
Article:
Heikkinen, M.E, Ruokonen, M, Alexander, Michelle Marie orcid.org/0000-0001-8000-3639 et al. (3 more authors) (2015) Relationship between wild greylag and European domestic geese based on mitochondrial DNA. Animal Genetics. 485–497. ISSN: 1365-2052

https://doi.org/10.1111/age.12319

ANIMAL GENETICS
Immunogenetics, Molecular Genetics and Functional Genomics

Relationship between wild greylag and European domestic geese based on mitochondrial DNA

Journal:	*Animal Genetics*
Manuscript ID:	AnGen-14-12-0366.R2
Manuscript Type:	Full Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Heikkinen, Marja; University of Oulu, Genetics and Physiology Unit Ruokonen, Minna; University of Oulu, Genetics and Physiology Unit Alexander, Michelle; University of York, Department of Archaeology Aspi, Jouni; University of Oulu, Genetics and Physiology Unit Pyhäjärvi, Tanja; University of Oulu, Genetics and Physiology Unit Searle, Jeremy; Cornell University, Department of Ecology and Evolutionary Biology
Keywords:	Anser anser, control region, domestication, genetic diversity, phylogeography

Relationship between wild greylag and European domestic geese based on mitochondrial DNA

Short running title:

Relationship between greylag and domestic geese

Marja E. Heikkinen[1], Minna Ruokonen[1]*, Michelle Alexander[2], Jouni Aspi[1], Tanja Pyhäjärvi[1], Jeremy B. Searle[3]

[1]Genetics and Physiology Unit, University of Oulu, Oulu, 90570, Finland. [2]Department of Archaeology, University of York, York, YO10 5DD, United Kingdom. [3]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853, USA.

*Deceased

Address for correspondence

Marja Heikkinen, Genetics and Physiology Unit, University of Oulu, Oulu, 90570, Finland.

E-mail: marja.e.heikkinen@oulu.fi

Fax: +358 8 344 064

Tel. +358 50 350 6001

Summary

The origins of the European domestic goose are uncertain. The available information comes from archaeological findings and historical literature but genetic evidence has hitherto been scarce. The domestic goose in Europe is derived from the greylag goose (*Anser anser*) but it is not known where the initial domestication took place and which of the two subspecies of greylag goose was ancestral. We aimed to determine the amount and geographical distribution of genetic diversity in modern populations of greylag geese as well as in different breeds of the domestic goose to make inferences on goose domestication. We studied DNA sequence variation in the mitochondrial control region of greylag geese from multiple populations across Europe and western Asia as well as specimens of domestic geese representing 18 modern breeds and individuals not belonging to any recognised breed. Our results show notable differences in genetic diversity between different greylag goose populations and the presence of six mitochondrial haplogroups which show a degree of geographical partitioning. The genetic diversity of the domestic goose is low, with 84% of sampled individuals having one of two major closely related haplotypes, suggesting that modern European domestic geese may derive from a narrow genetic base. The site of domestication remains unresolved, but domestic geese in Turkey were unusually diverse indicating the importance of further sampling in the vicinity of the eastern Mediterranean and the Near East. There appears to be past or ongoing hybridization between greylags and domestic geese in particular areas, consistent with field observations.

Keywords *Anser anser*, control region, domestication, genetic diversity, phylogeography

Introduction

The origins of the European domestic goose are uncertain and understudied. The available knowledge comes from archaeological findings and historical literature but genetic studies have hitherto been scarce. It is an undisputable fact that the European domestic goose is derived from the greylag goose (*Anser anser*) (Delacour 1954) but there have been only a few studies linking the genetics of domestic geese to their wild ancestry (Shi *et al.* 2006, Wang *et al.* 2010).

The greylag goose is a widespread Palearctic species (Fig. 1). Different sites of domestication have been suggested for the goose in Europe (see review by Albarella 2005). Based on archaeological evidence, southeastern Europe has been emphasised as the possible area of domestication, maybe through the actions of the ancient Greeks (Zeuner 1963). Crawford (1984) argues in favour of this and suggests that goose domestication started around 3000 BC. The Old Kingdom of Egypt (3rd millennium BC, circa 2686-2181 BC) has also been mentioned as a possible location for goose domestication, but this scenario has been considered less likely by Zeuner (1963). We think that it suffices to say that the European domestic goose was domesticated approximately 3000-5000 years ago, most likely in the vicinity of the eastern Mediterranean. This area considered broadly, i.e. including the Fertile Crescent, is a major area of domestication (Bruford *et al.* 2003). It should be noted that the Chinese also domesticated geese, but these derived from a different species, the swan goose (*Anser cygnoides*). Again, East Asia is a well-known domestication centre (Bruford *et al.* 2003).

More progress in understanding domestication of domestic geese can only be made through a detailed knowledge of wild greylag geese. The distribution of the greylag goose is fragmented,

likely due to human actions, but based on ringing data it can be subdivided into six rather well-defined biogeographic populations (Scott & Rose 1996) (Fig. 1). The first population consists of geese breeding in Iceland and wintering in Britain and Ireland. The second is a small, largely sedentary population in northwest Scotland but at least some geese of this population overwinter in England. The geese belonging to third population breed in Norway, Sweden, Denmark and western Germany. Their wintering area expands from the Netherlands to southern Spain and Morocco. The breeding area of the fourth population consists of Finland and northeast Sweden through the Baltic States to central Europe. The wintering area of these birds extends from central Europe to Tunisia and Algeria. The fifth population breeds and winters within the Black Sea region and Turkey. The sixth population breeds in western Siberia south to the Caspian Sea and winters to the south of the Caspian region, Iran and Iraq. The European breeding populations are well-monitored and the common trend seems to be that they have been increasing during recent decades. However, the current population status of eastern populations, those breeding around Black Sea and in southwestern Asia, is less well known, but it appears that there are stable with tens of thousands of individuals (Fox *et al.* 2010).

There are two recognised subspecies of greylag, the western, nominate form *A. a. anser* (Linnaeus, 1758) and the eastern form *A. a. rubrirostris* (Swinhoe, 1871). These two types are characterised by slight morphological differences. The eastern type is slightly larger and paler in tone and has a pink bill and cold pink legs in comparison to the western type's orange bill and flesh-coloured legs (Cramp & Simmons 1977). Geese of the eastern form are found in southeastern Europe eastwards (Fig. 1); however, the subspecies boundary is not well-defined and intermediate types are found in central and eastern Europe. Occasionally birds from Iceland, Scotland and Norway are classified as a third subspecies *A. a. sylvestris*, (Madge & Burn 1988) but this has not been widely accepted.

Delacour (1954) has stated that it was the western subspecies that was domesticated but more recent authors have suggested domestication of the eastern subspecies. The main support for the domestication of eastern subspecies comes from morphology. The domestic goose typically has a pink bill, which is in accordance with the morphology of eastern type (Harper 1972, Kear 1990).

To enhance understanding of goose domestication we studied the mitochondrial DNA (mtDNA) variability in modern wild greylag geese populations and domestic geese breeds found in Europe. Mitochondrial DNA has a maternal inheritance which simplifies phylogenetic interpretation. It has been used to trace domestication history of both mammals and birds; for example cattle (Loftus *et al.* 1994, Troy *et al.* 2001, Beja-Pereira *et al.* 2006, Achilli *et al.* 2008, Bollongino *et al.* 2012), sheep (Hiendleder *et al.* 2002, Tapio *et al.* 2006, Meadows *et al.* 2011), goat (Luikart *et al.* 2001, Naderi *et al.* 2008), dog (Savolainen *et al.* 2002, Pang *et al.* 2009, Thalmann *et al.* 2013), chicken (Fumihito *et al.* 1996, Liu *et al.* 2006, Miao *et al.* 2013) and turkey (Speller *et al.* 2010). We focused on the non-coding 'control region' (CR) as a genetic marker because its high substitution rate makes it suitable for phylogeographic and phylogenetic studies (Vigilant *et al.* 1989, Wenink *et al.* 1993).

