

Adaptive Evolution in an Avian Reproductive Protein: ZP3

Jennifer D. Calkins · Diana El-Hinn ·
Willie J. Swanson

Received: 7 May 2007 / Accepted: 17 August 2007 / Published online: 2 October 2007
© Springer Science+Business Media, LLC 2007

Abstract Proteins involved in reproduction appear to be evolving adaptively across taxa. This rapid evolution is thought to be the result of forces involved in sexual selection. One of the most often suggested of these forces is sexual conflict involving sperm competition and polyspermy avoidance. Bird species offer a unique opportunity to test this hypothesis since the avian egg coat tolerates physiological polyspermy, or the penetration of multiple sperm during fertilization, without negative effects on later development. Despite this, and the extensive amount of data gathered on sexual selection in birds, there are limited studies on the patterns of evolution of avian reproductive proteins. Here we present an analysis of the pattern of evolution of Zona Pellucida 3 (ZP3), a protein present on the avian egg coat. We found that, across several galliform and a single anseriform species, ZP3 appears to be diverging by positive adaptive evolution. In an exploratory analysis of portions of the gene in *Callipepla californica* we also found evidence of a selective sweep at the putative sperm binding region of the protein. In sum, ZP3 in birds, like reproductive proteins in other species, appears to be adaptively evolving. This result suggests that polyspermy avoidance is not sufficient to explain positive Darwinian selection in reproductive proteins across taxonomic groups.

Electronic supplementary material The online version of this article (doi:10.1007/s00239-007-9034-8) contains supplementary material, which is available to authorized users.

J. D. Calkins · D. El-Hinn · W. J. Swanson
Department of Genome Science, University of Washington,
Seattle, WA 98195, USA

J. D. Calkins (✉)
Department of Genome Science, University of Washington,
Foege Building S-250, Box 355065, Seattle, WA 98195, USA
e-mail: jdc4@u.washington.edu

Clearly, the inclusion of bird species in the study of reproductive proteins across taxa promises to add greatly to the discussion of the factors driving the widespread phenomenon of adaptive evolution in reproductive proteins.

Keywords ZP3 · ZPC · Galliformes ·
Avian reproduction · Adaptive evolution ·
Reproductive proteins · Fertilization · Polyspermy

Introduction

Very few evolutionary analyses have been conducted on the molecules involved in avian reproduction despite the large amount of data available regarding avian sexual behavior. Indeed, the study of birds has given rise to many of the hypotheses invoked to explain the pattern of adaptive evolution of reproductive proteins, which is observed in taxa as diverse as *Arabidopsis*, abalone and *Drosophila* (Darwin 1871; Andersson 1994; Clark et al. 2006). It is this fact, as well as the unusual nature of the avian fertilization system, with its tolerance of physiological polyspermy, that makes the study of avian reproductive proteins especially intriguing. Here we present one of the first studies of the evolutionary pattern of an avian egg surface protein, zona pellucida 3 (ZP3).

Darwinian adaptive evolution of reproductive proteins has been detected in invertebrates (sea urchin, abalone, *Drosophila*, crickets), vertebrates (mammals), and plants (*Arabidopsis*) (Clark et al. 2006 and references therein). A variety of hypotheses have been proposed to explain this apparently widespread phenomenon. These hypotheses primarily involve forces of sexual selection such as mate choice, intrasexual competition and sexual conflict. Conflict over polyspermy, in particular, is one of most

frequently discussed hypotheses for the rapid evolution of sperm and egg surface proteins (Gavrilets 2000; Swanson and Vacquier 2002; Haygood 2004). In this scenario, competition among sperm variants for more rapid binding with egg proteins leads to a decreased ability for the egg to protect itself from polyspermy. Since, in the vast majority of species, polyspermy results in egg death, females benefit from decreasing sperm efficiency through various blocks to polyspermy (Frank 2000; Tarin and Cano 2000). In contrast, because of anisogamy, the benefit to sperm of outcompeting the sperm cells of other males outweighs any costs of polyspermy to that male. This conflict between male and female fitness trajectories is predicted to result in a cyclical antagonistic coevolutionary process (Frank 2000).

Sexual conflict over polyspermy would not be expected to drive rapid evolution of egg surface proteins in birds, however, since they are one of a few taxa that tolerate physiological polyspermy (Tarin and Cano 2000; other taxa include some species of ctenophores and urodeles). In the avian fertilization system, the penetration of the outer membranes of the egg, including the inner perivitelline layer (IPVL) and plasma membrane, by multiple sperm does not result in a selective disadvantage to the egg, and may even ensure fertilization and proper development (Wishart 1987; Birkhead and Fletcher 1998). Thus, if polyspermy avoidance is the primary force driving rapid evolution in reproductive proteins, avian IPVL proteins should not show a pattern of divergence via adaptive evolution.

Rapid evolution of reproductive proteins may be observed in birds if forces other than polyspermy avoidance are acting. Sperm storage and regular multiple mating in many bird species create a scenario in which egg and sperm proteins might face pressures analogous to sexual selection at the organism level. For example, a female that mates with several males may possess eggs whose surface protein binds more readily with the sperm of those males with more similar proteins, thus resulting in assortative mate choice at the molecular level (Eberhard 1996; Palumbi 1999). Presently, the factors dictating the ultimate selection of a sperm nucleus to fuse with the egg are unknown, but it is possible that processes early on in the fertilization cascade influence this process.

