SCIENTIFIC REPORTS

OPEN

The Seminal fluid proteome of the polyandrous Red junglefowl offers insights into the molecular basis of fertility, reproductive ageing and domestication

Kirill Borziak[1,*], Aitor Álvarez-Fernández[2,*], Timothy L. Karr[3], Tommaso Pizzari[2] & Steve Dorus[1]

Seminal fluid proteins (SFPs) are emerging as fundamental contributors to sexual selection given their role in post-mating reproductive events, particularly in polyandrous species where the ejaculates of different males compete for fertilisation. SFP identification however remains taxonomically limited and little is known about avian SFPs, despite extensive work on sexual selection in birds. We characterize the SF proteome of the polyandrous Red junglefowl, *Gallus gallus*, the wild species that gave rise to the domestic chicken. We identify 1,141 SFPs, including proteins involved in immunity and antimicrobial defences, sperm maturation, and fertilisation, revealing a functionally complex SF proteome. This includes a predominant contribution of blood plasma proteins that is conserved with human SF. By comparing the proteome of young and old males with fast or slow sperm velocity in a balanced design, we identify proteins associated with ageing and sperm velocity, and show that old males that retain high sperm velocity have distinct proteome characteristics. SFP comparisons with domestic chickens revealed both qualitative and quantitative differences likely associated with domestication and artificial selection. Collectively, these results shed light onto the functional complexity of avian SF, and provide a platform for molecular studies of fertility, reproductive ageing, and domestication.

Seminal fluid (SF), the physiological, non-cellular component of an ejaculate, contains a diversity of molecules including peptides and proteins[1]. A growing body of empirical work has demonstrated the involvement of several such molecules in different stages of reproduction, including sperm fertilising efficiency and longevity, antimicrobial defences, female sperm utilisation, the coordination of ovulation and oviposition in females, maternal allocation, and female remating propensity[2]. Given their modulation of postcopulatory reproductive events, much research has focused on the role of seminal fluid proteins (SFPs) in polyandrous species, where females obtain ejaculates from multiple partners, prolonging sexual selection on males after copulation via sperm competition and cryptic female choice[1].

Increasing evidence indicates that SFPs can play a key role in postcopulatory sexual selection by promoting the fertilisation success of an ejaculate through distinct (although not necessarily mutually exclusive) functionalities. First, some SFPs can have a direct influence on the fertilising performance of self-sperm by modulating sperm motility and successful sperm storage within the female sperm storage organs. For example, mammalian SFPs may increase sperm velocity by accelerating the conversion of glucose into ATP[3]. In *Drosophila melanogaster*, the SFP Acp36DE facilitates post-copulatory sperm storage. Females mated to Acp36DE mutant males store only 15% of sperm and produce 10% of progeny relative to those mated to wild-type males[4]. Not surprisingly, Acp36DE mutant males are disadvantaged in sperm competition when competing against wild-type males with normal expression of Acp36DE[5]. Second, SFPs may have 'offensive' effects, by reducing the fertilising efficiency of

[1]Center for Reproductive Evolution, Department of Biology, Syracuse University, US. [2]Edward Grey Institute, Department of Zoology, University of Oxford, UK. [3]Drosophila Genetic Resource Center, Kyoto Institute of Technology, Saga Ippongi-cho, Ukyo-ku, Kyoto 616-8354, Japan. *These authors contributed equally to the work. Correspondence and requests for materials should be addressed to S.D. (email: sdorus@syr.edu)

competing (i.e. non-self) ejaculates[6], or 'defensive' effects which reduce the future risk of sperm competition, for example through the formation of mating plugs[7]. These 'defensive' effects can also be achieved by manipulating female responses to the male's advantage – e.g. by reducing female propensity to remate or accelerating female oviposition rate. Amongst *D. melanogaster* SFPs, Sex Peptide (Acp70A) delays female remating[8], while ovulin (Acp26Aa) enhances short term egg production by the female[9]. These functionalities, while adaptive to males within the context of postcopulatory sexual selection, can also have detrimental repercussions on the fitness of females, and have been put forth as potential examples of sexual conflict[10].

Consistent with their role in sexual selection, SFP genes often experience accelerated rates of adaptive evolution. For example, the *SEMG2* gene has experienced accelerated evolution in polyandrous species of primates in relation to monandrous species. This gene encodes a predominant component of the copulatory plug, a trait crucial to the fertilisation success of competing males participating in subsequent copulations[11]. Similarly, the molecular mass of the SFP SVSII, which is responsible for the formation of the copulatory plug in muroid rodents, correlates positively with the level of polyandry across species[12], and the gene encoding this protein is subjected to positive selection in lineages with higher levels of sperm competition[13]. Thus, polyandry can drive the rapid evolution of SFPs and this, in turn, can contribute to the establishment of reproductive barriers and, eventually, to speciation[14,15].

However, the study of SFPs remains taxonomically restricted and largely limited to populations that have long adapted to captivity or domestication. In particular, little is known about SFPs and their functional significance in wild birds (but see ref. 16), despite the fact that birds have long served as a paragon of sexual selection and more recent evidence that supports the influence of SF on reproductive fitness[17,18]. Avian SFP biology also has the potential to be unique because birds lack accessory tissues dedicated to the production of SFPs, such as the seminal vesicle in mammals and accessory gland present in some insects. Here, we begin to address this gap in knowledge by characterizing the SF proteome of a population of Red junglefowl (RJF), *Gallus gallus ssp*. The RJF is an appropriate model for the study of avian SF for the following reasons. First, RJF populations are typically polyandrous and the mechanisms of pre- and post copulatory sexual selection are reasonably well understood[19–21]. Second, because the RJF is the wild species that gave rise to the domestic chicken, *G. domesticus*[22], information on the physiology of chicken SF can be used to infer the functional significance of the RJF SF proteome, and, in turn, the RJF proteome can shed light on the influence of artificial selection associated with domestication. Domestication has had a profound impact on chicken reproductive biology[23], and information on domestic SFPs[24] provides a unique opportunity to investigate the molecular basis of reproductive divergence between domesticated and undomesticated species.

Physiological studies of domestic chickens indicate that the SF of ejaculates consists of two components: seminal plasma and transparent fluid. Seminal plasma is produced by the testis and vas deferens, and sperm are expected to be continuously exposed to these secretions as they leave the epididymis. Transparent fluid, a plasma-like secretion produced from the tumescent lymphatic folds surrounding the cloaca, has a similar composition to a dialysate of blood plasma, and is mixed with sperm upon ejaculation[25,26]. Empirical evidence from domestic chickens indicates that SF modulates aspects of sperm motility and fertilising efficiency[23,26,27], although the molecular nature of these effects remains unclear[23]. The role of SF in sperm motility is particularly significant, given that swimming velocity is a major determinant of fertilisation success under sperm competition[28,29]. Importantly, sperm velocity tends to decline as individual males age past their prime[30,31], and evidence from RJF populations suggests that this is mediated by changes in the antioxidant capacity of SF[32]. Collectively, these data suggest that SFPs may modulate aspects of male fertilising efficiency and reproductive ageing.

The aims of this study were to: (i) systematically characterize the SF proteome of RJF; (ii) identify proteome characteristics associated with male age and/or sperm swimming velocity; and (iii) compare and contrast these data with published SFP data in domestic chickens.

Results

The Red junglefowl SF proteome.
Tandem mass spectrometry analyses were conducted on SF from 12 male RJF, where each sample was the product of 3 pooled SF samples from the same male. Analysis of 144 protein fractions in total resulted in 200,622 high-quality spectral matches to peptides and the identification of 1,141 proteins (Table S1, Supporting Information). Proteins were required to be identified by no fewer than 2 unique peptides and of these proteins, 24.8% were identified in all 12 males and 55.9% were identified in at least half of the males surveyed (Fig. 1a). The proteomic complexity of RJF SF is therefore much greater than those in other well-studied invertebrates, such as *Drosophila* and mosquito, *Aedes aegypti*, where 138 and 93 ejaculated SFPs have been identified, respectively[33,34] but comparable in size to the SF proteome in humans[35].

To examine the contribution of transparent fluid to the RJF SF proteome, we investigated overlap with existing proteomic data from domestic chicken blood plasma[36]. This revealed that 52.1% (137 out of 263) of identified blood plasma proteins were present in RJF SF (Table S1, Supporting Information). These proteins were more commonly identified across most or all SF samples analysed than the remainder of the SF proteome (Fig. 1a) and a substantial number of these (46.7%; 64 out of 137) are amongst the most abundant 20% of SFPs. Gene ontology analysis of SF/blood plasma proteins identified numerous biological process enrichments (Table S1, Supporting Information), including response to stress (3.2E-19), defense response (3.26E-16), hemostasis (8.1E-12), regulation of peptidase activity (1.1E-12) and immune system process (8.4E-14). Predominant components of SF and blood plasma include serum albumin, transthyretin and ovotransferrin. Quantitative analysis indicated that 29% of SFPs by mass are accounted for by proteins also found in blood plasma and that on average these proteins are significantly greater in abundance than the non-blood plasma components of the SF proteome ($p < 0.0001$; Fig. 1b). These observations confirm that plasma secretions within the male reproductive track are a major component of RJF SF and likely represent the contribution of transparent fluid. To assess if these abundant SF components are evolutionarily conserved blood plasma proteins, we then compared our proteome to that of human
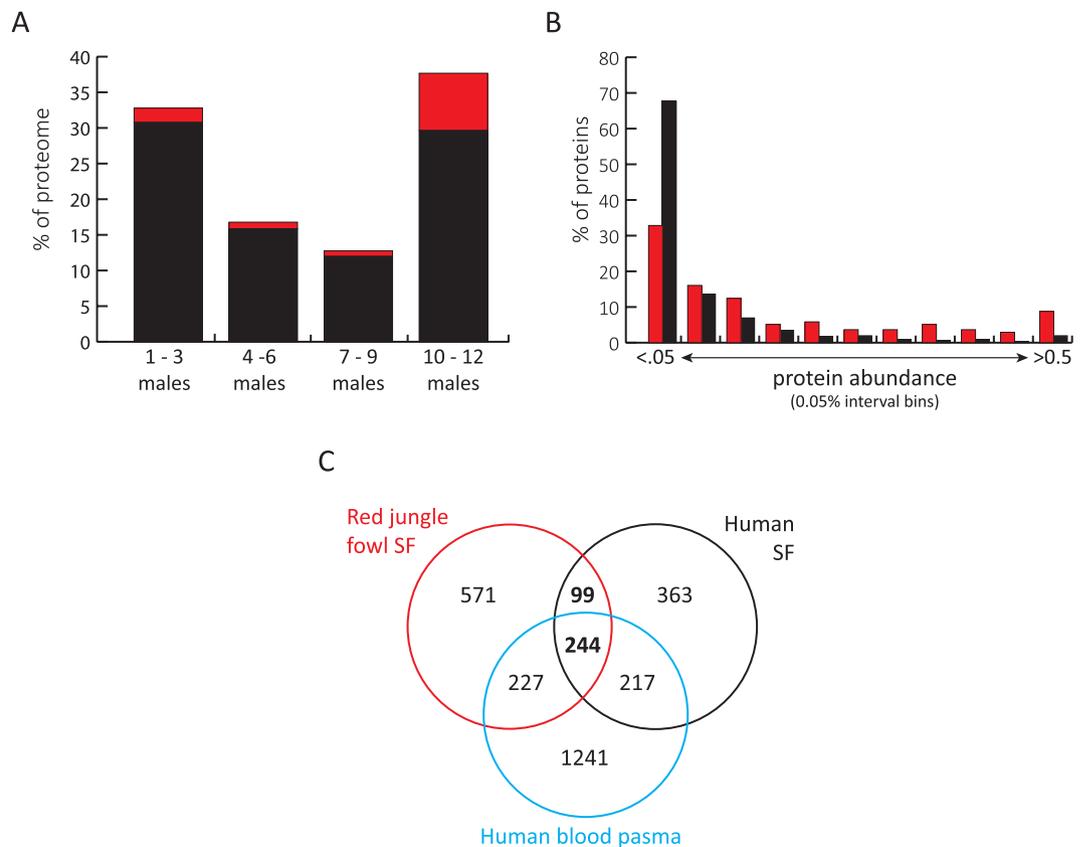
Figure 1. Red junglefowl seminal fluid protein identification and composition. (**A**) Histogram displaying the proportion of the seminal fluid proteome identified as a function of the number of males in which proteins were identified. Proteins also identified as components of blood plasma are indicated in red. (**B**) Histogram comparing the distribution of protein abundance amongst seminal fluid proteins identified (red) and not identified (black) as components of blood plasma. (**C**) Venn diagram displaying the protein overlap between Red junglefowl SF (red), human SF (black) and human blood plasma (blue). The vast majority of conserved SF proteins are also present in the blood plasma proteome.