To our knowledge this is the first time that the genetic relationship between wild greylag geese and their derivative, the European domestic goose, has been studied in a systematic way. We investigate: 1) the amount of genetic diversity in modern populations of greylags as well as in different breeds of domestic goose, 2) whether genetic evidence supports the biogeographical populations of greylag goose based on ringing data, 3) whether different subspecies are distinguishable based on maternal DNA, and finally, 4) which subspecies was domesticated and where.

Materials and methods

Samples

A total of 178 specimens of wild greylag geese were sampled throughout their distribution area (Fig. 1, Table S1). All the samples were collected between 1993 and 2011 (see also Appendix S1).

A wide range of different domestic geese were studied, concentrating on well-defined breeds as much as possible (Table 1). The non-breed individuals are designated as 'Domestic' followed by the country code for their sampling location. We included some samples of breeds that are likely to be crosses of the European and Chinese type, and one pure Chinese specimen. The total number of domestic geese studied was 101 (see also Appendix S1, Table S1).

Molecular methods

The biological materials used were muscle, blood and feathers. DNA was extracted using either Qiagen DNeasy blood and tissue kit or a modified isolation method for museum feathers/skins following Laird *et al.* (1991).

We amplified a 1249 bp sequence that contains the mitochondrial control region flanked by the complete tRNA-Glu gene upstream from control region and a partial sequence of the tRNA-Phe gene at the 3' end. This fragment is labelled 'CR' in this paper. In order to avoid amplifying Numts, we utilised mitochondria specific primers that amplify the whole CR in two overlapping fragments previously developed by Ruokonen *et al.* (2000) (see also Appendix S1). The 5' end was amplified with primers L16642 5' ACC CCA TAA TAC GGC GAA GGA TT and H411ANX 5' GTA GAG RAT TGT

TGT TAR GGT 3' and the 3' end was amplified with primers L336ANX 5' AAC ATG AAT GCT CYA GGA CCA C 3' and H1248X 5' CAR CTT CAG TGC CAT GCT TT 3'. This produced two fragments overlapping by 30 bp.

The PCR reactions were performed in either 25 or 50 µl total reaction. The reaction contained 1 unit of DyNAzyme II DNA Polymerase (Finnzymes Abgene), 1X reaction buffer, 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.5 µM each primer and 50-200 ng DNA. The PCR reactions were performed in Veriti™ Thermal Cycler (Applied Biosystems). The PCR conditions were as follows: 94°C for 3 min; 30 × (94°C for 30 s, 52-59°C for 30 s, 72°C for 1 min); and 72°C for 10 min. The PCR products were checked on an agarose gel and in some cases the appropriately sized band was isolated from the gel and the PCR product extracted using the GelElute agarose gel extraction protocol (5 Prime). This was carried out in suspected cases of Numts visible as multiple bands.

Sequencing of both strands was conducted using Big Dye Terminator v. 3.1 Cycle Sequencing Kit (Applied Biosystems). The sequencing PCR was performed with the same primers as the amplification PCR. Sequences were edited in CodonCode Aligner (CodonCode Corporation, www.codoncode.com) and aligned with Muscle (Edgar 2004) implemented in Mega6 (Tamura *et al.* 2013).

Analyses

Diversity Indices

The number of polymorphic sites and number of different haplotypes were calculated using DNAsp 5.10 (Librado & Rozas 2009). The haplotype diversity (*h*) and pairwise nucleotide diversity

(π, Nei 1987) were also estimated for every breed and greylag goose population with Arlequin 3.5.1. (Excoffier & Lischer 2010).

In order to compare sequence diversity between wild and domestic geese, these two were treated as two separate groups, and within each group the average number of pairwise differences between populations (πXY) was calculated for all population pairs with Arlequin 3.5.1.2. The πXY values were used to calculate the mean value within the group and the mean value was divided by the length of the sequence to generate the average pairwise difference per base pair for each group.

Population Structure Analyses

For an analysis of molecular variance (AMOVA) the greylag sequences were assigned to populations based on breed (domestic birds) or geographic location (wild birds) and the populations were assigned to two groups; wild and domestic. The non-breed individuals were excluded from this analysis. The amount of genetic variance among groups, among populations within groups and within populations was calculated with Arlequin 3.5.1.2. The sampling locations for wild birds in Finland were combined into two populations called Northern and Southern Finland (Fig. 1) since most locations had only one or two individuals sampled. The two sampling sites in Denmark were also combined (Fig. 1) because one site had only one individual.

SAMOVA (Spatial Analysis of MOlecular Variance) was used to define homogeneous and maximally differentiated groups. SAMOVA is based on a simulated annealing algorithm (SAA) that aims to find the composition of K groups of populations and their associated FCT value (the proportion of genetic variation, which is attributed to differences between groups of populations) (Dupanloup *et*

al. 2002). The FCT value is maximised when the number of maximally differentiated groups is found. At the same time FSC (the extent of which populations are differentiated within groups) should approach zero (Dupanloup *et al.* 2002, Rodríguez-Robles *et al.* 2010).

Phylogenetic analyses

For phylogenetic analyses jModeltest2 (Guindon & Gascuel 2003, Darriba *et al.* 2012) was used to determine the nucleotide substitution model that best fits the data. Both AIC (Akaike Information Criteria) and BIC (Bayesian Information Criterion) supported the Hasekawa-Kishino-Yano model (Hasegawa *et al.* 1985) with gamma distribution and invariant sites (HKY+G+I). Based on this information, phylogenetic trees were constructed, one based on maximum likelihood and the other one based on Bayesian inference. Numt and *A. albifrons albifrons* sequence were used as outgroups (GenBank accession numbers AF159970 and AF159958, respectively).

The maximum likelihood tree was constructed with Mega 6.05 and 1000 bootstrap replications were applied. The Bayesian tree was constructed with MrBayes (Huelsenbeck & Ronquist 2001, Ronquist & Huelsenbeck 2003). We ran 4 independent runs with 1 000 000 generations with the burn in of 30% and sample frequency of 1000. We considered the chains converged when the average standard deviation of split frequencies was <0.01 and Potential Scale Reduction Factor (PSRF) was 1 or very close to 1 (0.99). Tracer v.1.6 was used to confirm convergence (Rambaut et al. 2014). ESS values that were over 200 across the runs were accepted and the trace files were visually checked to confirm the convergence. To visualise the trees, FigTree 1.4.1. (http://tree.bio.ed.ac.uk/software/figtree/) was used.

A minimum spanning network was made with Hapstar (Teacher & Griffiths 2011) based on pairwise distances between different haplotypes calculated with Arlequin 3.5.1.2.

Results

Genetic variation in greylag and domestic goose populations

CR contained 39 variable sites and 36 parsimony informative sites. In total we found 44 haplotypes, of which 9 haplotypes were found in domestic geese and 38 in wild greylag geese. We found two major haplotypes for domestic geese comprising 84% of the domestic geese studied (haplotypes D3, 53 individuals, and D4, 32 individuals; Table 2). Three of the haplotypes we found in domestic geese were shared with wild individuals from Scotland, the Netherlands and Iran.

Among the wild greylags the haplotype diversity (*h*) was very high (>0.8) in the eastern populations in the well-represented populations in Iran (Gilan 1; and Fereydunkenar 0.92; Fig. 2, Table S1). Among the European populations, the Netherlands showed similarly high haplotype diversity. The Nordic populations showed moderate to low haplotype diversities, the haplotype diversity being the highest in Southern Finland (0.48). The populations in Scotland, Northern Finland and Greece were invariant. For all greylags the haplotype diversity was 0.86. The domestic breed with highest haplotype diversity was Tula (1.0, although only based on 2 individuals) followed by Brecon Buff, Steinbacher, Scania Goose, West of England, Danish Landrace, Embden, Russian Grey, Tufted Roman and Sebastopol. Czech, Landes and Öland Goose were invariant. Most of the breeds had two haplotypes. The haplotype diversity for all domestic geese was 0.57. The non-breed domestics had higher haplotype diversities than most of the breeds in general (Denmark 0.67; Russia 0.60, UK 0.57 and Turkey 0.78).

