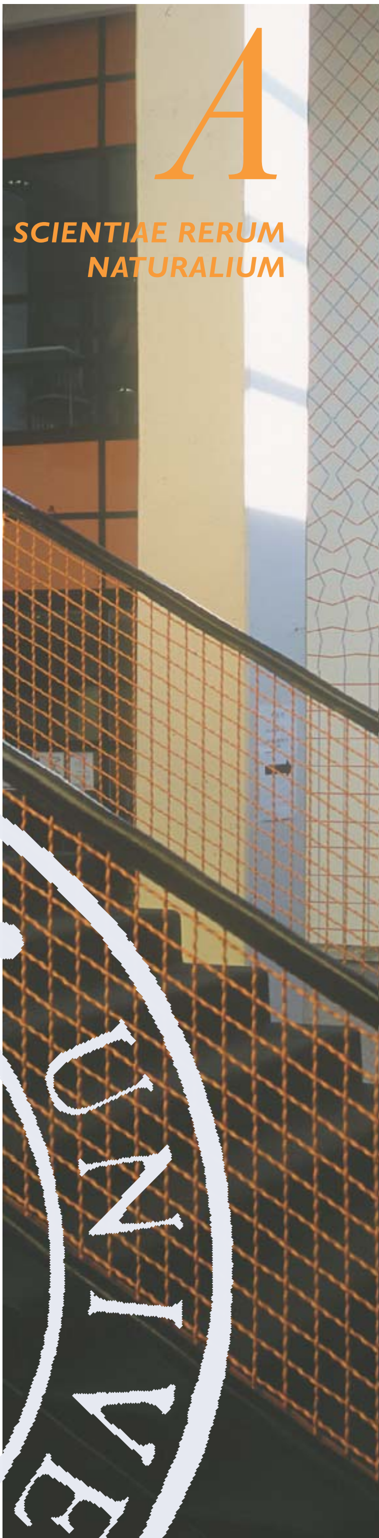
OULU 2017
A 692

ACTA UNIVERSITATIS OULUENSIS



A
SCIENTIAE RERUM
NATURALIUM

Marja Heikkinen

THE DOMESTICATION HISTORY OF THE EUROPEAN GOOSE

A GENOMIC PERSPECTIVE

*UNIVERSITY OF OULU GRADUATE SCHOOL;
UNIVERSITY OF OULU,
FACULTY OF SCIENCE*

ACTA UNIVERSITATIS OULUENSIS
A Scientiae Rerum Naturalium 692

MARJA HEIKKINEN

THE DOMESTICATION HISTORY OF THE EUROPEAN GOOSE
A genomic perspective

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Health and Biosciences of the University of Oulu for public defence in Kaljusensali (KTK112), Linnanmaa, on 20 June 2017, at 12 noon

UNIVERSITY OF OULU, OULU 2017

Acta Univ. Oul. A 692, 2017

Supervised by
Professor Jouni Aspi
Professor Jeremy Searle
Docent Tanja Pyhäjärvi
Docent Minna Ruokonen

Reviewed by
Assistant Professor Niclas Backström
Doctor Alain C. Frantz

Opponent
Professor Johanna Vilkki

ISBN 978-952-62-1596-9 (Paperback)
ISBN 978-952-62-1597-6 (PDF)

ISSN 0355-3191 (Printed)
ISSN 1796-220X (Online)

Cover Design
Raimo Ahonen

JUVENES PRINT
TAMPERE 2017

Heikkinen, Marja, The domestication history of the European goose. A genomic perspective
University of Oulu Graduate School; University of Oulu, Faculty of Science
Acta Univ. Oul. A 692, 2017
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

Animal domestication is a complex evolutionary process. Multiple forces influence the genetic variation of the species under domestication and leave their mark on the genome of the species. The European domestic goose is an economically and culturally important species, but knowledge about the domestication history of the species has been lacking. My doctoral thesis has focused on elucidating the genetic background of goose domestication using mitochondrial control region sequences and nuclear single nucleotide polymorphisms (SNPs). By comparing the patterns of genetic diversity observed in the greylag goose (*Anser anser*) and its descendant European domestic geese, I was able to conclude that genetic diversity has decreased in domestic geese following the domestication albeit being still relatively high. In addition, admixture of populations increased the genetic diversity in both greylag geese and domestic geese. The results also confirmed that greylag geese and domestic geese hybridise in certain locations. What is more, many breeds of European domestic geese shared a substantial amount of ancestry with Chinese domestic geese, domesticated from the swan goose (*Anser cygnoid*). While the timing and location of goose domestication remains unresolved, the results do not disagree with the suggested origin of domestication in the Eastern Mediterranean. More sampling in this region would be needed to further investigate the matter. Lastly, multiple regions in the goose genome have been targeted by selection which is likely to have contributed to phenotypic divergence of greylag and domestic geese, but the functional basis of these differences needs further investigation.

Keywords: Anser anser, domestication, European domestic goose, genetic diversity, greylag goose, hybridisation, population structure, selection

Heikkinen, Marja, Eurooppalaisen hanhen kesytyshistoria. Genominlaajuinen näkökulma
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Luonnontieteellinen tiedekunta
Acta Univ. Oul. A 692, 2017
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Eläinlajin kesyttäminen on monimutkainen evolutiivinen prosessi. Useat geneettiset tekijät vaikuttavat kesytettävän lajin perinnöllisen monimuotoisuuden määrään ja jättävät lajin perimään jälkensä. Eurooppalainen kesyhanhi on kulttuurillisesti ja taloudellisesti merkittävä laji, mutta tieto sen kesytyshistoriasta on puutteellista. Väitöskirjassani olen keskittynyt tutkimaan hanhen kesytyksen perinnöllistä taustaa käyttäen apuna mitokondrio-DNA:n kontrollialueen sekvenssejä ja yhden emäksen polymorfismeja. Kun vertailin perinnöllisen monimuotoisuuden jakautumista merihanhissa (*Anser anser*) ja eurooppalaisissa kesyhanhissa, pystyin toteamaan, että perinnöllinen monimuotoisuus on kesytyksen seurauksena vähentynyt kesyhanhissa, mutta se on edelleen suhteellisen korkeaa. Lisäksi risteytyminen muiden populaatioiden kanssa lisäsi perinnöllistä monimuotoisuutta sekä meri- että kesyhanhissa. Tulokset myös vahvistivat, että meri- ja kesyhanhet risteytyvät paikoitellen keskenään. Tämän lisäksi moniin eurooppalaisiin kesyhanhirotuihin on kohdistunut geenivirtaa kiinalaisesta kesyhanhesta, joka on kesytetty joutsenhanhesta (*Anser cygnoid*). Saadut tulokset vastaavat aiempia näkemyksiä, joiden mukaan hanhi kesytettiin Välimeren idänpuoleisilla alueilla, kanssa, mutta kesytyksen ajankohdan ja paikan tarkempi selvittäminen vaatii vielä lisätutkimuksia ja lisää näytteitä tältä alueelta. Lopuksi voidaan todeta, että useat alueet hanhen perimässä osoittivat merkkejä valinnasta, joka on todennäköisesti vaikuttanut meri- ja kesyhanhien välisiin fenotyyppisiin eroihin, mutta erojen funktionaalinen tausta vaatii lisätutkimuksia.

Asiasanat: Anser anser, eurooppalainen kesyhanhi, geneettinen monimuotoisuus, hybridisaatio, kesyttäminen, merihanhi, populaatiorakenne, valinta

A wild goose never laid a tame egg.
Gaelic proverb

In memory of Minna Ruokonen.

Acknowledgements

My journey towards PhD has been a rollercoaster ride in so many levels that my mind hasn't settled in yet. I can't believe I actually managed to pull this thing off. Hopefully, in retrospect, I can appreciate that I did. Nevertheless, I wouldn't have been able to do this on my own and I wish to express my gratitude to the people who contributed to my work one way or another.

I want to start off by thanking my supervisors. My first bittersweet thanks go to late Minna Ruokonen, who was the primus motor of this project. I wish you could have seen what came out of it. I certainly would have needed your expertise far longer than what was given to me because of your untimely death. Secondly, Jeremy, you have been on this journey with me from the start. There may have been an ocean between us, but luckily skype and emails have been invented. I've learnt so much from working with you and I will always cherish the time I spent in your lab in Cornell as one of the highlights of this journey. Tanja and Jouni, it took a while to find our way but we did find it. Jouni, I've known you from the day I was interviewed for this position and I feel you've always treated me as an equal. It means a lot to me, thank you. Last, but definitely not least, Tanja, I admire your enthusiasm over research and wealth of ideas. You haven't always been easy on me but I think it has made me a better scientist. I also want to thank you for pushing me towards "coding". I think I spent unreasonably long time resisting the idea, but in the end, you were right. Learning a bit of coding has made many things easier.

