Inferring allele frequency trajectories from ancient DNA indicates that selection on a chicken gene coincided with changes in medieval husbandry practices

Liisa Loog[1,2†], Mark G. Thomas[3], Ross Barnett[1], Richard Allen[1], Naomi Sykes[4], Ptolemaios D. Paxinos[5], Ophelie Lebrasseur[1], Keith Dobney[6,7], Joris Peters[5,8], Andrea Manica[2], Greger Larson[1*†], Anders Eriksson[9,2*†]

1 The Palaeogenomics and Bio-Archaeology Research Network, Research Laboratory for Archaeology, University of Oxford, Dyson Perrins Building, South Parks Road, Oxford OX1 3QY, UK

2 Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

3 Research Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK.

4 Department of Archaeology, University of Nottingham, Nottingham, NG7 5RD, UK

5 Department of Veterinary Sciences, Institute of Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, LMU Munich, Kaulbachstr. 37, D-80539 Munich, Germany

6 Department of Archaeology, School of Geosciences, University of Aberdeen, St. Mary's, Elphinstone Road, AB24 3UF, UK.

7 Department of Archaeology, Classics and Egyptology, University of Liverpool, 12-14 Abercromby Square, Liverpool L69 7WZ, UK.

8 SNSB, Bavarian State Collection of Anthropology and Palaeoanatomy, Karolinenplatz 2a, D-80333 Munich, Germany

9 Department of Medical & Molecular Genetics, King's College London, Guys Hospital, London SE1 9RT, UK.

[*] These authors contributed equally

[†]Corresponding authors: L.L. (liisa.loog@gmail.com), A.E. (anders.eriksson@kcl.ac.uk), G.L (greger.larson@arch.ox.ac.uk).

Abstract

Ancient DNA provides an opportunity to infer the drivers of natural selection by linking allele frequency changes to temporal shifts in environment or cultural practices. However, analyses have often been hampered by uneven sampling and uncertainties in sample dating, as well as being confounded by demographic processes. Here we present a Bayesian statistical framework for quantifying the timing and strength of selection using ancient DNA that explicitly addresses these challenges. We applied this method to time series data for two loci: TSHR and BCDO2, both argued to have undergone strong and recent selection in domestic chickens. The derived variant in TSHR, associated with reduced aggression to conspecifics and faster onset of egg laying, shows strong selection beginning around 1,100 years ago, coincident with archaeological evidence for intensified chicken production and documented changes in egg and chicken consumption. To our knowledge, this is the first example of pre-industrial domesticate trait selection in response to a historically attested cultural shift in food preference. For BCDO2, we find support for selection, but demonstrate that the recent rise in allele frequency could also have been driven by gene flow from imported Asian chickens during more recent breed formations. Our findings highlight that traits found ubiquitously in modern domestic species may not necessarily have originated during the early stages of domestication. In addition our results demonstrate the importance of precise estimation of allele frequency trajectories through time for understanding the drivers of selection.

Introduction

A comprehensive understanding of the evolutionary responses to changing selection pressures can be achieved by reconstructing allele frequency trajectories through time and linking them to changes in the ecological context in which the organism has evolved. This association has typically only been possible in longitudinal genomic studies of laboratory organisms with relatively short generation times, whose environments can be easily and rapidly manipulated (Kawecki et al. 2012). Linking episodes of selection to their drivers in natural systems is more challenging. The most common approach has been to study large sets of modern genetic data from closely related species or sub-populations using various statistical approaches (e.g. genome wide scans for selection) (Storz 2005). However, these approaches typically lack the power to reliably estimate the timing and strength of selection, especially over short evolutionary time scales. Consequently, this makes it challenging to associate past episodes of selection with specific drivers, especially when those drivers are not present in contemporary environments.

Ancient DNA data can provide direct information on changes in allele frequencies through time, and as such allows us to directly link past selection to contemporaneous ecological factors. However, ancient DNA sample sizes are typically small, and samples tend to be sparsely and unevenly distributed in space and time. Several methods exist for studying selection using ancient DNA (reviewed in Malaspinas 2016), but typically they lack the ability to model the confounding effects of gene flow – a process that could lead to over-estimation of selection coefficients (Mathieson et al. 2015). To overcome these challenges, we developed a Bayesian statistical framework that permits formal quantification of selection parameters using ancient allele frequency data, while explicitly taking into account both uncertainty in sample ages and gene flow from external sources and apply this framework to genotype data from domestic chicken.

The domestication of chickens (*Gallus gallus domesticus*) from wild junglefowl took place over a relatively short evolutionary timeframe (Rubin et al. 2010). In addition, the genetic architectures underlying many of the phenotypic traits that distinguish domestic populations are well understood (Eriksson et al. 2008; Rubin et al. 2010). As such, chickens present an ideal opportunity to correlate the timing of allele frequencies shifts with changes in human imposed selective pressures over the past several thousand years. To do so, we focussed on two loci: thyroid-stimulating hormone receptor (*TSHR*) and β-carotene dioxygenase 2 (*BCDO2*) previously identified as targets of selection in domesticated

chickens (Rubin et al. 2010).