blood plasma[37]. This revealed that the vast majority of SF proteins in chicken blood plasma (122 of 134) were present in the human plasma proteome and identified an additional 349 RJF SF proteins present in human blood plasma that have yet to be identified in the chicken plasma proteome. We note that the proteomic coverage of human blood, compiled from 91 high-quality MS/MS datasets, far exceeds that in chicken. Based on this observation it was then logical to determine if the prevalence of blood plasma proteins in the RJF might be a common feature of vertebrate ejaculates. A comparison of our SF proteome to that of humans, a dataset that was generated in a similar fashion using SF purified from whole ejaculates[35], identified 343 proteins in common, including 244 (71.1%) also present in the blood plasma proteome (Fig. 1c). Therefore, blood plasma proteins comprise a large fraction of SF proteins identified in whole ejaculates of RJF and humans, and accounts for most of the proteins identified in both species.

A diverse repertoire of immunity proteins function in reproduction and several impact fertility through direct or indirect influences on sperm[38]. A survey of the SF proteome revealed a wide array of molecules which function in immunity, such as (i) innate immunity proteins, including 10 proteins in the classical complement pathways (Table 1), (ii) a diverse set of adaptive immunity proteins, including all four annotated chicken immunoglobulin proteins, and (iii) the avian antimicrobial proteins gallinacin 9 and 10, which are highly abundant and comprise 0.73 and 0.93% of the total protein in SF, respectively (Table 2). It is noteworthy that several abundant immunity proteins are absent from the domestic chicken SF proteome (see below).

SF proteome and post-spermatogenic sperm maturation. Fowl sperm do not complete maturation in the epididymis[39] and, consistent with the idea of continued maturation in the vas deferens, research in domestic chickens has shown that sperm from the testis, epididymis, and three segments of the vas deferens experience a progressive increase in survival, fertilising ability, motility, acrosomal proteolytic activity, sperm penetration of the egg perivitelline layer, and binding capacity to the uterovaginal junction sperm storage tubules and infundibulum[40]. To explore the possibility that SF (or secreted vesicles therein; see below) might contribute to extragonadal maturation, we assessed the overlap between sperm proteins previously identified in a population of domestic broiler chickens[24] and two independent RJF SF datasets derived from SF purified using distinct protocols to control for potential sperm contamination (see Methods). The first dataset represents the combined proteome of 12

GI #	Protein Name	# of males	Abundance percentile	Not identified in domestic chicken*
46048795	vitronectin	12	93.2	
118098426	polyubiquitin-C	11	89.8	
363744121	complement C7	12	88.0	
363740281	complement C8 gamma chain	12	86.9	
363736454	complement factor H	12	86.4	X
363729287	complement C4-A	12	84.0	
513213068	complement C5	12	76.9	
363740226	macrophage migration inhibitory factor	12	76.1	
221325664	complement C6	12	82.2	
513220989	complement C2 isoform X2	10	71.6	X
57529999	toll-interacting protein	10	60.4	
433282973	complement factor I	12	59.5	
118094713	complement C8 alpha chain	11	56.2	X
118094715	complement C8 beta chain	2	22.0	X
326633216	soluble mannose-binding lectin	1	9.6	X

Table 1. Innate immunity proteins of the Red junglefowl SF proteome. *These differences may be attributable to methodological differences between this study and ref. 24 (see below).

GI #	Protein Name	Abundance Percentile	Not identified in domestic chicken*
Anti-microbial/viral proteins			
48976029	gallinacin- 10	99.3	
48675889	gallinacin-9	98.8	
Adaptive immunity proteins			
CAA30161.1	Immunoglobulin gamma chain	99.6	
513225677	alpha-2-macroglobulin	97.0	
AAB22614.2	Immunoglobulin alpha heavy chain	96.6	
AAA48833.1	Immunoglobulin heavy chain	94.1	
P01875.2	Immunoglobulin mu chain C	91.1	X
513200587	beta-2-microglobulin isoform X1	82.6	
52345405	Leukocyte surface antigen CD47	71.5	
52138689	Lymphocyte antigen 86	70.5	X
403224973	cathelicidin-B1	68.8	X
448261627	ATP synthase subunit beta, mitochondrial	55.3	
113206156	N-acetylmuramoyl-L-alanine amidase	50.1	
513211517	class II histocompatibility antigen, B-L beta chain	46.5	X
48675879	cathelicidin-1	26.6	X

Table 2. Anti-microbial and adaptive immunity proteins of the Red junglefowl SF proteome. *These differences may be attributable to methodological differences between this study and ref. 24 (see below).

RJF samples described above, of which 9.55% were also specifically identified in domestic chicken sperm (Table S1, Supporting Information). The second dataset, which included SF samples derived from 4 independent ejaculates from a single male, exhibited an almost identical overlap with sperm proteins (9.78%). Importantly, 78.9% of these SFPs also identified in sperm were present in both of our SF samples, strongly supporting a relationship between SF and sperm composition. This pattern is also evident in ref. 24, which identified a substantial number of proteins (n = 234) common to both domestic chicken SF and sperm. A survey of these 234 proteins revealed that 92.3% and 94.4% of these proteins were also identified in our RJF SF datasets, respectively. We then asked if these predicated sperm proteins identified in our SF dataset represent evolutionary conserved sperm components present in human[41] and mouse[42] sperm. Of the sperm proteins present in our dataset, 81 (74%) and 68 (62%) were also present in the human and mouse sperm proteomes, respectively, further supporting the potential importance of these proteins in sperm maturation and function.

Gene ontology analysis of SFPs also identified in chicken sperm further suggests a predicted role for SFPs in flagellum development. A significant enrichment of cytoskeletal (1.0E-22), microtubule (5.6E-15), cilia (2.5E-6) and flagella (7.2E-5) proteins was observed, including subunits of chaperonin containing TCP-1 complex, which is responsible for actin and tubulin folding. Enrichment of protein complex subunits involved in axoneme elongation were also identified, including proteasome core complex, alpha-subunit complex (1.3E-9) and microtubule organizing centre (5.1E-6). The identification of 12 flagellum proteins amongst the 284 protein identified in all 12

males (4.2%) exceeds the proportion of mouse (1 out of 69)[43] and *Drosophila* (3 out of 138)[33] SFPs with predicted subcellular localization in the flagellum. Although we cannot entirely rule out the possibility of sperm protein contamination during SF purification, the highly reproducible results obtained using alternative purification techniques supports the presence of integral sperm components in RJF SF.

Proteomic support for SF exosomes.   Sperm modification after spermatogenesis has been well documented in mammals, where sperm undergo complex modifications within the epididymis[44,45]. This process is governed, in part, by extracellular vesicles, termed epididymosomes, which are believed to deliver molecular components to sperm[46]. To explore the possibility of a similar mechanism in RJF, we investigated the presence of the most common protein markers of exosomes[47] within our SF dataset. This revealed the presence of 85.6% (60 out of 70) of the top exosome markers with one-to-one orthology between chickens and mammals (Table S1, Supporting Information). This included 7 members of the ras-related Rab protein family, which regulate vesicle formation, trafficking and membrane fusion, 6 annexin proteins, which are membrane scaffolding proteins that regulate vesicle formation, and 6 members of the 14-3-3 protein family, which regulate vesicle targeting through cytoskeleton interactions. Exosome markers in SF were also found to be significantly more abundant on average than the remainder of the proteome (non-parametric Kolmogorov-Smirnov test; $D = 0.523$; $p < 0.0001$), suggesting that exosome-like vesicles may be abundant in SF.

Proteomic changes associated with male age and sperm velocity.   We first performed PCA1 to summarise the variance in protein abundances across the four categories (Young Fast, Young Slow, Old Fast and Old Slow). The first dimension explained 95% of the variance in protein abundance across categories (Fig. 2a). The four categories loaded strongly with this dimension, indicating that it captures patterns of protein abundance variation that are consistent across the four categories (Fig. 2a; z-transformed protein abundance yields qualitatively identical results (data not shown)). Thus, dimension 1 captures the dynamic range of relative protein abundance within samples rather than absolute changes across samples. The values of the squared cosine for dimension 1 reflect the major contribution of abundant proteins to this dimension since the 30 most abundant proteins in the dataset have squared cosines scores between 0.83 and 1. The five proteins with the highest weighting on dimension 1 (i.e. the most abundant) were: serum albumin precursor, transthyretin precursor, fatty acid-binding protein brain, pendrin isoform X2, and gallinacin-10 preproprotein; collectively they explain more than 42% in the construction of dimension 1 (Table S2, Supporting Information). Conversely, for example, amylase (alpha 2A) is low in abundance and one of the lowest coordinates. Dimension 2 of PCA1 explained 3.23% of the variance in protein abundance across categories (Fig. 2a). Old Fast was the only category that had a significant positive loading with dimension 2, while Old Slow and Young Fast both had significant negative loadings (Table S3, Supporting Information), indicating that this dimension may reflect protein abundance variation that is distinct in the ejaculates of old males with fast sperm (Fig. 2a). The between-category PCA2 confirmed this pattern, revealing that Old Fast samples were tightly clustered together and away from the other categories in dimension 1, with clear separation from Old Slow and Young Fast (Fig. 2b). We also note that Young Slow samples occupy a unique space, particularly in relation to dimension 2, but do not cluster as tightly as Old Fast males. Finally, the NMDS analysis summarized the variance in a 2-dimensional ordination space with high fit with the original protein ranks (stress = 0.116; Table S3, Supporting Information). Once again, the Old Fast category emerged as different from the other three categories (NMDS dimension 1; Figure S3, Supplemental Information), confirming the results of PCA1 and PCA2.

