Bioinformatic Analysis of Whole Genome Sequencing Data

Detection of Selective Sweeps and Structural Changes

Khurram Maqbool
Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics
Uppsala

Doctoral Thesis
Swedish University of Agricultural Sciences
Uppsala 2014

Acta Universitatis agriculturae Sueciae
2014:58

Cover: Bioinformatic analysis of whole genome resequencing data from chicken, dog and pig populations. (photo: Khurram Maqbool)

ISSN 1652-6880
ISBN (print version) 978-91-576-8064-8
ISBN (electronic version) 978-91-576-8065-5

Print: SLU Service/Repro, Uppsala 2014

Bioinformatic analysis of whole genome sequencing data - Detection of Selective Sweeps and Structural Changes

Abstract

Evolution has shaped the life forms for billion of years. Domestication is an accelerated process that can be used as a model for evolutionary changes. The aim of this thesis project has been to carry out extensive bioinformatic analyses of whole genome sequencing data to reveal SNPs, InDels and selective sweeps in the chicken, pig and dog genome.

Pig genome sequencing revealed loci under selection for elongation of back and increased number of vertebrae, associated with the *NR6A1, PLAG1*, and *LCORL* genes. The scan for copy number variations (CNVs) revealed four duplications at the *KIT* locus associated with dominant white and belt colour phenotypes. Selective sweeps in the dog genome included genes involved in adaptation to a starch rich diet, fat metabolism and behavior. Identification of a selective sweep and a CNV in the *AMY2B* gene, which correlates with variation in α-amylase expression, along with selective sweeps in *MGAM* and *SGLT1,* in dogs revealed adaptation to a starch-rich diet after domestication. Analysis of chicken genome resequencing data identified hundreds of regions under selection shared among all domestic chicken and others that were specific for layers or broiler chickens and 68 fixed large deletions and 70 small InDels in domestic chicken populations.

Structural variations are changes in the genome that may affect the copy number of genes, their regulation or their coding sequence. Current methods utilize sequence information from either single sample or pair of samples to scan for CNVs across the genome. We developed a fast algorithm and a tool, MultiSV, to identify structural variations using short reads from massively parallel sequencing of multiple populations.

This thesis demonstrates the importance of structural variations as a factor contributing to phenotypic diversity in domestic animals and has revealed regions under strong selection during animal domestication and breeding. It also presents a new method for the identification of structural variations in populations using short reads from NextGen sequencers.

Keywords: Domestication, molecular evolution, chicken, dog, pig, next-generation sequencing, selective sweep, structural variations, CNV, duplication, deletion

Author's address: Khurram Maqbool, SLU, Department of Animal Breeding and Genetics, P.O. Box SE-597, 751 24 Uppsala, Sweden
E-mail: Khurram.Maqbool@slu.com

To my all teachers; mother, Zainab; wife, Shaista; daughters, Hafsa and Abeerah and memories of my late father Maqbool-Ul-Haq

The whole science is nothing more than a refinement of everyday thinking.
Albert Einstein

Contents

List of Publications 6

1 Introduction 9
1.1 Origin, complexity and variation of life 9
1.2 Domestication of animals 10
1.3 Genomics and next-generation sequencing 13
1.4 Selective sweeps 14
1.5 Structural variations 15

2 Aims of this Thesis 17

3 Current Research 19
3.1 Signatures of selection in the domestic pig genome (Paper I) 19
3.1.1 Background 19
3.1.2 Results and discussion 20
3.2 Signatures of selection in the dog genome (Paper II) 22
3.2.1 Background 22
3.2.2 Results and discussion 22
3.3 Signatures of selection in the chicken genome (Paper III) 25
3.3.1 Background 25
3.3.2 Results and discussion 26
3.4 Identification of structural variations in multiple populations (Paper IV) 28
3.4.1 Background 28
3.4.2 Method, Results and discussion 29

4 General Discussion and Future Perspectives 31

5 References 33

Acknowledgements 55

List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I Carl-Johan Rubin, Hendrik-Jan Megens, Alvaro Martinez Barrio, **Khurram Maqbool**, Shumaila Sayyab, Doreen Schwochow, Chao Wang, Örjan Carlborg, Patric Jern, Claus B. Jørgensen, Alan L. Archibald, Merete Fredholm, Martien A. M. Groenen, and Leif Andersson (2012). Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci USA*. 109, 19529–19536.

II Erik Axelsson, Abhirami Ratnakumar, Maja-Louise Arendt, **Khurram Maqbool**, Matthew Webster, Olof Liberg, Jon M. Arnemo, Åke Hedhammar and Kerstin Lindblad-Toh (2013). Whole genome resequencing reveals dog domestication genes and pathways including adaptation to a starch-rich diet. *Nature* 495, 360-364.

III **Khurram Maqbool**, Shumaila Sayyab, Alvaro Martinez Barrio, Bertrand Bed'hom, Michele Tixier-Boichard, Paul Siegel, Carl-Johan Rubin and Leif Andersson. Detecting signatures of selection in broiler and layer chickens using whole genome resequencing of pooled samples. (manuscript)

IV **Khurram Maqbool**, Lars Rönnegård and Leif Andersson. MultiSV: an R package for identification of structural variations in multiple populations based on whole genome resequencing. (manuscript)

Papers I-II are reproduced with the permission of the publishers.

Abbreviations

bp	base pair
Byr	billion years
CNV	copy number variations
DEL	deletion
F_{st}	fixation index
Gbp	gigabase (10^9 bases)
H_p	pooled heterozygosity
Indel	insertion or deletion
kbp	kilobase (1000 bases)
Mbp	megabase (10^6 bases)
Myr	million years
QTL	quantitative trait locus
RD	read depth
SNP	single nucleotide polymorphism
SV	structural variation
qPCR	quantitative polymerase chain reaction

1 Introduction

1.1 Origin, complexity and variation of life

Life appeared on earth around 3.5 Byr ago (Altermann & Kazmierczak, 2003) and then, with the passage of time, it evolved by accumulating genes, developing structures and processes and radiating to different life forms. The first eukaryotic cell developed 2.7 billion years ago, one billion years after the start of life on earth (Brocks *et al.*, 1999). After further evolution, the animals started to originate around 800 Myr to 1.7 Byr ago (Durham, 1978; Seilacher *et al.*, 1998). Those animals continued to evolve into complex as well as simple forms and around 600 Myr ago chordates started to emerge on earth (Ayala *et al.*, 1998).

Life took a turning point after 500-600 Myr ago in terms of spreading diversity and accumulation of variation with massive expansion of life forms (Welch *et al.*, 2005). Cell size increased along with an increase in diversity of living beings as they become large and complex with the passage of time (Carroll, 2001). With increasing complexity there is opportunity for an organism to use ecospace in a different manner than others have used it in the past. There are three trends of development clearly evident in the fossil record: development of multicellularity, increased size and diversification. Development of multicellularity likely happened independently many times. Once multicellularity was attained there followed larger and larger forms with a variety of new body plans and higher grades of complexity. This process led to development of body plans and explosive origin of metazoans during the Cambrian explosion (Welch *et al.*, 2005). A recent discovery suggests that the common ancestor of vertebrates, *Metaspriggina,* lived around 500 Myr ago (Morris & Caron, 2014). Birds and mammals started to diverge around 310 Myr

ago (Figure 1)(Kumar & Hedges, 1998; Hedges, 2002; Hillier *et al.*, 2004; Reisz & Müller, 2004; Alföldi *et al.*, 2011). Different orders of mammals started to diverge from each other into artiodactyla, carnivora, rodentia, primates, etc. Artiodactyla and primates started to diverge from each other around 79-97 Myr ago (Kumar & Hedges, 1998; Meredith *et al.*, 2011; Groenen *et al.*, 2012) while carnivores diverged from primates about 90-94 Myr ago (Archibald, 1996; Kumar & Hedges, 1998; Liu *et al.*, 2006).

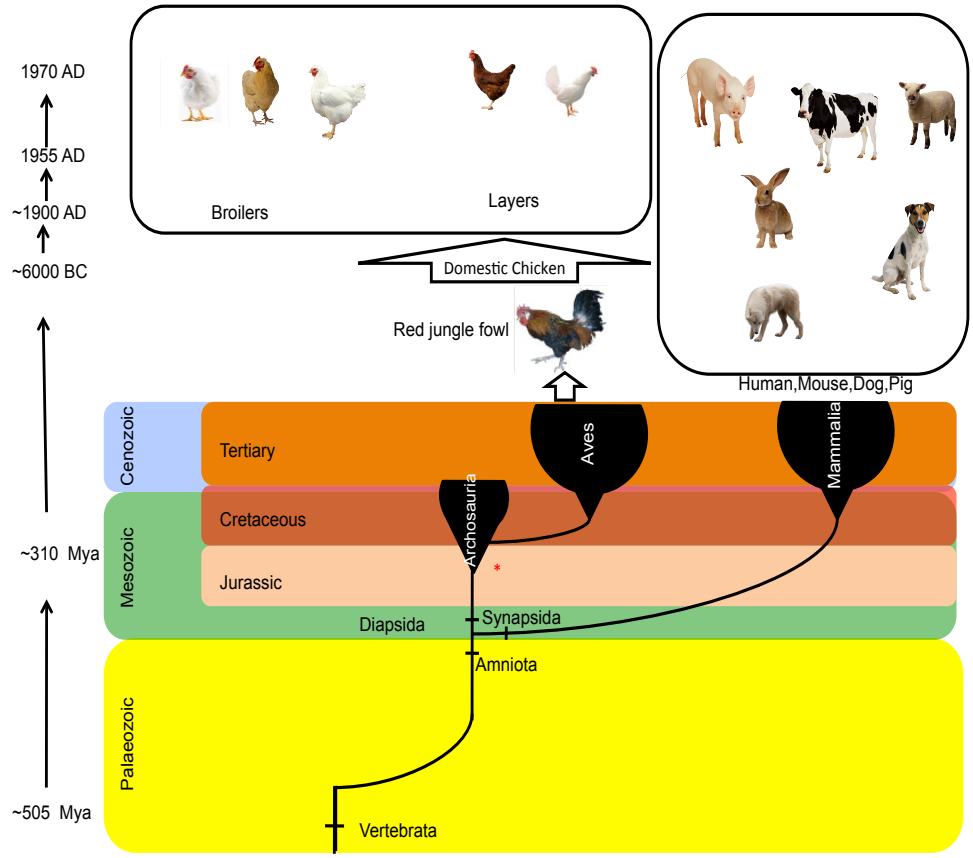

Figure 1. Origin and evolution of animals leading to domestication. Vertical axis represents approximate timeline of evolution. Asterisk indicates first bird, known to live about 150 Myr ago, modified from (Hillier *et al.*, 2004; Rubin *et al.*, 2010).

1.2 Domestication of animals

Evolution of life during 3.5 billion years produced the raw material in the form of enormous genetic variation as well as organisms well adapted to the

environment, life forms and processes around them. Humans utilized this genetic raw material to manipulate the evolution of some plants and animals according to our needs by artificial selection also known as domestication. Domestication is a process whereby a species comes in close contact to humans, which promotes genetic adaptation to the environment where humans control their reproduction and food or nutrient supply (Diamond, 2002). Allele frequency changes occur during domestication as a result of phenotypic selection, also called artificial selection, controlled by humans. Artificial selection can affect a wide array of phenotypic traits, such as: external morphology, internal morphology, physiology and development. These changes make the domestic animals differentiated from their wild ancestors having modified plumage or coat color, reduced size of brain, faster growth and reduced sense of fear (Jensen & Andersson, 2005; Jensen, 2006). The phenotypic variation in domestic animals provides an opportunity to understand genetic basis for phenotypic diversity (Andersson & Georges, 2004).