TSHR plays a role in growth, metabolic regulation and photoperiod control of reproduction in mammals and birds by stimulating the synthesis and release of thyroid hormones (Yoshimura et al. 2003; Hanon et al. 2008; Nakao et al. 2008; Follett 2015; Nishiwaki-Ohkawa and Yoshimura 2016). Modern chickens carry a derived recessive *TSHR* allele that is thought to cause loss of strict seasonal reproduction (Rubin et al. 2010) - a commonly observed difference between domesticated animals and their wild relatives (Belyaev 1979) - and shown to have a faster onset of egg laying at sexual maturity (Karlsson et al. 2016). This allele has also been directly associated with reduced aggressive behaviors toward conspecifics and decreased fear of humans (Karlsson et al. 2015). It has, therefore, been suggested that these traits could have been selected to increase egg yield and facilitate larger groups or higher bird densities in a domestic setting (Karlsson et al. 2015; Karlsson et al. 2016).

The *BCDO2* gene is expressed in the skin where it encodes an enzyme that cleaves colorful carotenoids into colorless apocarotenoids, and polymorphisms in the *BCDO2* gene have well-known effects on skin coloration in birds (Eriksson et al. 2008). Domestic chickens are known to possess both a dominant, ancestral *BCDO2* allele that results in white or grey skin, and a recessive derived allele that is associated with yellow skin (Eriksson et al. 2008) – the latter likely acquired by domestic chickens through admixture with grey jungle fowl (*Gallus sonneratii*) (Eriksson et al. 2008). The biological mechanism for selection at BCD02 is less clear than for TSHR locus, but it has been previously proposed that the yellow legs phenotype could have been used as a proxy for good nutritional status of chickens, or selected for purely cosmetic reasons (Eriksson et al. 2008).

Genome-wide comparison between domesticated chickens and their wild relatives provides a powerful method for identifying loci under selection since chickens were domesticated around 6,000 years ago (Peters et al. 2016). Given the ubiquity of the derived TSHR allele in modern domesticated chicken populations (Rubin et al. 2010), and its association with several hallmark domestication traits (Karlsson et al. 2015), it has been proposed that the derived TSHR allele was selected during the initial stages of the domestication process (Rubin et al. 2010; Karlsson et al. 2015). However, due to the low mutation and recombination rates in nuclear DNA, these studies lack the power to accurately estimate the timing and the strength of selection.

Flink et al. (2014) typed the TSHR and BCDO2 loci in archaeological chicken samples from Europe, spanning the last 2,200 years, to further examine the hypothesis of selection during early domestication. They observed significantly lower derived allele frequencies in their ancient samples relative to the modern populations, which led them to reject the early selection hypotheses for both loci in favor of more recent selection. Given the sparseness and uneven temporal distribution of their data, they did not attempt to quantify the timing or strength of selection. A further complication to the analysis of selection comes from recent gene flow in domesticated chickens. During modern European chicken breed formation (beginning 100 − 150 years ago), Asian chickens were imported and bred with local European stock to improve existing birds and create novel breeds (Dana et al. 2011; Flink et al. 2014; Lyimo et al. 2015). In this study, we applied our statistical framework to the ancient DNA data reported by Flink et al. (Flink et al. 2014) and 16 additional samples (fig. 1 and Materials and Methods), and used forward simulations to test if genetic drift and gene flow, without invoking selection, could explain the observed allele frequency changes.

New Approaches

A Bayesian framework to infer past episodic selection

We represented the European domesticated chickens as a randomly mating population (an assumption we test below) in which an allele is under selection, and that receives gene flow from an external source. To simplify the model, we assumed that the population size was large enough to ignore genetic drift on these time scales. This allowed us to treat allele frequency trajectories as deterministic, an assumption that we also tested. Combining the standard model for selection on a recessive trait with a term that corresponds to gene flow from an external population, we obtained a differential equation for the allele frequency trajectory of the selected allele,

$$\frac{df}{dt} = s(t)f^2(1-f) + m(t)(f_{\text{external}} - f), \qquad \text{(Eq 1)}$$

where f is the frequency of the hypothesized selected allele, s is the selection coefficient, m is the rate of gene flow and f_{external} is the frequency of hypothesized selected allele in the external source population. The initial value for $f(t)$ is the ancestral frequency ($f_{\text{ancestral}}$). The model has three additional parameters defining the starting time of selection ($t_{\text{selection}}$) and the beginning (g_{start}) and end (g_{end}) of gene flow from

the external source. The selection coefficient *s* is constant after the starting time and is zero before. Similarly, *m* is constant in the gene flow interval and is zero at other times.

Using only the genotype information to estimate all seven parameters of the model would be challenging. However, because chickens are a relatively well-studied species, we were able to specify the values of four parameters by using historic information and mitochondrial DNA data, thus greatly reducing model search space. Gene flow from Asia is historically well-documented (Dana et al. 2011; Flink et al. 2014; Lyimo et al. 2015), beginning around 250 years ago (g_{start} = 250) and continuing until the present day (g_{end} = 0). The rate of gene flow *m* was estimated using modern and ancient mitochondrial haplotype frequencies, taking advantage of the fact that all ancient European chickens (i.e. prior to the gene flow from Asia) have the same mitochondrial haplotype (Materials and Methods). The frequency of the derived allele in the Asian population at the time of the gene flow is currently unknown. To minimize the possibility of overestimating the role of selection in increasing the allele frequency over the effect of gene flow from Asia, we made a conservative assumption that $f_{external}$ = 0.99 (i.e. the selected allele is nearly fixed in Asian chickens).

The likelihood of the data is a function of the frequency curve of the selected allele in the population and is calculated as the product of the probabilities of all observed alleles:

$$L = \prod_i f(t_i)^{x_i}\big(1 - f(t_i)\big)^{1-x_i}. \qquad \text{(Eq. 2)}$$

Here *L* is the likelihood of the data, *f(t)* is the allele frequency at the time *t*, t_i is the age of the sample *i*, and x_i = 1 in case of observing the derived allele in sample *i*, and x_i = 0 when observing the wild type.