The complex stages of avian fertilization offer up a series of opportunities for sperm and egg proteins to interact, and for selection to dictate success or failure. The initial stage of fertilization involves the penetration of the IPVL, the structure analogous to the mammalian zona pellucida, by multiple sperm, most of which enter through the germinal disk (Tarin and Cano 2000). Approximately fifteen minutes after sperm penetrate the IPVL, the outer perivitelline layer (OPVL) is laid down, apparently

blocking further penetration by sperm (Bellairs et al. 1963). A proportion of sperm that have not acrosome-reacted are trapped by this outer layer (Tarin and Cano 2000). Untrapped sperm either fuse with the oolemma and form pronuclei, or are engulfed completely (Tarin and Cano 2000). Ultimately, only a single sperm pronucleus fuses with the egg pronucleus—the others migrate to the periphery and are degraded. Both the manner in which egg and sperm proteins act at these later stages of fertilization, and the actual mechanism by which the successful sperm nucleus is selected is unknown.

Five proteins homologous to zona pellucida proteins in other taxa have been identified in the avian egg. We analyzed the evolutionary patterns of one of these proteins, ZP3, the avian homologue of the mammalian egg coat protein ZP3 (Takeuchi et al. 1999). In birds, it is a primary component of the egg IPVL, along with ZP1 and ZPD; all three of which appear to be involved with sperm activation and entry (Bausek et al. 2004; Okumura et al. 2004). If the patterns of rapid evolution observed in mammalian ZP3 and other reproductive proteins arise out of protection against polyspermy threats, resulting from sperm competition, we expect to see no evidence of rapid evolution in the avian form of this protein. However, if other forces result in adaptive evolution of reproductive proteins, we would expect to see rapid evolution in avian ZP3 similar to the pattern seen in other taxa.

We tested whether ZP3 sequence evolution shows a pattern of adaptive change by comparing the nucleotide sequences among 15 galliform and a single anseriform species (Table 1). We used a maximum likelihood algorithm that uses patterns of synonymous and nonsynonymous ratios of substitution (d_N/d_S or ω) to test for rapid adaptive divergence in ZP3 sequences across taxa, and to predict the protein residues most likely to be under selection. We also obtained the sequence of two intronic/exonic regions for several individuals within a population of California quail (*Callipepla californica*) and tested for signs of recent selective pressures on this protein. The California quail is a free-ranging species likely to demonstrate normal levels of polymorphism, and is part of a larger study we are conducting on the reproductive behavior and molecular evolution of the *Callipepla* species group.

Methods

Data Collection

In addition to several avian sequences of the gene *ZP3* available on GenBank (*Gallus gallus* reference sequence NM_204389; see Table 1 for accession numbers), we

Table 1 Study species, sample source and accession number for the analyzed sequence

Species name	Sample source	Accession number
<i>Callipepla californica</i> , California quail	Blood	xxxxxxx
<i>C. gambelli</i> , Gambel's quail	Blood	xxxxxxx
<i>Oreortyx pictus</i> , mountain quail	Tissue	xxxxxxx
<i>Cairina moschata</i> , Muscovy duck	Tissue	xxxxxxx
<i>Gallus gallus</i> , chicken	GenBank	NM204389
<i>Coturnix japonica</i> , Japanese quail	GenBank	AB081506
<i>C. coturnix</i> , European quail	GenBank	AY630569
<i>Bonasa bonasia</i> , hazel grouse	GenBank	DQ004569
<i>Phasianus colchichus</i> , ring-necked pheasant	GenBank	AY630571
<i>Lagopus mutus</i> , rock ptarmigan	GenBank	DQ004568
<i>L. lagopus</i> , willow ptarmigan	GenBank	DQ004565
<i>Alectoris rufa</i> , red-legged partridge	GenBank	DQ004567
<i>Centrocercus urophasianus</i> , sage grouse	GenBank	AY630568
<i>Meleagris gallopavo</i> , turkey	GenBank	AY630570
<i>Perdix perdix</i> , grey partridge	GenBank	DQ004566
<i>Tetrao tetrix</i> , black grouse	GenBank	AY630572

sequenced *ZP3* for three galliform and one anseriform species not included in earlier analyses (California quail; Gambel's quail, *Callipepla gambelii*; mountain quail, *Oreortyx pictus*; and Muscovy duck, *Cairina moschata*). The Muscovy duck (anseriformes) represents the first nongallinaceous bird sequenced for this gene.