The nucleotide diversity (π) values among the wild greylag geese showed highest diversities in Iran and the Netherlands in which is line with the haplotype diversities (Fig. 2, Table S1). The Nordic populations showed much lower nucleotide diversities, being highest in Southern Finland and Denmark (0.0023 and 0.0022, respectively). The nucleotide diversity for all greylags was 0.0064. The nucleotide diversities were consistently lower in domestic geese than the wild greylags, ranging from 0 to 0.0045 at the highest. If we only consider different breeds, Brecon Buff and Tula had the highest nucleotide diversities (0.0008) followed by Tufted Roman, Scania Goose, Steinbacher, Danish Landrace, Sebastopol, West of England, Embden and Russian Grey. The nucleotide diversity for all domestic geese was and 0.00054. The non-breed domestics had generally higher nucleotide diversities than individual breeds (nucleotide diversity for non-breed individuals varied between 0.0005 and 0.0045) with Turkish domestics having the highest value (0.0045).

For some of the geese sampled (10 wild, 1 domestic), only part of the control region was successfully sequenced, the first hypervariable region (HVR1; Appendix S1). The nucleotide and haplotype diversities for HVR1 show very similar trends to the CR dataset (Table S1). These are the only analyses carried out with the HVR1 data. All other analyses refer to the CR dataset.

In addition to population specific haplotype (*h*) and nucleotide diversities (π), we estimated the average number of pairwise differences between populations (πXY) and used it to calculate average number of pairwise differences per base pair among wild and domestic geese. This value showed much higher sequence divergence for greylag geese than domestic geese (0.0075 and 0.0006, respectively).

Population Structure

In the AMOVA the differentiation between wild greylag and domestic geese explained 35.29% of the variation observed. The differentiation between populations within groups accounted for 33.93% of variation and within population variation was 30.78%.

The SAMOVA results indicate population structure that can be matched with geography. Maximising FCT combined with FSC=0, there were 5-8 groups (Table 3). When K=5, the Iranian populations together with Kazakhstan form a group that also includes the Netherlands. The Finnish and Norwegian populations also form their own group but the rest of the populations stay as separate entities. When K=6, the Eastern-Dutch group is split into two, the Dutch population being grouped with the individuals from Fereydunkenar, Iran and the other Iranian population from Gilan is grouped with Kazakhstan. When K = 7, Netherlands and Fereydunkenar split but the other groupings remain the same. When K = 8, all the populations form their own groups except for the Finnish and Norwegian which stay together. Dupanloup et al. (2002) have stated that the following conditions need to be met for SAMOVA to have a high success rate in accurately recovering groups: low gene flow between groups (Nm 0.01) and within group gene flow 1000 times that of between group gene flow. Therefore, based on FSC and FCT values, we calculated the gene flow within and between groups (Nm intra and Nm inter, respectively; Wright 1931, Slatkin 1985) as well as their relative magnitude (Table 3). Within group gene flow was only maximally 83.9 times higher than between group gene flow (at K = 5). Therefore the groupings cannot be considered as conclusive, and are for guidance only.

Phylogeny of greylag and domestic geese

The Bayesian and maximum likelihood trees and the minimum spanning network supported similar phylogenies with six haplogroups (Figs 3, S1, S2); here we describe the Bayesian tree has two main clades with a major split between haplogroup A and all other haplogroups (B-F). The geographic distributions of each of these haplogroups is shown in Fig. 4. Haplogroup A is best represented in Dutch wild greylags, half of which belong to this haplogroup. It is also found in Iran and to lesser degree in Southern Finland and Denmark. Haplogroup B is in one single individual from the Netherlands. Haplogroup C is only found in Fereydunkenar, Iran and Lake Kulykol in Kazakhstan. Almost all the haplotypes that were found in domestics were in the haplogroup D and will refer to as the 'domestic haplogroup' from now on. The wild greylags that belonged to this haplogroup were from the Netherlands, Scotland and Denmark. Haplogroup E was dominant in Finland and Norway. It was also found in Denmark, the Netherlands, Greece and Gilan, Iran. Haplogroup F was found in both Iranian localities and Kazakhstan but also almost 25% of geese from the Netherlands and one individual from Denmark had a haplotype that belonged to this haplogroup. Three domestic geese from Turkey had haplotypes that belong to group F.

Discussion

Genetic variation in greylag goose populations and subspecies distribution

The prerequisite for using genetic information to untangle goose domestication is to understand how genetic variation is distributed among wild greylag goose populations. The genetic patterns observed in wild greylag goose populations across Europe potentially reflect both natural post-glacial colonization (Hewitt 1999) and human-mediated translocations of wild geese (Rooth 1971). Moreover, our results indicate past and probably still ongoing hybridization between wild greylag geese and domestic geese in multiple locations.

The haplotype and nucleotide diversities of wild greylag goose were generally highest in southeastern populations and in the Netherlands. The population sizes in our sample sites in Iran and Kazakhstan are not known but the population wintering around the Caspian Sea is estimated to consist of more than 100 000 individuals (Fox *et al.* 2010) which is consistent with the high genetic diversity observed. The distribution of diversity may reflect survival of the species in southeastern areas during the Last Glacial Maximum and loss of haplotypes during postglacial colonisation northwards (Hewitt 1999).

The high genetic diversity in the Netherlands is not in line with postglacial colonisation reducing diversity towards the north. However, goose introductions were carried out in Belgium in 1954 and in the Netherlands in 1962 (Rooth 1971). The geese introduced in Belgium were *A. a. rubrirostris* but the origin of geese introduced in the Netherlands is not known. The *rubrirostris*-type birds hybridised with *A. a. anser* and during 1960s and 1970s there were multiple observations of heavy, pink-billed geese with characteristics of *rubrirostris* observed on Atlantic flyway (Kuijken & Devos 1996). The *rubrirostris* morphological characteristics have subsequently reduced but the original breeding colony has expanded along the Dutch border to northern parts of East Flanders since the early 1980s. At the same time on the Dutch side of the border, the Zeeuwsch-Vlaanderen region has become colonised naturally by greylag geese (Kuijken & Devos 1996, Madsen *et al.* 1999), creating plenty of opportunities for hybridisation. The eastern origin of birds within this area is also supported by SAMOVA, which groups the individuals from the Netherlands with those sampled in Iran and Kazakhstan (Table 3).

Since no morphological data were available, we did not have prior information about the subspecies status of greylag goose samples used in this study, except inferences based on sampling location. A previous study defined the subspecies based on mitochondrial CR (Ruokonen *et al.* 2000) but our results suggest that genetic separation of different subspecies is not absolute, given that most major haplogroups are found on both sides of the suggested subspecies boundary. However, if the potential effect of introductions in the Netherlands is taken into account, the geographic distribution of different haplogroups supports some separation between the eastern and western subspecies. Haplogroup A comprises half of the individuals in the Netherlands. Moreover, haplogroup F is also frequent in the Netherlands. Both of these haplogroups A and F are absent or rare in other populations except those in Iran and Kazakhstan. It seems plausible that haplogroups A and F are more typical for the eastern subspecies as well as haplogroup C which was only found in Kazakhstan and Fereydunkenar, Iran. Finding individuals from the Netherlands with haplotypes belonging to haplogroups A and F suggests that the individuals, which were introduced to Belgium and the Netherlands were indeed the *A. a. rubrirostris* type and carried these haplotypes, which thereafter became very common in the Netherlands when the population expanded. The admixed nature of the Dutch population would also explain the high genetic diversity in this population. Haplogroup E is primarily associated with the western subspecies of greylag, being dominant in Nordic areas and present in the Netherlands. However, this association is not absolute as E haplotypes have been found in Greece and Iran, which are areas associated with the eastern subspecies.

Our data does not give any genetic support on the third subspecies *A. a. sylvestris*. There was no strong genetic divergence between Norwegian and Finnish population as according to the SAMOVA, Finnish and Norwegian populations were the last ones to be separated as the number of

possible groups was increased (Table 3) and a single haplotype (E2, Table 2) was shared by almost all the individuals from Norway and Finland. Rather, there is genetic discontinuity along a South-North axis, separating the Norwegian and Finnish populations from the Scottish and Danish. We suggest that the Scottish and Danish populations are impacted by hybridisation with domestic geese and birds of the eastern subspecies, which does not appear to have happened in Norway and Finland.