Eq. (2) assumes that sample ages are precisely known. However, archaeological samples often come with some age uncertainty. To account for this, we replace the likelihood of each genotype (each factor in Eq. 2) with its posterior likelihood given the distribution of sample age from archaeological dating (Materials and Methods).

Results

Timing and strength of selection, and past allele frequency trajectories for *TSHR* and *BCDO2*

The remaining parameters of the model – the starting time of selection ($t_{selection}$), the selection coefficient (s) and the ancestral frequency of the selected allele ($f_{ancestral}$) – were inferred by making a full parameter sweep where we calculated the deterministic allele frequency trajectory for each parameter combination. We then calculated the posterior probability density distribution for each parameter by numerically integrating the likelihood of the data over the remaining two parameters (assuming a uniform prior for starting time of selection, ancestral frequency, and the logarithm of the selection strength). In addition, to visualize the inferred allele frequency trajectories, we calculated the posterior distribution of allele frequency through time by weighting the allele frequency trajectory of each parameter combination (sampled from their prior distributions) by the likelihood of the data given the parameters.

For the TSHR locus, we estimated a selection coefficient of 0.0049 (95% CI: 0.0030-0.0069, fig. S1), which would have generated a rapid increase in allele frequency (fig. 2a) starting 920 AD (95% CI: 290 AD – 1210 AD, fig. S1). We also inferred that the frequency of the TSHR derived allele prior to selection in the ancestral chicken population was 0.44 (95% CI: 0.34-0.54, fig. S1), similar to that estimated in a red jungle fowl captive zoo population (frequency 0.35, 95% CI:0.22-0.50 (Rubin et al. 2010)).

For the BCDO2 locus, we inferred a selection coefficient of 0.0036 (95% CI: 0.0001-0.0158, fig. S2), similar in magnitude to that estimated for the TSHR locus. Despite supporting selection, the inferred allele frequency trajectories corresponding to the best fitting parameter combinations show only a very marginal increase in the BCDO2 derived allele frequency prior to the time of gene flow (fig. 2b), therefore the observed rise in allele frequency can be largely attributed to recent Asian gene flow. This result is expected given the low inferred ancestral frequency of the derived allele ($f_{ancestral}$=0.14, 95% CI: 0.07–0.25, fig. S2), as the efficacy of selection for a recessive adaptive allele is highly sensitive to its initial frequency. As a consequence of minimal increase in allele frequency prior to the period of Asian gene flow, the starting time of selection is poorly resolved in this analysis (fig. S2), and, therefore, has little effect on the allele frequency trajectory.

Analyses of selection usually focus on estimating the strength of selection, and on rejecting the null model of genetic drift as an explanation of the data (Malaspinas 2016). However, selection coefficients can be difficult to interpret as the inferred values do not necessary reflect the true rate of change in

allele frequencies, especially for recessive alleles – we inferred similar selection coefficients for the TSHR and BCDO2 genes, but while the former shows a rapid increase in allele frequency, the latter results in only very marginal rate of change in allele frequencies. Thus, genotype data spanning an episode of selection, combined with statistical modeling, allow precise estimation of allele frequency trajectories though time (e.g. fig. 2). This in turn enables direct examination of both the timing as well as rate of adaptation, and greatly simplifies the biological interpretation of the effect of selection during the episode.

Testing the confounding effect of genetic drift

To test if the observed data could be explained by purely demographic factors, such as the rate of genetic drift and recent Asian gene flow, we simulated 10,000 stochastic allele frequency trajectories under this model (Materials and Methods). We then estimated the proportion of likelihood values greater than or equal to the values from the deterministic model of selection and gene flow (for a formal significance test), as well as the average likelihood of the trajectories (for a formal likelihood ratio test). To minimize the possibility of underestimating the role of drift in increasing the allele frequency over that of selection, we took a conservative approach and used BEAST (Drummond et al. 2012) to estimate the effective population size of the European chicken population based on mtDNA sequences from modern samples (Material and Methods). Since the male to female ratio in the ancient European chicken population is unknown, we assumed the size of the female effective population size as the effective populations size in our stochastic simulations.

For the TSHR locus, all of the stochastic trajectories from the drift model had a lower likelihood than the best fitting trajectory from the selection model. In addition, the formal likelihood ratio test estimates the selection model to be 3×10^{19} times more likely than the drift model. Consequently, the high frequency of the derived allele in the modern domesticated chicken populations cannot be explained by drift and gene flow alone, despite our conservative assumptions.

For the BCDO2 derived allele, the best fitting selection model explains the data only marginally better than the drift model. Out of the 10,000 simulated allele frequency trajectories, 366 have a likelihood value as high or higher than the best fitting selection model, and the selection model is nine times more

likely than the drift model. Thus, while it is possible to obtain the patterns observed in the ancient DNA under the drift model, it is rare and the selection model is still supported by the data.

Finally, to test the effect of gene flow on our ability to reject selection over genetic drift, we ran the deterministic and stochastic models without migration. As expected from the timing and strength of selection acting on the TSHR locus, removing gene flow from the stochastic model made little difference to our ability to reject the null-hypothesis of no selection (p < 0.0001). In contrast, selection for the BCDO2 locus became highly significant (s = 0.0060, p = 0.0005) in models without migration, confirming that unaccounted gene flow can cause spurious effects in selection analyses.