We used blood samples of California quail acquired from a population in Encinitas, CA between 1995 and 1998. For the diversity study we used tissue samples of mountain quail acquired from the University of Washington Burke Museum genetic resources collection (UWBM 66924 and 46830) and blood samples of Gambel's quail collected by J. Gee in Palm Desert, CA. We also used a tissue sample from a whole, frozen Muscovy duck (source: Grimaud Farms) purchased at Seattle's Finest Exotic Meats, Seattle WA. Although the sample was whole, we ensured that it was indeed duck by sequencing part of the CO1 region of mitochondrial DNA (mtDNA), using methods detailed for *ZP3* sequencing below, and compared the result against other sequences available on GenBank using BLAST. The CO1 region has been shown to be an effective species-specific identifier, or genetic barcode (Hebert et al. 2004). There is no recorded sequence of this gene for Muscovy duck but a BLAST search against all GenBank entries returned significant matches to two waterfowl species, *Anas crecca* and *Aix sponsa*, thus supporting the claim that this sample is from anseriformes.

We extracted DNA from the whole blood (California quail and Gambel's quail) and tissue samples (mountain quail and Muscovy duck) using the Puregene DNA purification kit (Gentra Systems, Minneapolis, MN). We used polymerase chain reaction (PCR) primers designed from the chicken sequence using PRIMER3 version 0.2 to amplify exons, 3' untranslated region (UTR) and 5' UTR (Rozen and Skaletsky 2000; for primer sequences, see Table 1 in the supplementary information). PCR products were diluted 5:1 and 5 μ l of the dilution was added to a cycle sequence reaction that included BigDye version 3.1. The products of the sequencing reaction were ethanol precipitated then run on an ABI 3100 sequencer.

Data Analysis

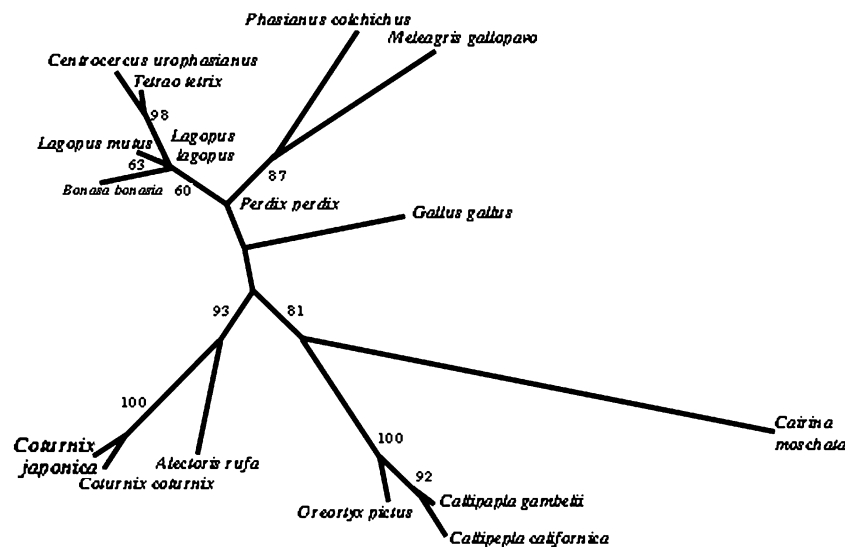
We used Sequencher 4.2 (Gene Codes, Ann Arbor, MI) to manually assemble sequences used in the divergence analysis, and visually checked the results using Se-AL version 2 (Rambaut 1996). We used a maximum likelihood species tree generated with DNAML version 3.5c (Felsenstein 1981) for analysis in CODEML (Fig. 1). This tree resembled other species trees generated for this group (Zink and Blackwell 1998; Kimball et al. 1999; Armstrong et al. 2001; Dimcheff et al. 2002; Dyke and Tuinen 2004; Crowe et al. 2006).

For the polymorphism study, we imported sequences into Pred, Phrap and polyPhred for base-calling, assembly and single nucleotide polymorphism (SNP) discovery (Nickerson et al. 1997; Ewing and Green 1998; Ewing et al. 1998). Assemblies were visualized using Consed (Gordon et al. 1998). We used PHASE to infer haplotypes from exported sequences (Stephens et al. 2001); and DnaSP version 4.0 to estimate population genetic parameters and to test for neutrality (Rozas and Rozas 1999).

(i) Divergence Among Species

The CODEML program in the PAML 3.15 package allows for maximum likelihood estimation of the d_N/d_S ratio (the ω parameter) at sites along a protein. As a d_N/d_S ratio greater than 1 indicates positive Darwinian selection, this analysis allows for the detection of particular sites or lineages subject to divergence via positive selection. Two models, a neutral model that limits ω to a value between 0 and 1 and a second model with selection that allows for an additional class of ω that can take on any positive value, are generated by CODEML, each with an associated likelihood value. These likelihood values are then compared and statistical significance is determined using a likelihood ratio test.

Fig. 1 Phylogenetic tree of studied species used in the CODEML analysis. The tree was generated using PHYLIP and a maximum-likelihood analysis. Branch lengths indicate evolutionary distance. Bootstrap values > 50 are included along corresponding branches



We analyzed the site-specific ω across ZP3 for the 16 species listed in Table 1 by comparing the following models: M1 to M2, M7 to M8 and M8a to M8. M1 is a neutral model that allows for two classes of codons, one with $\omega = 0$ and one with $\omega = 1$; M2 includes these classes along with a third class with a freely estimated ω value. M7 is a neutral model that includes a beta distribution of ω over the interval 0 to 1, while M8 incorporates this distribution with an additional class of codons assigned a freely estimated ω . Finally, M8a restricts the additional class of codons for M8 to one (Swanson et al. 2003). For all selection models, a Bayes empirical Bayes approach was used to calculate the posterior probabilities that sites with $\omega > 1.0$ are under positive selection (Yang et al. 2005). We conducted the tests without removing sequences with missing data and ran the CODEML analysis several times with different starting values of ω to ensure convergence.