Genetic variation in domestic geese populations

Our results suggest very low mitochondrial diversity for the European domestic goose. The haplotype and nucleotide diversity estimates are approximately the same order of magnitude as those published for Chinese domestic duck breeds (He *et al.* 2008, Qu *et al.* 2009) but less than the estimates for domestic chicken (Liu *et al.* 2006, Kanginakudru *et al.* 2008). We analysed 101 domestic geese and found only 9 different haplotypes, 7 of which belonged to our domestic haplogroup D. Moreover, 84% of individuals were divided between two major haplotypes D3 and D4. D3 is the most common and widespread haplotype and the central haplotype of a starburst of very closely related haplotypes among domestic geese (Fig. S2), of which D4 is one (separated by one nucleotide substitution). It is possible that this low mitochondrial diversity may be due to a small number of individuals that contributed to the founding population in the early stages of domestication. It is interesting that we found multiple haplotypes in Turkey that were not found anywhere else among domestic geese including two haplotypes that belonged to haplogroup F instead of haplogroup D where all the other domestic haplotypes belong. It can be expected that genetic diversity is highest in the domestication centre and decreases with increasing distance from there (Medugorac *et al.* 2009). Since it has been suggested that the goose was domesticated in the vicinity of the eastern Mediterranean, Turkey might be the sampling location nearest to the

origin of domestication. More sampling is needed around the eastern Mediterranean and the Near East to confirm this result. However, caution is needed in using modern breeds to interpret domestication events (Larson *et al.* 2012) and it is desirable to genotype archaeological samples closer to the time of domestication. It can at least be said that modern breeds of geese in Europe come from a narrow genetic base. Whether this dates back to the domestication event or a more recent derivation of modern breeds from a narrow stock, is uncertain.

We know of no sequence analysis comparing the control region in the greylag goose and the swan goose, but a study of mitochondrial DNA cleavage patterns has shown that Chinese and European type domestic geese have different restriction fragment length polymorphism profiles (Shi *et al.* 2006). We sequenced three breeds (Kholmogor, Steinbacher and Tula) that, based on the literature, are crossbreds between Chinese and European geese. All shared the *A. anser* type mitochondrial haplotype indicating that if these truly are crossbreeds, female European domestic geese must have mated with Chinese type ganders. However, our sample sizes are small (Kholmogor and Tula both had n=2 and Steinbacher n=3) and further studies using both mitochondrial and nuclear markers are needed. We also genotyped one individual that was reported to be of the Chinese type (Lavender Chinese) but which again had a European type mitochondrial sequence, presumably through cross-breeding.

Hybridization among wild and domestic goose

The two major domestic haplotypes D3 and D4 were found also in wild individuals from the Netherlands and Scotland. Haplotype D3, found in 53% of the domestic individuals, was present in four wild individuals from the Netherlands. The other major haplotype, D4, recorded in 32% of domestics, was shared with the Scottish wild greylags. This is most reasonably explained by recent

hybridization. Hybridizations between domestics and greylags have been observed in Belgium (Kuijken & Devos 1996) and in the Netherlands (J. Ottenburghs, pers. comm.) giving rise to phenotypically hybrid individuals.

The genetic composition of the Danish population is, however, more complicated. The modern day Danish individuals all belong to haplogroup D that, with the exception of three individuals, is found in all the domestic individuals that we analysed. The Danish geese do not share a haplotype with the domestic geese that we studied. However, the most common haplotype that we found in Denmark (D2) is only separated by a single nucleotide substitution from the most common haplotype that we found in domestic geese (D3)(Fig. S2). As for Scotland and the Netherlands, hybridisation between domestic and wild geese may be the explanation for the Danish result, but involving domestic geese with an unusual haplotype based on current sampling.

In summary, our study is significant for being the first large scale analysis of genetic diversity of greylag goose and the genetic relationships of its subspecies. It is also the first attempt to decipher the relationships between greylag goose and its derivative, the European domestic goose. These data show unexpectedly complex relationships between and within wild greylags and domestic geese which sets up a foundation for further studies. The initial data presented here would benefit from future sampling of additional individuals from the southeastern parts of the distribution of greylag goose and also the use of nuclear markers and particularly genomic data. The analysis of ancient DNA from archaeological goose specimens would also be beneficial in providing a temporal resolution to the question of domestication, as has been successfully carried out on other species (Achilli *et al.* 2008, Kimura *et al.* 2011, Ottoni *et al.* 2013, Thalmann *et al.* 2013).

Acknowledgements

We dedicate this paper to our co-author, Minna Ruokonen, who is sorely missed. We thank Saija Ahonen, Jenni Harmoinen, Matti Heino, Hannele Parkkinen and Laura Törmälä for the laboratory work. The following people, societies and institutes are acknowledged for providing samples: Tomas Aarvak, Nina Bulatova, Arne Follestad, Islam Gündüz, Abolghasem Khaleghizadeh, Maarten Loonen, Svetlana Pavlova, Jan Wójcik, Alan Leitch; The Royal Society for the Protection of Birds (RSPB), Annita Logotheti; Society for the Protection of Prespa, Mikael Olsson; Svenska Lanthönsklubben, Helle Palmø; Ministeriet for Fødevarer, Landbrug og Fiskeri, all the British goose breeders contacted through The Goose Club, Finnish Game and Fisheries Research Institute, Zoological Museum; Natural History Museum of Denmark. We also thank two anonymous reviewers for valuable comments on the earlier version of the manuscript. We thank CSC – IT Center for Science in Finland for providing computing resources. Lastly, we thank the Academy of Finland for funding.

Conflict of interest

The authors declare no conflict of interest.

Author Contributions

M.E.H, M.R. and J.B.S designed the study. M.R., T.P., J.A. and J.B.S. supervised the study. M.E.H. analysed the data. M.E.H. wrote the manuscript. M.A. assisted with domestic goose sample collection. All authors excluding M.R. have read and edited the manuscript.

References

Achilli, A., Olivieri, A., Pellecchia, M., Uboldi, C., Colli, L., Al-Zahery, N., Accetturo, M., Pala, M., Kashani, B.H. & Perego, U.A. 2008, "Mitochondrial genomes of extinct aurochs survive in domestic cattle", *Current Biology,* vol. 18, no. 4, pp. R157-R158.

Albarella, U. 2005, "Alternate fortunes? The role of domestic ducks and geese from Roman to Medieval times in Britain" in *Documenta Archaeobiologiae III. Feathers, Grit and Symbolism*, ed. Grupe, G. & Peters, J., pp. 249-258.

Beja-Pereira, A., Caramelli, D., Lalueza-Fox, C., Vernesi, C., Ferrand, N., Casoli, A., Goyache, F., Royo, L.J., Conti, S., Lari, M., Martini, A., Ouragh, L., Magid, A., Atash, A., Zsolnai, A., Boscato, P., Triantaphylidis, C., Ploumi, K., Sineo, L., Mallegni, F., Taberlet, P., Erhardt, G., Sampietro, L., Bertranpetit, J., Barbujani, G., Luikart, G. & Bertorelle, G. 2006, "The origin of European cattle: Evidence from modern and ancient DNA", *Proceedings of the National Academy of Sciences, USA,* vol. 103, no. 21, pp. 8113-8118.

Bollongino, R., Burger, J., Powell, A., Mashkour, M., Vigne, J.D. & Thomas, M.G. 2012, "Modern taurine cattle descended from small number of Near-Eastern founders", *Molecular Biology and Evolution,* vol. 29, no. 9, pp. 2101-2104.

Bruford, M., Bradley, D. & Luikart, G. 2003, "DNA markers reveal the complexity of livestock domestication", *Nature Reviews Genetics,* vol. 4, no. 11, pp. 900-910.

Cramp, S. & Simmons, K.E.L. (eds) 1977, *Handbook of the birds of Europe, the Middle East, and North Africa: the birds of the Western Palearctic. Vol. 1: Ostrich-Ducks.*, Oxford University Press, New York.

Crawford, R.D. 1984, "Goose" in *Evolution of domesticated animals*, ed. I.L. Mason, Longman, London and New York, pp. 345-349.

Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. 2012, "jModelTest 2: more models, new heuristics and parallel computing", *Nature Methods,* vol. 9, no. 8, pp. 772.

Delacour, J. 1954, *The waterfowl of the world. Vol. 1,* Country Life Ltd., London.