Sensitivity analyses

As discussed in the introduction, sparseness of samples and geographic patterns could confound analysis of past selection. To assess whether geographic structure in the data could affect the analysis, we first applied a standard Hardy-Weinberg test to ancient samples predating the Medieval times, and reject Hardy-Weinberg disequilibrium for both TSHR (χ^2=0.057, p=0.81) and BCD02 (χ^2=0.20, p=0.66) (Materials and Methods). This lack of structure is also consistent with the observation that European ancient chicken samples have the same mitochondrial haplotype (Flink et al. 2014). As an explicit test of spatial genetic and cultural differences across Europe, we restricted the data to locations from Northwest Europe and excluded samples from Mediterranean region (Materials and Methods and fig. S4). Analysis on this subset of the data yielded parameter ranges similar to the full dataset (table S9), indicating little or no effect of geographic structure on our estimates. The main difference between the two results is that the starting time and coefficient of selection are less well defined in the analyses with the restricted data compared to the analysis with the full dataset (tables S9 and figs. S5 and S6). For TSHR, we also observed an increased uncertainty of inferred allele frequency trajectory between 1100 AD and 1600 AD (fig. S7). This is mostly due to the fact that the analysis with the reduced dataset lacks most samples from this informative time period. However, the increase in allele frequency in the 10th century AD is still best supported, and we can confidently rule out no significant rise in allele frequency prior to 900AD due to the large number of samples predating the period (fig. S7).

To explore the effect of sparseness of ancient samples, we picked a parameter combination similar to the best supported scenario for TSHR and simulated 50 replicate datasets with sample sizes ranging

from 10 to 200 diploid individuals (Materials and Methods), and repeated the analysis for each dataset. Fig. S8 shows the confidence intervals for the three estimated parameters in each dataset. As expected, we observe a trend of increasingly wide confidence intervals with decreasing sample size. However, the medians of the posterior distributions are scattered near the true value, indicating little or no bias in estimating the parameters even for relatively small number of sample sizes.

As illustrated above, a large sample size is important for accurate reconstruction of past selection, but more fundamentally, it is important that the samples used in the analysis represents the underlying allele frequency curve, i.e. to have data before, during and after the change in allele frequency due to selection. Nevertheless, as also demonstrated in the analyses above, temporal clustering of the data does not necessarily cause biased estimates but will instead result in a wider range of parameter combinations as plausible explanations for the data.

Selection at the TSHR locus coincides with medieval changes in chicken husbandry

Our analysis indicates that selection on the *TSHR* derived allele began around 920 AD (290 – 1210 AD at 95% level). Intriguingly, this period coincides with major changes in chicken husbandry witnessed in several archaeological assemblages across Northwestern Europe (Boessneck et al. 1979; Heinrich 1987; Pieper and Reichstein 1995; Heinrich 2006; Hüster Plogmann 2006; Serjeantson 2006; Sykes 2007; Holmes 2014). Analysis of data from 184 English archaeological faunal assemblages and 104 German archaeological faunal assemblages demonstrate a substantial increase in the frequency of chicken remains from the Early to the High Middle Ages (fig. 3 and Materials and Methods) and higher proportions of adult hens in the flocks, presumably to increase egg production (Serjeantson 2006; Sykes 2007; Holmes 2014).

Discussion

This study demonstrates how inferring allele frequency trajectories through time, using ancient DNA in a flexible Bayesian inference framework, has numerous advantages over previous approaches to studying selection. Firstly, our method provides a means of formally accommodating small sample sizes, uneven temporal distribution of samples and uncertainties in sample ages. Secondly, it allows for episodic selection and gene flow to be explicitly modeled, thus enabling formal integration of independent lines of evidence related to migration, admixture rates, past population sizes and the expression of alleles

(dominant or recessive), all of which significantly enhance inferential power. Thirdly, the deterministic nature of our inference approach renders the method fast, flexible, and capable of handling any genotype or allele frequency data from samples before, during and after the hypothesized period of selection. Additionally, as demonstrated by our sensitivity analyses, the approach can also be used to (iteratively) identify the most informative time periods for selection analysis (the periods with the largest uncertainty in the reconstruction of the allele trajectory), and thereby be used as a tool for selecting samples for genetic sequencing.

Although determining the causal factors behind the inferred selection on the TSHR locus is beyond the scope of this paper, by combining ancient DNA with statistical modeling we can increase our ability to identify specific ecological or socio-cultural drivers behind selection. Rather than relying purely on knowledge of the biological function of genes to infer causes of selection, our method allows to also include factors such as known climatic or socio-cultural developments inferred, from geological, archeological or historical records that are broadly contemporaneous with the time of inferred allele frequency changes.

For example, our inferred timing of the onset of selection on the *TSHR* derived allele brackets a key period of changes in dietary preferences and chicken husbandry practices across northwestern Europe between the 9th and 12th centuries AD (fig. 3). The significant intensification of chicken and egg production, evident in the Medieval European archaeological record, has been linked to Christian fasting practices which forbade the consumption of meat from four-legged animals during fasting periods, but allowed for the consumption of birds, eggs and fish (Venarde 2011). These rules, which originated in the Benedictine Monastic Order, became widely adopted across Europe due to increasing political influence of the Catholic church and applied to all segments of society ca. 1000 AD (Sykes 2007). An increase in chicken consumption has also been observed in Viking sites in present-day Scandinavia and Northern Germany (Boessneck et al. 1979; Pieper and Reichstein 1995; Hüster Plogmann 2006), coinciding with Christianization of these areas (starting in the late 10th-11th and the mid-9th centuries AD respectively) (Näsman 1976; Sanmark 2004).