I want express my gratitude to my co-authors. Michelle, thank you for getting the domestic samples from the UK, and also for your work on the GBS data. Tom, you've always been very swift with your email responses when I have needed a bit of an advice. Islam, driving around outskirts of Turkey in a van with you remains to be one of the most memorable trips I have ever taken. Thank you for helping with the sampling; those Turkish samples were absolutely invaluable.

Huge thanks to Assistant Professor Niclas Backström and Doctor Alain C Frantz for pre-examining my thesis. You did wonderful job. I dare say my thesis benefitted from your constructive criticism a great deal. Thank you, Professor Johanna Vilkki, for agreeing to be my opponent. I look forward to our discussion.

I am grateful to my follow-up group, Professor Outi Savolainen, Doctor Helmi Kuittinen and Emeritus Jaakko Lumme, for our meetings. The meetings were a combination of hard science, encouragement and humor. I always dreaded them beforehand but felt relieved afterwards. I also want to thank Outi for the support. I think you got a fair share of my angst during our conversations over the years. I

imagine it wasn't always easy for you to relate to my struggles over science but I feel you tried to encourage and support me as best you could.

I want to thank my group members, past and present, for your help and company, Alina, Eeva, Hilde, Jenni, Johanna, Liisa, Matti, Sujeet, Veli-Matti and all the others I don't remember to mention. Seven years is a long time, people have come and gone. Special mention to Johanna, I'm not sure how much I actually managed to teach you while supervising your Master's thesis but I definitely learnt a lot from the experience. You did great and I wish you best of luck with your PhD studies.

I'm grateful to the people in our current unit and in the previous biology department for helping out and being part of this journey. Heidi A, I think you've shared my joy and angst of being a PhD student like no one else. Thank you for your friendship and the bizarre "cultural activities" we've participated to. Thanks to the plant genetics people, Tiina, Ulla, Anu, Päivi, Sonja, Jaakko, Tuomas H, Yongfeng, Komlan and Esa. Special thanks to Tuomas T and Jaro who shared an office with me for a while. Thank you Lumi. Thank you Esa Hohtola for helping out with a subproject that ended up being a dead end, but definitely not because of you. It was a learning curve for me.

Thank you Soile Alatalo, Hannele Parkkinen and Laura Törmälä for your help in the lab. Thank you to all the hunters, goose-breeders and other people who helped me to get the samples. I also want to acknowledge Emil Aaltonen Foundation and Oulun Läänin Talousseuran Maataloussäätiö for the personal grants awarded to me.

Many thanks to my Cornell crew, Adam, Alex, Amanda, Ardern, Frida, Heidi E, Rodrigo, Soraia et al. for making my stay in Ithaca so fun. Amanda, the house parties at your place were epic. You're the only person I know with a room full of reptiles. A big shout out to Heidi E, thanks to you my poor pool playing skills were at their finest during my time in Ithaca. You're the coolest. Thank you, Rodrigo, I know you were busy but you always had the time and patience to help me out with various aspects of GBS.

Thanks to my biologist friends from my days in Jyväskylä, Lily, Sanni, Jenni and others. Thanks to my non-biologist friends, Iina, Tanja, Maija and others. Iina, you're my oldest friend. Our friendship has lasted for two decades and I've been really bad at keeping touch with you lately but I haven't forgotten you. Thank you for being there for me and using your professional skills for proofreading of this thesis. I take a full responsibility for the mistakes in the acknowledgements. Thanks to people involved in the Irish Festival of Oulu for the craic.

Finally, my warmest thanks to my family and relatives for being there for me and for reminding me there's life outside of academia as well. Kiitos äiti ja isä, kaikesta. My sisters, Suvi and Sini, we're all the same but still different. I'm lucky to be part of our trio. My "brothers in law", Vesa and Ilkka, thank you. Thank you to my nieces, Eerika, Neea and Inka, for being you. I am thankful to my maternal grandmother, Saima, for always being interested in my work, although, I don't think you ever really got what I was doing. I wish I had had the patience to explain it to you more carefully before you passed away. I thank my godmother Tarja for the encouragement and for serving me tea, when I'm visiting Kuhmo. Thank you to my fur niece and nephew, Pippa and Hemmo who are always up for cuddles. Yes, Sini, I know, they're not your kids but this is my thesis and I get to write what I want. Thank you to my Chihuahua boys in Kuhmo for the excitement and joy that my arrival brings forth in you. I like to think I'm on your top 3 favorite persons list. Otto, my canine soulmate, home doesn't feel the same when you're not there anymore.

Oulu, May 2017 Marja Heikkinen

Abbreviations

A	adenine
AMOVA	analysis of molecular variance
BCE	before current era
bp	base pair
C	cytosine
DNA	deoxyribonucleic acid
FDR	false discovery rate
F_{ST}	fixation index: measure of genetic differentiation
G	guanine
GBS	genotyping-by-sequencing
h	haplotype diversity
HVR1	hypervariable region 1
K	number of clusters
Mb	megabase
MCMC	Markov chain Monte Carlo
mtDNA	mitochondrial DNA
N_e	effective population size
Nm	number of migrants
NUMT	nuclear copy of mitochondrial DNA
PCA	principal component analysis
π	nucleotide diversity
SAMOVA	spatial analysis of molecular variance
SNP	single nucleotide polymorphism
T	thymine

List of original articles

This thesis is based on the following publications, which are referred throughout the text by their Roman numerals:

I Heikkinen ME, Ruokonen M, Alexander M, Aspi J, Pyhäjärvi T & Searle JB (2015) Relationship between wild greylag and European domestic geese based on mitochondrial DNA. Animal Genetics 46(5): 485–497.

II Heikkinen ME, Ruokonen M, White TA, Alexander M, Gundüz I, Dobney KM, Aspi J, Searle JB* & Pyhäjärvi T* (2017) Genomic analysis reveals a spectrum of hybrid background in European domestic geese and their wild progenitor (*Anser anser*). Manuscript.

III Heikkinen ME, Aspi J, Pyhäjärvi T* & Searle JB* (2017) Becoming domestic: genomic signatures of selection comparing European domestic geese and their wild progenitor. Manuscript.

*Equal contribution

Author contributions

	I	II	III
Original idea	MH, MR, JS	MH, MR, KD, TP, JA, JS	MH, TP, JA, JS
Data collection	MH, MR, MA	MH, MR, MA, IG	MH
Laboratory work	MH	MH, MA, TW	MH
Data analyses	MH, TP, JA	MH, TP, JA	MH, TP
Manuscript preparation	MH, TP, JA, JS	MH, TP, JA, JS	MH, TP, JA, JS

Marja Heikkinen (MH), Minna Ruokonen (MR), Jeremy Searle (JS), Tanja Pyhäjärvi (TP), Jouni Aspi (JA), Keith Dobney (KD), Michelle Alexander (MA), Tom White (TW), Islam Gundüz (IG)

Table of contents

1 Introduction

Arguably, not many things have shaped human history as much as the domestication has. The transition that started in the early Holocene around 11 000 years ago from nomadic hunter-gatherer lifestyle to more settled existence of agriculture was vital for the development of human societies as we know them now (Brown, Jones, Powell, & Allaby, 2009; Gupta, 2004).

The first steps of this transition took place in multiple places independently one of which was Southwest Asia and especially the area called the Fertile Crescent. There, the management of plant populations started and transformed these into many of the domestic crops that are a staple of our diet even today (Brown et al., 2009; Diamond, 2002; Doebley, Gaut, & Smith, 2006).

At the same time there were the first attempts at harnessing animals to serve humans. The cultivation of plants and animal husbandry offered something that could not be obtained by hunting and gathering: a stable source of food all year round (Brown et al., 2009). This allowed people to settle, to establish villages and towns that further developed into intricate societies with people occupying different trades that mark the modern civilizations; farmers, teachers, religious leaders, artisans etc. (Gupta, 2004). The history of humankind took an important step forward with the help of domestic plants and animals. How this came to be has puzzled researchers around the world for years.

Domestication is a complex, long-standing process with a beginning but no end, as for any evolutionary process (Larson & Burger, 2013). It is constantly changing and responding to environmental signals leading to adaptations that enable the species to live and reproduce in the current environmental conditions. In the case of domestic species, the environment is mostly controlled by humans. Still, the evolution of species, wild or domestic, is bound to be limited by boundaries that are set by genetics. The potential of a species to respond to change depends on its genetic diversity. There are three factors that have the most potential to impact a species under domestication by affecting its genetic diversity: inbreeding, drift and selection (Price, 1984). Each of them leaves their mark on the genome of the species, which can be used to trace the history of the species.