Dupanloup, I., Schneider, S. & Excoffier, L. 2002, "A simulated annealing approach to define the genetic structure of populations", *Molecular Ecology,* vol. 11, no. 12, pp. 2571-2581.

Edgar, R.C. 2004, "MUSCLE: multiple sequence alignment with high accuracy and high throughput", *Nucleic Acids Research,* vol. 32, no. 5, pp. 1792-1797.

Excoffier, L. & Lischer, H.E. 2010, "Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows", *Molecular Ecology Resources,* vol. 10, no. 3, pp. 564-567.

Fox, A.D., Ebbinge, B.S., Mitchell, C., Heinicke, T., Aarvak, T., Colhoun, K., Clausen, P., Dereliev, S., Faragó, S. & Koffijberg, K. 2010, "Current estimates of goose population sizes in western Europe, a gap analysis and an assessment of trends", *Ornis Svecica,* vol. 20, no. 3-4, pp. 115-127.

Fumihito, A., Miyake, T., Takada, M., Shingu, R., Endo, T., Gojobori, T., Kondo, N. & Ohno, S. 1996, "Monophyletic origin and unique dispersal patterns of domestic fowls", *Proceedings of the National Academy of Sciences, USA,* vol. 93, no. 13, pp. 6792-6795.

Guindon, S. & Gascuel, O. 2003, "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood", *Systematic Biology,* vol. 52, no. 5, pp. 696-704.

Harper, J. 1972, "The tardy domestication of the duck", *Agricultural History,* vol. 46, no. 3, pp. 385-389.

Hasegawa, M., Kishino, H. & Yano, T. 1985, "Dating of the human-ape splitting by a molecular clock of mitochondrial DNA", *Journal of Molecular Evolution,* vol. 22, no. 2, pp. 160-174.

He, D.Q., Zhu, Q., Chen, S.Y., Wang, H.Y., Liu, Y.P. & Yao, Y.G. 2008, "A homogenous nature of native Chinese duck matrilineal pool", *BMC Evolutionary Biology,* vol. 8, no. 298.

Hewitt, G.M. 1999, "Post-glacial re-colonization of European biota", *Biological Journal of the Linnean Society,* vol. 68, no. 1-2, pp. 87-112.

Hiendleder, S., Kaupe, B., Wassmuth, R. & Janke, A. 2002, "Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies", *Proceedings of the Royal Society B: Biological Sciences,* vol. 269, no. 1494, pp. 893-904.

Huelsenbeck, J.P. & Ronquist, F. 2001, "MRBAYES: Bayesian inference of phylogenetic trees", *Bioinformatics (Oxford, England),* vol. 17, no. 8, pp. 754-755.

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,* vol. 8, no. 174.

Kear, J. 1990, *Man and Wildfowl,* T & AD Poyser Ltd, London.

Kimura, B., Marshall, F.B., Chen, S., Rosenbom, S., Moehlman, P.D., Tuross, N., Sabin, R.C., Peters, J., Barich, B., Yohannes, H., Kebede, F., Teclai, R., Beja-Pereira, A. & Mulligan, C.J. 2011, "Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication", *Proceedings of the Royal Society B-Biological Sciences,* vol. 278, no. 1702, pp. 50-57.

Kuijken, E. & Devos, K. 1996, "The status of the Greylag Goose *Anser anser* in Flanders, Belgium", *Wetlands International Goose Specialist Group Bulletin,* , no. 8, pp. 26-28.

Laird, P., Zijderveld, A., Linders, K., Rudnicki, M., Jaenisch, R. & Berns, A. 1991, "Simplified mammalian DNA isolation procedure", *Nucleic Acids Research,* vol. 19, no. 15, pp. 4293-4293.

Larson, G., Karlsson, E.K., Perri, A., Webster, M.T., Ho, S.Y.W., Peters, J., Stahl, P.W., Piper, P.J., Lingaas, F., Fredholm, M., Comstock, K.E., Modiano, J.F., Schelling, C., Agoulnik, A.I., Leegwater, P.A., Dobney, K., Vigne, J., Vilà, C., Andersson, L. & Lindblad-Toh, K. 2012, "Rethinking dog domestication by integrating genetics, archeology, and biogeography", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 109, no. 23, pp. 8878-8883.

Librado, P. & Rozas, J. 2009, "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data", *Bioinformatics (Oxford, England),* vol. 25, no. 11, pp. 1451-1452.

Liu, Y.Y., Wu, G.G., Yao, Y.Y., Miao, Y.Y., Luikart, G.G., Baig, M.M., Beja-Pereira, A.A., Ding, Z.Z., Palanichamy, M.G.M. & Zhang, Y.Y. 2006, "Multiple maternal origins of chickens: out of the Asian jungles", *Molecular Phylogenetics and Evolution,* vol. 38, no. 1, pp. 12-19.

Loftus, R., MacHugh, D., Bradley, D., Sharp, P. & Cunningham, P. 1994, "Evidence for two independent domestications of cattle", *Proceedings of the National Academy of Sciences, USA,* vol. 91, no. 7, pp. 2757-2761.

Luikart, G., Gielly, L., Excoffier, L., Vigne, J., Bouvet, J. & Taberlet, P. 2001, "Multiple maternal origins and weak phylogeographic structure in domestic goats", *Proceedings of the National Academy of Sciences, USA,* vol. 98, no. 10, pp. 5927-5932.

Madge, S. & Burn, H. 1988, *Wildfowl: an identification guide to the ducks, geese and swans of the world,* Christopher Helm, London.

Madsen, J., Cracknell, G. & Fox, T. 1999, *Goose populations of the Western Palearctic: a review of status and distribution,* National Environmental Research Institute Rønde.

Meadows, J.R., Hiendleder, S. & Kijas, J.W. 2011, "Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel", *Heredity,* vol. 106, no. 4, pp. 700-706.

Medugorac, I., Medugorac, A., Russ, I., Veit-Kensch, C.E., Taberlet, P., Luntz, B., Mix, H.M. & Foerster, M. 2009, "Genetic diversity of European cattle breeds highlights the conservation value of traditional unselected breeds with high effective population size", *Molecular Ecology,* vol. 18, no. 16, pp. 3394-3410.

Miao, Y., Peng, M., Wu, G., Ouyang, Y., Yang, Z., Yu, N., Liang, J., Pianchou, G., Beja-Pereira, A. & Mitra, B. 2013, "Chicken domestication: an updated perspective based on mitochondrial genomes", *Heredity,* vol. 110, no. 3, pp. 277-282.

Naderi, S., Rezaei, H.R., Pompanon, F., Blum, M.G., Negrini, R., Naghash, H.R., Balkiz, O., Mashkour, M., Gaggiotti, O.E., Ajmone-Marsan, P., Kence, A., Vigne, J.D. & Taberlet, P. 2008, "The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 105, no. 46, pp. 17659-17664.

Nei, M. 1987, *Molecular evolutionary genetics,* Columbia University Press, New York.

Ottoni, C., Girdland Flink, L., Evin, A., Geörg, C., De Cupere, B., Van Neer, W., Bartosiewicz, L., Linderholm, A., Barnett, R., Peters, J., Decorte, R., Waelkens, M., Vanderheyden, N., Ricaut, F., Çakirlar, C., Çevik, Ö, Hoelzel, A.R., Mashkour, M., Mohaseb Karimlu, A.F., Sheikhi Seno, S., Daujat, J., Brock, F., Pinhasi, R., Hongo, H., Perez-Enciso, M., Rasmussen, M., Frantz, L., Megens, H., Crooijmans, R., Groenen, M., Arbuckle, B., Benecke, N., Strand Vidarsdottir, U., Burger, J., Cucchi, T., Dobney, K. & Larson, G. 2013, "Pig domestication and human-mediated dispersal in Western Eurasia revealed through ancient DNA and geometric morphometrics", *Molecular Biology and Evolution,* vol. 30, no. 4, pp. 824-832.

Pang, J.F., Kluetsch, C., Zou, X.J., Zhang, A.B., Luo, L.Y., Angleby, H., Ardalan, A., Ekstrom, C., Skollermo, A., Lundeberg, J., Matsumura, S., Leitner, T., Zhang, Y.P. & Savolainen, P. 2009, "mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves", *Molecular Biology and Evolution,* vol. 26, no. 12, pp. 2849-2864.