Religiously inspired dietary preferences are not the only factors potentially affecting medieval poultry consumption. The observed increase in chicken production could also have been driven by urbanization

11

and population growth (Bartlett 1994), facilitated by the widespread introduction of the more efficient agricultural practices including the three-field system with crop rotation and new tools such as the heavy plough (White 1967; Andersen et al. 2016) as well as favorable climatic conditions (Mann et al. 2009) in the High Middle Ages. Although heavily reliant on the agricultural land surrounding the towns and the cities (Keene 1995), the urban population likely secured some of its own needs by keeping single animals such as goats, pigs and small flocks of chickens in the domestic space in order to have predictable access to milk, eggs and occasionally meat. These factors could, individually or in combination, explain the observed medieval increase in poultry consumption, which would most likely have resulted in intensified production of poultry. This could have been achieved either through increasing chicken flock sizes or (in an urban setting) flock densities, which could have been facilitated by selection on reduced aggression within the flocks (Karlsson et al. 2015) or faster onset of egg laying at sexual maturity (Karlsson et al. 2016), both demonstrated effects of the derived TSHR allele.

For BCDO2 we also find some support of selection. However, our analysis suggests that the significant rise in allele frequency between ancient and modern non-commercial chicken breeds can be attributed to recent gene flow from Asia. It is also possible that the commercial chicken lines (excluded from this study) have experienced additional strong selection at the BCDO2 locus. This is supported by observation that the frequency of the derived BCDO2 allele is much higher in commercial chicken lines compared to the local chicken breeds used in this study (Rubin et al. 2010).

To our knowledge, the spread of the TSHR derived allele is the first example of a pre-industrial domesticate trait selection in response to a historically attested cultural shift in food preference. This result supports the view (Allaby et al. 2008) that domestication is an ongoing process where numerous traits that have been traditionally associated with domestication underwent selection long after the initial domestication phases, and that domestic populations have experienced highly variable selective regimes in response to shifting cultural preferences. More generally, humans and their domestic plants and animals have undergone radical shifts in allele frequencies as they have each responded to alterations in natural and artificial selective pressures over the past 15,000 years. The ability to infer allele frequency trajectories and the timing and strength of selection episodes, based upon time-stamped genetic data, allows for those episodes to be correlated with specific ecological and cultural drivers, especially for drivers that are not present in contemporary populations. This in turn can reveal

both the processes and causative mechanisms responsible for generating the patterns of genomic variation in humans and their co-dependent domestic plants and animals found today.

Materials and Methods

<u>Ancient genotypes</u>

We use genotype data from two loci, TSHR (position 43,250,347 on chromosome 5) and BCDO2 (position 6,273,428 on chromosome 24), both argued to have undergone selection in domestic chickens. The two loci are typed, respectively, in 59 and 30 non-modern West Eurasian and Moroccan chickens, temporally raging approximately from 2,200 years ago to present day. The genotypes of ancient West Eurasian individuals as well as their approximate ages are taken from Flink et al. (2014). We increased the size of dataset for both TSHR and BCDO2 locus by 14 and 4 individuals respectively. These additional samples were genotyped using the same protocol as reported in Flink et al. (2014) (Supplementary Information for details). Six ancient samples (four samples out of the 14 samples newly presented in this study) were directly radiocarbon dated (table S7). The rest of the samples were dated by stratigraphic association. In order to minimize the possibility of intrusion from more recent periods all indirectly dated samples originate from sites with no following occupational layers (e.g. no building structures or archaeological strata). See table S2 and S3 for list of all samples used in the analyses, as well as their genotypes at the two loci and ages, for TSHR and BCDO2 respectively.

<u>Modern allele frequencies</u>

The modern allele frequencies for TSHR locus is calculated from the total frequency of selected allele in sampled 167 non-industrial European chickens, representing 24 breeds. Data presented in (Rubin et al. 2010) (see table S4 for breeds included and the TSHR allele frequencies in those breeds).

As the allele frequencies at the BCDO2 locus in modern populations are unknown we use the phenotypes of 18 modern non-industrial European chicken breeds, reported in European collaboration project on chicken biodiversity database (AVIANDIV: http://w3.tzv.fal.de/aviandiv) (table S5) for breeds included, typical phenotypes and estimated average population sizes). As the hypothesized selected BCDO2 allele is recessive, the birds from breeds reported to be yellow legged were considered to carry two copies of the selected allele. For birds from breeds reported to be predominantly white legged the frequency of the proposed selected allele was assumed to be 0.016, based on the estimates of

13

frequency in the white legged European breeds in (Eriksson et al. 2008) (see table S6 for the breeds included) and assuming Hardy-Weinberg equilibrium.

To factor in the uncertainty rising from the fact that the BCDO2 allele counts are not independent observations but estimates themselves, we used bootstrapping at the level of breeds to estimate the uncertainty in the population wide frequency. We then approximate this distribution with a binomial distribution by fitting the mean and the variance of the binomial distribution to the bootstrapped distribution, yielding an effective sample size of 18.4 individuals. This approximation allows us to give the estimate an equal weight to the ancient samples when calculating the likelihoods in the model.

FIXED PARAMETERS:

Timing of gene flow from Asia

The starting time of gene flow from Asia is historically well documented (Dana et al. 2011; Flink et al. 2014; Lyimo et al. 2015) and as a result, is fixed to start at 250 years ago and lasts until the present.