The wild boar (*Sus scrofa*) emerged around 5.3 to 3.5 Myr ago in South East Asia and was domesticated around 10,000 years ago in different locations of Europe and Asia (Giuffra *et al.*, 2000; Larson *et al.*, 2005; Ottoni *et al.*, 2012). The size of the current pig genome assembly (SS10.2) is 2.8 Gbp (Groenen *et al.*, 2012), which is similar in size to the size of other mammalian genomes, including human (Lander *et al.*, 2001; Venter *et al.*, 2001), dog (Lindblad-Toh *et al.*, 2005), mouse (Chinwalla *et al.*, 2002), cattle (Elsik *et al.*, 2009) and horse (Wade *et al.*, 2009). However, the pig genome contains less repetitive DNA compared to other mammalian genomes (Groenen *et al.*, 2012). A comparison of orthologous sequences from human, mouse and pig revealed that there is a higher degree of sequence identity between sequences of pig and human compared to sequence similarity between human and mouse. Furthermore, purifying selection has been more efficient in pig compared to humans (Wernersson *et al.*, 2005). The pig genome is organized in 18 pair of autosomes and two sex chromosomes (2n=38) (Ellegren *et al.*, 1994; Archibald *et al.*, 1995) which is a less than some other mammals for example dogs (2n=78), cattle (2n = 60), goats (2n=50) and horses (2n=64). However, central Asian wild boars carry 36 chromosomes due to the fusion of chromosomes 16 and 17, while western European wild boars also carry 36 chromosome due to the fusion of chromosomes 15 and 17 (Troshina *et al.*, 1985).

The dog (*Canis familiaris*) was domesticated from grey wolf (Vilà *et al.*, 1997; Lindblad-Toh *et al.*, 2005). The process of dog domestication may have started around 11,000 to 16,000 years ago (Freedman *et al.*, 2014). Early dog domestication and recent breed development generated bottlenecks causing about sixteen times reduction in population size (Freedman *et al.*, 2014) that

resulted in long linkage disequilibrium blocks within breeds and short linkage disequilibrium blocks across breeds (Figure 2).

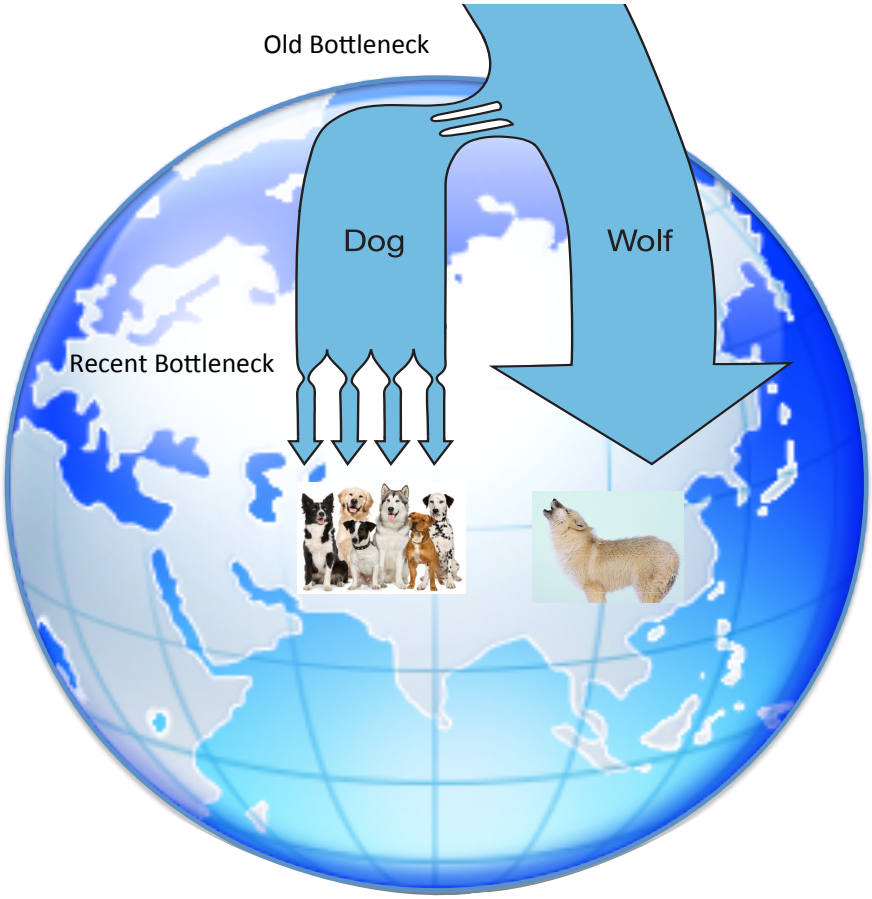

Figure 2. The origin and evolution of domestic dogs (modified from Lindblad-Toh *et al.*, 2005)

Chicken was the first bird and the first agricultural animal to have its genome sequenced (Hillier *et al.*, 2004). The size of chicken genome is around 1 Gbp and is similar in size and structure to the genome assemblies of zebra finch (Warren *et al.*, 2010) and large ground-finch (Rands *et al.*, 2013), but almost one third of the size of mammalian genomes. However, the number of genes in the chicken genome is similar to that in the human genome (Hillier *et al.*, 2004; Clamp *et al.*, 2007). The chicken genome is organized in 38 pair of autosomes and two sex chromosomes. The autosomes are further classified into five pairs of macro-, five pairs of intermediate-, and 28 pairs of microchromosomes (Burt, 2005; Schmid *et al.*, 2005). The microchromosomes have higher GC content and

gene density, shorter intronic regions and lower microsatellite density when compared to macrochromosomes (McQueen *et al.*, 1996; Primmer *et al.*, 1997; Hillier *et al.*, 2004). The repetitive sequences in microchromosomes make it difficult to assemble them; hence the current chicken genome assembly (*Gallus gallus,* ICGSC Gallus_gallus-4.0/galGal4) contains only 34 chromosomes. The female chicken is heterogametic carrying one Z and one W chromosome, whereas the male is homogametic with two Z chromosomes, unlike mammalian species where males are heterogametic with one X and one Y chromosome. Sex chromosomes in birds show similar size differences with one smaller W and one larger Z, similar to smaller Y and larger X chromosomes in mammals. However, the sex chromosomes of birds and mammals are not homologous (Fridolfsson *et al.*, 1998), but may have evolved from neighboring regions on a common ancestral chromosome (Smith & Voss, 2007). Regions from chicken chromosomes Z and 4 and from human chromosomes 9, 4, X, 5, and 8 were linked in a common ancestor (Smith & Voss, 2007).

Domestic chicken (*Gallus gallus domesticus*) was domesticated from Red jungle fowl (*Gallus gallus*) in South East Asia around 6000BC (Figure 1). They provide a major source of protein for human consumption in the form of eggs and meat (Burt, 2002). In the beginning of the 20th century, domestic chicken was split into two major breeds, layers and broilers, used for egg and meat production, respectively (Dodgson, 2007).

1.3 Genomics and next-generation sequencing

We are in the middle of a technological revolution that will transform all areas of biology. The very rapid development of new methods for DNA sequencing (so called NextGen sequencing) will make the genome and transcriptome from any species accessible for research. The possibility to screen populations for essentially all polymorphic loci in the genome is a dream coming true for geneticists. Just a few years ago there were only a handful of major genome centers in the world (Broad Institute in Boston, WashU genome center in St. Louis, Baylor college in Houston, Sanger center in the UK and the Beijing Genome Institute in China) that could carry out whole genome sequencing of a vertebrate genome. This situation has changed completely with the development of NextGen sequencers such as Roche 454, Illumina Solexa and Applied Biosystem SOLiD. This technological revolution has reduced the cost for sequencing, making whole genome sequencing at population scale feasible and it has reduced the amount of time required to perform all the analysis (Mardis, 2008).

The next big challenge after sequencing the human genome (Lander *et al.*, 2001; Venter *et al.*, 2001) was to characterize variations in the genome to obtain more in-depth understanding about function and structure of the genome (Pennisi, 2007), which requires more samples to be sequenced. The Sanger sequencing technique, which was used to sequence the human genome, is not only expensive, but also time consuming because of low throughput, which makes it impracticable to sequence whole genomes for multiple samples (Mardis, 2008; Shendure & Ji, 2008). The solution is to perform massive parallel sequencing, also called next generation (Next-Gen) sequencing (Wang *et al.*, 2008; Wheeler *et al.*, 2008; Pushkarev *et al.*, 2009; Rands *et al.*, 2013). Multiple samples can be sequenced to identify SNPs (Koboldt *et al.*, 2009; Li *et al.*, 2009; Bansal *et al.*, 2010; McKenna *et al.*, 2010). The SNP data can then be used to identify selective sweeps (Nielsen *et al.*, 2005; Tang *et al.*, 2007; Rubin *et al.*, 2010; Boitard *et al.*, 2012, 2013; Clément *et al.*, 2013) and structural variations (Koboldt *et al.*, 2009, 2012) and to study population demography and structure (Pritchard *et al.*, 2000; Falush *et al.*, 2003; Patterson *et al.*, 2006, 2012; Alexander *et al.*, 2009; Li & Durbin, 2011). Additionally, read depth and mate pair distance information can be utilized to identify structural variations.

1.4 Selective sweeps

Single nucleotide polymorphisms (SNPs) constitute the most common form of genetic variation. Although, most of the genetic variation remains selectively neutral, selection operates on SNPs that affects fitness, which may lead to fixation of favorable variants and hence reduced or no genetic variation in the selected region, also called selective sweeps. The beneficial variant gets selected and the alleles adjacent to the selected allele become enriched through genetic hitchhiking (Figure 3) (Smith & Haigh, 1974; Kaplan *et al.*, 1989; Barton, 2000; Fay & Wu, 2000; Andolfatto, 2001). This leads to loss of heterozygosity (Hudson *et al.*, 1987; Kim & Stephan, 2002; Jensen *et al.*, 2005), higher linkage disequilibrium (Przeworski, 2002; Kim & Nielsen, 2004; Kimura *et al.*, 2007), increase of rare alleles (Fu, 1997) and a high frequency of derived alleles (Fay & Wu, 2000).

It is a challenging task to carry out population genetic studies based on whole genome sequencing because it is costly to sequence many individuals in order to identify SNPs in populations and compute allele frequencies for each SNP. A cost effective way is to use pooled DNA sequencing of many individual from populations (Rubin *et al.*, 2010; Kofler *et al.*, 2011, 2012; Boitard *et al.*, 2012, 2013; Clément *et al.*, 2013; Gautier *et al.*, 2013). The reads covering

polymorphic positions can be used to compute allele frequencies that can then be used to compute heterozygosity in windows to identify sweep regions.

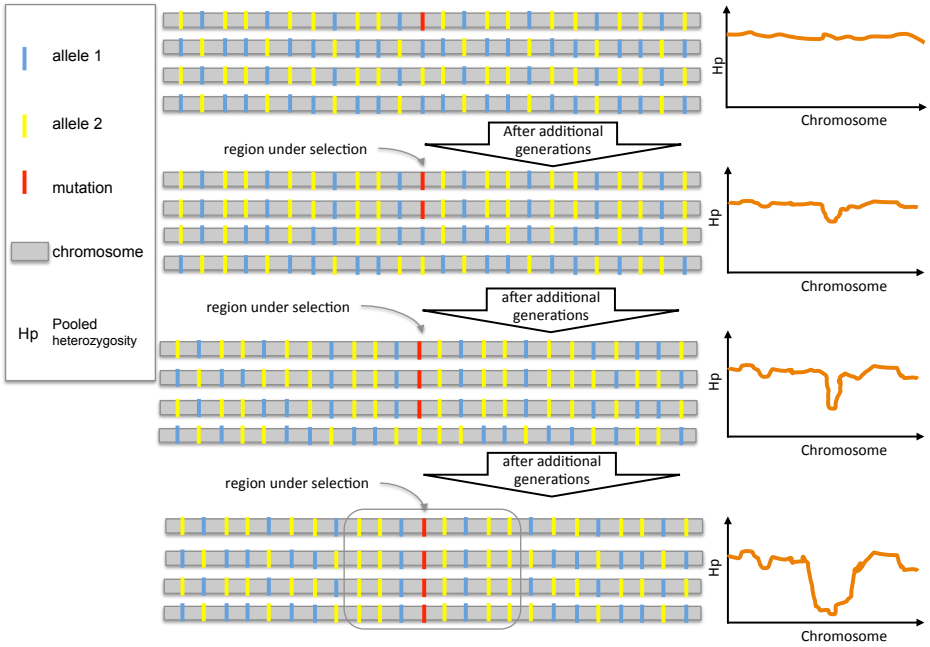

Figure 3. A beneficial mutation arises and the region gets selected. The frequency of the selected allele increases after few generations of selection and the frequency of alleles adjacent to the selected allele also increase due to linkage. The continuous process of selection leads to loss of heterozygosity in the region under selection. This produces initially long haplotypes having reduced heterozygosity. Recombination reduces the size of haplotype blocks into smaller blocks that are confined to the region under selection. We can scan for such regions to identify selective sweeps in the genome.