PCA3 summarized variation in the profile of individual SFPs across the 12 samples. This analysis identified 11 dimensions, of which the first 3 explained 51.2% of the cumulative variance (Table S3, Supporting Information). Importantly, dimension 1 (eigen value = 219.56, 26.14% of variance explained) captures a gradient in sperm velocity, discriminating Old Slow from Old Fast samples (Fig. 3a). Consistent with this, two of the three old slow males occupy negative coordinates in dimension 1 while all three old fast males occupy positive coordinates in this dimension (Table S2, Supporting Information). We next analyzed dimension 1 weightings of proteins exhibiting significant abundance differences within the single replicated old male (#308) who was sampled with both Fast and Slow sperm velocity. This revealed a striking association between the distribution of dimension 1 weightings and directional changes in protein abundance between Fast and Slow samples from this male (Fig. 3b). To explore the functional basis for dimension 1 weighting associations, we compared the weightings of blood plasma and exosome marker proteins identified above. This revealed a strong disparity between these functional classes, including (i) an enrichment for blood plasma proteins on the negative side of this dimension (Old Slow), (ii) a deficit of blood plasma protein on the positive side (Old Fast) and (iii) the inverse pattern for exosome-related components (Fig. 3c). Consistent with this observation, a substantial number of blood plasma proteins were significantly more abundant in the Slow sample of the replicated Old male (#308) relative to his Fast sample. Interestingly, a somewhat opposite pattern for differentially abundant blood plasma proteins is observed for the young male (#H28) sampled twice (i.e. both Slow and Fast sperm velocity; data not shown). Consistent with the colocalization of Old Fast and Young Slow on dimension 1, 26 of 43 blood plasma proteins are more abundant in the H28 Fast sample. Interestingly, these proteins are largely distinct from the blood plasma proteins identified as differentially abundant in male #308. Taken together, these observations suggest that variation in sperm velocity across males is determined by the abundance of blood plasma and exosome proteins in SF. The detection of distinct sets of proteins in the young and old males outlined above further suggests that the proteins and molecular pathways governing sperm velocity may be age-specific.

Two-way ANOVAs of individual proteins identified 9 cases in which, after controlling for false discovery rates, a protein was differentially expressed with respect to male age, sperm velocity or their interaction (Table 3; top section). Of these 9 proteins, five are proteins more abundant in Old Fast samples. Interestingly, patterns of differential protein expression associated with sperm velocity across males were recapitulated in changes within
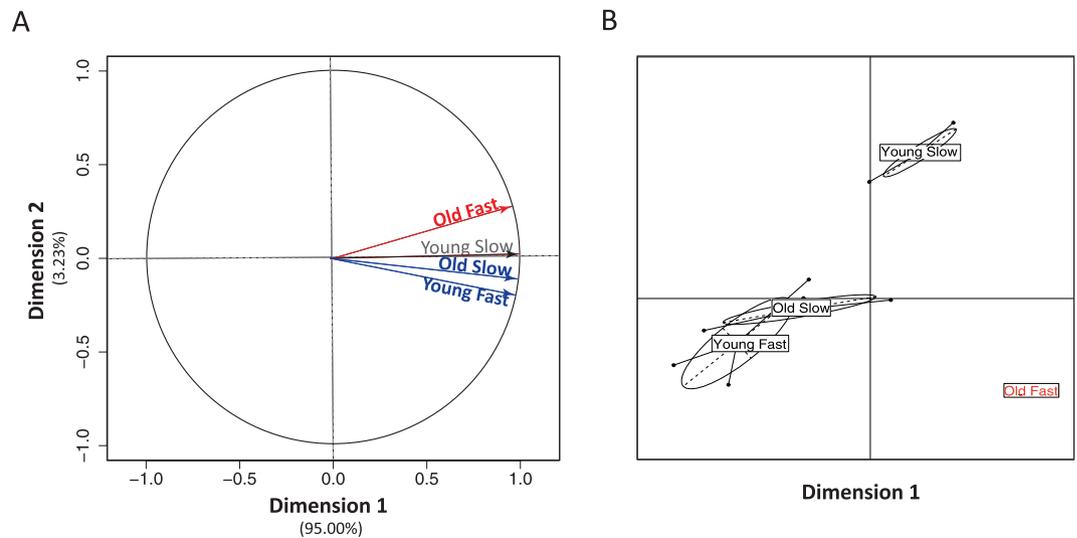
A

B

Figure 2. Multivariate ordination of seminal fluid protein abundance with respect to male age and sperm velocity. (**A**) Principle Component Analysis (PCA1) variables factor map. The direction of the arrows represents the relative loadings of the four categories (Old Fast, Old Slow, Young Fast, Young Slow) on the first and second dimension. All four categories have strong positive loadings on dimension 1; Old Fast has a significant positive loading and Old Slow and Young Fast have significant negative loadings on dimension 2, respectively. (**B**) Between group PCA (PCA2) biplot. First and second dimension loadings for the four categories of males are displayed, as well as the extent of within category variance (ovals) and loadings of individuals within each category (points at the end of solid lines).

the two males that were replicated twice and represented in both the Slow and Fast categories. This included three proteins up-regulated in the Old Fast category (fibulin 2, proto-oncogene protein kinase ROS and receptor-type tyrosine phosphatase eta precursor) and two proteins exhibiting differential abundance between the Young Fast and Young Slow categories (beta-galactoside-binding lecting and inosine-5′-monophosphate dehydrongenase 2). A total of 12 proteins were also identified exclusively and consistently amongst males of a single age (1), velocity (1) or age-velocity category (10) (Table 3; bottom section). Of these, 8 were found in the category Old Fast, which is consistent with the distinct nature of this category as identified by the PCA and NMDS analyses (above). The number of proteins identified exclusively in Old Fast males is unique within the dataset, as the identification of multiple proteins in a specific subset of 3 males was otherwise rare. Lastly, we re-ran PCA1 of the 'inclusive' proteome (above) excluding those proteins identified here by ANOVA and restricted to specific age-velocity categories. This analysis yielded nearly identical results, indicating that the loading of Old Fast samples on dimension 2 of PCA1 is not solely driven by a subset of SFPs, but reflects a general property of SF proteome profiles.

Comparative analysis of the RJF and domestic chicken SF proteome. A direct proteomic comparison, based on the reanalysis of raw MS/MS data from domestic chicken SF samples, revealed that 94.4% of domestic SFPs were identified in our RJF proteome. 31 proteins were identified as unique to the domestic SF proteome (Table S4, Supporting Information), while 585 RJF proteins were absent from the domestic proteome (Fig. 4a). Domestic SFP identification was significantly enriched amongst more abundant RJF proteins ($X^2 = 459$; $p < 0.0001$), indicative of limited coverage of low abundance proteins in the domestic chicken (Fig. 4b). Given the more comprehensive nature of the RJF proteome, proteins uniquely identified in the domestic chicken are likely to represent *bona fide* SF proteome differences between domestic chickens and RJF. Notable amongst these 31 proteins are a significant enrichment in cell recognition proteins (e.g. Cadm1 and Arsa). The relevance of these proteins to fertility is indicated by mammalian phenotypes, including the role of Arsa as an acrosome vesicle zona pellucida binding protein and sperm maturation defects and infertility in Cadm1 mutants. Gst3, a detoxifying enzyme and potential mediator of ROS damage to sperm, was also identified uniquely in the domestic chicken. The substantial overlap (>93%) between the RJF and domestic chicken proteomes amongst the most abundant protein class (Fig. 4b) supports the possibility that some highly abundant proteins not identified in the domestic chicken may be RJF-specific (Table S4, Supporting Information). Gene Ontology analyses of predicted RJF-specific proteins reveals a significant enrichment of proteins predicted to be involved in humoral immunity (3.6E-4) and complement activation (5.7E-3), amongst a diverse set of immunological process categories. Complement factor H, C8 gamma and immunoglobulin mu chain are of particular note as putative RJF-specific SFPs given their high abundance and consistent identification across individuals. Epididymal secretory protein E1, which has been implicated in mammalian sperm maturation, is also noteworthy.