(ii) Polymorphism Within the Population

To describe ZP3 polymorphism we used DnaSP version 4.0 to analyze patterns of diversity in exons and introns adjacent to residues identified as being under selection by CODEML (Rozas and Rozas 1999). We examined nucleotide diversity (π) and tested for significant deviations from neutral frequency spectrums using Tajima's D, and Fu and Li's D^* and F^* tests. Tajima's D allows for the comparison of the number of segregating sites and the allele frequencies at these sites. Excess rare alleles indicate a selective sweep whereas excess intermediate frequency variants indicate balancing selection (Tajima 1989). We used data obtained from Gambel's quail as an outgroup.

Fu and Li's D^* and F^* test, like Tajima's D, examine the pattern of allele frequency relative to the number of

segregating sites. A negative D^* and/or F^* indicates a high frequency of "singleton" or rare alleles that results from either a selective sweep or demographic effects. A positive D^* and/or F^* indicates a lack of singletons as a result of balancing selection, demographic effects, or, in some cases, a selective sweep (Fu and Li 1993).

As an exploratory analysis, we also examined values across a sliding window of the Tajima's D and the Fu and Li's D^* and F^* tests.

Results

The chicken ZP3 gene consists of 3634 nucleotides and nine exons coding for a protein 446 amino acids long (Fig. 2). We acquired sequences for 16 avian species covering 91% of the total target sequence. Four of the sequences downloaded from GenBank (red-legged partridge, *Alectoris rufa*; hazel grouse, *Bonasa bonasia*; rock ptarmigan, *Lagopus mutus*; willow ptarmigan, *Lagopus lagopus*) included a single nucleotide insertion in the ninth exon, at nucleotide position 1191 of the coding region. To ensure a conservative analysis of the data, we excluded the sequences from this base pair to the end of the gene in these species.

Divergence Among Species

Three models were compared using results from CODEML (Table 2). The neutral model M1 was not significantly different from the selection model M2 ($-2\Delta\ln L = 3.92$, $df = 2$, $p > 0.10$). However, in the more sensitive tests, the selection model M8 fit the data significantly better than the neutral models M7 or M8a (M7 vs. M8, $2\Delta\ln L = 15.6$, $df =$

2, $p < 0.01$; M8a vs. M8, $-2\Delta\ln L = 4.4$, $df = 1$, $p < 0.05$). Detailed results from all models, including putative selected sites, as determined through the Bayes empirical Bayes analysis for models M2 and M8, are presented in Table 2. Most of the putative selected sites clustered at the N and C termini of the protein in a pattern similar to that found in mammalian ZP3 (Swanson et al. 2001; Fig. 2).

Polymorphism Within Population

Using samples from a population of California quail, we examined the patterns of variation for the region around and including exons 1 and 8 of ZP3 using the Tajima's D and Fu and Li's D^* and F^* tests. For exon 1, no test showed a significant deviation from expected values, although all test statistics were positive ($N = 27$). For exon 8, all test statistics were negative and, although Tajima's D was not significant, Fu and Li's D^* and F^* values deviated significantly from those expected for a model of neutrality ($N = 19$; Table 3). A sliding window analysis of these test statistics in the regions under focus showed positive values for these test statistics in coding and noncoding regions around exon 1 (genomic sequence, 1258–1357; midpoint,

1307; Tajima's $D = 2.11$; $p < 0.05$; Fu and Li, NS) within the coding region. The same type of analysis resulted in negative values for Tajima's D for exon 8 and the surrounding noncoding regions (genomic sequence, 3181–3280; midpoint, 3230; Tajima's $D = -1.88$; $p < 0.05$); Fu and Li's D^* and F^* tests yielded negative values for a larger region around this exon (Table 4).

Discussion

The results of this study are the first to demonstrate positive adaptive evolution in an avian reproductive protein. Indeed, ZP3 is one of only two avian proteins found to be adaptively evolving (see also Cepelitis and Ellegren 2004). The divergence analysis found a pattern of adaptive change in ZP3, especially at the 5' and 3' ends of the genes. Consistent with this pattern across species, the California quail population we analyzed also indicated a selective sweep at the protein ends, as predicted by a Bayes empirical analysis.

A previous analysis of ZP3 sequence evolution in birds found no evidence of positive Darwinian selection along the gene (Berlin and Smith 2005). This earlier study included 12 species and, although the statistical tests approached significance, the authors concluded that selection was not influencing ZP3. Both phylogenetic branch length and total sample size influence the resolution of CODEML's estimation of positive selection, and developers recommend that a minimum of 16 species are included in analysis with CODEML (Anisimova et al. 2001). It is therefore not surprising that our study, which included a larger number of species and more taxonomic diversity, showed significant positive selection acting on ZP3.