In this thesis, I have examined the domestication history of the European domestic goose and its ancestor, the greylag goose (*Anser anser*), making inferences on forces that have shaped their genetic diversity and what consequences they have had on the genome of the domestic goose and its wild relative.

1.1 Animal domestication

All the definitions of animal, and plant, domestication recognise the involvement of a relationship between humans and the target animal or plant populations (Zeder, 2006). From a population genetic perspective, this relationship can be regarded as selection promoting adaptation to an agro-ecological niche created by humans, and at later stages of the process, to human preferences (Larson et al., 2014).

Domestic animals serve many purposes in human life. They are kept for food and products like fur and feathers. They are companions and beasts of burden. They have offered and still offer a means of transportation. In addition to their economic value, domestic animals also have been important for religious reasons. In many ways, they are indispensable for human societies.

Although the first domesticated species was the dog, the shift from hunting to active management of animal populations only happened in the different parts of Fertile Crescent. The first animals that were domesticated in this area were sheep and goat about 11 000 years ago, shortly followed by domestication of pig about 10 500 years ago and cattle about 10 000 years ago (Larson & Fuller, 2014; Zeder, 2008). Since then, a wide variety of animals, including mammals, birds, fish and insects, have been domesticated on nearly every continent.

Despite the number of domestic animals kept today, not all species are suitable for domestication. There are behavioural traits that make some species better candidates for domestication than others (Price, 1984; Zeder, 2012). The traits can be grouped into five main categories that affect 1) social structure, 2) sexual behaviour, 3) parent-young interactions, 4) feeding behaviour and habitat choice, and 5) response to humans. It is quite obvious that the last group, the response to human stimuli, is a key factor for successful domestication. The selection for tameness and low reactivity is a universal feature of animal domestication (Zeder, 2012).

There are a variety of ways in which animal becomes domesticated. These can be generalised into three pathways (Larson & Fuller, 2014; Zeder, 2012). Firstly, the commensal pathway which starts by establishment of a commensal relationship with humans that, later on, develops into a mutually beneficial association. A classic example of this is the domestication of the dog which is thought to have started when wolves scavenged on human refuse (Axelsson et al., 2013). Secondly, the prey pathway, which is by far the most common type of pathway to domestication. As the name suggests, these animals were hunted by humans for their meat and, over time, the hunting strategies changed to herd-management strategies that lead to a domestic relationship controlled by humans. The first domesticates in the Fertile Crescent are thought to have followed this pathway.

Thirdly, the directed pathway which, unlike the other two, involves an intentional goal of domesticating a species for some specific resource or a set of resources. It is likely that this pathway originated after humans were already familiar with domestic animals that had followed either of the other two pathways. The domestic horse is likely to have followed this pathway as it seems that horse domestication started in the western Eurasian steppe and spread across Eurasia with extensive repeated capture of wild females to maintain or grow the domestic herd (Warmuth et al., 2012).

1.2 Genetics of domestication

1.2.1 Neutral genetic variation

The factors that affect neutral genetic variation in populations undergoing domestication and even after the domestic status has been attained are genetic drift and inbreeding. They are random in a sense that they do not discriminate between different alleles in the way that selection does.

Rather than the actual number of individuals in the population (the census size, N), the effective number of individuals that contribute to the next generation determines the amount of random genetic drift of the population. The effective population size (N_e) is the number of individuals in the idealised Wright-Fisher population that retains the same amount of genetic variation and experiences equally much genetic drift as an actual population irrespective of census size (Wright, 1931). Genetic drift results from limited N_e and leads to random fluctuations in allele frequencies. Large random changes in allele frequencies may ensue when N_e suddenly drops causing a bottleneck, which leads to surviving individuals representing only a random subset of genotypes present in the original population (Mayr, 1954). The reduced N_e increases genetic drift which in turn decreases heterozygosity due to random fixation and loss of alleles (Kimura & Crow, 1964; Wright, 1931). Initially low frequency deleterious alleles may increase in frequency in the population just by chance because of the bottleneck. On the other hand, beneficial alleles can be lost for the same reason. When N_e is small, drift can surpass selection (Robertson, 1962) which is why drift has more drastic effects on small populations compared to large populations. When N_e is small, selection cannot remove deleterious variants nor promote beneficial ones. Drift has a prominent influence on species under domestication because domestication is usually associated with two bottlenecks, one in the beginning of domestication and another one when the modern breeds are created (Tanksley & McCouch, 1997; Wang, Xie, Peng, Irwin, & Zhang, 2014).

Inbreeding, the mating between close relatives, is usually a consequence of small N_e when mating between relatives cannot be avoided, and it may also result from non-random mating (Keller & Waller, 2002). Domestic animals do not usually have a choice in terms of their mate because their mating is governed by their owner and the owner's interests dictate which individuals are mated to create the desired outcome i.e. different breeds. This causes non-random mating between limited numbers of individuals, which may lead to inbreeding (Keller & Waller, 2002). Once the breeds have been obtained, the mating between individuals belonging to the same gene pool that constitutes the breed ensures that mating remains non-random and facilitates inbreeding (Leroy & Baumung, 2011). Inbreeding leads to decreased heterozygosity (Wright, 1921), and it may cause inbreeding depression when deleterious alleles are expressed in homozygous individuals (Wright, 1977). Thus, the accumulation of deleterious alleles caused by genetic drift and inbreeding resulting from small N_e pose a serious threat to domestic animals (Marsden et al., 2016). Moreover, when combined with small N_e, natural selection has reduced power to purge deleterious alleles (Lynch, Conery, & Burger, 1995).

1.2.2 Selection

Selection, unlike inbreeding and drift, is not a random process. Selection acting on populations under domestication can be divided into natural and artificial selection. In nature, natural selection occurs when the different genotypes of the same locus are not equally good in terms of fitness. When this happens and one genotype is beneficial over the other(s), selection can increase the frequency of the advantageous allele (positive selection) or work against the deleterious allele (negative selection) by decreasing its frequency or by removing it completely from the population (Nielsen, 2005). A slightly different form of selection is the balancing selection which can also be seen as a form of positive selection. The balancing selection differs from directional selection in the sense that it increases or maintains variability within the population by promoting several alleles in the locus instead of just one via heterozygote advantage or frequency-dependent selection (Charlesworth, 2006). However, long-term balancing selection appears to be rather uncommon (Asthana, Schmidt, & Sunyaev, 2005; Bubb et al., 2006; Charlesworth, 2006; Wiuf, Zhao, Innan, & Nordborg, 2004). Natural selection is the mechanism that allows populations to adapt to the changes that occur over time in the environment in which they live. In the case of domestication, these adaptations enable the population under domestication to adapt to the environment provided by humans. On the other hand, selection also eliminates individuals that are incapable of living and breeding under human management. As mentioned before, not all species are equally suitable for domestication.

Domestication may also lead to relaxation of natural selection when traits that are essential for survival in nature become less important in captivity (Larson & Fuller, 2014; Price, 1984; Wiener & Wilkinson, 2011). For instance, the behavioural traits that affect the ability to find food and shelter are under strong selective pressure in wild animals, but domestic species are usually provided with these by humans. For this reason, domestic animals may show more variability in these traits than their wild counterparts (Larson & Fuller, 2014; Price, 1984).

A form of selection unique to domestication and sometimes used as a synonym for selective breeding of domestic animals, is the artificial selection (Driscoll, Macdonald, & O'Brien, 2009). The artificial selection differs from natural selection in the sense that humans decide what is beneficial and what is not and take over the decision as to which individuals contribute to the next generation in the hope of creating the desired phenotype. The artificial selection is a conscious but not necessarily intentional process (Driscoll et al., 2009) in comparison to the selective breeding which is used to create e.g. different breeds. The artificial selection can be combined with inbreeding to maintain or increase the frequency of particular traits.

Populations may experience different selection pressures on different traits, which is reflected in the distribution of their neutral genetic variation. Selection creates differentiation in allele frequencies between populations with respect to neutral alleles due to genetic hitchhiking in individual populations (Kaplan, Hudson, & Langley, 1989; Maynard Smith & Haigh, 1974). The process in which neutral genetic variation is reduced due to its linkage on locus under selection is known as a selective sweep (Barton, 1998; Nielsen et al., 2005). The genetic structure of populations and the degree to which populations are differentiated from each other can be estimated with F-statistics (Weir & Cockerham, 1984; Wright, 1949, 1965). This differentiation can then be used to detect regions in the genome that have been targeted by selection (Beaumont & Nichols, 1996; Foll & Gaggiotti, 2008; Lewontin & Krakauer, 1973).