1.5 Structural variations

Structural variations play an important role in the divergence between closely related groups of organisms (Britten *et al.*, 2003) and contribute to phenotypic variation in domestic animals (Andersson, 2012) and humans (Feuk *et al.*, 2006).

Whole genome next-generation sequencing data can be used in different ways to identify structural variations. Among those methods, read depth based methods can be applied to the data from almost all sequencing platforms to identify deletions and duplications (Figure 4). Mate pairs are two reads from the same clone and the distance between them span over few kbp. Analysis of distances between mate pairs can be used to identify insertions, deletions,

inversions and translocations. The mapping positions of the mate pairs can be compared to the expected map distance and the regions where the distances between mates differ significantly indicate the presence of structural variation or an error in the genome assembly.

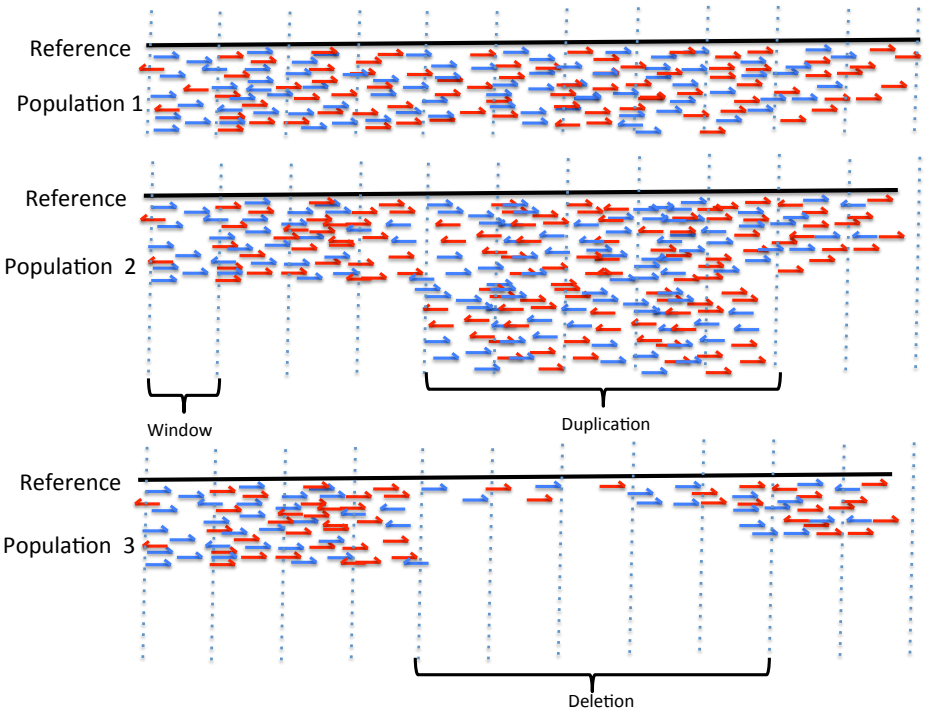

Figure 4. Read depth comparison among three hypothetical populations. Blue and red lines represent mate pairs. The mate pairs are aligned to a reference genome, which is represented by a black line. Read depth analysis is performed in windows represented by vertical blue dotted lines. Population 1 shows uniform read depth indicating that there is no evidence of structural variation. Population 2 shows increase in read depth indicating presence of a duplication and population 3 shows reduction in read depth indicating presence of a deletion.

2 Aims of this Thesis

The aim of this thesis was to reveal SNPs, short insertions and deletions, and structural rearrangements in the chicken, pig and dog genome by carrying out extensive bioinformatics analysis of whole genome sequencing data. We searched for signatures of selection in these data and cross-referenced with data on the location of quantitative trait loci (QTL) and expression data with the ultimate goal to identify causative mutations for phenotypic traits.

The step-wise aims of this thesis have been to:

I. Identify the most common allele in each population studied at all polymorphic loci in the chicken, pig and dog genome
II. Identify regions of the chicken, pig and dog genomes that have been under positive selection
III. Identify causative mutations that has been selected during chicken, pig and dog domestication
IV. Develop a method to identify structural variations from sequencing data obtained from multiple populations

3 Current Research

3.1 Signatures of selection in the domestic pig genome (Paper I)

3.1.1 Background

The wild boar (*Sus scrofa*) diverged into different subspecies 2.7 to 4.5 million years after their origin around 3.5 to 5.3 Myr ago (Giuffra *et al.*, 2000; Larson *et al.*, 2005; Groenen *et al.*, 2012; Ottoni *et al.*, 2012). The domestication of pig started independently in Asia and Europe from different subspecies of wild boars giving rise to considerable morphological and physiological differences when compared with their ancestors (Giuffra *et al.*, 2000; Kijas & Andersson, 2001; Larson *et al.*, 2005; Fang & Andersson, 2006).

One example of the parallel domestication of pigs from wild boars in Asia and Europe is independent missense mutations in *melanocortin receptor 1* (*MC1R*) causing black coat color that are present on different haplotypes originating from the two geographic regions (Kijas *et al.*, 1998, 2001; Fang *et al.*, 2009). Mutations in *MC1R* are associated with coat color variation in other animals including horse (Marklund *et al.*, 1996), cattle (Klungland *et al.*, 1995), fox (Våge *et al.*, 1997), sheep (Våge *et al.*, 1999), dog (Newton *et al.*, 2000) and chicken (Kerje *et al.*, 2003).

Selection for high protein and low fat content during the past 60 years resulted in accumulation of mutations in *Ryanodine receptor 1* (*RYR1*) (Fujii *et al.*, 1991), *PRKAG3*, which encodes 5'-AMP-activated protein kinase subunit gamma-3 (Milan *et al.*, 2000) and *insulin-like growth factor 2* (*IGF2*) (Van Laere *et al.*, 2003).

Here our focus was to identify selective sweeps and structural variations in the pig genome that may have played an important role during domestication.

3.1.2 Results and discussion

We generated pooled DNA sequences from 20 European wild boars, 78 domestic pigs and 14 F2 intercross animals from a wild boar, Large White intercross and aligned the reads to the SS10.2 pig reference genome sequence. We identified a total of 7.35 million SNPs, 863 fixed deletions and 1619 CNVs. We used allele frequencies at each SNP locus to identify signatures of selection in European domestic pigs by scanning the genome to look for stretches of regions with reduced heterozygosity (Rubin *et al.*, 2010). We divided the genome in windows that were of size 150 kbp each overlapping 50% with neighboring windows. We then used Z-transformation (ZH_p) and used a threshold of four standard deviations away from the mean in the lower tail separately for autosomes and chromosome X. Overlapping windows were merged after applying the threshold. This approach identified 13 regions with a ZH_p of heterozygosity lower than −5, and 64 regions with a $ZH_p < -4$. Some of the putative sweeps occurred in regions containing QTLs for phenotypic traits including feed intake, growth, elongation of the back, number of vertebrae and body length. Some of these phenotypic difference between domestic pig and wild boars were also noticed by Charles Darwin (Darwin, 1868).

One of the sweep regions contained a gene, *melanocortin 4 receptor* (*MC4R*), that harbors a QTL for feed intake and growth (Kim *et al.*, 2000). Another sweep region, which showed the strongest signature of selection includes the *NR6A1* (*Nuclear Receptor 6 A1*) gene that harbors a major QTL affecting the numbers of vertebrae in pigs, and a missense mutation in NR6A1 has been proposed to be causative (Mikawa *et al.*, 2007). It is known that wild boars have 19 vertebrae, whereas European domestic pigs intensively selected for meat production have 21–23 (King & Roberts, 1960). The next two strongest selective sweeps were found in regions harboring previously identified major QTLs for body length (Andersson *et al.*, 1994; Andersson-Eklund *et al.*, 1998). These two regions contain *PLAG1* (*pleomorphic adenoma gene 1*), which has been associated with variation in height in humans (Gudbjartsson *et al.*, 2008) as well as with a major QTL for height in cattle (Karim *et al.*, 2011) and *LCORL* (*ligand dependent nuclear receptor corepressor-like*), which has been associated with human stature (Lango Allen *et al.*, 2010), as well as with body size in dogs (Vaysse *et al.*, 2011), cattle (Pryce *et al.*, 2011), and horses (Signer-Hasler *et al.*, 2012). We performed QTL analysis after genotyping the F_2 population from an intercross (Andersson *et al.*, 1994) using informative markers in the *LCORL* and *PLAG1* regions and observed that the two QTLs associated with *LCORL* and *PLAG1* acted additively with a combined effect of 5.3 cm in body length difference between the opposite homozygotes. Furthermore, phylogenetic analysis of the three loci (*NR6A1*, *LCORL*, and *PLAG1*) loci revealed a

European origin of the swept haplotype for all three loci. A total of 180 loci affecting variation in human height only explained 10% of the population variance (Lango Allen *et al.*, 2010), while only two loci, *LCORL* and *PLAG1,* explained 18.4% of the residual variance in body length in pigs in this wild boar intercross. It is likely that alleles with similar large effects are also segregating in human populations, but that each such variants explain a small portion of the population variance. Another particularly interesting candidate selective sweep overlaps *Osteocrin* (*OSTN*), whose secreted protein product (OSTN) was first identified as an inhibitor of osteoblast differentiation (Thomas *et al.*, 2003) and later OSTN was rediscovered as "Musclin" in a screen for skeletal muscle-derived secretory factors (Nishizawa *et al.*, 2004), which may be related to selection for an altered body composition and/or altered skeletal development.

We identified 72 derived nonsynonymous substitutions approaching fixation in domestic pigs. Three (*NR6A1*, *CCT8L2* and *MLL3*) of these were colocalized with putative sweep regions detected in this study, and two of these (*NR6A1* and *SERPINA6*) have been identified as candidate causative mutations. First in a missense mutation (Pro192Leu) in *NR6A1* alters the binding affinity of NR6A1 to its coreceptors and has been proposed to be the causative mutation for the QTL affecting number of vertebrae (Mikawa *et al.*, 2007). Second a missense mutation (Gly307Arg) in *SERPINA6* has been shown to affect cortisol-binding capacity and proposed to underlie a pleiotropic QTL affecting serum cortisol levels, fat deposition, and muscle content (Guyonnet-Dupérat *et al.*, 2006).