Although Berlin and Smith (2005) suggest that comparisons between M7 and M8 have a high risk of false positive results, they found no evidence of false positive results with M8 and M8a model comparisons. In our analysis, comparisons between M8 and M8a significantly indicated that ZP3 is under positive selection in birds. Further support for positive selection in avian ZP3 is provided by our analysis of variation patterns in a wild California quail population. In contrast, Berlin and Smith (2005) found no evidence of selection in their genetic analysis of captive breeder stock chickens (*G. gallus*), which, given their status as captive inbred populations, are likely to have lost any imprint of reproduction-based selection.

The observation that ZP3 is rapidly evolving in birds means that aves can be added to the growing list of those taxa with adaptively evolving reproductive proteins (Swanson and Vacquier 2002). This inclusion of birds

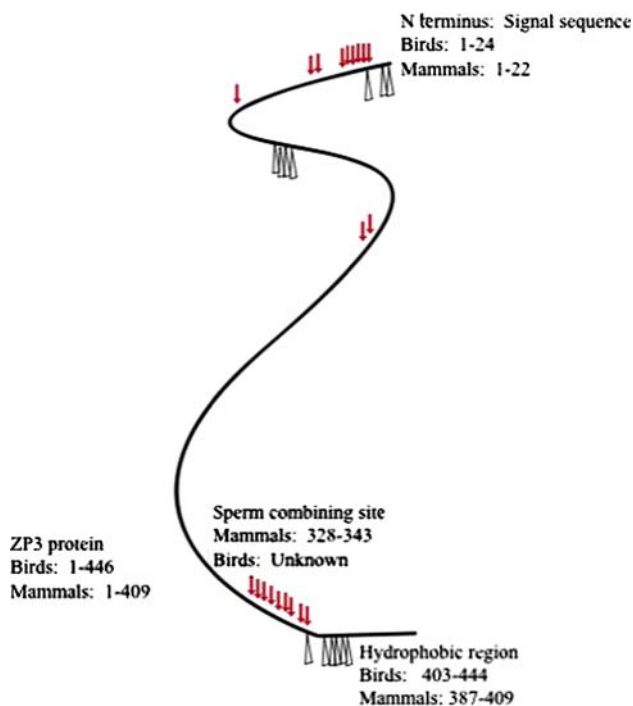


Fig. 2 Schematic of the ZP3 protein for both birds and mammals with mapped positively selected sites, as predicted by Bayes empirical analysis. Residues predicted to be under positive selection for birds are indicated by upward facing triangles below the protein. Residues predicted to be under positive selection in mammals (from Swanson et al. 2001) are represented by arrows above the protein (after Swanson et al. 2001)

Table 2 Results of CODEML analysis of ZP3 across species

Models compared	$-2\Delta\ln L$	Parameter estimates under selection model	Positively selected sites ^a
M1 & M2	3.84 (df = 2) $p > 0.10$	$p1: 0.92, \omega1: 0.05$ $p2: 0.06, \omega2: 1.00$ $p3: 0.01, \omega3: 3.45$	2, 7, 30, <u>119</u> , 120, 121, 396, 400, 405, 414, 436
M7 & M8	15.6 (df = 2) $p < 0.001$	$p0 = 0.96799$ $p = 0.29990$ $q = 3.60462$ ($p1 = 0.03201$) $\omega = 2.21505$	2, 7, 30 , <i>119</i> , 120 , 121 , 122, 381, <i>396</i> , 400 , <i>405</i> , <u>414</u> , <u>436</u>
M8a & M8	4.40 (df = 1) $p < 0.05$	$p0 = 0.96799$ $p = 0.29990$ $q = 3.60462$ ($p1 = 0.03201$) $\omega = 2.21505$	2, 7, 30 , <i>119</i> , 120 , 121 , 122, 381, <i>396</i> , 400 , <i>405</i> , <u>414</u> , <u>436</u>

Number of parameters for lnL value is shown in parentheses

^a Predicted by Bayes Empirical Bayes analysis (Yang et al. 2005); $p > 90\%$ for sites in bold, at $p > 80\%$ for sites in italics, and at $p > 70\%$ for underlined sites

Table 3 Patterns of polymorphism for the ZP3 gene in a population of *Callipepla californica*. Sites analyzed include exon and surrounding bases

Exon	N_{sites}	N_{ind}	π_{site}	k	θ_{seq}	θ_{site}	Tajima's D	Fu & Li's D	Fu & Li's F
1	762	27	0.008	6.15	5.3	0.007	1.4 NS	1.5 NS	1.3 NS
8	428	19	0.003	1.17	1.9	0.004	-1.119 NS	-2.72 $p < 0.05$	-2.60 $p < 0.05$

N_{ind} , number of individuals; k , average number of nucleotide differences (Rozas and Rozas 1999); NS, $p > 0.1$

Table 4 Sliding window values for Fu & Li's D^* and F^* for exon 8. Window values represent site within genomic sequence: range, midpoint. Exon range includes 3165–3254