A significant correlation in protein abundance was observed amongst the 556 proteins identified in both the domestic chicken and RJF proteomes (Fig. 4c), despite the systematic differences in protein abundance estimates revealed by comparing cumulative abundance distributions (Fig. 4d). Direct abundance comparisons would therefore be predicted to result in many low-to-moderate abundance proteins appearing to be of greater abundance in RJF and a smaller set of highly biased proteins as greater in abundance in the domestic chicken.
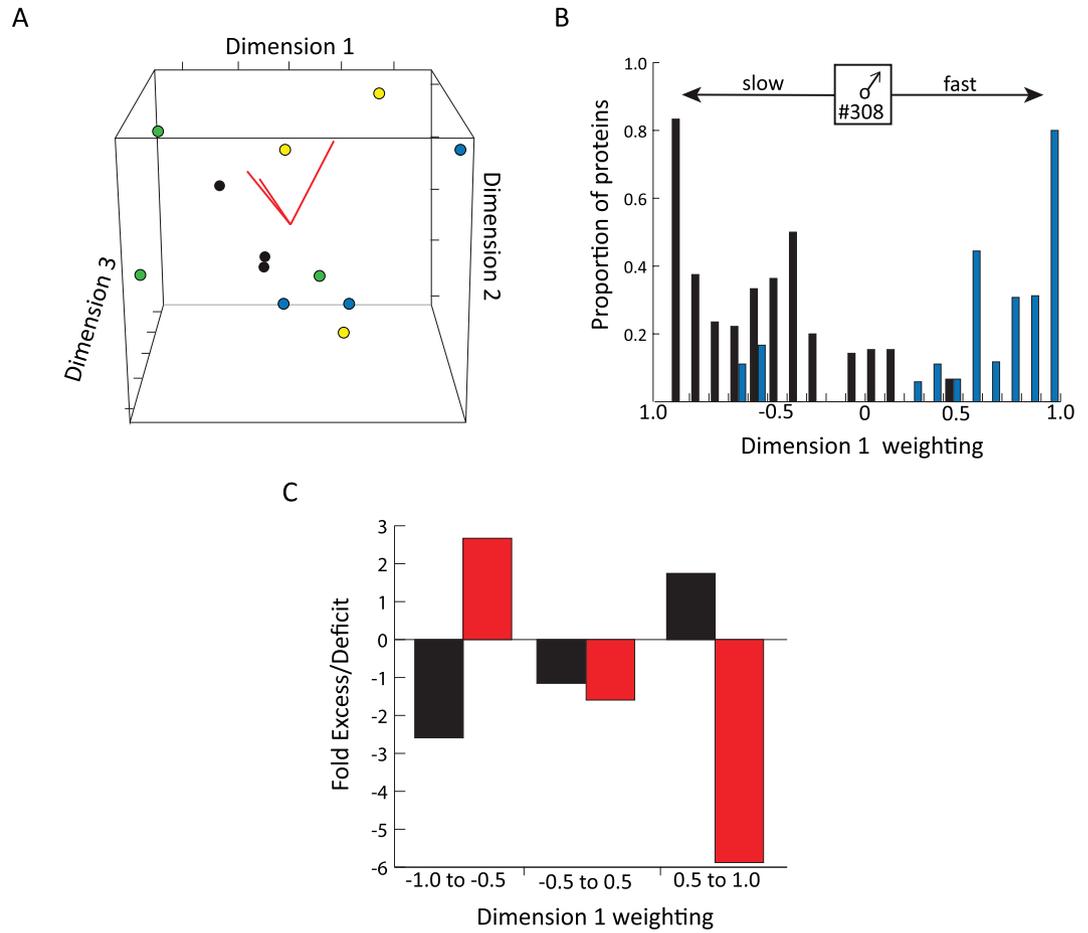
Figure 3. Seminal fluid protein abundance variation associated with sperm velocity in old males.
(**A**) 3-dimension visualization of Principle Component Analysis (PCA3), using the inclusive protein dataset as variables (Old Fast, blue; Old Slow; black; Young Fast; green; Young Slow, yellow). Dimension 1 captures variance associated with old male sperm velocity, including negative weightings for Old Slow and positive weightings for Old Fast. (**B**) Histogram displaying the distribution of first dimension weightings for proteins exhibiting significant abundance differences between the Slow and Fast sample of the replicated old male (#308) (abundance increase in fast sample, blue; abundance increase in slow sample, black). (**C**) Histogram displaying the excess or deficit of blood plasma proteins (red) and exosome markers (black) across first dimension weight ranges.

To minimize this issue, an approach was implemented that assessed the shift in both a protein's abundance rank and abundance estimate between species. This analysis successfully identified quantitative differences in a largely unbiased bidirectional pattern between taxa, including proteins with coherent functions (Table S4, Supporting Information). A significant enrichment was confirmed in the following biological processes: defense response (1.6E-4), response to stress (3.1E-4), inflammatory response (3.8E-3) and regulation of immune system process (3.9E-2). Three of these proteins (IgM, IgG and IgGFc-binding protein) were found to be quantitatively more abundant in RJF than domestic chicken, consistent with the identification of other immunity proteins putatively specific to the RJF SF proteome.

Discussion

This study presents a systematic SF proteome analysis of a non-domestic, polyandrous bird population and one of the first investigations of the signatures of male age and sperm quality in a SF proteome across a population. While the functions of SFPs have received substantial interest[2], much of this work has focused on a few lab-adapted or domestic study organisms. Information on SFPs is particularly patchy in birds, despite the importance of this group to the study of sexual selection. Information on avian SFPs is largely restricted to a recent study of a domestic broiler chicken population[24]. Our study contextualizes this earlier work, by conducting a more comprehensive analysis of a population that has not been exposed to artificial selection. Below we discuss the key aspects of: (i) the composition of the RJF SF proteome, (ii) SF proteome age and sperm velocity-dependent variation, and (iii) the signature of domestication and artificial selection.

In mammals and insects, specific reproductive tissues are responsible for the production of SFPs. Analogous dedicated tissues have yet to be identified in birds and many birds, such as passerines, produce little SF[26]. Fowl,

Category	GI#	Protein name	Comparison or Age-velocity category
Significant in FDR-corrected 2-way ANOVAs			
Age: Velocity	46048932	sodium-dependent phosphate transport protein 2B	OF > YF; OF > YS
Age: Velocity	71895709	glycine-tRNA ligase	OF > YF
Age: Velocity	45382907	receptor-type tyrosine-protein phosphatase eta precursor	OF > YF
Age: Velocity	513204938	fibulin-2	OF > OS; YS > OS
Age: Velocity	160333842	proto-oncogene tyrosine-protein kinase ROS precursor	OF > OS; OF > YF; OF > YS
Age: Velocity	45382785	beta-galactoside-binding lectin	YF > YS
Age	45383337	SPARC precursor	Y > O; YF > OS; YS > OS; YS > OF
Age	303227902	acetyl-CoA acetyltransferase, cytosolic	Y > O
Age	60302752	ras-related protein Rab-11B	Y > O; YF > OS
Identification restricted by age and/or velocity			
Age	56605916	protein tyrosine phosphatase type IVA 1	Old only
Age: Velocity	513237884	neuroblast differentiation-associated protein AHNAK-like	Old Fast only
Age: Velocity	45383816	protein S100-A6	Old Fast only
Age: Velocity	45384028	protein S100-A11	Old Fast only
Age: Velocity	118085711	acyl-CoA-binding protein	Old Fast only
Age: Velocity	160358333	neo-calmodulin	Old Fast only
Age: Velocity	326368309	guanine nucleotide-binding protein (gamma-10)	Old Fast only
Age: Velocity	363730875	myosin regulatory light chain 2,	Old Fast only
Age: Velocity	363745926	guanine nucleotide-binding protein subunit gamma-7-like	Old Fast only
Age: Velocity	478247057	microsomal glutathione S-transferase 3	Young Fast only
Age: Velocity	363742947	nitrilase homolog 1 isoformX1	Young Fast only
Velocity	71895387	inosine-5′-monophosphate dehydrogenase 2	Slow only

Table 3. SFPs differentially expressed with respect to male age, sperm velocity or age: velocity interaction. OF = Old Fast; YS = Young Slow; YF = Young Fast; O = Old; S = Slow.

on the other hand, produce substantial quantities of SF[23]. Our comparative analysis of the SF and blood plasma proteomes reveals that plasma-like secretions represent a major component of both fowl and human SF. Although the functional significance of this transparent fluid in fowls remains unclear, it may contain SFPs that contribute to ROS buffering and the protection of sperm from oxidative damage during prolonged periods of sperm storage in the male extragonadal reserves and in the female storage tubules. The present study also indicates that RJF SF proteome has a compositional complexity that exceeds that of insects and is likely to be more comparable to that of the characterized human SF proteome[35], indicative of diverse range of functions (see below). The substantive contribution of broadly expressed plasma proteins (with putative "housekeeping" roles), appears to stand in stark contrast to the highly specialized repertoire of male-specific SFPs in other species. However, our understanding of gene expression across both male and female reproductive tissues in birds is very limited. Additional proteomic and transcriptomic studies will be required to determine if avian SF also contains highly specialized, male-biased reproductive proteins and, if so, where they originate from in the male reproductive tract.

A diversity of innate and adaptive immunity proteins, as well as antimicrobial proteins, were identified in RJF SF. Immunity proteins are widespread in both male and female reproductive systems[38] and some SFPs are known to target microbes that pose a threat to sperm viability and clear pathogens introduced during copulation, thus protecting the egg from potential infections[48–52]. Evidence suggests that avian semen can harbour bacteria[53,54] and that the SF of some birds may have antibacterial properties[6,18,52]. Our results confirm that antimicrobial proteins gallinacin 9 and 10, of the defensin protein family, are amongst the most abundant proteins in the RJF SF proteome, a finding consistent with SF analysis in domestic chickens[24] and of particular relevance for promiscuous populations of RJF, where ejaculates host diverse bacterial communities (R. Eccleston & T. Pizzari, unpublished). Immunological proteins may also play a key role in genetic recognition, such as kin-biased interactions between ejaculates of different males or interactions between semen and the female reproductive tract. For example, SF can affect sperm fertilising efficiency differentially based on the genetic similarity between sperm and SF donor males in some taxa (e.g. refs 6 and 55). Polyandrous RJF females may also select against sperm of closely related males following mating, a response that appears to be associated with male-female similarity at the Major Histocompatibility Complex locus (MHC)[56]. Two proteins belonging to the MHC Class I (beta-2-microglobulin, and ATP synthase subunit beta, mitochondrial precursor), and one to the MHC Class II (class II histocompatibility antigen, B-L beta chain) were identified in our study. Both MHC Class I proteins are also present in the domestic chicken SF. Interestingly, the class II histocompatibility antigen, B-L beta chain was not present in domestic chicken SF (see discussion of domestication below). Our proteomic results are consistent with the proposition that MHC variation may provide a molecular cue for post-copulatory genetic recognition between rival ejaculates and between mates.