Qu, L., Liu, W., Yang, F., Hou, Z., Zheng, J., Xu, G. & Yang, N. 2009, "Origin and domestication history of Peking ducks deltermined through microsatellite and mitochondrial marker analysis", *Science in China Series C : Life Sciences,* vol. 52, no. 11, pp. 1030-5.

Rambaut, A., Suchard, M., Xie, D. & Drummond, A. 2014, *Tracer v1.6,* Available from http://beast.bio.ed.ac.uk/Tracer.

Rodríguez-Robles, J.A., Jezkova, T. & Leal, M. 2010, "Climatic stability and genetic divergence in the tropical insular lizard *Anolis krugi*, the Puerto Rican 'Lagartijo Jardinero de la Montaña'", *Molecular Ecology,* vol. 19, no. 9, pp. 1860-1876.

Ronquist, F. & Huelsenbeck, J.P. 2003, "MrBayes 3: Bayesian phylogenetic inference under mixed models", *Bioinformatics (Oxford, England),* vol. 19, no. 12, pp. 1572-1574.

Rooth, J. 1971, "The occurrence of greylag goose *Anser anser* in western parts of its distribution area", *Ardea,* vol. 59, no. 12, pp. 17-27.

Ruokonen, M., Kvist, L. & Lumme, J. 2000, "Close relatedness between mitochondrial DNA from seven *Anser* goose species", *Journal of Evolutionary Biology,* vol. 13, no. 3, pp. 532-540.

Savolainen, P., Zhang, Y., Luo, J., Lundeberg, J. & Leitner, T. 2002, "Genetic evidence for an East Asian origin of domestic dogs", *Science (New York, N.Y.),* vol. 298, no. 5598, pp. 1610-1613.

Scott, D. & Rose, P. 1996, *Atlas of Anatidae populations in Africa and Western Eurasia,* Wageningen: Wetlands International.

Shi, X., Wang, J., Zeng, F. & Qiu, X. 2006, "Mitochondrial DNA cleavage patterns distinguish independent origin of Chinese domestic geese and western domestic geese", *Biochemical Genetics,* vol. 44, no. 5-6, pp. 237-245.

Slatkin, M. 1985, "Gene flow in natural populations", *Annual Review of Ecology and Systematics,* vol. 16, pp. 393-430.

Speller, C.F., Kemp, B.M., Wyatt, S.D., Monroe, C., Lipe, W.D., Arndt, U.M. & Yang, D.Y. 2010, "Ancient mitochondrial DNA analysis reveals complexity of indigenous North American turkey domestication", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 107, no. 7, pp. 2807-2812.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. 2013, "MEGA6: Molecular evolutionary genetics analysis version 6.0", *Molecular Biology and Evolution,* vol. 30, no. 12, pp. 2725-2729.

Tapio, M., Marzanov, N., Ozerov, M., Cinkulov, M., Gonzarenko, G., Kiselyova, T., Murawski, M., Viinalass, H. & Kantanen, J. 2006, "Sheep mitochondrial DNA variation in European, Caucasian, and Central Asian areas", *Molecular Biology and Evolution,* vol. 23, no. 9, pp. 1776-1783.

Teacher, A. & Griffiths, D. 2011, "HapStar: automated haplotype network layout and visualization", *Molecular Ecology Resources,* vol. 11, no. 1, pp. 151-153.

Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V.J., Sawyer, S.K., Greenfield, D.L., Germonpre, M.B., Sablin, M.V., Lopez-Giraldez, F., Domingo-Roura, X., Napierala, H., Uerpmann, H.P., Loponte, D.M., Acosta, A.A., Giemsch, L., Schmitz, R.W., Worthington, B., Buikstra, J.E., Druzhkova, A., Graphodatsky, A.S., Ovodov, N.D., Wahlberg, N., Freedman, A.H., Schweizer, R.M., Koepfli, K.P., Leonard, J.A., Meyer, M., Krause, J., Paabo, S., Green, R.E. & Wayne, R.K. 2013, "Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs", *Science (New York, N.Y.),* vol. 342, no. 6160, pp. 871-874.

Troy, C.S., MacHugh, D.E., Bailey, J.F., Magee, D.A., Loftus, R.T., Cunningham, P., Chamberlain, A.T., Sykes, B.C. & Bradley, D.G. 2001, "Genetic evidence for Near-Eastern origins of European cattle", *Nature,* vol. 410, no. 6832, pp. 1088-1091.

Vigilant, L., Pennington, R., Harpending, H., Kocher, T.D. & Wilson, A.C. 1989, "Mitochondrial DNA sequences in single hairs from a southern African population", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 86, no. 23, pp. 9350-9354.

Wang, C.M., Way, T.D., Chang, Y.C., Yen, N.T., Hu, C.L., Nien, P.C., Jea, Y.S., Chen, L.R. & Kao, J.Y. 2010, "The origin of the White Roman Goose", *Biochemical Genetics,* vol. 48, no. 11-12, pp. 938-943.

Wenink, P.W., Baker, A.J. & Tilanus, M.G. 1993, "Hypervariable-control-region sequences reveal global population structuring in a long-distance migrant shorebird, the Dunlin (*Calidris alpina*)", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 90, no. 1, pp. 94-98.

Wright, S. 1931, "Evolution in Mendelian Populations", *Genetics,* vol. 16, no. 2, pp. 97-159.

Zeuner, F., E. 1963, *A History of Domesticated Animals,* Harper & Row, Publishers, New York and Evanston.

Table 1 Domestic breeds that were analysed for this study, including sampling locations and putative geographic origin of the breed as well as the species the breed was domesticated from.

Breed	Sampling location	Breed origin	Ancestral species
Brecon Buff	England & Wales, UK	South Wales, UK	*A. anser*
Buff	Wales, UK	Northern Europe	*A. anser*
Czech	Wales, UK	Czech Republic	*A. anser*
Danish Landrace	Denmark	Denmark	*A. anser*
Diepholzer	Wales, UK	Diepholz, Germany	*A. anser*
Embden	England & Wales, UK	Emden, Germany	*A. anser*
Emporda	Wales, UK	Catalunya, Spain	*A. anser*
Kholmogor	Moscow, Russia	Central Chernozem Region, Russia	*A. anser* x *A. cygnoides*[1]
Landes	England, UK	Landes region, France	*A. anser*
Lavender Chinese	Wales, UK	China	*A. cygnoides*
Russian Grey	Wales, UK	Unknown	*A. anser*
Scania goose	Sweden	Scania, Sweden	*A. anser*
Sebastopol	Wales, UK	Southeastern Europe, region around Black Sea	*A. anser*
Steinbacher	Wales, UK	Thuringen area, Germany	*A. anser* x *A. cygnoides*[1]
Tufted Roman	Wales, UK	Danube valley, Europe	*A. anser*
Tula	Moscow, Russia	Tula region, Russia	*A. anser* x *A. cygnoides*[1]
West of England	England & Wales, UK	West of England, UK	*A. anser*
Öland goose	Sweden	Öland, Småland & Northern Scania, Sweden	*A. anser*
Domestic DK	Denmark	Unknown	*A. anser*
Domestic UK	United Kingdom	Unknown	*A. anser*
Domestic FR	France	Unknown	*A. anser*
Domestic PO	Poland	Unknown	*A. anser*
Domestic RU	Russia	Unknown	*A. anser*
Domestic TR	Turkey	Unknown	*A. anser*

[1] http://www.ashtonwaterfowl.net/geese_two.htm

Table 2 Distribution of haplotypes across populations of greylag geese and domestic geese organised by haplogroups.