Window	D^*	F^*
3131–3230, 3180	-2.85*	-2.92*
3156–3255, 3205	-2.85*	-2.92*
3181–3280, 3230	-3.19*	-3.26*
3206–3305, 3250	-2.85*	-2.92*

* $p < 0.05$

amongst the species that show positive Darwinian evolution of reproductive proteins also indicates that conflict over polyspermy is not the sole force driving the pattern of evolution found in these proteins. These results do confirm the prediction that for most forces hypothesized to drive adaptive evolution in reproductive proteins, female-expressed reproductive proteins will evolve by positive Darwinian selection in the same manner as male-expressed proteins (Swanson and Vacquier 2002). Though a pattern of positive Darwinian evolution has been found for male-expressed proteins, many fewer studies have tested this pattern in female-expressed proteins (Swanson et al. 2001; Swanson and Vacquier 2002). Our results indicate that detailed studies of avian reproductive proteins will help elucidate the evolutionary forces driving the observed pattern of positive Darwinian selection in reproductive proteins across taxa.

The Bayes empirical analysis predicted that 13 sites of the ZP3 protein are under positive selection. These putative selected sites cluster at the N and C termini and correspond roughly to the sites predicted to be under selection in the orthologous mammalian ZP3 (Swanson et al. 2001; Fig. 2). Given this result, it is especially intriguing that, in mammals, the area of ZP3 predicted to bind sperm is localized at the C terminus, i.e., where several residues are predicted to be under selection in birds. The patterning of putative selected sites in birds indicates that avian ZP3 may function in a manner akin to mammalian ZP3, with the sperm-combining site in a similar location at the C terminus. These sites warrant further examination, especially since the function of individual residues along ZP3 in birds is unknown.

Glycosylation patterns may also indicate sites for further study since, in mammalian species, residues on the ZP3 protein that serve as potential glycosylation sites have been suggested to play a role in species-specific sperm-egg interactions. Three of the residue sites indicated to be under selection in birds—30, 120 and 381—are serines in some species and are therefore possible glycosylation sites. Other species analyzed here possess alternative amino acids at these sites, which potentially changes the glycosylation status and, possibly, the manner in which this protein interacts with its ligands.

In the analysis of variation patterns within a population of California quail, we observed a skewing towards rare derived alleles in the C terminus region, exon 8 and

flanking sequence, of ZP3, where many of these putative selected sites reside. Other reproductive proteins such as lysin in abalone and semenogelin 1 in chimpanzees show a pattern of reduced polymorphism within populations, indicating a selective sweep (Lee et al. 1995; Kingan et al. 2004). A selective sweep in the ZP3 C terminus region in California quail, coupled with the observation of divergence among species, may indicate strong selective pressure, such as that for species specificity. The N terminus of the protein, exon 1 and 2, and putative selected sites therein, showed a weak signal of excess intermediate frequency alleles. This pattern has been observed in a variety of species' reproductive proteins (e.g., sea urchin bindin, *Drosophila* Acps; Swanson and Vacquier 2002 and references therein) and has been predicted for female proteins by models of sexual conflict (Gavrilets and Waxman 2002).

The tests of selection we used for this study—Fu and Li's, and Tajima's—are vulnerable to demographic effects: excessive rare alleles may result from population growth while excessive intermediate alleles may signal population subdivision (Tajima 1989; Fu and Li 1993). Further tests of demographic patterns in the population will aid in sorting selective versus demographic effects. However, the difference in patterns observed on either end of ZP3, with excessive rare alleles on the 3' end and excessive intermediate alleles on the 5' end of the same gene, is less easily resolved through a demographic explanation than by a hypothesis that includes selection on these regions.

To explain why avian ZP3 is undergoing adaptive evolution we must invoke hypotheses other than that of sexual conflict via polyspermy avoidance. These alternative hypotheses include cryptic mate choice (including species specificity and reinforcement), sperm competition, and other forms of sexual conflict and pathogen avoidance. We do not explicitly test these hypotheses here, although our results indicate that further study of avian proteins is warranted. Comparative studies that exploit the wealth of avian-related reproductive data (e.g., testis size, mating system, extra pair copulation rates) will likely prove useful in discriminating among the various hypotheses. This approach has already been conducted in mammals, with intriguing results (Dorus et al 2005; Herlyn and Zischler 2007).

Similarly, the discovery and analysis of sperm proteins that interact with ZP3 will help to test the remaining hypotheses, which predict different evolutionary trajectories for the sperm and the egg proteins relative to one another. For example, if cryptic mate choice is driving the evolution of these proteins, ZP3 is expected to evolve more rapidly than its sperm cognate, while in a situation of sperm competition, we might expect the reverse (Civetta 2003). Sexual conflict would result in either the sperm

protein or ZP3 evolving more rapidly, depending on the point in the antagonistic cycle at which the proteins are sampled; however, if a surveyed population showed a pattern of one protein driving the diversification of the other, sexual conflict would be supported over the other two sexual selection hypothesis (Hayashi et al. 2007).