1.2.3 Hybridisation

Interbreeding between closely-related species, interspecific hybridisation, is a fairly common phenomenon even though the propensity for hybridisation varies between taxa (Mallet, 2005; Schwenk, Brede, & Streit, 2008). Among birds, Anseriformes or waterfowl consisting of ducks, swans and geese show the highest propensity for interspecific hybridisation. At least 41.6% of the species hybridise with other species (Grant & Grant, 1992) although more recent authors have stated that the number could be over 60% and even higher in captivity (Ottenburghs, van Hooft, van Wieren, Ydenberg, & Prins, 2016).

The hybridisation between domestic animals and their wild progenitors, or animals closely related to their wild progenitors, is also quite frequent. There are studies showing that after the initial domestication of pigs had happened in one place and they were transported to a new region, they mated with the local wild boars in the new region (Frantz et al., 2015; Ottoni et al., 2013). Hybridisations between wolves and dogs have also been observed in multiple places (Godinho et al., 2011; Hindrikson, Männil, Ozolins, Krzywinski, & Saarma, 2012; Kopaliani, Shakarashvili, Gurielidze, Qurkhuli, & Tarkhnishvili, 2014). A striking example of hybridisation comes from domestic chicken which originated from red junglefowl (*Gallus gallus*), but commonly carries a yellow skin phenotype acquired by hybridisation with grey junglefowl (*Gallus sonneratii*) (Eriksson et al., 2008).

1.3 Greylag goose

The greylag goose is the largest of the so-called grey geese of the genus *Anser*. It has a Palearctic distribution which due to human actions is now fairly disjointed (Rooth, 1971). It breeds at boreal and temperate latitudes across Europe and Asia and winters in Southern Europe and Northern Africa as well as in Southwest Asia, India and Southern China (Cramp & Simmons, 1977; Scott & Rose, 1996).

Morphologically and geographically, greylag geese are divided into two recognised subspecies, the western nominate form *A. a. anser* (Linnaeus, 1758) which is found in Europe, and the Eastern form *A. a. rubrirostris* (Swinhoe, 1871) which ranges from Western Asia eastwards (Cramp & Simmons, 1977). The subspecies boundary is not well defined and intermediate types are found in the Central and Eastern Europe (Scott & Rose, 1996). The two subspecies have some differences in their morphologies; the western form is slightly smaller in size and darker in tone than the eastern form. The colouration of their bill and legs also differ:

the western form has an orange bill and flesh-coloured legs, whereas the eastern one has a pink bill and cold pink legs (Cramp & Simmons, 1977).

It is also noteworthy that wild greylag goose introductions were carried out in Zwin, Belgium in 1954 and in Rottige Meenthe, the Netherlands in 1962 (Rooth, 1971). Geese introduced to Belgium were originally of the eastern *rubrirostris* type and geese with the characteristics of the *rubrirostris* were observed along the Atlantic flyway in the 1960's and 1970's. The eastern characteristics of the introduced geese have since become less evident following the blending with the local geese (Kuijken & Devos, 1996).

1.4 Goose domestication

The available information on goose domestication is scarce and mainly based on archaeological findings and historical literature. The domestic geese of the world derive from two different lineages. The European domestic geese, the subject of my thesis, derive from the greylag goose, whereas the Chinese domestic geese derive from the swan goose (*Anser cygnoid*) (Delacour, 1954).

Where and when these two geese were domesticated remains unclear, but several authors have suggested the south-eastern Europe as the site for domestication of the European domestic goose (Crawford, 1984; Zeuner, 1963). Zeuner (1963) states that the ancient Greeks domesticated the greylag goose and that geese were highly valued and regarded as sacred to Aphrodite in Greece and Asia Minor. The first reliable literary reference to domestic geese in Europe is found in Homer's Odyssey where Penelope is said to have had twenty geese. Geese were also extensively used by the Romans. However, more recent authors have pointed at Egypt as a strong candidate for the location of domestication (Albarella, 2005; Larson & Fuller, 2014), as archaeological evidence suggests that the goose was fully domesticated by the 18th Dynasty (1450–1341 BCE) (Albarella, 2005). Thus, it seems that the domestication of the European goose is most likely to have happened in the vicinity of the eastern Mediterranean which broadly speaking includes the Fertile Crescent. The swan goose, on the other hand, is likely to have been domesticated in East Asia (Larson & Fuller, 2014).

According to Larson & Fuller (2014), both types of domestic geese were fully domesticated about 2500 years ago but the domestication of the swan goose probably started earlier than the domestication of the greylag goose. It seems that the pathways to domestication were different for the two species. The domestication of the swan goose probably started as a commensal relationship, but the greylag goose was an object of intentional domestication preceded by a period of being hunted for meat.

Since domestication, geese have been reared for meat, eggs and fatty liver but they also provide secondary products like feathers (Albarella, 2005; Zeuner, 1963). In ancient times, geese were also sacrificial birds and they were considered sacred to Isis and Osiris in Egypt, to Aphrodite in Greece and to Priapus in Rome. The sacred geese of the Temple of Juno in Rome are said to have saved Rome from the invasion of Gauls with their alarm calls (Albarella, 2005). Nowadays, geese are also kept as pets.

Today, the Food and Agriculture Organization of the United Nations (FAO) has recognised 94 domestic goose breeds or genetic groups of geese around the world, but it is likely that there are more. Many of the breeds have little direct economic value because of their poor productive performance, and/or the small representation of the breed, and the limited geographical distribution. Most of the breeds have allegedly greylag goose ancestry (42), about a quarter have swan goose ancestry (23), and 10 are considered as a combinations of both types; for 19 breeds/lines, the ancestry is not known (Buckland & Guy, 2002).

1.5 Aims of the study

The European domestic goose is an economically and culturally important species, but how it became domestic remains unclear. It is known that the European domestic goose was domesticated from the greylag goose, and this is likely to have happened in the vicinity of eastern Mediterranean about 3000 years ago. It should be noted that not all the domestic geese of the world derive from the same species, but some of them were domesticated from the swan goose, in which case these are called the Chinese domestic geese. By using modern population genetics and genomics, I have addressed questions concerning the domestication history of the goose in Europe. My study is the first large scale study that addresses questions about the location, timing and genetic change in association with goose domestication. This doctoral thesis consists of three original papers (I-III) that aim to address the following questions:

1. What is the level and distribution of genetic diversity in modern greylag goose populations and how does it contrast with the genetic diversity observed in domestic geese?
2. What is the extent of hybridisation between domestic geese and greylag geese?

3. Has the Chinese domestic goose contributed to the European domestic goose?
4. Can we define the origin and timing of the goose domestication in Europe, given the data available?
5. What genes and/or genomic regions have been targeted by selection during the domestication history of the European domestic goose?

2 Materials and methods

This section briefly describes the materials and methods used. The full details are included in the original papers (I-III).

2.1 Sampling and DNA extraction

The sampling was performed with the aim of covering as much as possible of the geographic distribution and genetic diversity of the greylag goose and to get a representative picture of the genetic diversity present in the European domestic geese. Some individuals assumed to be Chinese domestic geese were also sampled as well as breeds of domestic geese that were reported to be hybrids between the European and Chinese domestic geese. The subsets of samples differed to some degree between different studies but were mostly overlapping. Greylag goose samples consisted of muscle tissues taken from hunted individuals, blood samples collected during ringing and feathers collected during the moulting period. All the greylags were sampled between 1993 and 2011. The domestic goose samples were obtained with the help of local goose breeders in Denmark, Sweden and the UK. Domestic samples were mostly feathers taken from living individuals and some blood samples collected specially for this study but also one muscle sample from a goose leg that was sold in a local grocery store in Oulu.

The paper I consisted of 178 greylag goose samples and 102 domestic goose samples (Fig 1, Table S1 in I). The studies II and III used a subset of the samples that were used in the study I, but also some samples that were not included in the study I (Table 1 in II). The number of greylag goose samples was 58 and the number of domestic goose samples was 75 in the study II. For analytical purposes, some samples that were included in the study II were excluded from the study III and those are described in the Materials and methods section in III. The number of samples that were analysed in the study III consisted, therefore, of 49 greylags and 51 domestic geese samples.