Population	Abbreviation for population	Haplogroup A	Haplogroup B	Haplogroup C	Haplogroup D	Haplogroup E	Haplogroup F
Orkney, Scotland (n=3)	SC				D4 (3)		
The Netherlands (n=46)	NL	A2 (13), A6 (4), A7 (2), A8 (1), A9 (1), A11 (1), A12 (1)	B1 (1)		D3 (4)	E8 (5), E9 (2)	F1 (8), F7 (1), F8 (1), F10 (1)
Vega, Norway (n=9)	VNO					E2 (8), E5 (1)	
Smøla, Norway (n=10)	SNO					E2 (7), E4 (3)	
Northern Finland (n=9)	NFI					E2 (9)	
Southern Finland (n=47)	SFI	A2 (2), A3 (2)				E1 (9), E2 (33), E6 (1)	
Denmark (n=20)	DK	A1 (1)			D1 (1), D2 (17)		F1 (1)
Greece (n=6)	GR					E7 (6)	
Fereydunkenar, Iran (n=9)	FIR	A2 (1), A5 (2)		C2 (1), C3 (2)			F2 (2), F3 (1)
Gilan, Iran (n=6)	GIR	A4 (1), A10 (1)				E3 (1), E6 (1)	F4 (1), F9 (1)
Lake Kulykol, Kazakhstan (n=2)	KZ			C1 (1)			F6 (1)
Brecon Buff (n=5)	BB				D3 (2), D4 (2), D6 (1)		
Buff (n=1)	BU				D3 (1)		
Sebastopol (n=8)	SEB				D3 (7), D5 (1)		
Czech (n=5)	CZE				D3 (5)		
Diepholzer (n=1)	DH				D3 (1)		
Domestic DK (n=3)	DDK				D3 (2), D4 (1)		
Domestic FR (n=1)	DFR				D3 (1)		
Domestic PO (n=1)	DPO				D4 (1)		
Domestic RU (n=5)	DRU				D3 (3), D4 (2)		
Domestic TR (n=11)	DTR				D3 (2), D8 (1), D9 (5)		F4 (2), F5 (1)
Domestic UK (n=7)	DUK				D3 (4), D4 (3)		
Embden (n=5)	EMB				D3 (1), D4 (4)		
Emporda (n=1)	EMP				D3 (1)		
Grey Landrace (n=13)	GL				D3 (9), D4 (4)		
Kholmogor (N=2)	KHO				D4 (2)		
Landes (n=2)	LD				D3 (2)		
Lavender Chinese (n=1)	LCH				D3 (1)		
Russian Grey (n=5)	RG				D3 (4), D4 (1)		
Scania Goose (n=5)	SG				D3 (3), D4 (2)		
Steinbacher (n=3)	ST				D3 (1), D4 (2)		
Tufted Roman (n=5)	TRO				D3 (1), D5 (4)		
Tula (n=2)	TUL				D3 (1), D7 (1)		
West of England (n=4)	WE				D3 (1), D4 (3)		
Oland Goose (n=5)	OG				D4 (5)		

Table 3 SAMOVA on greylag geese with group composition according to varying number of groups (K) and providing variance values (FSC, FST, FCT) as well as relative magnitude of intra- and inter-population gene flow.

K	Group composition	FSC	FST	FCT	Intra/inter
2	DK	0.326	0.724	0.591	2.994
	S-FI, N-FI, GR, KZ, S-NO, V-NO, G-IR, F-IR, NL, SC				
3	SC	0.306	0.710	0.583	3.167
	S-FI, N-FI, GR, KZ, S-NO, V-NO, G-IR, F-IR, NL				
	DK				
4	KZ	0.309	0.696	0.560	2.847
	DK				
	SC				
	S-FI, N-FI, GR, S-NO, V-NO, G-IR, F-IR, NL				
5	KZ, G-IR, F-IR, NL	0.015	0.569	0.562	83.949
	GR				
	S-FI, N-FI, S-NO, V-NO				
	DK				
	SC				
6	S-FI, N-FI, S-NO, V-NO	-0.016	0.558	0.565	-81.294
	DK				
	F-IR, NL				
	KZ, G-IR				
	SC				
	GR				
7	NL	-0.046	0.548	0.568	-30.186
	DK				
	F-IR				
	KZ, G-IR				
	S-FI, N-FI, S-NO, V-NO				
	GR				
	SC				
8	F-IR	-0.043	0.548	0.566	-31.963
	DK				
	NL				
	S-FI, N-FI, S-NO, V-NO				
	G-IR				
	KZ				
	GR				
	SC				
9	N-FI, S-NO, V-NO	-0.104	0.516	0.561	-13.537
	KZ				
	DK				
	GR				
	G-IR				
	S-FI				
	F-IR				
	NL				
	SC				
10	DK	-0.122	0.511	0.564	-11.904
	GR				
	N-FI, V-NO				
	NL				
	KZ				
	S-FI				
	G-IR				
	F-IR				
	S-NO				
	SC				

DK=Denmark; N-FI=Northern Finland; S-FI=Southern Finland; GR=Greece; F-IR=Fereydunkenar, Iran; G-IR=Gilan, Iran; NL=Netherlands; S-NO= Smøla, Norway; V-NO=Vega, Norway; KZ=Kazakhstan; SC=Scotland.

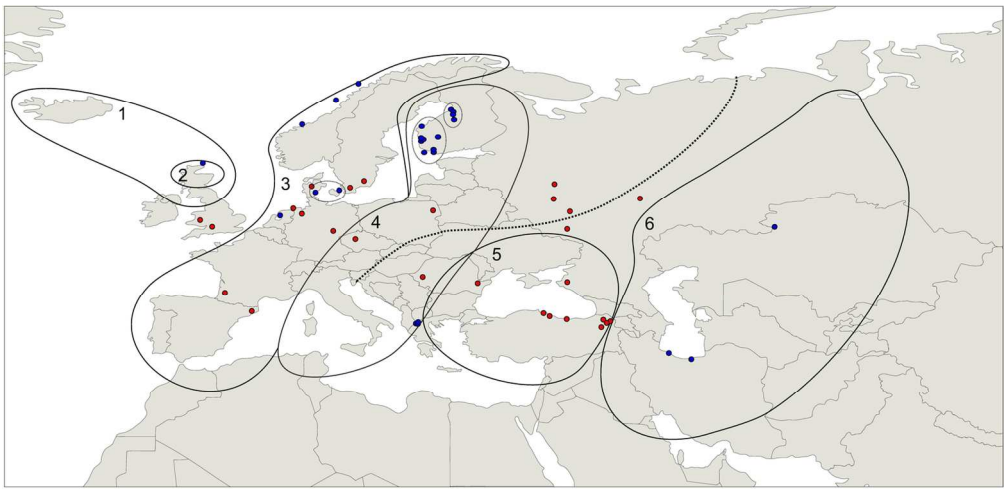

Figure 1 Map showing the greylag goose populations identified by Scott & Rose (1996) with numbering as described in the text. The approximate subspecies boundary is shown as a dashed line (modified from Scott & Rose 1996). Blue circles indicate the sampling locations for greylag geese; some have been grouped into 'Denmark', 'Southern Finland' and 'Northern Finland', as shown (Table S1). Red circles show the putative origin of domestic breeds or in case the origin is not known, the sampling location (Table 1).
142x68mm (300 x 300 DPI)