In many species, sperm are not mature or fertilisation competent following spermatogenesis and there is evidence that fowl sperm may also acquire fertilisation capacity progressively through their passage through the vas deferens[40,57–60]. Consistent with this idea, a comparison of two independently derived RJF SF proteomes with
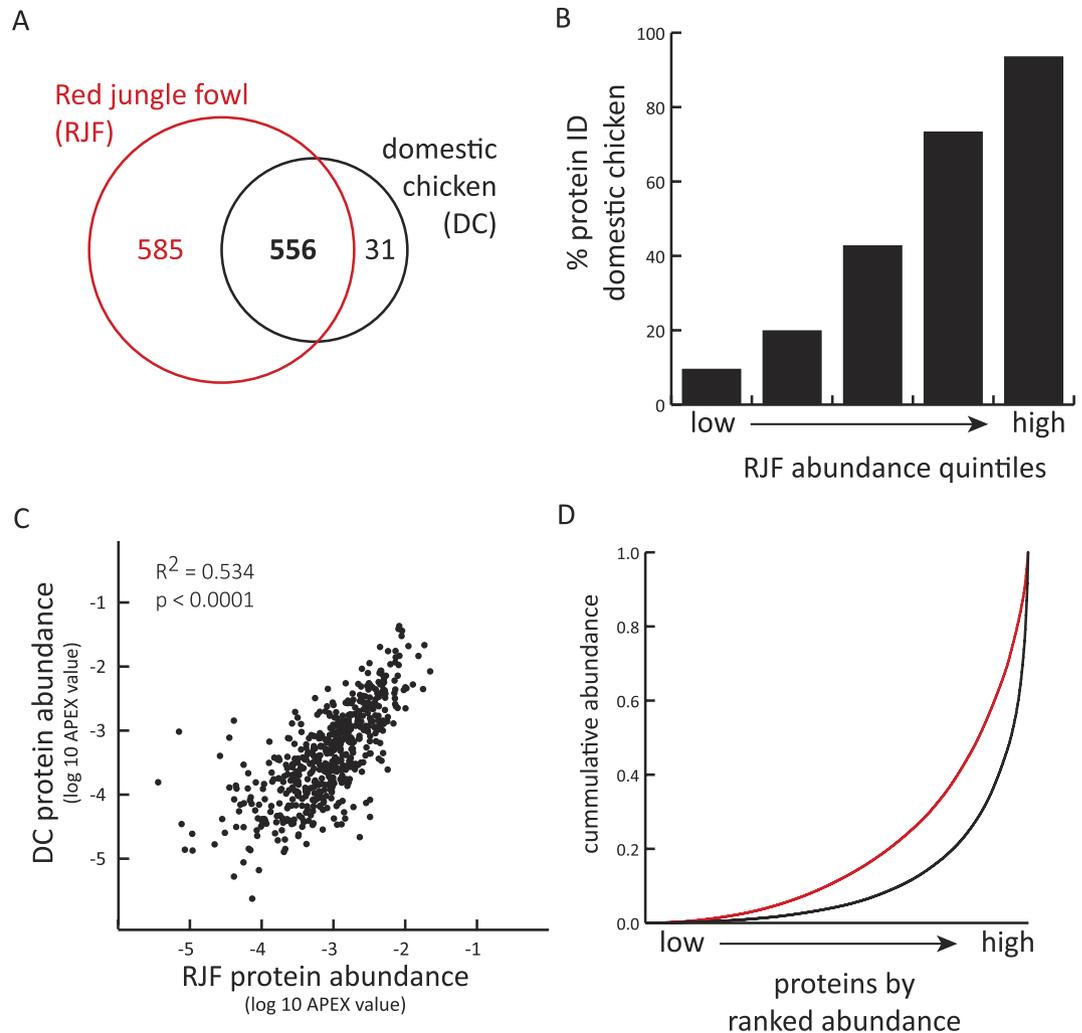
Figure 4. Comparison of the Red junglefowl and domestic chicken seminal fluid proteomes. (**A**) Venn diagram indicating the size and overlap between the seminal fluid proteomes of the Red junglefowl and domestic chicken. (**B**) Overlap in proteins identified in the domestic chicken in relation to protein abundance in the Red junglefowl proteome. (**C**) Linear regression identifies a significant correlation in the abundance of proteins identified in both the domestic chicken and Red junglefowl proteomes. Protein abundance is plotted as log transformed APEX values, which represent normalized estimates of protein abundance in each sample. (**D**) Cumulative protein abundance distribution of the domestic chicken (black) and Red junglefowl (red) proteomes.

the domestic chicken sperm proteome revealed considerable overlap and a significant enrichment of SFPs that are structural components of the flagellum. An alternative, non-mutually exclusive explanation of this overlap is that sperm proteins are released into SF as sperm cells age and degrade. While the present study cannot explicitly distinguish between these alternatives, a careful inspection of our proteomic data reveals the presence of numerous marker proteins that are commonly associated with exosomes. Exosomes produced in the epididymis of mammals are believed to play a prominent role in post-spermatogenic sperm maturation and their molecular cargo (*e.g.* protein or RNA) have been the focus of investigation[46]. A similar process has also been characterized in *Drosophila*, where exosomes produced in the secondary cells of the accessory gland bind and fuse with sperm[61]. Although it is not possible to make firm conclusions based on our data alone, our analysis is consistent with the continued maturation of sperm during transit, and points to the presence of secreted vesicles that could sustain the final stages of maturation in a manner analogous to those described in other systems. Members of the 14-3-3 protein family, which are diverse exosome markers in our dataset, influence mammalian sperm maturation and motility through their interactions with sperm-specific protein phosphatases[62]. Importantly, 14-3-3 eta and zeta are decreased in abundance in SF of subfertile domestic chickens[24] and 3 distinct protein phosphatases were identified in our study as differentially abundant based on age and velocity.

Sperm velocity is a key determinant of fertilising efficiency in domestic chickens[39], particularly under sperm competition when ejaculates with relatively high sperm mobility progressively achieve a competitive edge over female sperm storage time[28]. However, sperm velocity tends to decline as males age[30,31], and these age-dependent effects are associated with changes in SF composition[32]. This study set out to provide the first characterization of

the SF proteomic profiles associated with male age and sperm velocity in a wild species of bird. However, we urge caution in the interpretation of our results and stress that additional studies, using complementary quantitative techniques, will ultimately be required to achieve a refined understanding of SF proteomic variation.

Multivariate analyses distinguished the proteome of Old Fast samples from that of the other three categories of males. These results suggest that high sperm velocity in older males is associated with higher abundance of exosome markers, potentially linked to sperm maturation. Eight proteins were identified exclusively in the samples belonging to the three males of this category. Although it would be premature to draw any specific functional conclusions, many of these proteins are involved in calcium metabolism. For example, calmodulin, a calcium-binding messenger that controls sperm activation and hyperactivation in mammals, activates kinases in chicken sperm that influence both viability and motility[63,64]. It will be of great interest to determine if and how secreted calmodulin in SF might influence RJF sperm maturation and the acquisition of motility. Additionally, the S100 proteins are involved in calcium homeostasis and can function in concert with various growth factors following extracellular secretion[65]. Interestingly, neuroblast differentiation-associated protein AHNAK-like, also identified exclusively in Old Fast males, is implicated in calcium regulation and constitutes a calcium-dependent S100B target protein[66]. In domestic chickens, calcium constitutes a primary determinant of sperm mobility[67]. Excessive calcium uptake can result in mitochondrial degeneration, as was found to be the case in roosters with low sperm mobility[68]. It is possible that these proteins may ameliorate the decline in sperm velocity associated with ageing, through the mediation of calcium uptake. One possibility is that only some males may be able to retain high rates of sperm maturation in old age. Conversely, semen of Old Slow males (with low rates of sperm maturation) is characterized by relatively high amounts of blood plasma proteins, which likely originate from transparent fluid. These patterns could be due to two alternative mechanisms. First, plasma proteins may cause a reduction in sperm velocity. Second, given that many abundant plasma proteins have either buffering or antioxidant properties, it is tempting to speculate that ageing males with slow rates of sperm maturation may partially ameliorate the fertilizing efficiency of their ejaculates by preferentially investing in transparent fluid. The former mechanism would suggest that in the absence of elevated plasma proteins in their SF, the ejaculates of Old Slow males may display higher sperm velocity while the latter mechanism would predict even lower sperm velocity and/or faster rates of sperm mortality. At present the specific role of transparent fluid in fertilizing efficiency in fowl remains inconclusive[25], and more research is required to test these hypotheses. In light of the conserved contribution of blood plasma proteins to SF in fowl and humans, it is interesting to speculate that blood plasma protein abundance might also influence sperm quality characteristics across a wide range of vertebrates, possibly including humans.

Genetic homogenization during domestication of the chicken is likely to have resulted from a combination of intense artificial selection, life-history trade-offs and bottlenecks. The impact of this at key genomics regions, such as those important contributions to immunity and fertility[69], may underlie the well documented reductions in performance amongst domestic chickens and resultant impact on the poultry industry[70]. This loss of function is particularly pronounced in broiler breeders where extreme artificial selection for growth rate has resulted in a trade-off with traits associated with immunity and fertility[4,71]. Our comparative results, although preliminary, highlight divergence between the SF proteome of RJF and that of a population of broiler breeders. We note that methodological differences, including the analysis of larger quantities of proteins and substantially greater numbers of MS/MS runs per male in the current study, has resulted in a far greater depth of coverage in the RJF relative to the domestic chicken. Nonetheless, we show that predicted RJF-specific SFPs are functionally coherent, being more likely to be involved in immunological defenses, consistent with a loss of genetic diversity associated with domestication. A number of factors may underpin this selective disappearance in domestic chickens. For example, genes associated with immune responses may have been "lost" during domestication due to a trade-off with genes supporting different functionalities promoted by artificial selection (*e.g.* production traits), possibly catalyzed by the protection of domestic populations through artificial means (*e.g.* vaccines and antibiotics) which may relax selection on immunity genes and pathways[72,73].

The observation that several SFPs associated with male age and sperm velocity have yet to be identified in the domestic chicken SF proteome is also consistent with the notion that domestication may have led to reductions in male fertility through changes in SF composition. While morphological and behavioural traits, as well as sperm traits, have been evoked to explain reduced male broiler breeder fertility, the results of the present study indicate that SFPs may also play a key role. Additionally, in populations under natural conditions, such as in our study RJF population, males are sexually selected throughout their lifespan, which often extends past their prime[74]. Genes that encode SFPs conferring fertilisation efficiency late in life will be favoured by postcopulatory sexual selection. In domestic populations, which are typically female-biased and where males are culled early in life, sexual selection on late life fertilising efficiency will be relaxed, favoring the evolution of early reproductive senescence. Therefore, our study creates a platform for the identification of SFPs "swept away" by the domestication process, some of which might explain fertility deficits in domesticated chickens. Given that our study was based on semen samples obtained through manual abdominal massage, a related area of future research will be to establish the way in which individual males modulate allocation of different SFPs across natural ejaculates, which are notoriously plastic due to strategic ejaculate expenditure in response to multiple socio-sexual factors[75].

Materials and Methods

Semen collection and SF isolation.
We studied a population of RJF at the John Krebs Field Station (Oxfordshire), which was founded in 2006 from other captive populations. All sampling of birds was conducted in accordance with approved experimental protocols under Home Office licence (PPL 30/2931). In July-September 2013 semen samples were obtained from 16 sexually mature males by abdominal massage, which captures a representative sample of extragonadal semen reserves of a male[23]. Males were isolated from females for 48 hours before being massaged for the first time. Two age classes of males were selected: "Young" (1 year old, n = 7),

and "Old" (4, 5, and 7 years old, n = 9), which collectively represent a relevant dataset to detect age-dependent declines in fertility[30,76]. Males were kept sexually rested and isolated from females for 48 h. For each male, semen samples were collected every four days to allow full replenishment of extragonadal sperm reserves, between 12:00–17:00 h. In total, 157 semen samples were collected (77 from young, and 80 from old males). Immediately following collection, samples were placed in a water bath at RJF physiological temperature (41C). We then diluted 1 μl of semen into a culture medium containing 180 μl of Dulbecco's Modified Eagle's Medium (DMEM) and 20 μl of Standard Foetal Bovine Serum (FBS). We transferred 20 μl of this mixture to a microscope slide placed on a heated microscope stage set at 41C and connected to a digital video camera set at 10x magnification, and measured sperm average path velocity (VAP) using the Sperm Class Analyzer: SCA v. 3.0.3 (Microptics) set for chicken with the following specifications: capture 25 images/sec; tracking objects between 50 and 200 microns; Optics: Ph-; Chamber: Leja 10; Scale: 10x. A minimum of 200 individual sperm cells were tracked for each sample, across at least three different fields.