We conducted a preliminary test of evolution along lineages of hybridizing species, California and Gambel's quails, and European (*Coturnix coturnix*) and Japanese quails (*C. japonica*). We did not observe the pattern of increased rates of evolution along these lineages that would be expected under reinforcement. However, this test was not particularly sensitive and expanding the investigation of ZP3 to include allopatric and sympatric populations of hybridizing species will offer more conclusive evidence for or against the importance of reinforcement on the evolution of ZP3.

The impact reinforcement could have on the evolution of ZP3 is related to the intensity of species specificity or selectivity in sperm–egg interactions in birds (Vacquier et al. 1995). In a noncompetitive study of sperm–egg interactions, Stewart et al. (2004), found little evidence for differential penetration rates among con- and heterospecific crosses of egg and sperm, thus suggesting that species specificity is unlikely to drive the evolution of egg coat proteins in birds (Stewart et al. 2004; Birkhead and Brillard 2007). However, further studies examining the competition between different species' sperm to fertilize an egg are necessary to truly determine the importance of species-specific interactions in ZP3 evolution (Vacquier et al. 1995).

The influence of these sexual selection forces—cryptic mate choice, sperm competition and sexual conflict—on the evolution of ZP3 is affected by the amount of sperm that the egg itself encounters. This number is limited by the majority of the sperm in a male's ejaculate being ejected by the female bird prior to storage in the sperm storage tubules (SSTs). Ultimately, only 1–2% of more than a million sperm typically released during a single ejaculation are stored in the female's SSTs (Birkhead and Brillard 2007). Despite this limitation and the further reduction of representative sperm exiting the SSTs, more than 60 sperm can still reach and penetrate the egg at the level of the IPVL (Birkhead and Fletcher 1994). The exposure of the egg to more than a single sperm sets up the opportunity for some form of competition, to be resolved through cryptic choice by the egg or competition by the sperm. Therefore, some level of sexual selection acting on ZP3 on the avian egg coat is quite possible.

Ultimately, the discovery that avian reproductive proteins do evolve by positive Darwinian selection opens a large field of potential studies aimed at elucidating the specific causes of this evolutionary pattern. These studies

include functional analyses of ZP3, determination of the interacting sperm proteins and their evolutionary trajectories, in-depth population analyses of these proteins in hybrid zones, and determination of the avian mechanism of fertilization and final selection of a single sperm nucleus. Expanding the analysis of ZP3 across a broad taxonomic range of birds in light of the extensive knowledge about avian reproduction at the behavioral level will also help reveal the evolutionary relationship between forces occurring at the level of the organism and patterns occurring at the level of the molecule. Clearly, increasing our knowledge of avian reproduction at the cellular level will be invaluable to the general discussion of the forces that drive positive Darwinian evolution in reproductive proteins.

Acknowledgments We thank the Swanson lab and five anonymous reviewers for helpful comments on the manuscript. We also thank the Burke Museum's genetic resource and Dr. Jennifer Gee for access to samples. Blood samples for *Callipepla californica* were collected under the California Department of Fish and Game permit No. 6942 and the institutional animal care and use committee of the University of California, Irvine (permit no. 95-1664). JDC and WJS are supported by NIH grant HD42563 and NSF DEB-041022.