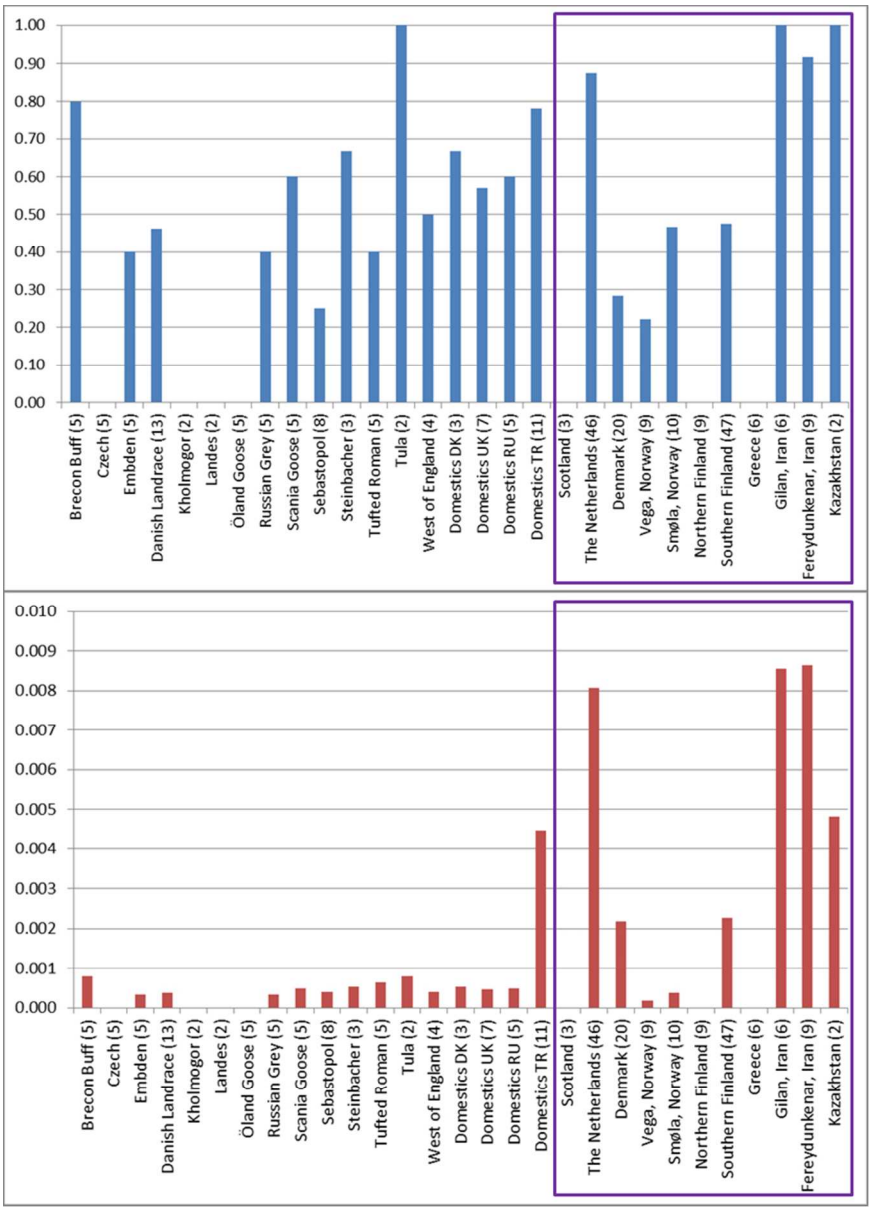

Figure 2 Haplotype diversity (h, above) and nucleotide diversity (π, below) for greylag and domestic goose populations. Number of specimens in parentheses. Greylag populations are inside the purple box.
140x194mm (150 x 150 DPI)

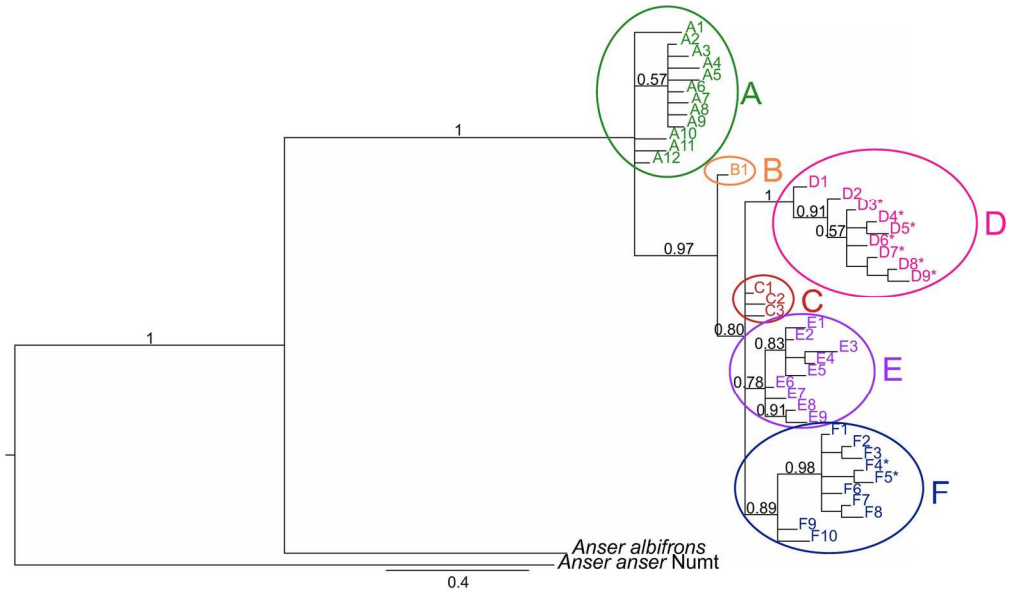

Figure 3 Bayesian tree of the haplotypes detected. Posterior probabilities above 0.5 are shown. Haplotypes shared by domestic geese are marked with an asterisk.
164x95mm (300 x 300 DPI)

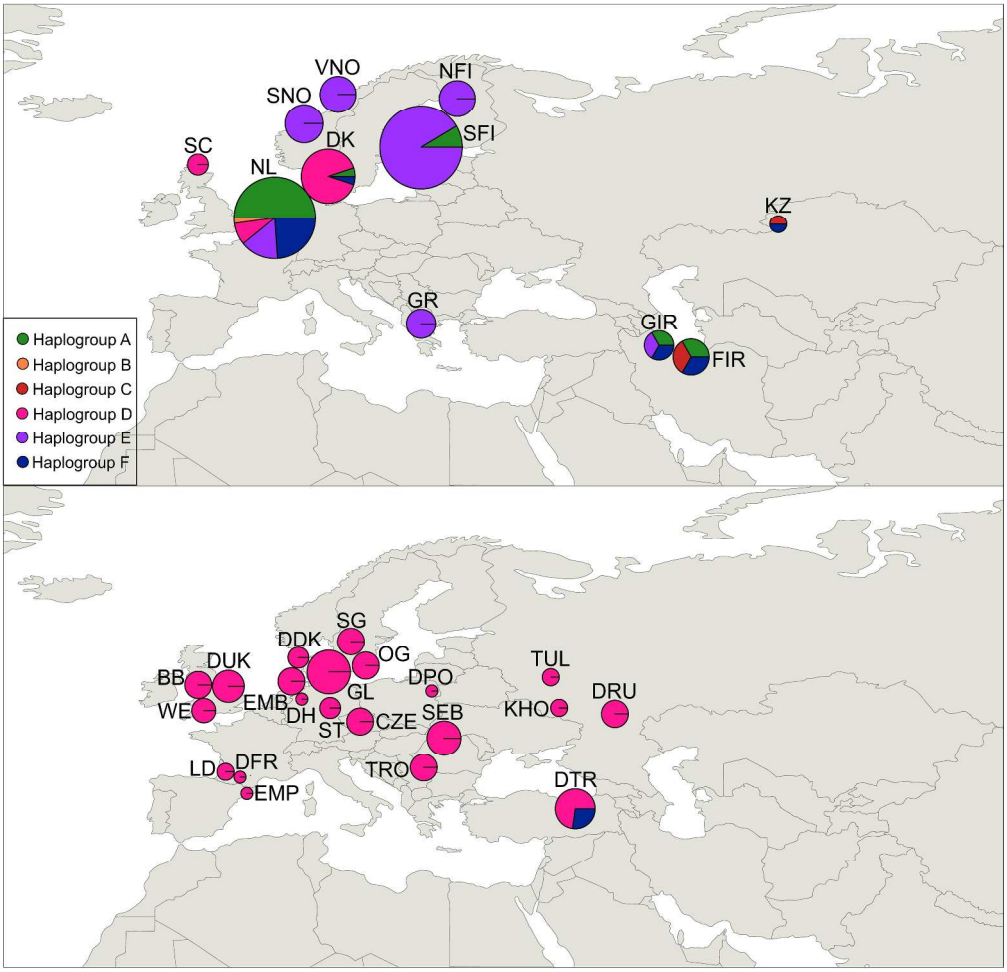

Figure 4 Haplotype distribution within greylag and domestic goose populations. The maps show the proportions of different haplotypes within each population at the haplogroup level for greylags (upper map) and for domestics (lower map). Abbreviations next to pie charts indicate population of origin (Table 3). The domestic goose breeds BU, LCH and RG are not mapped; the locations of origin of BU and RG are uncertain, LCH originated in Asia, off the map. Colour coding follows Figure 3.
274x264mm (300 x 300 DPI)