We used read depth signals to screen for deletions and CNVs from the pooled samples that showed large allele frequency differences between wild and domestic pigs. One of the selective sweeps, containing the Caspase 10 (*CASP10*) gene, overlapped an 8 kbp duplication in an intron of the gene. We also confirmed the presence of a previously known 450 kbp duplication at the *KIT* locus, which control white spotting in pigs (Moller *et al.*, 1996; Giuffra *et al.*, 2002). KIT is a tyrosine kinase receptor, and normal KIT signaling is required for development and survival of neural crest-derived melanoblasts (Chabot *et al.*, 1988). Three major *KIT* variants have been described in pigs: Dominant white (completely white), Patch (partially white), and Belt (white belt across forelegs). Mutations associated with Dominant white and Patch were previously reported, whereas no causative mutation has been identified for Belt (Giuffra *et al.*, 1999). We identified additional duplications, one 4.3 kbp duplication located about 100 kbp upstream of *KIT*, a second 23 kbp duplication located about 100 kbp downstream of *KIT* and a third 4.3 kbp duplication within the second duplication in Hampshire pigs, which have the Belt phenotype. The 4.3 kbp duplication is present in three to six copies in Hampshire pigs and is a causative candidate mutation because it overlaps with one of the most well-

conserved noncoding regions in the *KIT* region. This may constitute a regulatory element that becomes stronger with copy number expansion, as recently demonstrated for a melanocyte-specific enhancer located within the duplication causing graying with age in horses (Sundström *et al.*, 2012). These three *KIT* duplications in pigs are present within a large 450 kbp duplication, partly underlying the Dominant white phenotype in pigs.

3.2 Signatures of selection in the dog genome (Paper II)

3.2.1 Background

Dogs play an important role in the human society. It is uncertain, when and where dog domestication started (Pang *et al.*, 2009; Ovodov *et al.*, 2011). Dog domestication may have started either because humans used wolves in guarding and hunting or wolves coming close to human dwellings in southern East Asia, Middle East or Europe more than 16,000 years ago (Figure 1) (Vilà *et al.*, 1997; Savolainen *et al.*, 2002; Sharma *et al.*, 2004; Ostrander & Wayne, 2005; Verginelli *et al.*, 2005; Pang *et al.*, 2009; vonHoldt *et al.*, 2010; Ovodov *et al.*, 2011; Skoglund *et al.*, 2011; Thalmann *et al.*, 2013).

Domestic dogs, like other domestic animals, show phenotypic differences from its ancestor the grey wolf in size, color and behavior (Trut, 1999; Lindblad-Toh *et al.*, 2005; Trut *et al.*, 2009). The potential for uncovering genes and pathways underlying behavior makes the detection of domestication genes in dogs especially appealing.

Earlier studies have mainly focused on origin of dogs and their phylogenetic relationships (Vilà *et al.*, 1999; Verginelli *et al.*, 2005), here we focused to identify regions of selection and structural variations in the dog genome that may have played an important role during domestication.

3.2.2 Results and discussion

We used pooled DNA sequences from 12 wolves and 60 dogs and aligned the reads to the CanFam 2.0 dog reference genome sequence. The unique alignments were used to identify a total of 3.78 million SNPs, more than 506,000 small InDels and more than 26,000 CNVs. We then used allele counts and frequencies at each SNP locus to identify signatures of selection in dog genome by scanning the genome for stretches of regions with reduced heterozygosity (H_p) in dogs and high populations differentiation between dogs and wolves (F_{st}). To compute the H_p (Rubin *et al.*, 2010) and F_{st} (Weir & Cockerham, 1984), we divided the genome in 21,927 windows that were of size 200 kbp each overlapping 50% with neighboring windows. We then used Z-transformation of each statistic separately and used a threshold of five standard

deviations away from mean in lower tail for H_p and higher tail for F_{st}. Overlapping windows, after applying the thresholds, were merged separately for windows identified using H_p and F_{st} respectively. This approach identified 14 regions with low H_p and 35 regions with high F_{st}. The average lengths were 400 kb for the regions identified using H_p and 340 kb for the regions identified using F_{st}. We then combined the two approaches to finally identify 36 unique putative domestication regions in autosomes that contained 122 genes.

We used GOstat (Beißbarth & Speed, 2004) to identify genes enriched in biological processes. The gene ontology analysis identified that genes including *MBP*, *VWC2*, *SMO*, *TLX3*, *CYFIP1*, *SH3GL2*, *RNF103* and *YWHAH* were enriched mainly in 'nervous system development'. Additionally other genes with a function in the central nervous system included *GALR1*, *ARID1B*, *NKAIN2*, *CRYM*, *GPR139*, *FRMD6*, *GRIK3*, *VEZT*, *HIPK2*, *ACMSD* and *TCTN3*. This indicate that selection of genetic variants in developmental genes may have played an important role to modify behavior of dogs during domestication (Hare *et al.*, 2012). The second major category of biological processes with enriched genes was related to digestion, fatty acid and starch metabolism. The genes that were enriched include *AMY2B*, *MGAM*, *SGLT1*, *GP2*, *SGLT3*, *TAS2R38*, *ACSM5*, *ACSM2B*, *METAP2* and *FABP5*. This indicates that selection of variants in genes involved in metabolism may have played an important role in adaptation to food with high level of starch during dog domestication.

Main events in starch digestion involve conversion of starch to maltose by the activity of α-amylase in small intestine (Butterworth *et al.*, 2011), which is converted to glucose by the activity of maltase-glucoamylase (Nichols *et al.*, 2003; Mochizuki *et al.*, 2010). Transport of glucose across plasma membrane is facilitated by brush border protein, SGLT1 (Wright *et al.*, 2011) where it enters the bloodstream that distributes it throughout the body (Figure 5).

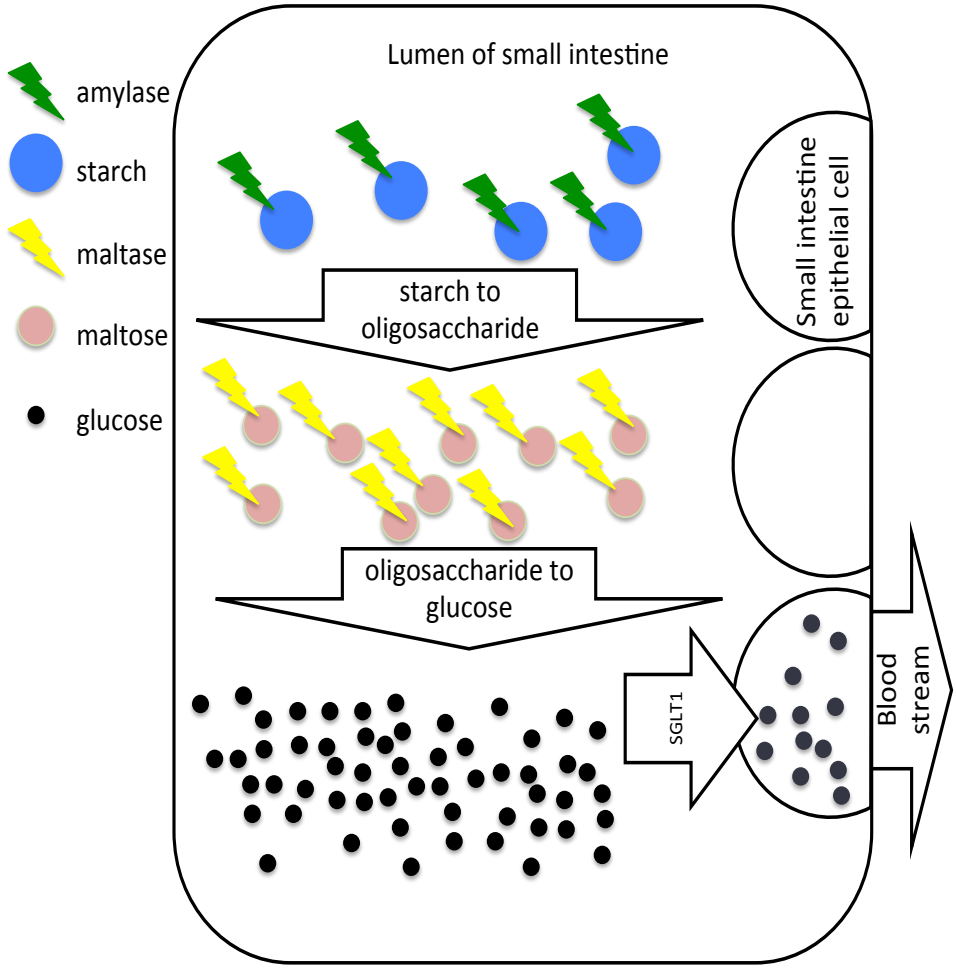

Figure 5. Starch metabolism pathway. Starch is first converted to maltose by the activity of α-amylase and maltose is converted to glucose by the activity of maltase-glucoamylase. Finally glucose is absorbed into the cells facilitated by brush border protein SGLT1.

We used read depth signals to scan for SVs that are fixed in the dog genome. The SVs were overlapped with domestication sweep regions. One of the domestication regions overlapped with one of the strongest signals of CNVs that were fixed in all dogs. The size of the CNV was around 8.8 kbp and the region was annotated using EST sequences to contain *AMY2B* gene. The identification of a CNV and selective sweep in α-amylase (*AMY2B*) along with selective sweeps in maltase-glucoamylase (*MGAM*) and *SGLT1* suggests that selection may have operated on all three stages of starch digestion during dog domestication.

We confirmed copy number differences between 136 wolves and 35 dogs using qPCR. The analysis revealed that all tested wolves were carrying only two copies while all tested dogs were carrying 4 to 30 copies of α-amylase gene. We further confirmed differences in amylase expression using qPCR for the α-amylase gene using pancreatic tissue from 9 dogs and 12 wolves to find out if variation in copy numbers is associated with variation in α-amylase expression. We observed that expression of the α-amylase gene in dogs was 28 times higher than the expression in wolves. We used frozen serum samples from 12 dogs and 13 wolves to test if the increase in *AMY2B copy* number is associated with increased amylase activity and found a 4.7 times higher activity in dogs than in wolves. Similar gene expression analysis and enzyme activity assay was performed to confirm differences in maltase activity in pancreas tissue from dogs and wolves, which revealed a 12 times higher *MGAM* expression in dogs and two times higher activity of MGAM in dogs when compared with wolves. The diet of most dogs today is starch rich while that of wolves is mainly carnivorous. However, there is variation of *AMY2B* copy numbers in dogs (Arendt *et al.*, 2014; Freedman *et al.*, 2014), for instance Dingoes carry only two copies and Huskies carry four copies of *AMY2B* while some wolves also carry greater than two copies of *AMY2B,* suggesting that the diet of early dogs, at the start of domestication, may have shared carnivorous diet with hunter-gatherers (Andersson, 2012; Freedman *et al.*, 2014).

We also genotyped SNPs identified in the maltase and SGLT1 sweeps in 71 dogs and 19 wolves. The dogs were selected from 38 different breeds and wolves from a worldwide distribution. Genotyping was performed to identify the most common haplotype associated with the maltase sweep and revealed that 55 dogs were homozygous and 13 were heterozygous for a specific haplotype whereas three dogs and all of the 19 wolves did not carry this particular haplotype.

In conclusion, out data suggest that selection at all three stages of starch digestion and for altered behavior may have played an important role during dog domestication.

3.3 Signatures of selection in the chicken genome (Paper III)

3.3.1 Background

The chicken is an extremely valuable organism. Firstly, it is a major source of animal protein worldwide (Burt 2005). Secondly, it is an important model organism for genome biology. We use it to study the mechanisms by which animal genomes evolve and function. Initially, the process of chicken domestication may have started by changing the behavior to make it possible to

breed them in captivity. Once the chickens were domesticated, divergent selection for egg and meat production led to the development of distinct populations of two types of chickens, egg layers and meat producing chicken (broilers). Furthermore, dual-purpose chickens and lines carrying specific phenotypic characteristics were also developed.