The DNA was extracted using the DNeasy Blood and Tissue Kit (QIAGEN) according to manufacturer's instructions with some modifications to the procedure when the DNA was extracted from feathers (I-III). Some of the feather samples were extracted with a method for museum feathers/skins following Laird et al. (1991) in I. An RNase treatment was included for the DNA extraction of samples used in the studies II and III.

2.2 Mitochondrial DNA

Vertebrate mitochondrial DNA (mtDNA) is a double-stranded, circular molecule about 16-18 kb in size (Shadel & Clayton, 1997). It is almost exclusively maternally inherited, haploid and non-recombining (Bruford, Bradley, & Luikart, 2003). Due to its uniparental inheritance, the N_e of mtDNA is only a quarter of the N_e of nuclear DNA. These characteristics have made it an ideal tool for the study of maternal lineages of the species, and it has been a popular choice for the study of animal domestication (e.g. Fumihito et al., 1996; Loftus, MacHugh, Bradley, Sharp, & Cunningham, 1994; Savolainen, Zhang, Luo, Lundeberg, & Leitner, 2002).

One particular region of mtDNA has been especially widely used due to its high substitution rate: the mitochondrial control region (Vigilant, Pennington, Harpending, Kocher, & Wilson, 1989; Wenink, Baker, & Tilanus, 1993). The mitochondrial control region is highly conserved in length varying from 1174 to 1179 bp between different *Anser* species (Ruokonen, Kvist, & Lumme, 2000), and it can be divided into three domains: a conservative region in the middle flanked by hypervariable regions at the 5' and 3' ends.

A 1249-bp sequence containing the whole mitochondrial control region flanked by a complete *tRNA-Glu* gene at the 5' end and the partial *tRNA-Phe* gene at the 3' end was amplified and sequenced in the study I with primers specified by Ruokonen et al. (2000).

2.2.1 Genetic diversity and phylogeny of mtDNA

The MtDNA diversity was studied in populations of greylag goose and domestic goose using the control region as the genetic marker and the details are given in I. Briefly, the genetic diversity was estimated based on the number of polymorphic sites and the number of different haplotypes within the whole data set along with the population level estimates of genetic diversity; nucleotide diversity (π) and haplotype diversity (h) (Nei, 1987). A group specific estimate of sequence divergence was also calculated for greylag geese and domestic geese. The hierarchical distribution of molecular variation among greylag geese and domestic geese was estimated with an analysis of molecular variance (AMOVA, Excoffier, Smouse, & Quattro, 1992) along with a spatial analysis of molecular variance (SAMOVA, Dupanloup, Schneider, & Excoffier, 2002) within greylag populations.

The nucleotide substitution model that best fit the data was determined to be the Hasegawa-Kishino-Yano model (Hasegawa, Kishino, & Yano, 1985). The rate heterogeneity between sites was included in the model with gamma distribution (Yang, 1994), and the proportion of invariant sites (Fitch, 1986; Fitch & Margoliash, 1967; Shoemaker & Fitch, 1989) were taken into an account. Therefore, the model used for the construction of phylogenetic trees was HKY+G+I. The phylogenetic relationships of haplotypes were determined using both Bayesian inference (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) and Maximum Likelihood methods (Felsenstein, 1973; Tamura, Stecher, Peterson, Filipski, & Kumar, 2013) as well as a minimum-spanning network of pairwise differences between different haplotypes (Prim, 1957; Teacher & Griffiths, 2011).

2.3 Single-nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs) are, as the name suggests, a single base-pair differences between individuals found in both mitochondrial and nuclear genomes of an organism. They occur throughout the genome being more common in non-coding regions, and they are mostly selectively neutral which makes them a good marker for detection of genetic relationships between individuals of varying genetic backgrounds.

The rise of next generation sequencing based methods has increased the genome-wide data available for population genomic studies. One such method is genotyping-by-sequencing (GBS, Elshire et al., 2011) which is based on high throughput sequencing of restriction site-associated DNA. This method was used to detect nuclear SNPs between populations of greylag geese and domestic geese for population genomic analyses performed in the studies II and III.

2.3.1 Reference genome

While access to a reference genome is not necessary for all genomic studies, it increases the number of analyses available for genome-wide data. The greylag goose genome remains to be sequenced, but the Chinese domestic goose (Zhedong breed) genome has been published (Lu et al., 2015). The Zhedong goose, being a descendant of the swan goose and thus a close relative of the greylag goose, was used as a reference for SNP calling and, therefore, in generating the SNP dataset analysed in the studies II and III. The reference genome alleles were also included in the neighbor-joining tree construction in the study II and the SNPs were annotated using the reference genome information in the study III.

2.3.2 Genetic diversity and population structure

The objective of the study II was to describe the amount and distribution of neutral genetic variation observed in greylag geese and domestic geese based on SNPs and to investigate the underlying population structure. Genetic diversities within greylag and domestic goose populations were inferred from expected heterozygosities that were calculated for each locus and population and averaged across loci. The difference in average expected heterozygosity between greylag geese and domestic geese was tested with Welch's t-test (Welch, 1938). The variance components across all loci and hierarchical F statistics for greylag geese and domestic geese were estimated with hierarchical locus-by-locus AMOVA (Excoffier et al., 1992).

Population clustering and structure was analysed using a Bayesian clustering method (Pritchard, Stephens, & Donnelly, 2000) which aims to find the optimal number of genetic clusters in the given dataset by taking into account the individual genotypes and estimating the allele frequencies in populations. The method assumes that loci are in linkage equilibrium and each population is in the Hardy-Weinberg equilibrium. The second method to analyse the population structure was principal component analysis (PCA, Menozzi, Piazza, & Cavalli-Sforza, 1978; Patterson, Price, & Reich, 2006) which reduces the multidimensional data to components that retain most of the variation observed in the data. The analyses were performed on the whole dataset but within greylag geese and domestic geese samples as well.

In addition, a phylogenetic tree of individual relationships was generated based on a pairwise distance matrix between individuals (Saitou & Nei, 1987).

2.3.3 Selection

In contrast to the studies I and II which concentrated on neutral genetic variation, the study III focused on detecting signs of selection in the goose genome that are associated with the domestication process. Two F_{ST} outlier based methods (Beaumont & Nichols, 1996; Excoffier, Hofer, & Foll, 2009; Foll & Gaggiotti, 2008; Lewontin & Krakauer, 1973) were used to detect SNPs that are likely to have been under selection at one point or another during the process of goose domestication. Both methods use differences in allele frequencies to detect F_{ST} outliers. The first method uses coalescent simulations to determine whether the observed F_{ST} values can be considered as outliers. The underlying population structure is taken into an account with the use of hierarchical island model. The second method is Bayesian and based on a multinomial-Dirichlet model. This method decomposes the selection into population and locus-specific components, and it estimates the probability of a neutral model vs. a model involving selection to detect F_{ST} outliers.

3 Results and discussion

3.1 Genetic diversity

The mitochondrial data in the study I and nuclear SNPs in the study II showed that genetic diversity is lower in the European domestic geese than in the greylag geese. The nucleotide and haplotype diversities in the greylag geese were 0.0064 and 0.86, respectively, whereas they were 0.00054 and 0.57 in the domestic geese (Figure 2 in I). Moreover, 84% of the sampled domestic geese had one of two major haplotypes. The sequence divergence was also higher for the greylag geese than for the domestic geese, 0.0075 vs. 0.0006. Even though the level of genetic diversity was lower in the domestic geese, it was still relatively high, which has also been observed in other domestic species (Wiener & Wilkinson, 2011). A few greylags from the Netherlands and Scotland shared haplotypes with the domestic geese suggesting a hybridisation with the domestic geese. There were geographical differences in the distribution of genetic variation among the greylag geese, the eastern populations being more variable than the western populations. An exception to this was the Dutch population which showed a genetic diversity comparable to those observed in the eastern populations of the greylag goose in Iran and Kazakhstan, but this can be explained by the goose introductions that were carried out in the Western Europe in 1950's and 1960's (Rooth, 1971).

The expected heterozygosity that was calculated for each locus and population and averaged across loci in the study II showed that the greylag geese had a significantly higher average expected heterozygosity than the European domestic geese, 0.146 and 0.096, respectively (Welch Two Sample t-test, degrees of freedom (df) = 10.594, *p*-value = $3.91e^{-05}$, see also Table 1 and Figure 2 in II). Admixture with other populations increased the average expected heterozygosity in both the greylag goose and the domestic goose populations. The greylag goose populations in the Netherlands and Turkey showed high admixture with the domestic geese, and their average expected heterozygosities were also higher than what was observed in other populations of greylag geese, although the difference was not significant. The goose introductions in the Western Europe are likely to have contributed to the high genetic diversity measured in the Dutch population along with the hybridisation with the domestic geese, both of which were also observed in the study I. The trend of admixture increasing the average expected heterozygosity was also observed in the domestic geese where admixture with the Chinese domestic geese increased diversity in the populations that were most admixed. This difference was also significant.

