ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

SHORT COMMUNICATION doi:10.1111/j.1365-2052.2009.01974.x

Variation in chicken populations may affect the enzymatic activity of lysozyme

T. Downing*,[1], C. O'Farrelly[†], A. K. Bhuiyan[‡], P. Silva[§], A. N. Naqvi[¶], R. Sanfo, R.-S. Sow[††], B. Podisi[‡‡], O. Hanotte[§§] and D. G. Bradley***

*Smurfit Institute of Genetics, Trinity College, University of Dublin, Dublin, Ireland. [†]School of Biochemistry and Immunology, Trinity College, University of Dublin, Dublin, Ireland. ‡Department of Animal Breeding and Genetics, Bangladesh Agricultural University, Mymensingh, Bangladesh. [§]Department of Animal Science, University of Peradeniya, Peradeniya, Sri Lanka. [¶]PARC, Animal Sciences Division, Islamabad, Pakistan. **INERA, Agriculture and Environment, Ouagadougou, Burkina Faso. [††]Institut Sénégalais de Recherches Agricoles, Dakar, Senegal. [‡‡]Department of Agricultural Research, CTA, Gaborone, Botswana. [§§]School of Biology, University of Nottingham, Nottingham, United Kingdom NG7 2RD and International Livestock Research Institute (ILRI), PO Box 30709, Nairobi 00100, Kenya

Summary

The chicken *lysozyme* gene encodes a hydrolase that has a key role in defence, especially *in ovo*. This gene was resequenced in global chicken populations [red, grey, Ceylon and green jungle fowl (JF)] and related bird species. Networks, summary statistics and tests of neutrality indicate that although there is extensive variation at the gene, little is present at coding sites, with the exception of one non-synonymous site. This segregating site and a further fixed non-synonymous change between red JF and domestic chicken populations are spatially close to the catalytic sites of the enzyme and so might affect its activity.

Keywords chicken, domestication, jungle fowl, lysozyme, population genetics.

Lysozyme is an important part of innate chicken defence systems: it hydrolyses peptidoglycan and chitodextrin, both of which are components of gram positive bacterial cell membranes (Holler *et al.* 1975a,b). The enzyme reaches its highest concentration in egg-white, where innate immune mechanisms are vital because the adaptive immune system is not fully developed (Sippel *et al.* 1978; Ask *et al.* 2007). Certain immune genes that determine susceptibility to infection have been shown to be subject to selective forces in the chicken, such as *Mx* (Li *et al.* 2006; Hou *et al.* 2007; Berlin *et al.* 2008), *MHC-B* (Worley *et al.* 2008), *IL1B* (Downing *et al.* 2009a) and *IL-4Rα* (Downing *et al.* 2009b). Therefore it is possible that *lysozyme* has been subject to similar evolutionary pressure during its evolution. Here, we explore diversity present at the gene by resequencing it in chicken populations and related species, and find one non-synonymous substitution segregating at an intermediate frequency. Tests indicate that this site and one other non-synonymous change fixed between red jungle fowl (JF) and chicken are spatially close to the key catalytic sites.

Village chicken samples from three Asian (Bangladesh, Pakistan and Sri Lanka) and four African (Burkina Faso, Botswana, Kenya and Senegal) populations numbering 70 in total were acquired from the International Livestock Research Institute (Kenya). Samples from red, grey and Ceylon JF from Wallslough Farm (Ireland); green JF, bamboo partridge and grey francolin from the Californian Academy of Sciences; and 20 commercial broilers from Manor Farms (Ireland) were also surveyed. More details of the samples are listed elsewhere (Downing *et al.* 2009b). The DNA was isolated from the samples using a phenol-chloroform extraction following a proteinase K digestion.

The UCSC (http://genome.ucsc.edu), Ensembl (http://www.ensembl.org) and GenBank MAP VIEW (http://www.ncbi.nlm.nih.gov/projects/mapview/) browsers were used to map the gene structure. PCR primer sequences (Table S1) were designed using PRIMER3 (http://frodo.wi.mit.edu) and were created by VHBio (http://www.vhbio.com). Five amplicons covering 3726 bp at *lysozyme* were successfully amplified by PCR (Table S2) for the 96 samples (Fig. S1). Resequencing, sequence assembly and haplotype reconstruction followed the protocols in Appendix S1.