The pooled sequencing method was previously applied to study signatures of selection and identify fixed deletions in chicken genome (Rubin *et al.*, 2010). It involved whole genome sequencing, to a depth of four to five fold coverage, of pooled samples representing eight populations of domestic chicken and red jungle fowls from zoo populations, as well as the single red jungle fowl female previously used to establish the draft genome assembly for chicken (Hillier *et al.*, 2004). The study revealed 7.5 million single nucleotide polymorphisms (SNPs). The scan for fixed deletions revealed more than 1,000 deletions occurring at a high frequency in at least in one of the sequenced populations and the scan for selective sweeps identified 21 loci with very strong signatures of selection in all domestic chicken populations.

Here we present results from further exploring genetic variants and loci having undergone positive selection in domestic chicken by analysis of more samples and carrying out deeper sequencing.

The data have been used to estimate nucleotide diversity in each population and to identify signatures of selection and structural variations in all domestic, broiler and layer populations.

3.3.2 Results and discussion

We resequenced the genomes of two wild and 12 domestic chicken populations and aligned the reads to the galGal 4.0 chicken reference genome assembly. The unique alignments were used to identify a total of 13.62 million SNPs. We computed pooled heterozygosity (H_p) (Rubin *et al.*, 2010) in all domestic, broiler and layer populations and F_{st} (Weir & Cockerham, 1984) between all domestic and wild, and broiler and layer populations to identify signatures of selection. This revealed 344 selective sweeps in all domestic, 118 in broiler and 29 in layer chicken populations.

The amount of whole genome sequencing data based on short reads from multiple populations generated using different platforms, in public databases are rapidly increasing. It is challenging to combine next generation sequencing short read data from different sequencing platforms and even from different versions of the same platform. The sequences from different sequencing platforms and versions of the sequencing platforms differ, including but not limited to read length and read type. The read type includes single fragments, mate pair and pair end, etc. The reads from different sequencing platforms, different hardware and

software versions have variable sequencing quality scores, different read orientations and sequencing error rates. Furthermore, the different platforms sometimes have specific biases regarding how efficiently they sequence GC-rich DNA and certain other DNA motifs, which result in read depth also differing significantly, creating problems to obtain unbiased test statistics in a study that combines these data. It becomes extremely challenging to combine and analyze the data in a single study. However, this can be done by identification of biases in the data and development of methods that could remove those biases.

We have combined whole genome sequencing data from two different sequencing platforms, SOLiD and Illumina. Furthermore, short reads from SOLiD platform are obtained from different types of fragment libraries, which includes 35 bp single fragments, 50 bp mate pairs and 100 bp mate pairs.

The analyzed data included eight populations each sequenced to 4X-7X coverage using 35bp SOLiD single reads from a previous study (Rubin *et al.*, 2010). In the present study we increased sequence depth for three of those populations and added one more broiler population by generating 50+50 bp SOLiD mate pair libraries. Additionally, we added one wild and three domestic populations by generating 100+100 bp Illumina paired end libraries. All the reads were mapped to the chicken reference genome assembly (ICGSC Nov. 2011Gallus_gallus-4.0). SOLiD reads were mapped using SOLiD BioScope software v.1.3.1 accepting six mismatches and 100 hits per reads while reads from Illumina sequencing platform were mapped to the same reference genome assembly using BWA (Li & Durbin, 2009) also accepting six mismatches. A total of 13.62 million high-quality SNPs were called using GATK v.2.3.6 (McKenna *et al.*, 2010) after quality score normalization. Only biallelic SNPs were included in all analyses, as they comprised 99.29 percent of all the SNPs.

We obtained unbiased estimates of nucleotide diversity by introducing depth filter for each sample. This also involved computation of effective nucleotides in each population. Effective nucleotides constitute the total number of nucleotides with sequence coverage of at least two reads. Finally, nucleotide diversity was computed by introducing a correction factor for population sizes. This was accomplished using the following formula ($\pi = \Sigma H / (\Sigma S \times (1 - 1/2n))$, where ΣH is the total number of heterozygous positions detected using two randomly selected reads for each position, ΣS is the total number of nucleotide positions with at least two reads and n is the number of individuals included in the pool. We estimated nucleotide diversities (π) in the range of 0.1-0.4 for commercial populations of chickens and captive populations of red junglefowl.

We computed pooled heterozygosity (Hp) using average allele frequencies and Fst using allele counts accounting for sample size differences for different contrasts of wild, domestic, layer and broiler chicken populations. We used a

probability distribution function to compute the probability of each 10 kbp window as deviation from the average Hp and Fst. The top convincing putative selective sweeps in all domestic chicken populations, also detected in the previous study, include but are not limited to the regions harboring *TSHR* (thyroid-stimulating hormone receptor) and *VSTM2A* (*V-set and transmembrane domain containing 2A*) genes. Moreover, we identified a putative selective sweep that was not detected in the previous study, which overlaps the 3' end of *ARID1B* encoding AT rich interactive domain 1B (SWI1-like). Furthermore, we identified regions with the most pronounced differentiation between broiler and layers located on chromosome 5. One region overlapped the *NELL1* (*NEL-like 1*) gene while the other was in the interval between *SPRED1*(*sprouty-related EVH1 domain-containing protein 1*) and *MEIS1* (*Meis homeobox 2*).

We identified putative fixed deletions by combining all the read depth from the 12 chicken populations divided into three overlapping groups, (1) all domestic (2) all broilers and (3) all layers and scanned for regions with no read coverage. The probability of each deletion was computed based on median read depth from wild populations combined. This revealed a total of 68 fixed deletions. The deletions shared by all domestic, layer or broiler populations do not delete coding regions in the chicken genome. One of the deletions fixed in all domestic chicken populations was found in intron 7 of *TSHR* and another fixed deletion present in all broiler populations was found in an intron of *ARID1B*.

3.4 Identification of structural variations in multiple populations (Paper IV)

3.4.1 Background

Structural variations include regions of the genome that show copy number variations and chromosomal rearrangements. These may contribute to disease risk as well as to genome evolution (Völker *et al.*, 2010; Alkan *et al.*, 2011). Structural variations include deletions, duplications and insertions, inversions. Read depth (RD) comparisons give better resolution compared to mate pair based approaches to identify breakpoints of duplications and deletions. Current methods for analysis of CNVs from NGS data have limitations and present substantial computational and bioinformatic challenges (Alkan *et al.*, 2011; Teo *et al.*, 2012). Currently available tools that utilize information from RD identify SVs in scans of single samples or pairwise samples. This limits their ability to accommodate complex study designs.

The use of free and open source software in any field of science, particularly biology enables us to apply, modify and enhance computational methods in different experimental settings. R (R Development Core Team, 2011) is a free and open source, interpreted programming language and a software environment for statistical computing and graphics. It runs on a wide variety of UNIX platforms and similar systems including FreeBSD and Linux (including Fedora, Ubuntu, etc.), Mac OS X and Microsoft Windows, which makes the method implementation available to many users running different operating systems.

We developed a novel algorithm that provides an approach for detection of structural variations based on linear modeling and implemented as a package (MultiSV) in R (this thesis).

3.4.2 Method, Results and discussion

We applied a linear mixed model to identify SVs using RD from multiple samples that have been sequenced using any current next generation sequencing method. MultiSV can be used with almost no modifications and it can scan multiple genomes with same ploidy level from any organism.

The installation of MultiSV requires first installation of R (R Development Core Team, 2011), which can be downloaded from http://cran.r-project.org. Additionally, RStudio (Racine, 2012) can also optionally be installed. MultiSV is freely available under LGPL license from http://cran.r-project.org/web/packages/MultiSV/index.html. This also includes an extensive documentation. The following R syntax would install and load MultiSV in R (“>” indicates only the prompt and is not part of the code).

```
> install.packages("MultiSV")
```

MultiSV package includes example data for two pig chromosomes, chr8 and chr9, obtained using SOLiD sequencing from seven pig populations (Rubin *et al.*, 2012). The R syntax below can be used to run of MultiSV on pig sequencing data for chr8 and chr9, which gives output of the structural variations in gff (generic feature format).

```
> library(MultiSV)
> MultiSVExample(MultiSVData)
```

We used a previously described duplication of the *KIT* region in pigs and a known duplication affecting the *AMY2B* gene in dogs to demonstrate the function of MultiSV. For this, we used pig whole genome resequencing data (Rubin *et al.*, 2012) and data from six dog pools and one wolf pool (Axelsson *et*

al., 2013). We can also use MultiSV to compare RD differences from multiple male and female samples to identify scaffolds that belong to sex chromosome in genome assemblies in which scaffolds yet have not been assigned to autosomes and sex chromosomes.

MultiSV can handle data obtained from multiple sequencing platforms and works with model as well as non-model organisms and keeps performing where other tools (e.g. CNVnator, CNV-seq etc.) crash due to multiple reasons, including but not limited to lower memory allocation, very long computation time due to inefficient use of iterations and incompatibility with data structure.

4 General Discussion and Future Perspectives

In this study we identified regions under selection during domestication as well as structural variations using whole genome resequencing of pooled samples from different populations of pig, dog and chicken. This provided insight on the genetic basis for phenotypic evolution during domestication of animals. We associated the mutations and genes with phenotypes by performing additional expression, functional and QTL analysis where applicable. For instance, we performed assays to quantify CNVs and gene expression, serum amylase and maltase activity in dogs, and QTL analysis to study selective sweep candidates in more detail in pigs. Additionally, we also searched the literature as well as QTL databases to learn more about previous knowledge about the genes and chromosomal regions identified in these screens. This revealed convincing associations between putative sweeps and structural variations and the striking phenotypic differences between wild and domestic populations of pig, dog and chicken. The analyses performed in paper I, II and III revealed that artificial selection targeted genes that are mainly involved in metabolism, growth, morphology and behavior. Hence, we have demonstrated that domestic animals can efficiently be used as models for deciphering complex phenotype–genotype relationships.

We have shown that alleles with large phenotypic effects have played a significant role, and strong directional selection may result in the evolution of loci that differ by multiple mutational steps from wild-type alleles. For instance in Paper I, we identified multiple structural changes near *KIT* in the pig genome, which explain variation in coat color in pigs. The *Dominant white* allele is due to at least three causative mutations. The intermediate forms were associated with a subset of the mutations present in *Dominant white*, *Belt* and *Patch*. Additionally, three loci (*NR6A1*, *PLAG1* and *LCORL*) show large phenotypic effects, which probably played a major role to increase body length and number

of vertebrae in pigs. Similarly, Paper II revealed that selection during domestication of dogs at three genes (*AMY2B*, *MGAM* and *SGLT1*) was important to enable dogs to share food with humans, food which is rich in starch.

The analyses performed in Paper I, II, and III only scratch the surface of the selective history of domestication by identifying causal variations underlying few of the selective sweeps in domestic animals. Domestic animals provide an opportunity to further study the loci by developing experimental crosses and populations, which make them an excellent model for deciphering complex phenotype-genotype relationships. Performing similar studies would be very difficult if not impossible in humans (Sabeti *et al.*, 2007). More work is needed to identify causative mutations for sweeps with larger sample size to improve resolution by continuing addition of expanded data sets. This can be achieved by sampling individuals from different breeds representing worldwide distribution and performing a deeper sequencing of each one of them. Moreover, improving the functional annotation of the genes and regulatory factors to elucidate functional role of the causative mutations selected for during domestication, by detailed transcriptome profiling and RNAseq analysis would also enhance the understanding about the mechanisms by which these interact in the genome and underlie physiological, behavioral and morphological changes. It further requires development of computational methods to efficiently perform analysis over the large amount of data generated using Next-generation sequencing. In Paper IV, we suggested a method and provided a tool, MultiSV, for the identification of structural variations in multiple populations based on whole genome resequencing. The better resolution of data from deeper sequencing of multiple populations and development of improved methods to analyze such data would enable us to identify most common alleles across the whole populations with unbiased and precise estimates of allele frequencies, the causal variants and breakpoints of selective sweeps and structural variations.

