

972 **B**

973

974 **Supplementary Materials File 1**

975 **Table S1.** Parameters used in the MaxQuant analysis.

976 **Table S2.** Full output from MaxQuant analysis.

977 **Table S3.** Proteins found in the zebra finch sperm proteome.

978 **Table S4.** Molecular function GO analysis of proteins in the zebra finch sperm proteome.

979 **Table S5.** Cellular component GO analysis of proteins in the zebra finch sperm proteome.

980 **Table S6.** Biological Process GO analysis of proteins in the zebra finch sperm proteome.

981 **Table S7.** PANNZER GO annotation and gene descriptions for the 86 proteins not assigned

982 Ensembl IDs.

983 **Table S8.** dN/dS values for the zebra finch sperm proteome against flycatcher, *Ficedula albicollis*.

984 **Table S9.** dN/dS values for the zebra finch sperm proteome against Chicken, *Gallus gallus*

985 **Table S10.** dN/dS values for the zebra finch sperm proteome against flycatcher, *Ficedula albicollis*,

986 entire genome. Only 1:1 high confidence orthologues.

987 **Figure S1.** Network analysis of proteins in the ZfSP for (A) Molecular function, (B) Cellular

988 component, and (C) Biological Process.

989 **Figure S2.** Distribution of median dN/dS for the ZfSP compared to the zebra finch genomic

990 background.

991 **Figure S3.** Protein interaction network centered on the gene product ADAM2.