Address for correspondence

D. G. Bradley, Smurfit Institute of Genetics, Trinity College, University of Dublin, Dublin, Ireland.
E-mail: dbradley@tcd.ie

[1]Present address: Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom CB10 1SA

Accepted for publication 14 August 2009

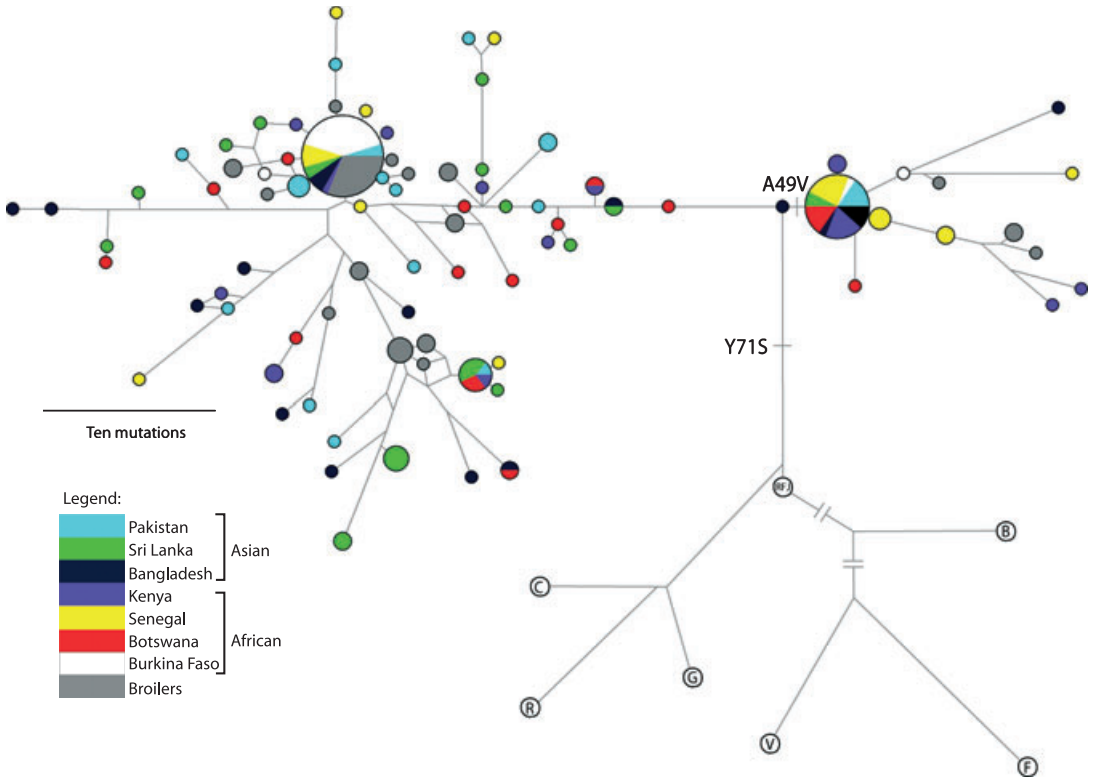

Figure 1 Median-joining phylogenetic network of chicken and outgroup haplotypes. Populations are denoted in the legend. Branch lengths are proportional to the number of mutational differences between haplotypes. The outgroup samples are represented by the colourless nodes; their branch lengths are considerably reduced to show the details of the chicken population network. V represents the green JF sequences; F the grey francolin; B the bamboo partridge; G the grey JF; C the Ceylon JF; R the red JF; and RJF the red JF genome sequence.

Of 59 SNPs discovered among domestic chicken sequences at this gene, only one was a non-singleton coding SNP (cSNP). This cSNP was a non-synonymous substitution (p.Ala49Val) at base 1398 in exon 2 and defined the two most numerous haplotypes of 84 observed in a median-joining network constructed using Network 4.510 (http://www.fluxus-technology.com) (Fig. 1). In this network, the red JF genome sequence was the most closely related JF genotype to the chicken samples. When only cSNPs were used to construct a network (Fig. S2), this substitution alone separated the two principal alleles that are present in all eight chicken populations. Substitution p.Tyr71Ser at site 1464 was the solitary coding sequence distinguishing the red JF genome sequence and the chicken haplotypes. The grey, red and Ceylon JF were separated from the reference genome sequence by a single synonymous SNP at base 1699. Segregating cSNPs were of particular interest because of the extensive coding sequence conservation.