5 References

Alexander, D. H., Novembre, J. & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19(9), 1655–1664.

Alföldi, J., Di Palma, F., Grabherr, M., Williams, C., Kong, L., Mauceli, E., Russell, P., Lowe, C. B., Glor, R. E., Jaffe, J. D., Ray, D. A., Boissinot, S., Shedlock, A. M., Botka, C., Castoe, T. A., Colbourne, J. K., Fujita, M. K., Moreno, R. G., ten Hallers, B. F., Haussler, D., Heger, A., Heiman, D., Janes, D. E., Johnson, J., de Jong, P. J., Koriabine, M. Y., Lara, M., Novick, P. A., Organ, C. L., Peach, S. E., Poe, S., Pollock, D. D., de Queiroz, K., Sanger, T., Searle, S., Smith, J. D., Smith, Z., Swofford, R., Turner-Maier, J., Wade, J., Young, S., Zadissa, A., Edwards, S. V., Glenn, T. C., Schneider, C. J., Losos, J. B., Lander, E. S., Breen, M., Ponting, C. P. & Lindblad-Toh, K. (2011). The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* 477(7366), 587–591.

Alkan, C., Coe, B. P. & Eichler, E. E. (2011). Genome structural variation discovery and genotyping. *Nature Reviews Genetics* 12(5), 363–376.

Altermann, W. & Kazmierczak, J. (2003). Archean microfossils: a reappraisal of early life on Earth. *Research in microbiology* 154(9), 611–617.

Andersson, L. (2012). How selective sweeps in domestic animals provide new insight into biological mechanisms. *Journal of Internal Medicine* 271(1), 1–14.

Andersson, L. & Georges, M. (2004). Domestic-animal genomics: deciphering the genetics of complex traits. *Nature Reviews Genetics* 5(3), 202–212.

Andersson, L., Haley, C. S., Ellegren, H., Knott, S. A., Johansson, M., Andersson, K., Andersson-Eklund, L., Edfors-Lilja, I., Fredholm, M., Hansson, I. & Et, A. (1994). Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* 263(5154), 1771–1774.

Andersson-Eklund, L., Marklund, L., Lundström, K., Haley, C. S., Andersson, K., Hansson, I., Moller, M. & Andersson, L. (1998). Mapping quantitative trait loci for carcass and meat quality traits in a wild boar x Large White intercross. *Journal of Animal Science* 76(3), 694–700.

Andolfatto, P. (2001). Adaptive hitchhiking effects on genome variability. *Current Opinion in Genetics & Development* 11(6), 635–641.

Archibald, A. L., Haley, C. S., Brown, J. F., Couperwhite, S., McQueen, H. A., Nicholson, D., Coppieters, W., Weghe, A. V. de, Stratil, A., Winterø, A. K., Fredholm, M., Larsen, N. J., Nielsen, V. H., Milan, D., Woloszyn, N., Robic, A., Dalens, M., Riquet, J., Gellin, J., Caritez, J.-C., Burgaud, G., Ollivier, L., Bidanel, J.-P., Vaiman, M., Renard, C., Geldermann, H., Davoli, R., Ruyter, D., Verstege, E. J. M., Groenen, M. a. M., Davies, W., Høyheim, B., Keiserud, A., Andersson, L., Ellegren, H., Johansson, M., Marklund, L., Miller, J. R., Dear, D. V. A., Signer, E., Jeffreys, A. J., Moran, C., Tissier, P. L., Muladno, Rothschild, M. F., Tuggle, C. K., Vaske, D., Helm, J., Liu, H.-C., Rahman, A., Yu, T.-P., Larson, R. G. & Schmitz, C. B. (1995). The PiGMaP consortium linkage map of the pig (Sus scrofa). *Mammalian Genome* 6(3), 157–175.

Archibald, J. D. (1996). Fossil Evidence for a Late Cretaceous Origin of "Hoofed" Mammals. *Science* 272(5265), 1150–1153.

Arendt, M., Fall, T., Lindblad-Toh, K. & Axelsson, E. (2014). Amylase activity is associated with AMY2B copy numbers in dog: implications for dog domestication, diet and diabetes. *Animal Genetics* n/a–n/a.

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

Ayala, F. J., Rzhetsky, A. & Ayala, F. J. (1998). Origin of the metazoan phyla: Molecular clocks confirm paleontological estimates. *Proceedings of the National Academy of Sciences* 95(2), 606–611.

Bansal, V., Harismendy, O., Tewhey, R., Murray, S. S., Schork, N. J., Topol, E. J. & Frazer, K. A. (2010). Accurate detection and genotyping of SNPs utilizing population sequencing data. *Genome Research* 20(4), 537–545.

Barton, N. H. (2000). Genetic hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences* 355(1403), 1553–1562.

Beißbarth, T. & Speed, T. P. (2004). GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* 20(9), 1464–1465.

Boitard, S., Kofler, R., Françoise, P., Robelin, D., Schlötterer, C. & Futschik, A. (2013). Pool-hmm: a Python program for estimating the allele frequency spectrum and detecting selective sweeps from next generation sequencing of pooled samples. *Molecular Ecology Resources* 13(2), 337–340.

Boitard, S., Schlötterer, C., Nolte, V., Pandey, R. V. & Futschik, A. (2012). Detecting Selective Sweeps from Pooled Next-Generation Sequencing Samples. *Molecular Biology and Evolution* 29(9), 2177–2186.

Britten, R. J., Rowen, L., Williams, J. & Cameron, R. A. (2003). Majority of divergence between closely related DNA samples is due to indels. *Proceedings of the National Academy of Sciences of the United States of America* 100(8), 4661–4665.

Brocks, J. J., Logan, G. A., Buick, R. & Summons, R. E. (1999). Archean Molecular Fossils and the Early Rise of Eukaryotes. *Science* 285(5430), 1033–1036.

Burt, D. w. (2002). Applications of biotechnology in the poultry industry. *World's Poultry Science Journal* 58(01), 5–13.

Burt, D. W. (2005). Chicken genome: Current status and future opportunities. *Genome Research* 15(12), 1692–1698.

Butterworth, P. J., Warren, F. J. & Ellis, P. R. (2011). Human α-amylase and starch digestion: An interesting marriage. *Starch - Stärke* 63(7), 395–405.

Carroll, S. B. (2001). Chance and necessity: the evolution of morphological complexity and diversity. *Nature* 409(6823), 1102–1109.

Chabot, B., Stephenson, D. A., Chapman, V. M., Besmer, P. & Bernstein, A. (1988). The proto-oncogene c-kit encoding a transmembrane tyrosine kinase receptor maps to the mouse W locus. *Nature* 335(6185), 88–89.

Chinwalla, A. T., Cook, L. L., Delehaunty, K. D., Fewell, G. A., Fulton, L. A., Fulton, R. S., Graves, T. A., Hillier, L. W., Mardis, E. R., McPherson, J. D., Miner, T. L., Nash, W. E., Nelson, J. O., Nhan, M. N., Pepin, K. H., Pohl, C. S., Ponce, T. C., Schultz, B., Thompson, J., Trevaskis, E., Waterston, R. H., Wendl, M. C., Wilson, R. K., Yang, S.-P., An, P., Berry, E., Birren, B., Bloom, T., Brown, D. G., Butler, J., Daly, M., David, R., Deri, J., Dodge, S., Foley, K., Gage, D., Gnerre, S., Holzer, T., Jaffe, D. B., Kamal, M., Karlsson, E. K., Kells, C., Kirby, A., Kulbokas, E. J., Lander, E. S., Landers, T., Leger, J. P., Levine, R., Lindblad-Toh, K., Mauceli, E., Mayer, J. H., McCarthy, M., Meldrim, J., Mesirov, J. P., Nicol, R., Nusbaum, C., Seaman, S., Sharpe, T., Sheridan, A., Singer, J. B., Santos, R., Spencer, B., Stange-Thomann, N., Vinson, J. P., Wade, C. M., Wierzbowski, J., Wyman, D., Zody, M. C., Birney, E., Goldman, N., Kasprzyk, A., Mongin, E., Rust, A. G., Slater, G., Stabenau, A., Ureta-Vidal, A., Whelan, S., Ainscough, R., Attwood, J., Bailey, J., Barlow, K., Beck, S., Burton, J., Clamp, M., Clee, C., Coulson, A., Cuff, J., Curwen, V., Cutts, T., Davies, J., Eyras, E., Grafham, D., Gregory, S., Hubbard, T., Hunt, A., Jones, M., Joy, A., Leonard, S., Lloyd, C., Matthews, L., McLaren, S., McLay, K., Meredith, B., Mullikin, J. C., Ning, Z., Oliver, K., Overton-Larty, E., Plumb, R., Potter, S., Quail, M., Rogers, J., Scott, C., Searle, S., Shownkeen, R., Sims, S., Wall, M., West, A. P., Willey, D., Williams, S., Abril, J. F., Guigó, R., Parra, G., Agarwal, P., Agarwala, R., Church, D. M., Hlavina, W., Maglott, D. R., Sapojnikov, V., Alexandersson, M., Pachter, L., Antonarakis, S. E., Dermitzakis, E. T., Reymond, A., Ucla, C., Baertsch, R., Diekhans, M., Furey, T. S., Hinrichs, A., Hsu, F.,

Karolchik, D., Kent, W. J., Roskin, K. M., Schwartz, M. S., Sugnet, C., Weber, R. J., Bork, P., Letunic, I., Suyama, M., Torrents, D., Zdobnov, E. M., Botcherby, M., Brown, S. D., Campbell, R. D., Jackson, I., Bray, N., Couronne, O., Dubchak, I., Poliakov, A., Rubin, E. M., Brent, M. R., Flicek, P., Keibler, E., Korf, I., Batalov, S., Bult, C., Frankel, W. N., Carninci, P., Hayashizaki, Y., Kawai, J., Okazaki, Y., Cawley, S., Kulp, D., Wheeler, R., Chiaromonte, F., Collins, F. S., Felsenfeld, A., Guyer, M., Peterson, J., Wetterstrand, K., Copley, R. R., Mott, R., Dewey, C., Dickens, N. J., Emes, R. D., Goodstadt, L., Ponting, C. P., Winter, E., Dunn, D. M., Niederhausern, A. C. von, Weiss, R. B., Eddy, S. R., Johnson, L. S., Jones, T. A., Elnitski, L., Kolbe, D. L., Eswara, P., Miller, W., O'Connor, M. J., Schwartz, S., Gibbs, R. A., Muzny, D. M., Glusman, G., Smit, A., Green, E. D., Hardison, R. C., Yang, S., Haussler, D., Hua, A., Roe, B. A., Kucherlapati, R. S., Montgomery, K. T., Li, J., Li, M., Lucas, S., Ma, B., McCombie, W. R., Morgan, M., Pevzner, P., Tesler, G., Schultz, J., Smith, D. R., Tromp, J., Worley, K. C. & Green, E. D. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420(6915), 520–562.

Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M. F., Kellis, M., Lindblad-Toh, K. & Lander, E. S. (2007). Distinguishing protein-coding and noncoding genes in the human genome. *Proceedings of the National Academy of Sciences* 104(49), 19428–19433.

Clément, J. A. J., Toulza, E., Gautier, M., Parrinello, H., Roquis, D., Boissier, J., Rognon, A., Moné, H., Mouahid, G., Buard, J., Mitta, G. & Grunau, C. (2013). Private Selective Sweeps Identified from Next-Generation Pool-Sequencing Reveal Convergent Pathways under Selection in Two Inbred Schistosoma mansoni Strains. *PLoS Negl Trop Dis* 7(12), e2591.

Darwin, C. (1868). *The Variation of Animals and Plants Under Domestication*. John Murray.

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

Dodgson, J. B. (2007). The Chicken Genome: Some Good News and Some Bad News. *Poultry Science* 86(7), 1453–1459.