Rate of gene flow from Asia

The rate of gene flow was estimated using modern and ancient mitochondrial haplogroup frequencies: Based on all samples currently available, all European chickens prior to historically documented gene flow from Asia belonged to mitochondrial haplogroup E (Flink et al. 2014). Today 90% of European chickens belong to haplogroup E while the rest are assigned to haplogroups A, B and C. Haplogroup E is also most common among Asian chicken populations, however only 30% of the sampled Asian Chickens belong to this haplogroup. Assuming that the change in the frequency of haplogroup E is the direct result of gene flow from Asia and that this gene flow has not been sex biased, we can estimate that approximately 15% of the modern European chickens have Asian origin. As a result, the rate of gene flow per year in the model is assumed to be equal to the proportion of European chickens with Asian ancestry (15%) divided by the length of the admixture period (250 years).

Allele frequencies in the source population

The frequency of the selected allele in the Asian population at the time of the gene flow is unknown. Therefore we used a conservative estimate of ($f_{external}$ = 1) i.e. the selected allele is fixed in Asian chickens. We chose this conservative approach to minimize the possibility of overestimating the role of selection at increasing the allele frequency over the effect of gene flow from Asia. As a result, lower frequency in Asian population during the period of the gene flow could result in later estimates of the starting times of selection and higher estimates of selection coefficients in the two loci considered. However, for TSHR we observe that selection brings the frequency to near fixation before the onset of gene flow, and therefore the precise frequency of the derived allele in Asia is not very important for the allele frequency trajectory. For BCDO2, a lower frequency of the derived allele in Asia could lead to slightly higher levels of selection.

MODEL FITTING

In order to infer starting time of selection, selection coefficient and ancestral frequency of a selected allele, we perform a full sweep of those parameters and calculate the deterministic allele frequency trajectory for each parameter combination. The deterministic allele frequency curves were calculated using *lsoda* function in R package *deSolve* (Soetaert et al.).

Value ranges considered for estimated parameters:

Parameter values considered for starting time of selection are between 0 to 2000 years ago in uniformly spaced steps of ten years. For the selection coefficient, we considered 100 uniformly spaced points between 10^{-4} and 1 on a log scale. Parameter values considered for ancestral frequency of a selected allele ranged from 0.01 to 1 in 100 steps. We derive the marginal likelihoods for each of the parameters (starting time of selection, selection coefficient and the starting time of selection) by averaging over the likelihoods of the remaining two parameters

Incorporating sample age uncertainty

Following the Bayesian principle, we treat the age of each sample as a random variable with uniform distribution over its age range (see tables S2 and S3 for age ranges), and calculate the posterior likelihood of the sample data given this distribution (numerically approximating the continuous distribution within each interval by ten equally spaced time points)

TEST OF SIGNIFICANCE

Stochastic Simulations

In the model described above the change in the allele frequency is assumed to be a result of selection and gene flow, and that the effective population size is large enough to ignore genetic drift. To test if selection is needed to explain the change in the allele frequency through time we explicitly test whether the selection models are statistically different from models including drift and gene flow alone. We do this by simulating allele frequency trajectories with drift, under the assumption of no selection but the same gene flow parameter values as in the deterministic allele frequency trajectories above.

We use the *sde.sim* function in the R package *sde* (Iacus 2008) to simulate 10,000 allele frequency trajectories using the stochastic differential equation corresponding to Eq. 1 without selection and including genetic drift in a population of size N_e:

$$\frac{df}{dt} = m(t)(f_{external} - f) + \sqrt{f(1-f)/2N_e}\xi(t), \text{(Eq S1)}$$

To match the selection model described above, ancestral allele frequency is sampled from a uniform distribution over the interval [0,1]. The population size N_e is sampled from the posterior distribution of population size estimates from BEAST (Drummond et al. 2012) (see below), with a median value of $1.8*10^5$ and 95% HPD interval ranging from 26000 to $4.6*10^5$. Since this population size estimate is based on a non-recombining and maternally inherited locus, the total effective population size should be four times that. However, as male to female ratio in ancient European chicken population is unknown, we take a conservative approach and use the minimum effective population size, which roughly corresponds to the size of the female effective population size. Therefore, we multiply the posterior distribution of population size by factor of two to get an estimate of a minimum effective population size. This approach is conservative since the power of genetic drift to influence allele frequencies decreases with increasing population size. To formally evaluate whether the model with selection provides a significantly better fit for the data we calculate the proportion of log likelihood values from the stochastic simulations that are as high or higher than the log likelihood of the selection model.

Chicken population size estimation using BEAST

We estimated an order of magnitude effective population size for chickens through time using Bayesian Phylogenetic tool BEAST (Drummond et al. 2012) and a 200 base pairs long mitochondrial control region fragment from 194 modern European chickens and 39 ancient European individuals with known ages and a uniform prior for constant population size ranging [0, 10^{-7}], (for the full BEAST input file see Supplementary File 1).

SENSITIVITY ANALYSES

Hardy-Weinberg analysis

We selected samples dating to before 300AD to test for deviation from Hardy-Weinberg equilibrium. Due to the large sample size (n=37 for TSHR and n=24 for BCD02) we used the standard χ^2 test (with one degree of freedom) to investigate whether the genotype frequencies are binomially distributed.