Analysis of variation at different levels of population structure with Arlequin (Schneider *et al.* 2000) using AMOVA (Excoffier *et al.* 1992) show this high allelic diversity (shown in Fig. 1) was partitioned within the populations (90.588%, $P < 10^{-5}$) and among populations (9.412%, $P < 10^{-5}$), but not among the continents.

Significantly high allelic variation was observed at the gene: this was supported by coalescent simulations incorporating recombination using DnaSP 4.0 (Rozas & Rozas 1999; Rozas *et al.* 2003) that evaluated the degree of deviation from neutrality of the observed data for a number of statistics, including the haplotype diversity (Hd = 0.923, P = 0.006; Depaulis & Veuille 1998) and Fu's F_S (−34.48, P = 0.008; Table S3; Fu 1997). A relative deficit of singletons shown by the positive Fu and Li's D (1.42, P = 0.022; Fu & Li 1993) and F (1.34, P = 0.016) suggest that such alleles are not the cause of the elevated allele variation. A positive Tajima's D (0.618; Tajima 1989) and negative Fay and Wu's H (−14.08, P = 0.002; Fay & Wu 2000) supported a trend of the elevated variation around the two most numerous haplotypes in the network diagrams (Fig. 1), which appeared to centre on substitution p.Ala49Val. Although a significantly high minimum number of recombination events (R_M = 25, $P < 0.001$) (Hudson & Kaplan 1985) suggested that some new haplotypes created by recombination were preserved (Tables S4 & S5), these recombinants were not maintained at non-synonymous sites (Fig. S2).

The ratio of the relative rate of non-synonymous mutations (d_N) to the relative rate of synonymous mutations (d_S)

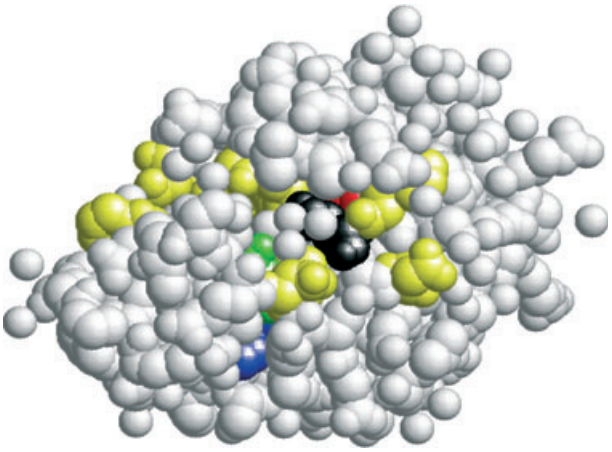

Figure 2 The three-dimensional structure of chicken lysozyme (Protein Data Bank http://www.rcsb.org/pdb/explore/explore.do?structureId=3B6L; Michaux *et al.* 2008), displayed using RasMol. The positions of Ala49 (red) and Tyr71 (blue) are shown relative to the catalytic sites (Glu53, black; Asp70, green) and the catalytic cleft in yellow (Phe52, Asn55, Asn62, Gln75, Asn77, Trp80, Ile116, Asp119, Ala125 and Trp126). Ala49 and Asp53 are in helix 2. Tyr71 is located in sheet strand 2.

in the protein-coding portion of the gene was calculated as d_N/d_S (ω) for models using the codeml implementation of the PAML 3.15 package (Yang 1997). Departures from neutrality occur when $\omega > 1$ or $\omega < 1$, which indicate an excess of non-synonymous or synonymous changes respectively (Yang 2002) (Figure S3). Site specific models estimate ω for each site across the whole coding sequence for a neutral model (M7, $0 \geq \omega \leq 1$) and a variable model (M8) that allows for $\omega > 1$ as well as $0 \geq \omega \leq 1$ (Yang 1997). A likelihood ratio test performed between these models for the seven resequenced species shows that M8 was significantly more likely than M7 according to a chi-square distribution ($P = 5 \times 10^{-4}$, Table S6). Using a random sites Bayes empirical Bayes model, a significantly high posterior probability of $\omega > 1$ ($P > 0.95$) indicated positive selection at candidate sites 57, 70, 72 and 96 (Table S7; Nielsen & Yang 1998; Yang *et al.* 2005).