It should be noted that the geographical differences in the levels of genetic variation that were observed in the greylag goose populations in the study I were not observed in the nuclear SNPs in the study II. The average expected heterozygosities were more equal in all the greylag populations, excluding those in the Netherlands and Turkey, than the nucleotide and haplotype diversities observed in the study I. This, however, can be explained by sex-biased dispersal where female greylag geese return to breed in their natal area, whereas males disperse further (Nilsson & Persson, 2001). This has also been observed in other goose species e.g. bean goose (Honka et al., 2017) and lesser white-fronted goose (Ruokonen, Aarvak, Chesser, Lundqvist, & Merilä, 2010).

In terms of possible domestication location, an interesting observation was that the Turkish domestic geese showed the highest genetic diversity among domestic geese at the mitochondrial level and they also had haplotypes that were not observed in any other domestic population (Table 2 in I). The northern Turkey domestic population in the study II also had higher average expected heterozygosity than what was observed in domestic populations in general if the admixed populations were excluded. The genetic diversity is expected to be highest in the domestication centre (Medugorac et al., 2009), and the high genetic diversity in Turkey could reflect the vicinity of goose domestication centre.

3.2 Population structure

In terms of population structure, both the mitochondrial sequences and SNPs suggested that the greylag geese and European domestic geese populations are clearly diverged from each other (F_{ST} 0.268, study II). The mitochondrial haplotypes formed a domestic clade separate from haplotypes observed in the greylag geese (Figure 3 and Figure S1 in I), and, although a few greylags had domestic haplotypes, they are most likely reflecting a local hybridisation with domestic geese. The nuclear loci told a similar story: the greylag geese, European domestic geese and Chinese domestic geese formed separate clusters (Figures 3-5 in II), but it was also evident that there is hybridisation between greylag geese and domestic geese, especially in the Netherlands and Turkey. However, the possibility of ancestral variation present in modern Turkish greylags should not be excluded. What appears as hybridisation between the greylag and domestic geese may actually be ancestral variation that dates back to the time of the domestication given the close proximity to the possible domestication location in the eastern Mediterranean. Furthermore, many of the European domestic geese shared ancestry with Chinese domestic geese, and this implies an intentional or accidental cross-breeding of the two types.

Even though the genetic separation of different subspecies was not clearly observed from mtDNA in the study I, there was some population structure within greylag geese based on SNPs suggesting the divergence of the two subspecies to some degree (Figures 5-7 in II). This can be expected as the subspecies occupy different breeding and wintering areas, although the subspecies boundary is not clear and somewhat intermediate individuals are observed in central and eastern Europe (Scott & Rose, 1996). The European populations of greylag geese appeared to be genetically more uniform than the eastern populations in Iran and Kazakhstan that seemed to have more diverse ancestry (Figure 6 in II) which corresponds with the higher genetic diversity observed in the eastern populations in the study I. The distribution area of greylag goose extends all the way to Asia, and the information concerning the easternmost greylags is very scarce (Fox et al., 2010; Madsen, Cracknell, & Fox, 1999). More thorough sampling of greylag geese would be needed to fully resolve its population structure.

The European domestic geese also showed some structure. The most interesting observation, aside from the admixture with Chinese domestic geese, was that the European breeds and the Turkish domestic geese were genetically distinct, the separation being visible in the neighbor-joining tree and in PCA (Figures 5 & 9, respectively, in II). A possible explanation for this could be that the European breeds were derived from the same relatively homogeneous gene pool but the Turkish domestic geese have been culturally and geographically isolated and appear distinctive for this reason (Larson et al., 2012).

3.3 Selection

Based on phenotypic differences between greylag geese and domestic geese, different selection pressures have impacted them. This is further demonstrated by the results from the F_{ST} outlier analyses in the study III. The two methods that were used yielded a different set of SNPs putatively under selection, but the overlap between them was high (Figure 3 in III). Systematic differences in allele frequencies between greylag geese were observed, and some of them were linked to potentially interesting genes (Tables 2-3 in III). One particularly interesting region was found from scaffold NW_013185806.1 which had six F_{ST} outliers along a 1 Mb long stretch of sequence upstream from a gene called LINGO2. LINGO2 has been associated with essential tremor and Parkinson disease in humans (Vilariño-Güell et al., 2010), but its association to body mass has also been observed (Rask-Andersen, Almén, Lind, & Schiöth, 2015). Since domestic geese are generally heavier than greylag geese, it is possible to speculate, until further evidence has been gathered, that LINGO2 has an effect on body mass in geese. Some breed specific changes in allele frequencies were also observed, and especially the Czech breed stood out in that respect (Table 3 in III). Further investigations would be needed to study if these are real signs of breed-specific selection and what genes have been affected or whether they are chance effects affecting allele frequencies in specific breeds.

A downside associated with reduced representation techniques for genome scans is the small proportion of genome they actually cover. A recent study by Lowry et al. (2017) found that many RADseq-based genome scans are likely to miss loci under selection because the density of markers obtained with these methods is not high enough to detect local adaptations especially in species with short regions of linkage disequilibrium. Therefore, the length of regions of linkage disequilibrium should be evaluated in the future studies when investigating the effect of selection on the goose genome.

4 Conclusions

Until now, the information concerning the domestication history of the European goose has been scarce and has mostly relied on archaeological findings and historical literature accompanied with a few genetic studies using mostly mitochondrial DNA. This study is the first attempt to illuminate the genome-wide effects of domestication on the well-known but little studied poultry species, the European domestic goose.

Firstly, in comparison to its ancestor, the greylag goose, the genetic diversity of the European domestic goose is lower. The same trend was observed in both mitochondrial and nuclear markers. This is expected, and it has been observed in many other domestic animals. However, the amount of genetic diversity did not only vary between the greylag geese and domestic geese but also within them. Multiple analyses suggested that the two subspecies of greylag goose are genetically distinct and that the European domestic geese also showed some population structure. Admixture with other populations increased the observed genetic diversity in both greylag geese and domestic geese.

Secondly, the study confirmed that greylag geese and domestic geese can hybridise. However, hybridisations seem to happen only locally and the domestic gene pool is not freely spreading to other populations of greylag geese outside the place of hybridisation. It was also noted that many European domestic goose breeds shared a relatively large portion of their ancestry with the Chinese domestic geese.

Thirdly, due to hybridisations between the greylag geese and the European domestic geese and because the Chinese domestic geese have contributed to the genetic diversity of European domestic geese, the timing and location of the domestication of the European goose remains unresolved. Nevertheless, my findings do not contradict the assumed origin of the domestic geese in south-eastern Europe and/or in Egypt. In this respect, the eastern Mediterranean including the Fertile Crescent remains the strongest candidate for the centre of goose domestication. More samples of the greylag geese from this region would be required to study this in detail. A more thorough sampling of greylag geese would also enable us to determine which of the subspecies was domesticated or whether both subspecies have been involved in the domestication. Ancient samples of geese might also prove beneficial.

Lastly, differential selection pressures have played a role in the divergence of the greylag geese and the European domestic geese. The differences in allele frequencies between the greylag geese and the domestic geese suggested that multiple regions in the goose genome have been the targets of selection, but the functional significance of them would need more investigation.

In conclusion, this study has increased our understanding of the genetic background of the domestication of the European goose. While many questions concerning the domestication history of the European goose remain unanswered, this study offers a valuable starting point for further studies.

References

Albarella, U. (2005). Alternate fortunes? The role of domestic ducks and geese from Roman to Medieval times in Britain. In G. Grupe & J. Peters (Eds.), *Documenta Archaeobiologiae 3. Feathers, grit and symbolism; Birds and humans in the ancient Old and New Worlds.* (pp. 249–258). Rahden: Verlag Marie Leidorf.

Asthana, S., Schmidt, S., & Sunyaev, S. (2005). A limited role for balancing selection. *Trends in Genetics*, *21*(1), 30–32.

Axelsson, E., Ratnakumar, A., Arendt, M.-L., Maqbool, K., Webster, M. T., Perloski, M., … Lindblad-Toh, K. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature*, *495*(7441), 360–364.

Barton, N. H. (1998). The effect of hitch-hiking on neutral genealogies. *Genetical Research*, *72*(2), 123–133.

Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B*, *263*(1377), 1619–1626.