North-Western subset of the data (sensitivity to potential geographic structure)

We generated a reduced data set covering North-West Europe by excluding ancient samples from Turkey, Greece and Morocco, resulting in 42 ancient samples for TSHR locus and 30 ancient samples for BCDO2 locus (fig. S4) for the temporal distribution of these samples). For consistency, all data points from Italy and Spain (BCDO2) and Italy and Israel (TSHR) were also excluded from the calculation of modern allele frequencies (but this did not affect the modern frequencies).

Simulated data (sensitivity to sample size)

We chose a combination of parameters similar to that found for TSHR (starting time of selection=1000AD, selection coefficient=0.005 and ancestral frequency=0.4). For simplicity, we assume no gene flow from Asia (m = 0). We simulated 10 replicate data sets each of 10, 20, 50, 100 and 200 ancient diploid samples (and 150 modern samples). Simulated ancient samples were given randomly chosen dates (uniformly distributed from year 0 to 2000AD) and 100 years age uncertainty (on both sides of the assigned age). For each individual we generated random diploid genotypes according to the derived allele frequency given by the model at the sample age.

ARCHAEOLOGICAL CHICKEN REMAINS FREQUENCY CALCULATION

We used data from Appendix Ia in Sykes (2007) on the faunal number of identified specimens (NISP) from 184 English archaeological faunal assemblages and compiled information (Supplementary File 2) on faunal NISP from 104 German archaeological faunal assemblages to estimate the relative frequency of chicken remains at each of the four time periods for English and German sites respectively. We calculated the relative frequency by dividing the number of identified chicken specimens by the total faunal NISP for the same time period (fig. 3 and table S1) The 95% confidence intervals were calculated using the normal distribution approximation (justified by the large sample sizes).

The framework described above is implemented in the statistical environment R. The R code is available to download from GitHub (https://github.com/LiisaLoog/Selection-Framework) and from L.L upon request.

Author Contributions: M.G.T. & G.L. initiated the project; L.L., M.G.T. & A.E. designed the research; A.E. & L.L. designed the statistical framework; L.L. implemented the framework and performed statistical analysis with input from A.E.; R.B. & R.A. generated the new data; J.P. provided sample material; P.D.P compiled the archaeological faunal data from German sites; O.L. researched information about historic chicken population sizes; N.S., K.D. & J.P. interpreted the TSHR results within an archaeological and historical context; L.L., M.G.T., A.M., G.L. & A.E. wrote the paper with input from N.S., K.D. & J.P.

References

Allaby RG, Fuller DQ, Brown TA. 2008. The genetic expectations of a protracted model for the origins of domesticated crops. Proc. Natl. Acad. Sci. 105:13982–13986.

Andersen TB, Jensen PS, Skovsgaard CV. 2016. The heavy plow and the agricultural revolution in Medieval Europe. J. Dev. Econ. 118:133–149.

Bartlett R. 1994. The Making of Europe: Conquest, Colonization and Cultural Change 950 - 1350. Penguin UK

Belyaev DK. 1979. Destabilizing selection as a factor in domestication. J. Hered. 70:301–308.

Boessneck J, Driesch A von den, Stenberger L. 1979. Eketorp: Befestigung und Siedlung auf Öland, Schweden : die Fauna. Almqvist & Wiksell International

Dana N, Megens H-J, Crooijmans RPMA, Hanotte O, Mwacharo J, Groenen M a. M, van Arendonk J a. M. 2011. East Asian contributions to Dutch traditional and western commercial chickens inferred from mtDNA analysis. Anim. Genet. 42:125–133.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29:1969–1973.

Eriksson J, Larson G, Gunnarsson U, Bed'hom B, Tixier-Boichard M, Strömstedt L, Wright D, Jungerius A, Vereijken A, Randi E, et al. 2008. Identification of the Yellow Skin Gene Reveals a Hybrid Origin of the Domestic Chicken. PLoS Genet. [Internet] 4. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2265484/

Flink LG, Allen R, Barnett R, Malmström H, Peters J, Eriksson J, Andersson L, Dobney K, Larson G. 2014. Establishing the validity of domestication genes using DNA from ancient chickens. Proc. Natl. Acad. Sci. 111:6184–6189.

Follett BK. 2015. "Seasonal changes in the neuroendocrine system": Some reflections. Front. Neuroendocrinol. 37:3–12.

Hanon EA, Lincoln GA, Fustin J-M, Dardente H, Masson-Pévet M, Morgan PJ, Hazlerigg DG. 2008. Ancestral TSH Mechanism Signals Summer in a Photoperiodic Mammal. Curr. Biol. 18:1147–1152.

Heinrich D. 1987. Untersuchungen an mittelalterlichen Fischresten aus Schleswig: Ausgrabung Schild 1971-1975. Wachholtz

Heinrich D. 2006. Die Fischreste aus dem Hafen von Haithabu - Handaufgelesene Funde. In: Untersuchungen an Skelettresten von Tieren aus dem Hafen von Haithabu. Neumünster: Wachholtz. p. 157–194.

Holmes M. 2014. Animals in Saxon and Scandinavian England: Backbones of Economy and Society. Sidestone Press

Hüster Plogmann H. 2006. Untersuchungen an den Skelettresten von Säugetieren und Vögeln aus dem Hafen von Haithabu. In: Schietzel K, editor. Berichte über die Ausgrabungen in Haithabu. Vol. 35. Neumünster: Wachholtz. p. 25–156.