Immediately following VAP measurement, we used the remainder of a sample to separate sperm from SF by centrifugation at 3,500 g for 1 minute at 4C. The SF supernatant was transferred to a sterile tube for storage at −20C. This protocol was designed to purify SF from sperm cells, while reducing damage (i.e. flagellum or midpiece separation) and protein contamination from damaged cells. To conduct an independent assessment of the presence of integral sperm components in SF in samples in which whole sperm were confirmed to be absent by microscopy, we purified semen samples from four separate ejaculates of a single male using a more extensive centrifugation protocol. This included centrifugation at 600 g for 10 minutes at 4C, and, then, centrifuging the supernatant twice more at 600 g for 10 minutes and 3,000 g for 30 minutes at 4C[24]. The absence of sperm cells and flagella in this SF samples was confirmed by phase contrast microscopy. Mass spectrometry analyses of all samples were conducted using the same methodology (see below).

To determine SF proteomic profiles associated with male age and sperm VAP, we created four categories: young males with high VAP (Young Fast), young males with low VAP (Young Slow), old males with high VAP (Old Fast), and old males with low VAP (Old Slow). To do so, we measured the average VAP of each male monthly (Fig. S5a, Supporting Information), and selected the three samples with the highest VAP for each of the three young males with the highest average monthly VAP, and the three samples with the lowest VAP for each of the three young males with the lowest average monthly VAP. We performed a similar selection among the old males. Two males were represented in two different categories (Old Fast - Old Slow, and Young Fast - Young Slow, respectively) as they produced highest and lowest average VAP in two separate months. Pearson correlations within these two males that were replicated twice and represented in both the Slow and Fast categories, indicated that these males had a high overall consistency in their SF proteomic profile (correlation coefficients of 0.85 for the Old male, and 0.84 for the Young male). Therefore, we worked on a dataset of 12 samples from 10 males. This selection resulted in significant differences in VAP across the Young and Old, and the Slow and Fast categories (Fig. S5b, Supporting Information).

Protein preparation and 1-dimensional SDS-PAGE. Solubilization buffer (4% Chaps, 7 M Urea, and 2 M Thiourea) was added to each SF sample and protein concentrations were quantified using the Quick Start Bradford assay and a spectraMax M2 microplate reader. Bovine Serum Albumin (BSA) was used as a standard (scale from 0–2 mg/ml) and samples were sonicated using a bioruptor to disrupt residual protein interactions. To account for variation between ejaculates, 15 μg of protein were combined from each of three ejaculates per male and diluted appropriately to obtain equal protein concentrations across samples from each of the 12 samples. 45 μg of total protein from each sample was separated for the maximum duration possible (without running small proteins off the gel) on 4–12% Nupage Bis-Tris gels running on an XCell SureLock Mini-Cell PowerEase 200 system (Life-Technologies). Gels were then fixed in 45% methanol and 1.0% acetic acid for 1 h and immediately stained with Coomassie (0.1%w/v Coomassie, 34% methanol, 17%w/v ammonium sulfate, and 0.5% acetic acid). Gels were transferred to a gel slicer where each lane was cut horizontally into 12 slices, resulting in total of 144 samples for MS/MS analysis. To reduce spectra derived from the highly abundant SFP serum albumin and increase proteomic coverage, a narrow region centered on the predominant serum albumin band was not included amongst the fractionated gel slices. Gel slices were subjected to proteolytic digestion using an Automated Preparation Station (Perkin Elmer). Proteins were first reduced (1 hr at 37 °C in 15.4 mg dithiothreitol in 10 ml 100 mM ammonium bicarbonate solution) and alkylated (1 hr at room temperature in 102 mg iodoacetamide in 10 ml 100 mM ammonium bicarbonate solution) followed by digestion with trypsin (500 ng/μl, Promega) at 37C for 16 h.

Tandem mass spectrometry analysis. Tandem mass spectrometry (MS/MS) experiments were performed using a Dionex Ultimate 3000 RSLC nanoUPLC (Thermo Fisher Scientific Inc, Waltham, MA, USA) system and a QExactive Orbitrap mass spectrometer (Thermo Fisher Scientific Inc, Waltham, MA, USA). Separation of peptides was performed by reverse-phase chromatography at a flow rate of 300 nL/min and a Thermo Scientific reverse-phase nano Easy-spray column (Thermo Scientific PepMap C18, 2 μm particle size, 100A pore size, 75 μm i.d. × 50 cm length). Peptides were loaded onto a pre-column (Thermo Scientific PepMap 100 C18, 5 μm particle size, 100A pore size, 300 μm i.d. × 5 mm length) from the Ultimate 3000 autosampler with 0.1% formic acid for 3 minutes at a flow rate of 10 μL/min. After this period, the column valve was switched to allow elution of peptides from the pre-column onto the analytical column. Solvent A was water +0.1% formic acid and solvent B was 80% acetonitrile, 20% water +0.1% formic acid. The linear gradient employed was 2–40% B in 30 minutes. The LC eluant was sprayed into the mass spectrometer by means of an Easy-spray source (Thermo Fisher Scientific Inc.). All *m/z* values of eluting ions were measured in an Orbitrap mass analyzer, set at a resolution of 70000. Data dependent scans (Top 20) were employed to automatically isolate and generate fragment ions by higher energy

collisional dissociation (HCD) in the quadrupole mass analyser and measurement of the resulting fragment ions was performed in the Orbitrap analyser, set at a resolution of 17500. Peptide ions with charge states of 2+ and above were selected for fragmentation. Post-run, the data was processed and peak-lists generated using Protein Discoverer (version 1.3, ThermoFisher). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD00005160.

Peptide identification and protein annotation. Raw data from each tandem mass spectrometry (MS/MS) run was analyzed by X!Tandem[77] and Comet[78] against the complete *Gallus gallus* refseq protein database (June 2013 release, 32134 proteins), supplemented with characterized, but yet to be annotated, immunoglobulin sequences. A fragment ion mass tolerance of 0.40 Da and a parent ion tolerance of 10.0 PPM were used. Iodoacetamide derivative of cysteine was specified as a fixed modification, whereas oxidation of methionine was specified as a variable modification. Peptides were allowed up to two missed trypsin cleavage sites. All downstream analyses were conducted using the Trans-Proteomic Pipeline (TPP v4.7 POLAR VORTEX rev 1[79]). False Discovery Rates (FDRs) were estimated with a randomized decoy database using PeptideProphet, employing accurate mass binning model and the nonparametric negative distribution model. X!Tandem and Comet PeptideProphet results were combined using iProphet, to provide more accurate and conservative peptide identification probabilities. Peptide identifications were accepted if they could be established at greater than 95.0% iProphet probability and protein assignations were accepted if they could be established at greater than 99.0% probability. Proteins that contained identical peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. The same analysis pipeline was used to reanalyze available raw MS/MS data from fractionated domestic chicken SF samples[24]. SF proteomes were compared to the chicken blood plasma proteome[36], the exosome ExoCarta high quality protein marker dataset[47], human blood plasma[37], human sperm[41], mouse epididymal sperm[42] and human seminal plasma[35] using orthology relationships in Ensembl Genes 84 (Galgal4). Composition of protein datasets were analyzed using the DAVID Bioinformatics Resource (6.7)[80,81], including statistical analyses of functional enrichment using the *G. gallus* genome as the background dataset for comparisons.

Protein quantitation and analysis. After consideration of serum albumin depletion options and the potential impact of this highly abundant protein on various quantitative approaches, a semi-quantitative spectral counting approach was employed. Protein quantitation was conducted using the APEX Quantitative Proteomics Tool[82]. Fifty SFPs previously identified using the same MS/MS methodology (data not shown) were utilized for the training dataset based on having the highest numbers of spectral counts per protein and the highest protein identification probabilities. The 35 physicochemical properties available in the APEX tool were used for prediction of peptide detection/non-detection in the construction of a training dataset file. Protein probabilities (O_i) were computed using the Random Forest classifier algorithm trained with the dataset generated in the previous step. APEX protein abundances per male and as a combined dataset were calculated using the protXML file generated by the ProteinProphet. The same approach was used to calculate Apex scores for the domesticated *G. gallus* SF proteome[24]. Given the skew in the distribution of peptide matches for the lower coverage domestic chicken dataset (see Results), proteins exhibiting abundance differences between subspecies were identified as those with the greatest abundance rank shift between proteomes, in conjunction with quantitative differences based on their Apex estimate.

Analysis of male age and sperm swimming velocity (VAP). In order to establish SF proteomic profiles associated with male age (Old and Young) and sperm quality (Slow and Fast), we combined multivariate ordinal approaches with the analysis of individual proteins. First, we asked whether variation in the proteomic profile of individual samples could be explained by the sperm velocity and/or the male age of a sample. We therefore summarized variation in APEX-corrected protein spectral counts across the four specified categories (i.e. Old Slow, Old Fast, Young Slow, Young Fast). To do this, Principal Component Analysis, as implemented by the FactoMine R package, was used to compare the protein spectral counts across the four categories (PCA1) after normalization (i.e. equal total number of spectral counts across all males). We focused on an 'inclusive' dataset, which included all the proteins (840) identified in a minimum of 3 and up to all 12 samples (a "0" was entered when no spectra were detected for a protein in a given sample). Because each age/velocity category was comprised of 3 males, we chose a minimum of 3 males to identify situations in which a protein was expressed in all samples of a category but nowhere else. For each protein, we calculated the average spectral count for each of the four categories of males. We positioned these categories of males in a variables factor map to visualize their correlations with the two main dimensions of the PCA. To refine the interpretation of the variables factor map, we calculated for each protein over the four dimensions: (a) its coordinates on the dimension; (b) its squared cosine, which measure the quality of the projection of the protein on the dimension; (c) and its contribution (%) to the construction of the dimension. We confirmed these results using between-groups PCA (PCA2)[83] using the ade4 R package, which distinguishes variance within and across categories. This optimizes group variability, instead of between sample variability, by transforming the abundance data by averaging across the individuals belonging to each age/speed group. The protein and sample weightings, measuring distances between the proteins and samples, respectively, do not change from PCA1. Four groups were used for the PCA2 analysis, representing each of the age by velocity categories. Because PCA can be sensitive to transformation, we verified the robustness of these results with the results from Nonmetric multidimensional scaling (NMDS) in the Vegan R package, using the 840 proteins of the 'inclusive' dataset. NMDS is a non-eigenvector technique that summarizes the variance in a N-dimensional ordination space that best represents the differences in protein ranks[84]. Unlike the PCA, NMDS ranks the proteins based on their abundance, therefore eliminating potential problems (e.g., sensitivity to transformation) associated

with the use of absolute distances between their spectral counts. NMDS iteratively repositioned the proteins into the ordination space to minimize the disagreement (i.e. stress) between their original distances and their final dimensional representation. We plotted the relationship between the original protein distances and their dissimilarities in the final two-dimensional ordination space using a Shepard stress plot.