References

- Andersson M (1994) Sexual selection. Princeton University Press, Princeton, NJ
- Anisimova M, Bielawski JP, Yang Z (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol* 18: 1585–1592
- Armstrong MH, Braun EL, Kimball RT (2001) Phylogenetic utility of avian ovomucoid intron G: a comparison of nuclear and mitochondrial phylogenies in galliformes. *Auk* 118: 799–804
- Bausek N, Ruckenbauer HH, Pfeifer S, Schneider WJ, Wohlrab F (2004) Interaction of sperm with purified native chicken ZP1 and ZPC proteins. *Biol Reprod* 71: 684–690
- Bellairs R, Harkness M, Harkness RD (1963) The vitelline membrane of the hen's egg: a chemical and electron microscopical study. *J Ultrastruct Res* 8: 339–359
- Berlin S, Smith NGC (2005) Testing for adaptive evolution of the female reproductive protein ZPC in mammals, birds and fishes reveals problems with the M7-M8 likelihood ratio test. *BMC Evol Biol* 5: 65
- Birkhead TR, Fletcher F (1994) Numbers of spermatozoa attached to and penetrating perivitelline layers of Japanese Quail eggs. *Auk* 111: 997–1000
- Birkhead TR, Fletcher F (1998) Sperm transport in the reproductive tract of female zebra finches (*Taeniopygia guttata*). *J Reprod Fertil* 114: 141–145
- Birkhead TR, Brillard J-P (2007) Reproductive isolation in birds, postcopulatory, prezygotic barriers. *Trends Ecol Evol* 22: 266–272
- Ceplitis H, Ellegren H (2004) Adaptive molecular evolution of HINTW, a female specific gene in birds. *Mol Biol Evol* 24: 249–254
- Civetta A (2003) Shall we dance or shall we fight? Using DNA sequence data to untangle controversies surrounding sexual selection. *Genome* 46: 925–929
- Clark NL, Aargaard JE, Swanson WJ (2006) Evolution of reproductive proteins from animals and plants. *Reproduction* 131: 11–22
- Crowe TM, Rauri CKB, Bloomer P, Mandiwana TG, Hedderston TAJ, Randi E, Pereira SL, Wakeling J (2006) Phylogenetics, biogeography and classification of, and character evolution in, gamebirds (Aves: Galliformes): effects of character exclusion, data partitioning and missing data. *Cladistics* 22: 495–532
- Darwin C (1871) The descent of man, and selection in relation to sex. Murray, London
- Dimcheff DE, Drovetski SV, Mindell DP (2002) Phylogeny of Tetraoninae and other galliform birds using mitochondrial 12S and ND2 genes. *Mol Phylogenet Evol* 24: 203–215
- Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT (2004) Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nat Genet* 36: 1326–1329
- Dyke GJ, van Tuinen M (2004) The evolutionary radiation of modern birds (Neornithes): reconciling molecules, morphology, and the fossil record. *Zool J Linn Soc* 141: 153–177
- Eberhard W (1996) Female control. Princeton University Press, Princeton, NJ
- Ewing B, Green P (1998) Basecalling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8: 186–194
- Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8: 175–185
- Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 17: 368–376
- Frank SA (2000) Sperm competition and female avoidance of polyspermy mediated by sperm-egg biochemistry. *Evol Ecol Res* 2: 613–625
- Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. *Genetics* 133: 693–709
- Gavrilets S (2000) Rapid evolution of reproductive barriers driven by sexual conflict. *Nature* 403: 886–889
- Gavrilets S, Waxman D (2002) Sympatric speciation by sexual conflict. *Proc Natl Acad Sci USA* 99: 10533–10538
- Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. *Genome Res* 8: 195–202
- Hayashi T, Vose M, Gavrilets S (2007) Genetic differentiation by sexual conflict. *Evolution* 61: 516–529
- Haygood R (2004) Sexual conflict and protein polymorphism. *Evolution* 58: 1414–1423
- Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. *PLoS Biol* 2: 1657–1663 (DOI:10.1371/journal.pbio.0020312)
- Herlyn H, Zischler H (2007) Sequence evolution of the sperm ligand zonadhesin correlates negatively with body weight dimorphism in primates. *Evolution* 61: 289–298
- Kimball RT, Braun EL, Zwartjes PW, Crowe TM, Ligon JD (1999) A molecular phylogeny of the pheasants and partridges suggests that these lineages are not monophyletic. *Mol Phylogenet Evol* 11: 38–54
- Kingan SB, Tator M, Rand DM (2004) Reduced polymorphism in the chimpanzee semen coagulating protein, semenogelin I. *J Mol Evol* 57: 159–169
- Lee Y-H, Ota T, Vacquier VD (1995) Positive selection is a general phenomenon in the evolution of abalone sperm lysin. *Mol Biol Evol* 12:231–238
- Nickerson DA, Tobe VO, Taylor SL (1997) Polyphred: Automating the detection and genotyping of single nucleotide substitutions using fluorescence based resequencing. *Nucleic Acids Res* 25: 2745–2751
- Okumura H, Kohno Y, Iwata Y, Mori H, Aoiki N, Sato C, Kitajima K, Nadano D, Matsuda T (2004) A newly identified zona pellucida

- glycoprotein, ZPD, and dimeric ZP1 of chicken egg envelope are involved in sperm-egg interaction. *Biochem J* 384: 191–199
- Palumbi SR (1999) All males are not created equal: Fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proc Natl Acad Sci USA* 96: 12632–12637
- Rambaut A (1996) *Se-Al: Sequence alignment editor*. Oxford. <http://www.evolve.zoo.ox.ac.uk>
- Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15: 174–175
- Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132: 365–386
- Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68: 978–989
- Stewart SG, Bausek N, Wohlrab F, Schneider WJ, Horraks AJ, Wishart GJ (2004) Species specificity in avian sperm: perivitelline interaction. *Comp Biochem Physiol A Mol Integr Physiol* 137: 657–663
- Swanson WJ, Vacquier VD (2002) Reproductive protein evolution. *Annu Rev Ecol Syst* 33: 161–179
- Swanson WJ, Yang Z, Wolfner MF, Aquadro CF (2001) Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc Natl Acad Sci USA* 98: 2509–2514
- Swanson WJ, Nielsen R, Yang Z (2003) Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol* 20: 18–20
- Takeuchi Y, Nishimura K, Aoki N, Adachi T, Sato C, Kitajima K, Matsuda T (1999) A 42-kDa glycoprotein from chicken egg-envelope, an avian homolog of the ZPC family glycoproteins in mammalian zona pellucida. *Eur J Biochem* 260: 736–742
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595
- Tarin JJ, Cano A (Eds) (2000) *Fertilization in protozoa and metazoan animals*. Springer, Berlin
- Vacquier VD, Swanson WJ, Hellberg ME (1995) What have we learned about sea urchin sperm binding? *Dev Growth Differ* 37: 1–10
- Wishart GJ (1987) Regulation of the length of the fertile period in the domestic fowl by numbers of oviductal spermatozoa. *J Reprod Fertil* 80: 493–498
- Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22: 1107–1118
- Zink RM, Blackwell RC (1998) Molecular systematics of the Scaled Quail complex (genus *Callipepla*). *Auk* 115: 394–403