Durham, J. W. (1978). The Probable Metazoan Biota of the Precambrian as Indicated by the Subsequent Record. *Annual Review of Earth and Planetary Sciences* 6(1), 21–42.

Ellegren, H., Chowdhary, B. P., Johansson, M., Marklund, L., Fredholm, M., Gustavsson, I. & Andersson, L. (1994). A primary linkage map of the porcine genome reveals a low rate of genetic recombination. *Genetics* 137(4), 1089–1100.

Elsik, C. G., Tellam, R. L. & Worley, K. C. (2009). The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. *Science* 324(5926), 522–528.

Falush, D., Stephens, M. & Pritchard, J. K. (2003). Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* 164(4), 1567–1587.

Fang, M. & Andersson, L. (2006). Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. *Proceedings of the Royal Society B: Biological Sciences* 273(1595), 1803–1810.

Fang, M., Larson, G., Soares Ribeiro, H., Li, N. & Andersson, L. (2009). Contrasting Mode of Evolution at a Coat Color Locus in Wild and Domestic Pigs. *PLoS Genet* 5(1), e1000341.

Fay, J. C. & Wu, C.-I. (2000). Hitchhiking Under Positive Darwinian Selection. *Genetics* 155(3), 1405–1413.

Feuk, L., Carson, A. R. & Scherer, S. W. (2006). Structural variation in the human genome. *Nature Reviews Genetics* 7(2), 85–97.

Freedman, A. H., Gronau, I., Schweizer, R. M., Ortega-Del Vecchyo, D., Han, E., Silva, P. M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., Beale, H., Ramirez, O., Hormozdiari, F., Alkan, C., Vilà, C., Squire, K., Geffen, E., Kusak, J., Boyko, A. R., Parker, H. G., Lee, C., Tadigotla, V., Siepel, A., Bustamante, C. D., Harkins, T. T., Nelson, S. F., Ostrander, E. A., Marques-Bonet, T., Wayne, R. K. & Novembre, J. (2014). Genome Sequencing Highlights the Dynamic Early History of Dogs. *PLoS Genet* 10(1), e1004016.

Fridolfsson, A.-K., Cheng, H., Copeland, N. G., Jenkins, N. A., Liu, H.-C., Raudsepp, T., Woodage, T., Chowdhary, B., Halverson, J. & Ellegren, H. (1998). Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proceedings of the National Academy of Sciences* 95(14), 8147–8152.

Fu, Y.-X. (1997). Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection. *Genetics* 147(2), 915–925.

Fujii, J., Otsu, K., Zorzato, F., Leon, S. de, Khanna, V. K., Weiler, J. E., O'Brien, P. J. & MacLennan, D. H. (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* 253(5018), 448–451.

Gautier, M., Foucaud, J., Gharbi, K., Cézard, T., Galan, M., Loiseau, A., Thomson, M., Pudlo, P., Kerdelhué, C. & Estoup, A. (2013). Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Molecular Ecology* 22(14), 3766–3779.

Giuffra, E., Evans, G., Törnsten, A., Wales, R., Day, A., Looft, H., Plastow, G. & Andersson, L. (1999). The Belt mutation in pigs is an allele at the Dominant white (I/KIT) locus. *Mammalian Genome* 10(12), 1132–1136.

Giuffra, E., Kijas, J. M. H., Amarger, V., Carlborg, Ö., Jeon, J.-T. & Andersson, L. (2000). The Origin of the Domestic Pig: Independent Domestication and Subsequent Introgression. *Genetics* 154(4), 1785–1791.

Giuffra, E., Törnsten, A., Marklund, S., Bongcam-Rudloff, E., Chardon, P., Kijas, J. M. H., Anderson, S. I., Archibald, A. L. & Andersson, L. (2002). A large duplication associated with dominant white color in pigs originated by homologous recombination between LINE elements flanking KIT. *Mammalian Genome* 13(10), 569–577.

Groenen, M. A. M., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.-J., Li, S., Larkin, D. M., Kim, H., Frantz, L. A. F., Caccamo, M., Ahn, H., Aken, B. L., Anselmo, A., Anthon, C., Auvil, L., Badaoui, B., Beattie, C. W., Bendixen, C., Berman, D., Blecha, F., Blomberg, J., Bolund, L., Bosse, M., Botti, S., Bujie, Z., Bystrom, M., Capitanu, B., Carvalho-Silva, D., Chardon, P., Chen, C., Cheng, R., Choi, S.-H., Chow, W., Clark, R. C., Clee, C., Crooijmans, R. P. M. A., Dawson, H. D., Dehais, P., De Sapio, F., Dibbits, B., Drou, N., Du, Z.-Q., Eversole, K., Fadista, J., Fairley, S., Faraut, T., Faulkner, G. J., Fowler, K. E., Fredholm, M., Fritz, E., Gilbert, J. G. R., Giuffra, E., Gorodkin, J., Griffin, D. K., Harrow, J. L., Hayward, A., Howe, K., Hu, Z.-L., Humphray, S. J., Hunt, T., Hornshøj, H., Jeon, J.-T., Jern, P., Jones, M., Jurka, J., Kanamori, H., Kapetanovic, R., Kim, J., Kim, J.-H., Kim, K.-W., Kim, T.-H., Larson, G., Lee, K., Lee, K.-T., Leggett, R., Lewin, H. A., Li, Y., Liu, W., Loveland, J. E., Lu, Y., Lunney, J. K., Ma, J., Madsen, O., Mann, K., Matthews, L., McLaren, S., Morozumi, T., Murtaugh, M. P., Narayan, J., Truong Nguyen, D., Ni, P., Oh, S.-J., Onteru, S., Panitz, F., Park, E.-W., Park, H.-S., Pascal, G., Paudel, Y., Perez-Enciso, M., Ramirez-Gonzalez, R., Reecy, J. M., Rodriguez-Zas, S., Rohrer, G. A., Rund, L., Sang, Y., Schachtschneider, K., Schraiber, J. G., Schwartz, J., Scobie, L., Scott, C., Searle, S., Servin, B., Southey, B. R., Sperber, G., Stadler, P., Sweedler, J. V., Tafer, H., Thomsen, B., Wali, R., Wang, J., Wang, J., White, S., Xu, X., Yerle, M., Zhang, G., Zhang, J., Zhang, J., Zhao, S., Rogers, J., Churcher, C. & Schook, L. B. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491(7424), 393–398.

Gudbjartsson, D. F., Walters, G. B., Thorleifsson, G., Stefansson, H., Halldorsson, B. V., Zusmanovich, P., Sulem, P., Thorlacius, S., Gylfason, A., Steinberg, S., Helgadottir, A., Ingason, A., Steinthorsdottir, V., Olafsdottir, E. J., Olafsdottir, G. H., Jonsson, T., Borch-Johnsen, K., Hansen, T., Andersen, G., Jorgensen, T., Pedersen, O., Aben, K. K., Witjes, J. A., Swinkels, D. W., Heijer, M. den, Franke, B., Verbeek, A. L. M., Becker, D. M., Yanek, L. R., Becker, L. C., Tryggvadottir, L., Rafnar, T., Gulcher, J., Kiemeney, L. A., Kong, A., Thorsteinsdottir, U. & Stefansson, K. (2008). Many sequence variants

affecting diversity of adult human height. *Nature Genetics* 40(5), 609–615.

Guyonnet-Dupérat, V., Geverink, N., Plastow, G. S., Evans, G., Ousova, O., Croisetière, C., Foury, A., Richard, E., Mormède, P. & Moisan, M.-P. (2006). Functional Implication of an Arg307Gly Substitution in Corticosteroid-Binding Globulin, a Candidate Gene for a Quantitative Trait Locus Associated With Cortisol Variability and Obesity in Pig. *Genetics* 173(4), 2143–2149.

Hare, B., Wobber, V. & Wrangham, R. (2012). The self-domestication hypothesis: evolution of bonobo psychology is due to selection against aggression. *Animal Behaviour* 83(3), 573–585.

Hedges, S. B. (2002). The origin and evolution of model organisms. *Nature Reviews Genetics* 3(11), 838–849.

Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., Bork, P., Burt, D. W., Groenen, M. A. M., Delany, M. E., Dodgson, J. B., Map, G. fingerprint, Assembly, S. A., Chinwalla, A. T., Cliften, P. F., Clifton, S. W., Delehaunty, K. D., Fronick, C., Fulton, R. S., Graves, T. A., Kremitzki, C., Layman, D., Magrini, V., McPherson, J. D., Miner, T. L., Minx, P., Nash, W. E., Nhan, M. N., Nelson, J. O., Oddy, L. G., Pohl, C. S., Randall-Maher, J., Smith, S. M., Wallis, J. W., Yang, S.-P., Mapping, Romanov, M. N., Rondelli, C. M., Paton, B., Smith, J., Morrice, D., Daniels, L., Tempest, H. G., Robertson, L., Masabanda, J. S., Griffin, D. K., Vignal, A., Fillon, V., Jacobbson, L., Kerje, S., Andersson, L., Crooijmans, R. P. M., Aerts, J., Poel, J. J. van der, Ellegren, H., Sequencing, cDNA, Caldwell, R. B., Hubbard, S. J., Grafham, D. V., Kierzek, A. M., McLaren, S. R., Overton, I. M., Arakawa, H., Beattie, K. J., Bezzubov, Y., Boardman, P. E., Bonfield, J. K., Croning, M. D. R., Davies, R. M., Francis, M. D., Humphray, S. J., Scott, C. E., Taylor, R. G., Tickle, C., Brown, W. R. A., Rogers, J., Buerstedde, J.-M., Wilson, S. A., Libraries, O. sequencing and, Stubbs, L., Ovcharenko, I., Gordon, L., Lucas, S., Miller, M. M., Inoko, H., Shiina, T., Kaufman, J., Salomonsen, J., Skjoedt, K., Wong, G. K.-S., Wang, J., Liu, B., Wang, J., Yu, J., Yang, H., Nefedov, M., Koriabine, M., deJong, P. J., Annotation, A. and, Goodstadt, L., Webber, C., Dickens, N. J., Letunic, I., Suyama, M., Torrents, D., Mering, C. von, Zdobnov, E. M., Makova, K., Nekrutenko, A., Elnitski, L., Eswara, P., King, D. C., Yang, S., Tyekucheva, S., Radakrishnan, A., Harris, R. S., Chiaromonte, F., Taylor, J., He, J., Rijnkels, M., Griffiths-Jones, S., Ureta-Vidal, A., Hoffman, M. M., Severin, J., Searle, S. M. J., Law, A. S., Speed, D., Waddington, D., Cheng, Z., Tuzun, E., Eichler, E., Bao, Z., Flicek, P., Shteynberg, D. D., Brent, M. R., Bye, J. M., Huckle, E. J., Chatterji, S., Dewey, C., Pachter, L., Kouranov, A., Mourelatos, Z., Hatzigeorgiou, A. G., Paterson, A. H., Ivarie, R., Brandstrom, M., Axelsson, E., Backstrom, N., Berlin, S., Webster, M. T., Pourquie, O., Reymond, A., Ucla, C., Antonarakis, S. E., Long, M., Emerson, J. J.,

Betrán, E., Dupanloup, I., Kaessmann, H., Hinrichs, A. S., Bejerano, G., Furey, T. S., Harte, R. A., Raney, B., Siepel, A., Kent, W. J., Haussler, D., Eyras, E., Castelo, R., Abril, J. F., Castellano, S., Camara, F., Parra, G., Guigo, R., Bourque, G., Tesler, G., Pevzner, P. A., Smit, A., Management, P., Fulton, L. A., Mardis, E. R. & Wilson, R. K. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432(7018), 695–716.

Hudson, R. R., Kreitman, M. & Aguadé, M. (1987). A Test of Neutral Molecular Evolution Based on Nucleotide Data. *Genetics* 116(1), 153–159.