The positions of the catalytic sites (53 and 70), catalytic cleft and polymorphic sites (49 and 71) clustered closely in a three-dimensional model of chicken lysozyme displayed in RASMOL 2.7.4.2 (http://www.openrasmol.org/software/rasmol/; Fig. 2). Using the Euclidean distance between the α-carbon atoms in the protein database file, the length between sites 49 and 53 was 5.58 Å, which was substantially smaller than average; only 4.3% of site pairs were closer. Sites 70 and 71 were separated by 3.78 Å; only 0.9% of pairs and 10.9% of adjacent pairs were closer. The proximity of these amino acids to the catalytic sites was likely to be of significance because single site changes and interactions can alter the activity and stability of this enzyme (Klein-Seetharaman *et al.* 2002; Zhou *et al.* 2007), implying that mutations at sites 49 and 71 spatially affected the catalytic sites. Certain sets of amino acid substitutions in the lysozyme peptide have been shown to be compensatory, even though they were located at the core of the molecule (Wilson *et al.* 1992); this could be possible for variants at sites 49 and 71.

A T-Coffee (Notredame *et al.* 2000) alignment of lysozyme protein sequences for human, zebra finch, turkey and birds from the family Phasianidae [chicken (*Gallus gallus*); copper (*Syrmaticus soemmerringii*), kalij (*Lophura leucomelanos*) and golden pheasant (*Chrysolophus pictus*); bobwhite (*Colinus virginianus*) and Japanese quail (*Coturnix japonica*)] showed that all Phasianidae had a different amino acid (Ala) at site 49 compared with the zebra finch and human (Leu) (Fig. S4). Site 71 was conserved in all samples (Tyr), except the domestic chicken (Ser), indicating that while site 49 has evolved in the avian lineage, site 71 appeared to be altered in chicken alone.

The considerable conservation at coding regions suggested p.Ala49Val and p.Tyr71Ser were likely to have function relevance (Table S8). p.Ala49Val was segregating in all eight domestic chicken populations, an indication that it was actively being maintained at high frequencies in each. Site 71 was different in chickens (Ser) compared with the sequenced red JF genome and all other birds (Tyr), and adjacent sites 70 and 72 appear to be subject to positive selection between species (Table S7). Diversity among the chicken populations was distributed around substitution 49; this may be a result of latent admixture following domestication and ongoing selective processes stimulated by novel pathogenic challenges.

The signature of high allelic diversity observed here is reminiscent of the previous work on variation at chicken mtDNA (Liu *et al.* 2006), *MHC-B* (Worley *et al.* 2008; O'Neill *et al.* 2009), *Mx* (Berlin *et al.* 2008), *IL1B* (Downing *et al.* 2009a) and *IL-4Rα* (Downing *et al.* 2009b), suggesting that this may be the result of the complex population history of the chicken during domestication. Although the main source of chicken genetic variation is the red JF (Fumihito *et al.* 1994; International Chicken Genome Sequencing Consortium 2004), multiple domestications (Fumihito *et al.* 1996; Liu *et al.* 2006) and genetic introgressions of other JF into chicken populations (Nishibori *et al.* 2005; Eriksson *et al.* 2008; Silva *et al.* 2008) suggest that diverse alleles may have been introduced during domestication. The impact of human trade, migration and selection for novel characteristics is likely to have created a widespread intermixing of red JF subspecies with chicken, the result of which may be the high haplotype diversity observed in many studies (Granevitze *et al.* 2007; Oka *et al.* 2007; Bao *et al.* 2008; Kanginakudru *et al.* 2008; Muchadeyi *et al.* 2008; Berthouly *et al.* 2009). Though the elevated allele variation may be a relic of chicken domestication, this does not exclude the proposal of pathogen-driven selective pressure, which might explain the continued persistence of the divergent alleles in modern chicken populations.