Brown, T. A., Jones, M. K., Powell, W., & Allaby, R. G. (2009). The complex origins of domesticated crops in the Fertile Crescent. *Trends in Ecology & Evolution*, *24*(2), 103–109.

Bruford, M. W., Bradley, D. G., & Luikart, G. (2003). DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics*, *4*(11), 900–910.

Bubb, K. L., Bovee, D., Buckley, D., Haugen, E., Kibukawa, M., Paddock, M., … Olson, M. V. (2006). Scan of human genome reveals no new loci under ancient balancing selection. *Genetics*, *173*(4), 2165–2177.

Buckland, R., & Guy, G. (Eds.). (2002). Goose production. In *FAO Animal Production and Health Paper - 154*. Rome: Food and Agriculture Organization of the United Nations.

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, *2*(4), e64.

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.* New York: Oxford University Press.

Crawford, R. D. (1984). Goose. In I. L. Mason (Ed.), *Evolution of domesticated animals* (pp. 345–349). London and New York: Longman.

Delacour, J. (1954). *The waterfowl of the world : Vol. 1, The magpie goose, whistling ducks, swans and geese, sheldgeese and shelducks* (2nd ed.). London: Country Life.

Diamond, J. (2002). Evolution, consequences and future of plant and animal domestication. *Nature*, *418*(6898), 700–707.

Doebley, J. F., Gaut, B. S., & Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell*, *127*(7), 1309–1321.

Driscoll, C. A., Macdonald, D. W., & O’Brien, S. J. (2009). From wild animals to domestic pets, an evolutionary view of domestication. *Proceedings of the National Academy of Sciences USA*, *106*(suppl. 1), 9971–9978.

Dupanloup, I., Schneider, S., & Excoffier, L. (2002). A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, *11*(12), 2571–2581.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, *6*(5), e19379.

Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Strömstedt, L., … Andersson, L. (2008). Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genetics*, *4*(2), e1000010.

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, *103*(4), 285–298.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), 479–491.

Felsenstein, J. (1973). Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics*, *25*(5), 471–492.

Fitch, W. M. (1986). An estimation of the number of invariable sites is necessary for the accurate estimation of the number of nucleotide substitutions since a common ancestor. *Progress in Clinical and Biological Research*, *218*, 149–159.

Fitch, W. M., & Margoliash, E. (1967). A method for estimating the number of invariant amino acid coding positions in a gene using cytochrome c as a model case. *Biochemical Genetics*, *1*(1), 65–71.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, *180*(2), 977–993.

Fox, A. D., Ebbinge, B. S., Mitchell, C., Heinicke, T., Aarvak, T., Colhoun, K., … van der Jeugd, H. (2010). Current estimates of goose population sizes in western Europe, a gap analysis and an assessment of trends. *Ornis Svecica*, *20*(3–4), 115–127.

Frantz, L. A. F., Schraiber, J. G., Madsen, O., Megens, H.-J., Cagan, A., Bosse, M., … Groenen, M. A. M. (2015). Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nature Genetics*, *47*(10), 1141–1148.

Fumihito, A., Miyake, T., Takada, M., Shingu, R., Endo, T., Gojobori, T., … Ohno, S. (1996). Monophyletic origin and unique dispersal patterns of domestic fowls. *Proceedings of the National Academy of Sciences USA*, *93*(13), 6792–6795.

Godinho, R., Llaneza, L., Blanco, J. C., Lopes, S., Álvares, F., García, E. J., … Ferrand, N. (2011). Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. *Molecular Ecology*, *20*(24), 5154–5166.

Grant, P. R., & Grant, B. R. (1992). Hybridization of bird species. *Science*, *256*(5054), 193–197.

Gupta, A. K. (2004). Origin of agriculture and domestication of plants and animals linked to early Holocene climate amelioration. *Current Science*, *87*(1), 54–59.

Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, *22*(2), 160–174.

Hindrikson, M., Männil, P., Ozolins, J., Krzywinski, A., & Saarma, U. (2012). Bucking the trend in wolf-dog hybridization: first evidence from Europe of hybridization between female dogs and male wolves. *PLoS ONE*, *7*(10), e46465.

Honka, J., Kvist, L., Heikkinen, M. E., Helle, P., Searle, J. B., & Aspi, J. (2017). Determining the subspecies composition of bean goose harvests in Finland using genetic methods. *European Journal of Wildlife Research*, *63*(1), 19.

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, *17*(8), 754–755.

Kaplan, N. L., Hudson, R. R., & Langley, C. H. (1989). The "hitchhiking effect" revisited. *Genetics*, *123*(4), 887–899.

Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, *17*(5), 230–241.

Kimura, M., & Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics*, *49*(4), 725–738.

Kopaliani, N., Shakarashvili, M., Gurielidze, Z., Qurkhuli, T., & Tarkhnishvili, D. (2014). Gene flow between wolf and shepherd dog populations in Georgia (Caucasus). *Journal of Heredity*, *105*(3), 345–353.

Kuijken, E., & Devos, K. (1996). The status of the greylag goose *Anser Anser* in Flanders, Belgium. *Wetlands International Goose Specialist Group Bulletin*, (8), 26–28.

Laird, P. W., Zijderveld, A., Linders, K., Rudnicki, M. A., Jaenisch, R., & Berns, A. (1991). Simplified mammalian DNA isolation procedure. *Nucleic Acids Research*, *19*(15), 4293.

Larson, G., & Burger, J. (2013). A population genetics view of animal domestication. *Trends in Genetics*, *29*(4), 197–205.

Larson, G., & Fuller, D. Q. (2014). The evolution of animal domestication. *Annual Review of Ecology, Evolution, and Systematics*, *45*, 115–136.

Larson, G., Karlsson, E. K., Perri, A., Webster, M. T., Ho, S. Y. W., Peters, J., … Lindblad-Toh, K. (2012). Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proceedings of the National Academy of Sciences USA*, *109*(23), 8878–8883.

Larson, G., Piperno, D. R., Allaby, R. G., Purugganan, M. D., Andersson, L., Arroyo-Kalin, M., … Fuller, D. Q. (2014). Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences USA*, *111*(17), 6139–6146.

Leroy, G., & Baumung, R. (2011). Mating practices and the dissemination of genetic disorders in domestic animals, based on the example of dog breeding. *Animal Genetics*, *42*(1), 66–74.

Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, *74*(1), 175–195.

Loftus, R. T., MacHugh, D. E., Bradley, D. G., Sharp, P. M., & Cunningham, P. (1994). Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences USA*, *91*(7), 2757–2761.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, *17*(2), 142–152.

Lu, L., Chen, Y., Wang, Z., Li, X., Chen, W., Tao, Z., … Wang, J. (2015). The goose genome sequence leads to insights into the evolution of waterfowl and susceptibility to fatty liver. *Genome Biology*, *16*(1), 89.

Lynch, M., Conery, J., & Burger, R. (1995). Mutation accumulation and the extinction of small populations. *The American Naturalist*, *146*(4), 489–518.

Madsen, J., Cracknell, G., & Fox, T. (Eds.). (1999). *Goose populations of the Western Palearctic. A review of status and distribution.* Wetlands International Publication No. 48, Wageningen, The Netherlands: Wetlands International. Rønde, Denmark: National Environmental Research Institute.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*(5), 229–237.

Marsden, C. D., Ortega-Del Vecchyo, D., O'Brien, D. P., Taylor, J. F., Ramirez, O., Vilà, C., … Lohmueller, K. E. (2016). Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proceedings of the National Academy of Sciences USA*, *113*(1), 152–157.

Maynard Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research*, *23*(1), 23–35.

Mayr, E. (1954). Change of genetic environment and evolution. In J. Huxley, A. C. Hardy, & E. B. Fox (Eds.), *Evolution as a Process* (pp. 157–180). London: Allen & Unwin.

Medugorac, I., Medugorac, A., Russ, I., Veit-Kensch, C. E., Taberlet, P., Luntz, B., … Förster, M. (2009). Genetic diversity of European cattle breeds highlights the conservation value of traditional unselected breeds with high effective population size. *Molecular Ecology*, *18*(16), 3394–3410.

Menozzi, P., Piazza, A., & Cavalli-Sforza, L. (1978). Synthetic maps of human gene frequencies in Europeans. *Science*, *201*(4358), 786–792.

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

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, *39*, 197–218.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Research*, *15*(11), 1566–1575.

Nilsson, L., & Persson, H. (2001). Natal and breeding dispersal in the Baltic greylag goose *Anser anser*. *Wildfowl*, *52*(52), 21–30.