Second, we performed a third PCA (PCA3), using the 840 proteins of the 'inclusive' dataset, to characterize the 12 samples from the point of view of these proteins (i.e. the spectral count of each SFP is a variable). Thus, PCA3 uses the spectral count of individual protein as variables and summarises this variation across the 12 samples, in a way similar to a PCA of ecological communities in which the abundance of each species (i.e. SFP spectral count) is a variable and the PCA summarises variation in species abundances across a number of sites (i.e. samples). This analysis explores whether any of the main dimensions of the PCA summarizes biologically relevant differences in the proteomics profiles associated with male age and/or sperm quality, by positioning the 12 samples of individual males in the multidimensional space. The rationale of this approach is that, should individuals belonging to a given category have different loadings on a particular dimension than individuals belonging to a different category, the functional analysis of the proteins differentially expressed in these categories will shed light on the biological significance of that dimension. As in PCA1, a "0" was entered when no spectra were detected for a protein in a given sample.

Finally, we complemented the multivariate approach outlined above with an in-depth analysis on each individual protein of the inclusive dataset. To do so, we conducted a two-way ANOVA, with "male age" and "sperm quality" as fixed factors, and "spectral counts" as a response variable. We conducted the Shapiro test for each individual protein. Those proteins that were not normally distributed ($p < 0.05$) were log 10 transformed (Log_{10}(corrected spectral counts $+ 0.001$)). The model fitted in R was: model<-aov (spectral~age*velocity). If the 2-term interaction was not significant, we repeated the procedure by fitting a new model: model1<-aov (spectral~age $+$ velocity). When the ANOVA detected significant effects across the four categories, we performed pair-wise comparisons to determine which categories differed from each other. We used t-tests with a 2-tailed distribution assuming unequal variances (heteroscedasticity) to perform pairwise comparisons of the spectral counts across the 8 possible comparisons, and used the Benjamini-Hochberg formula to control the false discovery rate over the multiple comparisons. This approach identified genes that were differentially up- or down-regulated with respect to male age, sperm velocity or both. In addition, we made use of the fact that two males were represented in two different categories (Old Fast - Old Slow, and Young Fast - Young Slow, respectively). We used Pearson correlations to measure the strength of the relationship of the spectral counts between the two measurements of each male (that is, when belonging to each of the two categories) to quantify overall intra-male consistency in SF proteome, and to identify individual proteins with abundance changes associated with changes in categories that deviated from the overall trend (i.e. proteins with the largest residual).

References

1. Perry, J. C., Sirot, L. & Wigby, S. The seminal symphony: how to compose an ejaculate. *Trends Ecol Evol* **28,** 414–422, doi: 10.1016/j.tree.2013.03.005 (2013).
2. McGraw, L. A., Suarez, S. S. & Wolfner, M. F. On a matter of seminal importance: Insights & Perspectives. *BioEssays* **37,** 142–147, doi: 10.1002/bies.201400117 (2015).
3. Storey, B. T. Mammalian sperm metabolism: oxygen and sugar, friend and foe. *Int J Dev Biol* **52,** 427–437, doi: 10.1387/ijdb.072522bs (2008).
4. Neubaum, D. M. & Wolfner, M. F. Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. *Genetics* **153,** 845–857 (1999).
5. Chapman, T., Neubaum, D. M., Wolfner, M. F. & Partridge, L. The role of male accessory gland protein Acp36DE in sperm competition in *Drosophila melanogaster*. *Proc Biol Sci* **267,** 1097–1105, doi: 10.1098/rspb.2000.1114 (2000).
6. den Boer, S. P. A., Baer, B. & Boomsma, J. J. Seminal Fluid Mediates Ejaculate Competition in Social Insects. *Science* **327,** 1506–1509, doi: 10.1126/science.1184709 (2010).
7. Dean, M. D. Genetic disruption of the copulatory plug in mice leads to severely reduced fertility. *PLoS Genet* **9,** e1003185, doi: 10.1371/journal.pgen.1003185 (2013).
8. Wolfner, M. F. Tokens of love: functions and regulation of *Drosophila* male accessory gland products. *Insect Biochem Mol Biol* **27,** 179–192 (1997).
9. Herndon, L. A. & Wolfner, M. F. A *Drosophila* seminal fluid protein, Acp26Aa, stimulates egg laying in females for 1 day after mating. *Proc Natl Acad Sci USA* **92,** 10114–10118 (1995).
10. Wigby, S. & Chapman, T. Sex peptide causes mating costs in female *Drosophila melanogaster*. *Curr Biol* **15,** 316–321, doi: 10.1016/j.cub.2005.01.051 (2005).
11. Dorus, S., Evans, P. D., Wyckoff, G. J., Choi, S. S. & Lahn, B. T. Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nature Genetics* **36,** 1326–1329, doi: 10.1038/ng1471 (2004).
12. Ramm, S. A., McDonald, L., Hurst, J. L., Beynon, R. J. & Stockley, P. Comparative proteomics reveals evidence for evolutionary diversification of rodent seminal fluid and its functional significance in sperm competition. *Mol Biol Evol* **26,** 189–198, doi: 10.1093/molbev/msn237 (2009).
13. Ramm, S. A., Oliver, P. L., Ponting, C. P., Stockley, P. & Emes, R. D. Sexual selection and the adaptive evolution of mammalian ejaculate proteins. *Mol Biol Evol* **25,** 207–219, doi: 10.1093/molbev/msm242 (2008).
14. Gavrilets, S. Rapid evolution of reproductive barriers driven by sexual conflict. *Nature* **403,** 886–889, doi: 10.1038/35002564 (2000).
15. Swanson, W. J. & Vacquier, V. D. The rapid evolution of reproductive proteins. *Nature reviews. Genetics* **3,** 137–144, doi: 10.1038/nrg733 (2002).
16. Finseth, F. R., Bondra, E. & Harrison, R. G. Selective constraint dominates the evolution of genes expressed in a novel reproductive gland. *Mol Biol Evol* **31,** 3266–3281, doi: 10.1093/molbev/msu259 (2014).
17. Rowe, M. & McGraw, K. J. Carotenoids in the seminal fluid of wild birds: interspecific variation in fairy-wrens. *The Condor* **110,** 694–700 (2008).
18. Rowe, M., Czirjak, G. A., McGraw, K. J. & Giraudeau, M. Sexual ornamentation reflects antibacterial activity of ejaculates in mallards. *Biology letters* **7,** 740–742, doi: 10.1098/rsbl.2011.0276 (2011).
19. Zuk, M., Johnson, K., Thornhill, R. & Ligon, J. D. Mechanisms of Female Choice in Red Jungle Fowl. *Evolution* **44,** 477–485, doi: Doi 10.2307/2409430 (1990).

20. McDonald, G. C. & Pizzari, T. Why patterns of assortative mating are key to study sexual selection and how to measure them. *Behav Ecol Sociobiol* **70,** 209–220, doi: 10.1007/s00265-015-2041-7 (2016).

21. Pizzari, T. & Birkhead, T. R. The sexually-selected sperm hypothesis: sex-biased inheritance and sexual antagonism. *Biol Rev Camb Philos Soc* **77,** 183–209 (2002).

22. Fumihito, A. *et al.* Monophyletic origin and unique dispersal patterns of domestic fowls. *Proc Natl Acad Sci USA* **93,** 6792–6795 (1996).

23. Etches, R. *Reproduction in poultry.* (CAB International, 1996).

24. Labas, V. *et al.* Qualitative and quantitative peptidomic and proteomic approaches to phenotyping chicken semen. *J Proteomics* **112,** 313–335, doi: 10.1016/j.jprot.2014.07.024 (2015).

25. Lake, P. & Wishart, G. In *Reproductive Biology of Poultry* (eds Cunningham, F. J., Lake, P. E. & Hewitt, D.) 151–160 (Harlow, 1984).

26. Fujihara, N. Accessory reproductive fluids and organs in male domestic birds. *World Poultry Science Journal* **48,** 39–56 (1992).

27. Cornwallis, C. K. & O'Connor, E. A. Sperm: seminal fluid interactions and the adjustment of sperm quality in relation to female attractiveness. *Proc Biol Sci* **276,** 3467–3475, doi: 10.1098/rspb.2009.0807 (2009).

28. Pizzari, T., Worley, K., Burke, T. & Froman, D. P. Sperm competition dynamics: ejaculate fertilising efficiency changes differentially with time. *BMC Evol Biol* **8,** 332, doi: 10.1186/1471-2148-8-332 (2008).

29. Wishart, G. In *Biology of Breeding Poultry, Poultry Science Symposium Series* (ed Hocking, P. M.) (CAB International, 2009).

30. Cornwallis, C. K., Dean, R. & Pizzari, T. Sex-specific patterns of aging in sexual ornaments and gametes. *Am Nat* **184,** E66–78, doi: 10.1086/677385 (2014).

31. Dean, R. *et al.* Male reproductive senescence causes potential for sexual conflict over mating. *Curr Biol* **20,** 1192–1196, doi: 10.1016/j.cub.2010.04.059 (2010).

32. Noguera, J., Dean, R., Isaksson, C., Velando, A. & Pizzari, T. Age-specific oxidative status and the expression of pre- and postcopulatory sexually selected traits in male red junglefowl, *Gallus gallus*. *Ecology and Evolutin* **2,** 2155–2167 (2012).

33. Findlay, G. D., Yi, X., MacCoss, M. J. & Swanson, W. J. Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. *PLoS biology* **6** (2008).

34. Sirot, L. K. *et al.* Towards a semen proteome of the dengue vector mosquito: protein identification and potential functions. *PLoS Negl Trop Dis* **5,** e989, doi: 10.1371/journal.pntd.0000989 (2011).

35. Pilch, B. & Mann, M. Large-scale and high-confidence proteomic analysis of human seminal plasma. *Genome Biol* **7,** R40, doi: 10.1186/gb-2006-7-5-r40 (2006).

36. Ma, Y. *et al.* Towards an animal model of ovarian cancer: cataloging chicken blood proteins using combinatorial peptide ligand libraries coupled with shotgun proteomic analysis for translational research. *OMICS* **18,** 280–297, doi: 10.1089/omi.2013.0164 (2014).

37. Farrah, T. *et al.* A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol Cell Proteomics* **10,** M110 006353, doi: 10.1074/mcp.M110.006353 (2011).