Acknowledgements

This work is supported by Government of Ireland Department of Agriculture FIRM grant 04/R+D/D/295. We offer grateful thanks to Sarah Connell, Andrew T. Lloyd (Trinity College, University of Dublin) and David J. Lynn (Simon Fraser University, Canada) for help with DNA extractions and useful comments on the paper. We thank the Department of Ornithology and Mammalogy, Californian Academy of Sciences (San Francisco, USA), Manor Farms (Co. Monaghan, Ireland) and Donal Campion, Wallslough Farm (Co. Kilkenny, Ireland) for bird samples, as well as Kieran Meade and Ronan Shaughnessy for help with sample collection (Trinity College, University of Dublin).

References

Ask B., van der Waaij E.H., Glass E.J. *et al.* (2007) Modelling immunocompetence development and immunoresponsiveness to challenge in chicks. *Poultry Science* **86,** 1336–50.

Bao W., Chen G., Li B. *et al.* (2008) Analysis of genetic diversity and phylogenetic relationships among red jungle fowls and Chinese domestic fowls. *Science in China Series C: Life Sciences* **51,** 560–8.

Berlin S., Qu L., Li X. *et al.* (2008) Positive diversifying selection in avian Mx genes. *Immunogenetics* **60,** 689–97.

Berthouly C., Leroy G., Van T.N. *et al.* (2009) Genetic analysis of local Vietnamese chickens provides evidence of gene flow from wild to domestic populations. *BMC Genetics* **10,** 1.

Depaulis F. & Veuille M. (1998) Neutrality tests based on the distribution of haplotypes under an infinite-site model. *Molecular Biology and Evolution* **15,** 1788–90.

Downing T., Lynn D.J., Connell S. *et al.* (2009a) Contrasting evolution of diversity at two disease-associated chicken genes. *Immunogenetics* **61,** 303–14.

Downing T., Lynn D.J., Connell S. *et al.* (2009b) Bioinformatic detection and population-level validation of selection at the chicken *interleukin 4 receptor alpha chain* gene. *BMC Evolutionary Biology* **9,** 136.

Eriksson J., Larson G., Gunnarsson U. *et al.* (2008) Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genetics* **4,** e1000010.

Excoffier L., Smouse P.E. & Quattro J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131,** 479–91.

Fay J.C. & Wu C.I. (2000) Hitchhiking under positive Darwinian selection. *Genetics* **155,** 1405–13.

Fu Y.X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147,** 915–25.

Fu Y.X. & Li W.H. (1993) Statistical tests of neutrality of mutations. *Genetics* **133,** 693–709.

Fumihito A., Miyake T., Takada M. *et al.* (1994) Monophyletic origin and unique dispersal patterns of domestic fowls. *Proceedings of the National Academy of Sciences of the United States of America* **93,** 6792–5.

Fumihito A., Miyake T., Sumi S. *et al.* (1996) One subspecies of the red junglefowl (*Gallus gallus gallus*) suffices as the matriarchic ancestor of all domestic breeds. *Proceedings of the National Academy of Sciences of the United States of America* **91,** 12505–9.

Granevitze Z., Hillel J., Chen G.H. *et al.* (2007) Genetic diversity within chicken populations from different continents and management histories. *Animal Genetics* **38,** 576–83.

Holler E., Rupley J. & Hess G. (1975a) Productive and unproductive lysozyme-chitosaccaride complexes. Equilibrium measurements. *Biochemistry* **14,** 1088–94.

Holler E., Rupley J. & Hess G. (1975b) Productive and unproductive lysozyme-chitosaccaride complexes. Kinetic investigations. *Biochemistry* **14,** 2377–85.

Hou Z.C., Xu G.Y., Su Z. *et al.* (2007) Purifying selection and positive selection on the myxovirus resistance gene in mammals and chickens. *Gene* **396,** 188–95.

Hudson R.R. & Kaplan N.L. (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111,** 147–64.

International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432,** 695–716.

Kanginakudru S., Metta M., Jakati R.D. & Nagaraju J. (2008) Genetic evidence from Indian red jungle fowl corroborates multiple domestication of modern day chicken. *BMC Evolutionary Biology* **8,** 174.

Klein-Seetharaman J., Oikawa M., Grimshaw S.B. *et al.* (2002) Long-range interactions within a nonnative protein. *Science* **295,** 1719–22.

Li X.Y., Qu L.J., Yao J.F. *et al.* (2006) Skewed allele frequencies of an *Mx* gene mutation with potential resistance to avian influenza virus in different chicken populations. *Poultry Science* **85,** 1327–9.