Iacus SM. 2008. Simulation and Inference for Stochastic Differential Equations With R Examples. Springer NY

Karlsson A-C, Fallahshahroudi A, Johnsen H, Hagenblad J, Wright D, Andersson L, Jensen P. 2016. A domestication related mutation in the thyroid stimulating hormone receptor gene (TSHR) modulates photoperiodic response and reproduction in chickens. Gen. Comp. Endocrinol. 228:69–78.

Karlsson A-C, Svemer F, Eriksson J, Darras VM, Andersson L, Jensen P. 2015. The Effect of a Mutation in the Thyroid Stimulating Hormone Receptor (TSHR) on Development, Behaviour and TH Levels in Domesticated Chickens. PLOS ONE 10:e0129040.

Kawecki TJ, Lenski RE, Ebert D, Hollis B, Olivieri I, Whitlock MC. 2012. Experimental evolution. Trends Ecol. Evol. 27:547–560.

Keene D. 1995. Small towns and the metropolis : the experience of medieval England. In: Duvosquel JM, Thoen E, Verhulst AE, editors. Peasants & townsmen in medieval Europe: studia in honorem Adriaan Verhulst. Gent: Snoeck-Ducaju & Zoon. p. 223–238.

Lyimo CM, Weigend A, Msoffe PL, Hocking PM, Simianer H, Weigend S. 2015. Maternal genealogical patterns of chicken breeds sampled in Europe. Anim. Genet. 46:447–451.

Malaspinas A-S. 2016. Methods to characterize selective sweeps using time serial samples: an ancient DNA perspective. Mol. Ecol. 25:24–41.

Mann ME, Woodruff JD, Donnelly JP, Zhang Z. 2009. Atlantic hurricanes and climate over the past 1,500 years. Nature 460:880–883.

Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, et al. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. Nature 528:499–503.

Nakao N, Ono H, Yamamura T, Anraku T, Takagi T, Higashi K, Yasuo S, Katou Y, Kageyama S, Uno Y, et al. 2008. Thyrotrophin in the pars tuberalis triggers photoperiodic response. Nature 452:317–322.

Näsman U. 1976. Introduction to the Descriptions of Eketorp-I, -II and –III. In: Eketorp – Fortification and Settlement on Öland/Sweden. Stockholm: Almqvist & Wiksell

Nishiwaki-Ohkawa T, Yoshimura T. 2016. Molecular basis for regulating seasonal reproduction in vertebrates. J. Endocrinol.:JOE-16-0066.

Peters J, Lebrasseur O, Deng H, Larson G. 2016. Holocene cultural history of Red jungle fowl (Gallus gallus) and its domestic descendant in East Asia. Quat. Sci. Rev. 142:102–119.

Pieper H, Reichstein H. 1995. Untersuchungen an Skelettresten von Vögeln aus dem mittelalterlichen Schleswig. In: Vogel V, editor. Ausgrabungen in Schleswig, Berichte und Studien. Vol. 11. p. 9–114.

Rubin C-J, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S, et al. 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464:587–591.

Sanmark A. 2004. Power and Conversion: A Comparative Study of Christianization in Scandinavia. Department of Archaeology and Ancient History, Uppsala University, Occasional Papers in Archaeology

Serjeantson D. 2006. Birds: food and a mark of status. In: Woolgar CM, Serjeantson D, Waldron T, editors. Food in Medieval England: Diet and Nutrition. Oxford University Press. p. 131–147. Available from: http://eprints.soton.ac.uk/342285/

Soetaert K, Petzoldt T, Setzer RW. Solving Differential Equations in R: Package deSolve. Available from: https://core.ac.uk/display/6287976/tab/similar-list

Storz JF. 2005. INVITED REVIEW: Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol. Ecol. 14:671–688.

Sykes NJ. 2007. The Norman Conquest: A Zooarchaeological Perspective. British Archaeological Reports.International Series 1656.

Venarde BL. 2011. The Rule of Saint Benedict. Harvard University Press

White LJ. 1967. The Life of the Silent Majority. In: Hoyt RS, editor. Life and Thought in the Early Middle Ages. Minneapolis: University of Minnesota Press. p. 85–100.

Yoshimura T, Yasuo S, Watanabe M, Iigo M, Yamamura T, Hirunagi K, Ebihara S. 2003. Light-induced hormone conversion of T4 to T3 regulates photoperiodic response of gonads in birds. Nature 426:178–181.

Figure legends

Fig. 1. Observed allele counts through time for the TSHR locus (a) and the BCDO2 locus (b). Wild type alleles in ancient samples are represented by yellow dots and derived alleles by red dots. Modern allele proportions are shown as solid bars to the right of each panel.

Fig. 2. Posterior distribution of the derived allele frequency as a function of time for the TSHR locus (a) and the BCDO2 locus (b). The likelihoods are color coded (see color bar) and are shown relative to the maximum likelihood in the plot.

Fig. 3. Proportion of chicken remains (by number of identified specimens) from 184 English (top) and 104 German (bottom) archaeofaunal assemblages (Material and Methods)

Fig. 1. Observed allele counts through time for the TSHR locus (a) and the BCDO2 locus (b). Wild type alleles in ancient samples are represented by yellow dots and derived alleles by red dots. Modern allele proportions are shown as solid bars to the right of each panel.

Fig. 2. Posterior distribution of the derived allele frequency as a function of time for the TSHR locus (a) and the BCDO2 locus (b). The likelihoods are color coded (see color bar) and are shown relative to the maximum likelihood in the plot.

Fig. 3. Proportion of chicken remains (by number of identified specimens) from 184 English (top) and 104 German (bottom) archaeofaunal assemblages (Material and Methods)