Ottenburghs, J., van Hooft, P., van Wieren, S. E., Ydenberg, R. C., & Prins, H. H. T. (2016). Hybridization in geese: a review. *Frontiers in Zoology*, *13*(1), 20.

Ottoni, C., Flink, L. G., Evin, A., Geörg, C., De Cupere, B., Van Neer, W., … Larson, G. (2013). Pig domestication and human-mediated dispersal in western Eurasia revealed through ancient DNA and geometric morphometrics. *Molecular Biology and Evolution*, *30*(4), 824–832.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*(12), e190.

Price, E. O. (1984). Behavioral aspects of animal domestication. *The Quarterly Review of Biology*, *59*(1), 1–32.

Prim, R. C. (1957). Shortest connection networks and some generalizations. *Bell System Technical Journal*, *36*(6), 1389–1401.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Rask-Andersen, M., Almén, M. S., Lind, L., & Schiöth, H. B. (2015). Association of the LINGO2-related SNP rs10968576 with body mass in a cohort of elderly Swedes. *Molecular Genetics and Genomics*, *290*(4), 1485–1491.

Robertson, A. (1962). Selection for heterozygotes in small populations. *Genetics*, *47*(9), 1291–1300.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572–1574.

Rooth, J. (1971). The occurrence of the greylag goose *Anser anser* in the Western part of its distribution area. *Ardea*, *59*, 17–27.

Ruokonen, M., Aarvak, T., Chesser, R. K., Lundqvist, A.-C., & Merilä, J. (2010). Temporal increase in mtDNA diversity in a declining population. *Molecular Ecology*, *19*(12), 2408–2417.

Ruokonen, M., Kvist, L., & Lumme, J. (2000). Close relatedness between mitochondrial DNA from seven *Anser* goose species. *Journal of Evolutionary Biology*, *13*(3), 532–540.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425.

Savolainen, P., Zhang, Y., Luo, J., Lundeberg, J., & Leitner, T. (2002). Genetic evidence for an East Asian origin of domestic dogs. *Science*, *298*(5598), 1610–1613.

Schwenk, K., Brede, N., & Streit, B. (2008). Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philosophical Transactions of the Royal Society of London B*, *363*(1505), 2805–2811.

Scott, D. A., & Rose, P. M. (1996). *Atlas of Anatidae populations in Africa and Western Eurasia*. Wetlands International Publication No 41, Wageningen, The Netherlands: Wetlands International.

Shadel, G. S., & Clayton, D. A. (1997). Mitochondrial DNA maintenance in vertebrates. *Annual Review of Biochemistry*, *66*, 409–435.

Shoemaker, J. S., & Fitch, W. M. (1989). Evidence from nuclear sequences that invariable sites should be considered when sequence divergence is calculated. *Molecular Biology and Evolution*, *6*(3), 270–289.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725–2729.

Tanksley, S. D., & McCouch, S. R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science*, *277*(5329), 1063–1066.

Teacher, A. G. F., & Griffiths, D. J. (2011). HapStar: automated haplotype network layout and visualization. *Molecular Ecology Resources*, *11*(1), 151–153.

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 USA*, *86*(23), 9350–9354.

Vilariño-Güell, C., Wider, C., Ross, O. A., Jasinska-Myga, B., Kachergus, J., Cobb, S. A., … Farrer, M. J. (2010). LINGO1 and LINGO2 variants are associated with essential tremor and Parkinson disease. *Neurogenetics*, *11*(4), 401–408.

Wang, G.-D., Xie, H.-B., Peng, M.-S., Irwin, D., & Zhang, Y.-P. (2014). Domestication genomics: evidence from animals. *Annual Review of Animal Biosciences*, *2*, 65–84.

Warmuth, V., Eriksson, A., Bower, M. A., Barker, G., Barrett, E., Hanks, B. K., … Manica, A. (2012). Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proceedings of the National Academy of Sciences USA*, *109*(21), 8202–8206.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, *29*(3/4), 350.

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 USA*, *90*(1), 94–98.

Wiener, P., & Wilkinson, S. (2011). Deciphering the genetic basis of animal domestication. *Proceedings of the Royal Society of London B*, *278*(1722), 3161–3170.

Wiuf, C., Zhao, K., Innan, H., & Nordborg, M. (2004). The probability and chromosomal extent of trans-specific polymorphism. *Genetics*, *168*(4), 2363–2372.

Wright, S. (1921). Systems of mating. II. The effects of inbreeding on the genetic composition of a population. *Genetics*, *6*(2), 124–143.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(2), 97–159.

Wright, S. (1949). The genetical structure of populations. *Annals of Eugenics*, *15*(1), 323–354.

Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, *19*(3), 395–420.

Wright, S. (1977). *Evolution and the genetics of populations. Vol. 3: experimental results and evolutionary deductions.* Chicago, IL: University of Chicago Press.

Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution*, *39*(3), 306–14.

Zeder, M. A. (2006). Central questions in the domestication of plants and animals. *Evolutionary Anthropology: Issues, News, and Reviews*, *15*(3), 105–117.

Zeder, M. A. (2008). Domestication and early agriculture in the Mediterranean Basin: origins, diffusion, and impact. *Proceedings of the National Academy of Sciences USA*, *105*(33), 11597–11604.

Zeder, M. A. (2012). The domestication of animals. *Journal of Anthropological Research*, *68*(2), 161–190.

Zeuner, F. E. (1963). *A history of domesticated animals*. New York and Evanston: Harper Row Publishers.

Original articles

- I Heikkinen ME, Ruokonen M, Alexander M, Aspi J, Pyhäjärvi T & Searle JB (2015) Relationship between wild greylag and European domestic geese based on mitochondrial DNA. Animal Genetics 46(5): 485–497.
- II Heikkinen ME, Ruokonen M, White TA, Alexander M, Gundüz I, Dobney KM, Aspi J, Searle JB & Pyhäjärvi T (2017) Genomic analysis reveals a spectrum of hybrid background in European domestic geese and their wild progenitor (*Anser anser*). Manuscript.
- III Heikkinen ME, Aspi J, Pyhäjärvi T & Searle JB (2017) Becoming domestic: genomic signatures of selection comparing European domestic geese and their wild progenitor. Manuscript.

Reprinted with permission from John Wiley and Sons (I).

Original publications are not included in the electronic version of the dissertation.

ACTA UNIVERSITATIS OULUENSIS
SERIES A SCIENTIAE RERUM NATURALIUM

676. Ronkainen, Katri (2016) Polyandry, multiple mating and sexual conflict in a water strider, Aquarius paludum
677. Pulkkinen, Elina (2016) Chemical modification of single-walled carbon nanotubes via alkali metal reduction
678. Runtti, Hanna (2016) Utilisation of industrial by-products in water treatment : carbon-and silicate-based materials as adsorbents for metals and sulphate removal
679. Suoranta, Terhi (2016) Advanced analytical methods for platinum group elements : applications in the research of catalyst materials, recycling and environmental issues
680. Pesonen, Janne (2016) Physicochemical studies regarding the utilization of wood- and peat-based fly ash
681. Kelanti, Markus (2016) Stakeholder analysis in software-intensive systems development
682. Ahmad, Muhammad Ovais (2016) Exploring Kanban in software engineering
683. Mustonen, Kaisa-Riikka (2016) Climate change and boreal rivers : predicting present-day patterns and future changes in hydrological regime and its effects on river communities
684. Karppinen, Pasi (2016) Studying user experience of health behavior change support systems : a qualitative approach to individuals' perceptions of web-based interventions
685. Sarja, Jari (2016) Developing technology pushed breakthroughs : defining and assessing success factors in ICT industry
686. Taušan, Nebojša (2016) Choreography modeling in embedded systems domain
687. Ylänne, Henni (2017) Herbivory control over tundra carbon storage under climate change
688. Siira, Olli-Pekka (2017) Developmental features of lacustrine basins on the uplift coast of the Bothnian Bay
689. Singh, Sujeet Kumar (2017) Conservation genetics of the Bengal tiger (*Panthera tigris tigris*) in India
690. Annanperä, Elina (2017) Managing technology-based service innovations in emerging wellness business ecosystems
691. Hens, Hilde (2017) Population genetics and population ecology in management of endangered species

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/

UNIVERSITY OF OULU P.O. Box 8000 FI-90014 UNIVERSITY OF OULU FINLAND

ACTA UNIVERSITATIS OULUENSIS

UNIVERSITY of OULU
OULUN YLIOPISTO

ISBN 978-952-62-1596-9 (Paperback)
ISBN 978-952-62-1597-6 (PDF)
ISSN 0355-3191 (Print)
ISSN 1796-220X (Online)