Liu Y.P., Wu G.S., Yao Y.G. *et al.* (2006) Multiple maternal origins of chickens: out of the Asian jungles. *Molecular Phylogenetics and Evolution* **38,** 12–9.

Michaux C., Pouyez J., Wouters J. & Privé G.G. (2008) Protecting role of cosolvents in protein denaturation by SDS: a structural study. *BMC Structural Biology* **8,** 29.

Muchadeyi F.C., Eding H., Simianer H. *et al.* (2008) Mitochondrial DNA D-loop sequences suggest a Southeast Asian and Indian origin of Zimbabwean village chickens. *Animal Genetics* **39,** 615–22.

Nielsen R. & Yang Z. (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148,** 929–36.

Nishibori M., Shimogiri T., Hayashi T. *et al.* (2005) Molecular evidence for hybridization of species in the genus *Gallus* except for *Gallus varius*. *Animal Genetics* **36,** 367–75.

Notredame C., Higgins D.G. & Heringa J. (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* **302,** 205–17.

O'Neill A.M., Livant E.J. & Ewald S.J. (2009) The chicken *BF1* (classical MHC class I) gene shows evidence of selection for diversity in expression and in promoter and signal peptide regions. *Immunogenetics* **61,** 289–302.

Oka T., Ino Y., Nomura K. *et al.* (2007) Analysis of mtDNA sequences shows Japanese native chickens have multiple origins. *Animal Genetics* **38,** 287–93.

Rozas J. & Rozas R. (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15,** 174–5.

Rozas J., Sanchez-DelBarrio J.C., Messeguer X. *et al.* (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19,** 2496–7.

Schneider S., Roessli D. & Excoffier L. (2000) *Arlequin: A Software for Population Genetics Data Analysis. Ver 2.000*. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva, Switzerland.

Silva P., Guan X., Ho-Shing O. *et al.* (2008) Mitochondrial DNA-based analysis of genetic variation and relatedness among Sri Lankan indigenous chickens and the Ceylon junglefowl (*Gallus lafayeti*). *Animal Genetics* **40,** 1–9.

Sippel A.E., Land H., Lindenmaier W. *et al.* (1978) Cloning of chicken lysozyme structural gene sequences synthesized *in vitro*. *Nucleic Acids Research* **5,** 3275–94.

Tajima F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123,** 585–95.

Wilson K.P., Malcolm B.A. & Matthews B.W. (1992) Structural and thermodynamic analysis of compensating mutations within the core of chicken egg white lysozyme. *Journal of Biological Chemistry* **267,** 10842–9.

Worley K., Gillingham M., Jensen P. *et al.* (2008) Single locus typing of MHC class I and class II B loci in a population of red jungle fowl. *Immunogenetics* **60,** 233–47.

Yang Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* **13,** 555–6.

Yang Z. (2002) Inference of selection from multiple species alignments. *Current Opinion in Genetics and Development* **12,** 688–94.

Yang Z., Wong W.S. & Nielsen R. (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology & Evolution* **22,** 1107–18.

Zhou R., Eleftheriou M., Royyuru A.K. *et al.* (2007) Destruction of long-range interactions by a single mutation in lysozyme. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 5824–9.

Supporting Information

Additional supporting information may be found in the online version of this article.

Figure S1 *Lysozyme* gene structure.

Figure S2 Median-joining network of chicken population haplotypes for coding SNPs only.

Figure S3 Neighbour-joining phylogeny constructed using codeml.

Figure S4 A multiple sequence alignment of the active portion of protein-coding sequences.

Table S1 Sets of primer pair sequences and their associated optimal PCR parameters.

Table S2 PCR cycle programme used for each pair of primers.

Table S3 Gene data, summary statistics and tests of neutrality from DnaSP.

Table S4 Genotypes at SNP sites polymorphic in chicken samples.

Table S5 Recombination according the percentage GC content, Hudson's R and R_M and Kelly's $Z_n s$ per kb from DnaSP.

Table S6 Generated PAML parameters used and output for significant test results for the major coding allele.

Table S7 Sites potentially under positive selection according to BEB analysis of PAML M8 results.

Table S8 Predicted functional impacts of different non-synonymous SNPs on the protein product.

Appendix S1 Additional Methodology.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.