38. Dorus, S., Skerget, S. & Karr, T. L. Proteomic discovery of diverse immunity molecules in mammalian spermatozoa. *Syst Biol Reprod Med* **58,** 218–228, doi: 10.3109/19396368.2012.700442 (2012).

39. Froman, D. P. & Kirby, J. D. Sperm mobility: phenotype in roosters (*Gallus domesticus*) determined by mitochondrial function. *Biol Reprod* **72,** 562–567, doi: 10.1095/biolreprod.104.035113 (2005).

40. Ahammad, M. U. *et al.* Maturational changes in the survivability and fertility of fowl sperm during their passage through the male reproductive tract. *Anim Reprod Sci* **128,** 129–136, doi: 10.1016/j.anireprosci.2011.09.010 (2011).

41. Wang, G. *et al.* In-depth proteomic analysis of the human sperm reveals complex protein compositions. *J Proteomics* **79,** 114–122, doi: 10.1016/j.jprot.2012.12.008 (2013).

42. Chauvin, T. *et al.* A Systematic analysis of a deep mouse epididymal sperm proteome. *Biology of Reproduction* **87,** 141–141, doi: 10.1095/biolreprod.112.104208 (2012).

43. Dean, M. D. *et al.* Identification of ejaculated proteins in the house mouse (*Mus domesticus*) via isotopic labeling. *BMC genomics* **12** (2011).

44. Liu, X., Jin, S. H., Liu, X. X., Wang, W. J. & Liu, F. J. Proteome profiling of the sperm maturation milieu in the rhesus monkey (*Macaca mulatta*) epididymis. *Reprod Fertil Dev*, doi: 10.1071/RD14322 (2014).

45. Skerget, S., Rosenow, M. A., Petritis, K. & Karr, T. L. Sperm proteome maturation in the mouse epididymis. *PLoS One* **10,** e0140650, doi: 10.1371/journal.pone.0140650 (2015).

46. Girouard, J., Frenette, G. & Sullivan, R. Compartmentalization of proteins in epididymosomes coordinates the association of epididymal proteins with the different functional structures of bovine spermatozoa. *Biol Reprod* **80,** 965–972, doi: 10.1095/biolreprod.108.073551 (2009).

47. Keerthikumar, S. *et al.* ExoCarta: A web-based compendium of exosomal cargo. *J Mol Biol* **428,** 688–692, doi: 10.1016/j.jmb.2015.09.019 (2016).

48. Lung, O., Kuo, L. & Wolfner, M. F. *Drosophila* males transfer antibacterial proteins from their accessory gland and ejaculatory duct to their mates. *J Insect Physiol* **47,** 617–622 (2001).

49. Otti, O., Naylor, R. A., Siva-Jothy, M. T. & Reinhardt, K. Bacteriolytic activity in the ejaculate of an insect. *Am Nat* **174,** 292–295, doi: 10.1086/600099 (2009).

50. Avila, F. W., Sirot, L. K., LaFlamme, B. A., Rubinstein, C. D. & Wolfner, M. F. Insect Seminal Fluid Proteins: Identification and Function. *Annual Review of Entomology* **56,** 21–40, doi: 10.1146/annurev-ento-120709-144823 (2011).

51. Peng, Y., Grassl, J., Millar, A. H. & Baer, B. Seminal fluid of honeybees contains multiple mechanisms to combat infections of the sexually transmitted pathogen *Nosema apis*. *Proc Biol Sci* **283,** doi: 10.1098/rspb.2015.1785 (2016).

52. Rowe, M., Czirják, G. A., Lifjeld, J. T. & Graudeau, M. Lysozyme-associated bactericidal activity in the ejaculate of a wild passerine. *Biological Journals of the Linnean Society* **109,** 92–100 (2013).

53. Westneat, D. & Rambo, T. Copulation exposes female Red-winged Blackbirds to bacteria in male semen. *Journal of Avian Biology* **31,** 1–7 (2000).

54. White, J. *et al.* Sexually transmitted bacteria affect female cloacal assemblages in a wild bird. *Ecology Letters* **13,** 1515–1524 (2010).

55. Rudolfsen, G., Serrano, J. & Folstad, I. Own, but not foreign seminal fluid inhibits sperm activation in a vertebrate with external fertilization. *Frontiers in Ecology and Evolution*, **92** (2015).

56. Løvlie, H., Gillingham, M. A., Worley, K., Pizzari, T. & Richardson, D. S. Cryptic female choice favours sperm from major histocompatibility complex-dissimilar males. *Proc Biol Sci* **280,** 20131296, doi: 10.1098/rspb.2013.1296 (2013).

57. Munro, S. Motility and fertilizing capacities of fowl sperm in the excretory ducts. *Proceedings of the Society for Experimental Biology and Medicine* **33,** 255–257 (1935).

58. Howarth, B. Fertilizing ability of cock spermatozoa from the testis, epididymis and vas deferens following intramagnal insemination. *Biology of Reproduction* **28,** 586–590 (1983).

59. Esponda, P. & Bedford, J. M. Surface of the rooster spermatozoa changes in passing through the Wolffian duct. *Journal of Experimental Zoology* **234,** 441–449 (1985).

60. Ahammad, M. U. *et al.* Maturational changes in motility, acrosomal proteolytic activity, and penetrability of the inner perivitelline layer of fowl sperm, during their passage through the male genital tract. *Theriogenology* **76,** 1100–1109, doi: 10.1016/j.theriogenology.2011.05.017 (2011).

61. Corrigan, L. *et al.* BMP-regulated exosomes from *Drosophila* male reproductive glands reprogram female behavior. *The Journal of Cell Biology* **206**, 671–688, doi: 10.1083/jcb.201401072 (2014).
62. Fardilha, M. *et al.* Protein phosphatase 1 complexes modulate sperm motility and present novel targets for male infertility. *Mol Hum Reprod* **17**, 466–477, doi: 10.1093/molehr/gar004 (2011).
63. Nguyen, T. M., Combarnous, Y., Praud, C., Duittoz, A. & Blesbois, E. Ca2+/Calmodulin-dependent protein kinase kinases (CaMKKs) effects on AMP-Activated Protein Kinase (AMPK) Regulation of Chicken Sperm Functions. *PLoS One* **11**, e0147559, doi: 10.1371/journal.pone.0147559 (2016).
64. Nguyen, T. M., Duittoz, A., Praud, C., Combarnous, Y. & Blesbois, E. Calcium channels in chicken sperm regulate motility and the acrosome reaction. *FEBS J* **283**, 1902–1920, doi: 10.1111/febs.13710 (2016).
65. Donato, R. *et al.* Functions of S100 proteins. *Curr Mol Med* **13**, 24–57 (2013).
66. Gentil, B. J. *et al.* The giant protein AHNAK is a specific target for the calcium- and zinc-binding S100B protein: potential implications for Ca2+ homeostasis regulation by S100B. *J Biol Chem* **276**, 23253–23261, doi: 10.1074/jbc.M010655200 (2001).
67. Froman, D. P. & Feltmann, A. J. Fowl (*Gallus domesticus*) sperm motility depends upon mitochondrial calcium cycling driven by extracellular sodium. *Biol Reprod* **72**, 97–101, doi: 10.1095/biolreprod.104.033209 (2005).
68. Froman, D. P., Wardell, J. C. & Feltmann, A. J. Sperm mobility: deduction of a model explaining phenotypic variation in roosters (*Gallus domesticus*). *Biol Reprod* **74**, 487–491, doi: 10.1095/biolreprod.105.046755 (2006).
69. Callaway, E. Chicken project gets off the ground. *Nature* **509**, 546, doi: 10.1038/509546a (2014).
70. Laughlin, K. In *Biology of Breeding Poultry* (ed Hocking, P. M.) 10–26 (CABI, 2009).
71. Hocking, P., Maxwell, M., Robertson, G. & Mitchell. M. Welfare assessment of modified rearing programmes for broiler breeders. *British Poultry Science* **42**, 424–432 (2001).
72. Decuypere, E. *et al.* Ethical, genetic and physiological perspectives, and suggestions for solutions. *British Poultry Science* **51**, 569–579 (2010).
73. Husby, A., Ekblom, R. & Qvarnstrom, A. Let's talk turkey: immune competence in domestic and wild fowl. *Heredity* (*Edinb*) **107**, 103–104, doi: 10.1038/hdy.2010.176 (2011).
74. Collias, N. & Collias, E. Social organization of a red junglefowl, *Gallus gallus*, population related to evolution theory. *Animal Behaviour* **51**, 1337–1354 (1996).
75. Pizzari, T., Cornwallis, C. K., Løvlie, H., Jakobsson, S. & Birkhead, T. R. Sophisticated sperm allocation in male fowl. *Nature* **426**, 70–74, doi: 10.1038/nature02004 (2003).
76. Dean, R., Nakagawa, S. & Pizzari, T. The risk and intensity of sperm ejection in female birds. *Am Nat* **178**, 343–354, doi: 10.1086/661244 (2011).
77. Craig, R. & Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466–1467, doi: 10.1093/bioinformatics/bth092 (2004).
78. Eng, J. K., Jahan, T. A. & Hoopmann, M. R. Comet: an open-source MS/MS sequence database search tool. *Proteomics* **13**, 22–24, doi: 10.1002/pmic.201200439 (2013).
79. Deutsch, E. W. *et al.* A guided tour of the Trans-Proteomic Pipeline. *Proteomics* **10**, 1150–1159, doi: 10.1002/pmic.200900375 (2010).
80. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44–57, doi: 10.1038/nprot.2008.211 (2009).
81. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1–13, doi: 10.1093/nar/gkn923 (2009).
82. Braisted, J. C. *et al.* The APEX Quantitative Proteomics Tool: generating protein quantitation estimates from LC-MS/MS proteomics results. *BMC Bioinformatics* **9**, 529, doi: 10.1186/1471-2105-9-529 (2008).
83. Krzanowski, J. Between-groups comparison of principal components. *Jounr of the American Statistical Association* **74**, 703–707 (1979).
84. Paliy, O. & Shankar, V. Application of multivariate statistical techniques in microbial ecology. *Molecular Ecology* **25**, 1032–1057, doi: 10.1111/mec.13536 (2016).

Acknowledgements

Author Contributions

T.K., T.P. A.A.-F. and S.D. conceived and designed the study. A.A.-F. and T.P. collected samples for analysis. K.B., A.A.-F., T.P. and S.D. analysed and interpreted the data. K.B., A.A.-F, T.P. and S.D. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Borziak, K. *et al.* The Seminal fluid proteome of the polyandrous Red junglefowl offers insights into the molecular basis of fertility, reproductive ageing and domestication. *Sci. Rep.* **6**, 35864; doi: 10.1038/srep35864 (2016).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.