Jensen, J. D., Kim, Y., DuMont, V. B., Aquadro, C. F. & Bustamante, C. D. (2005). Distinguishing Between Selective Sweeps and Demography Using DNA Polymorphism Data. *Genetics* 170(3), 1401–1410.

Jensen, P. (2006). Domestication—From behaviour to genes and back again. *Applied Animal Behaviour Science* 97(1), 3–15 (International Society for Applied Ethology (ISAE) Special Issue 2004 International Society for Applied Ethology (ISAE) Special Issue 2004 A Selection of Papers from the 38th International Congress of the ISAE, Helsinki, Finland, August 2004).

Jensen, P. & Andersson, L. (2005). Genomics Meets Ethology: A New Route to Understanding Domestication, Behavior, and Sustainability in Animal Breeding. *AMBIO: A Journal of the Human Environment* 34(4), 320–324.

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

Karim, L., Takeda, H., Lin, L., Druet, T., Arias, J. A. C., Baurain, D., Cambisano, N., Davis, S. R., Farnir, F., Grisart, B., Harris, B. L., Keehan, M. D., Littlejohn, M. D., Spelman, R. J., Georges, M. & Coppieters, W. (2011). Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nature Genetics* 43(5), 405–413.

Kerje, S., Lind, J., Schütz, K., Jensen, P. & Andersson, L. (2003). Melanocortin 1-receptor (MC1R) mutations are associated with plumage colour in chicken. *Animal Genetics* 34(4), 241–248.

Kijas, J. M. H. & Andersson, L. (2001). A Phylogenetic Study of the Origin of the Domestic Pig Estimated from the Near-Complete mtDNA Genome. *Journal of Molecular Evolution* 52(3), 302–308.

Kijas, J. M. H., Moller, M., Plastow, G. & Andersson, L. (2001). A Frameshift Mutation in MC1R and a High Frequency of Somatic Reversions Cause Black Spotting in Pigs. *Genetics* 158(2), 779–785.

Kijas, J. M. H., Wales, R., Törnsten, A., Chardon, P., Moller, M. & Andersson, L. (1998). Melanocortin Receptor 1 (MC1R) Mutations and Coat Color in Pigs. *Genetics* 150(3), 1177–1185.

Kim, K. S., Larsen, N., Short, T., Plastow, G. & Rothschild, M. F. (2000). A missense variant of the porcine melanocortin-4 receptor (MC4R) gene is associated with fatness, growth, and feed intake traits. *Mammalian Genome* 11(2), 131–135.

Kim, Y. & Nielsen, R. (2004). Linkage Disequilibrium as a Signature of Selective Sweeps. *Genetics* 167(3), 1513–1524.

Kim, Y. & Stephan, W. (2002). Detecting a Local Signature of Genetic Hitchhiking Along a Recombining Chromosome. *Genetics* 160(2), 765–777.

Kimura, R., Fujimoto, A., Tokunaga, K. & Ohashi, J. (2007). A Practical Genome Scan for Population-Specific Strong Selective Sweeps That Have Reached Fixation. *PLoS ONE* 2(3), e286.

King, J. W. B. & Roberts, R. C. (1960). Carcass length in the bacon pig; its association with vertebrae numbers and prediction from radiographs of the young pig. *Animal Science* 2(01), 59–65.

Klungland, H., Vage, D. I., Gomez-Raya, L., Adalsteinsson, S. & Lien, S. (1995). The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. *Mammalian Genome* 6(9), 636–639.

Koboldt, D. C., Chen, K., Wylie, T., Larson, D. E., McLellan, M. D., Mardis, E. R., Weinstock, G. M., Wilson, R. K. & Ding, L. (2009). VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* 25(17), 2283–2285.

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L. & Wilson, R. K. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* 22(3), 568–576.

Kofler, R., Betancourt, A. J. & Schlötterer, C. (2012). Sequencing of Pooled DNA Samples (Pool-Seq) Uncovers Complex Dynamics of Transposable Element Insertions in Drosophila melanogaster. *PLoS Genet* 8(1), e1002487.

Kofler, R., Pandey, R. V. & Schlötterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27(24), 3435–3436.

Kumar, S. & Hedges, S. B. (1998). A molecular timescale for vertebrate evolution. *Nature* 392(6679), 917–920.

Van Laere, A.-S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., Archibald, A. L., Haley, C. S., Buys, N., Tally, M., Andersson, G., Georges, M. & Andersson, L. (2003). A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* 425(6960), 832–836.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R.,

Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.-F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., Bastide, M. de la, Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H.-C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G. R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F. A., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S.-P., Yeh, R.-F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Patrinos, A. & Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature* 409(6822), 860–921.

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., Willer, C. J., Jackson, A. U., Vedantam, S., Raychaudhuri, S., Ferreira, T., Wood, A. R., Weyant, R. J., Segrè, A. V., Speliotes, E. K., Wheeler, E., Soranzo, N., Park, J.-H., Yang, J., Gudbjartsson, D., Heard-Costa, N. L., Randall, J. C., Qi, L., Vernon Smith, A., Mägi, R., Pastinen, T., Liang, L., Heid, I. M., Luan, J., Thorleifsson, G., Winkler, T. W., Goddard, M. E., Sin Lo, K., Palmer, C., Workalemahu, T., Aulchenko, Y. S., Johansson, Å., Carola Zillikens, M., Feitosa, M. F., Esko, T., Johnson, T., Ketkar, S., Kraft, P., Mangino, M., Prokopenko, I., Absher, D., Albrecht, E., Ernst, F., Glazer, N. L., Hayward, C., Hottenga, J.-J., Jacobs, K. B., Knowles, J. W., Kutalik, Z., Monda, K. L., Polasek, O., Preuss, M., Rayner, N. W., Robertson, N. R., Steinthorsdottir, V., Tyrer, J. P., Voight, B. F., Wiklund, F., Xu, J., Hua Zhao, J., Nyholt, D. R., Pellikka, N., Perola, M., Perry, J. R. B., Surakka, I., Tammesoo, M.-L., Altmaier, E. L., Amin, N., Aspelund, T., Bhangale, T., Boucher, G., Chasman, D. I., Chen, C., Coin, L., Cooper, M. N., Dixon, A. L., Gibson, Q., Grundberg, E., Hao, K., Juhani Junttila, M., Kaplan, L. M., Kettunen, J., König, I. R., Kwan, T., Lawrence, R. W., Levinson, D. F., Lorentzon, M., McKnight, B., Morris, A. P., Müller, M., Suh Ngwa, J., Purcell, S., Rafelt, S., Salem, R. M., Salvi, E., Sanna, S., Shi, J., Sovio, U., Thompson, J. R., Turchin, M. C., Vandenput, L., Verlaan, D. J., Vitart, V., White, C. C., Ziegler, A., Almgren, P., Balmforth, A. J., Campbell, H., Citterio, L., De Grandi, A., Dominiczak, A., Duan, J., Elliott, P., Elosua, R., Eriksson, J. G., Freimer, N. B., Geus, E. J. C., Glorioso, N., Haiqing, S., Hartikainen, A.-L., Havulinna, A. S., Hicks, A. A., Hui, J., Igl, W., Illig, T., Jula, A., Kajantie, E., Kilpeläinen, T. O., Koiranen, M., Kolcic, I., Koskinen, S., Kovacs, P., Laitinen, J., Liu, J., Lokki, M.-L., Marusic, A., Maschio, A., Meitinger, T., Mulas, A., Paré, G., Parker, A. N., Peden, J. F., Petersmann, A., Pichler, I., Pietiläinen, K. H., Pouta, A., Ridderstråle, M., Rotter, J. I., Sambrook, J. G., Sanders, A. R., Oliver Schmidt, C., Sinisalo, J., Smit, J. H., Stringham, H. M., Bragi Walters, G., Widen, E., Wild, S. H., Willemsen, G., Zagato, L., Zgaga, L., Zitting, P., Alavere, H., Farrall, M., McArdle, W. L., Nelis, M., Peters, M. J., Ripatti, S., van Meurs, J. B. J., Aben, K. K., Ardlie, K. G., Beckmann, J. S., Beilby, J. P., Bergman, R. N., Bergmann, S., Collins, F. S., Cusi, D., den Heijer, M., Eiriksdottir, G., Gejman, P. V., Hall, A. S., Hamsten, A., Huikuri, H. V., Iribarren, C., Kähönen, M., Kaprio, J., Kathiresan, S., Kiemeney, L., Kocher, T., Launer, L. J., Lehtimäki, T., Melander, O., Mosley Jr, T. H., Musk, A. W., Nieminen, M. S., O'Donnell, C. J., Ohlsson, C., Oostra, B., Palmer, L. J., Raitakari, O., Ridker, P. M., Rioux, J. D., Rissanen, A., Rivolta, C., Schunkert, H., Shuldiner, A. R., Siscovick, D. S., Stumvoll, M., Tönjes, A., Tuomilehto, J., van Ommen, G.-J., Viikari, J., Heath, A. C., Martin, N. G., Montgomery, G. W., Province, M. A., Kayser, M., Arnold, A. M.,

Atwood, L. D., Boerwinkle, E., Chanock, S. J., Deloukas, P., Gieger, C., Grönberg, H., Hall, P., Hattersley, A. T., Hengstenberg, C., Hoffman, W., Mark Lathrop, G., Salomaa, V., Schreiber, S., Uda, M., Waterworth, D., Wright, A. F., Assimes, T. L., Barroso, I., Hofman, A., Mohlke, K. L., Boomsma, D. I., Caulfield, M. J., Adrienne Cupples, L., Erdmann, J., Fox, C. S., Gudnason, V., Gyllensten, U., Harris, T. B., Hayes, R. B., Jarvelin, M.-R., Mooser, V., Munroe, P. B., Ouwehand, W. H., Penninx, B. W., Pramstaller, P. P., Quertermous, T., Rudan, I., Samani, N. J., Spector, T. D., Völzke, H., Watkins, H., Wilson, J. F., Groop, L. C., Haritunians, T., Hu, F. B., Kaplan, R. C., Metspalu, A., North, K. E., Schlessinger, D., Wareham, N. J., Hunter, D. J., O'Connell, J. R., Strachan, D. P., Wichmann, H.-E., Borecki, I. B., van Duijn, C. M., Schadt, E. E., Thorsteinsdottir, U., Peltonen, L., Uitterlinden, A. G., Visscher, P. M., Chatterjee, N., Loos, R. J. F., Boehnke, M., McCarthy, M. I., Ingelsson, E., Lindgren, C. M., Abecasis, G. R., Stefansson, K., Frayling, T. M. & Hirschhorn, J. N. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467(7317), 832–838.

Larson, G., Dobney, K., Albarella, U., Fang, M., Matisoo-Smith, E., Robins, J., Lowden, S., Finlayson, H., Brand, T., Willerslev, E., Rowley-Conwy, P., Andersson, L. & Cooper, A. (2005). Worldwide Phylogeography of Wild Boar Reveals Multiple Centers of Pig Domestication. *Science* 307(5715), 1618–1621.

Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14), 1754–1760.

Li, H. & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature* 475(7357), 493–496.

Li, R., Li, Y., Fang, X., Yang, H., Wang, J., Kristiansen, K. & Wang, J. (2009). SNP detection for massively parallel whole-genome resequencing. *Genome Research* 19(6), 1124–1132.

Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., Kamal, M., Clamp, M., Chang, J. L., Kulbokas, E. J., Zody, M. C., Mauceli, E., Xie, X., Breen, M., Wayne, R. K., Ostrander, E. A., Ponting, C. P., Galibert, F., Smith, D. R., deJong, P. J., Kirkness, E., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C.-W., Cook, A., Cuff, J., Daly, M. J., DeCaprio, D., Gnerre, S., Grabherr, M., Kellis, M., Kleber, M., Bardeleben, C., Goodstadt, L., Heger, A., Hitte, C., Kim, L., Koepfli, K.-P., Parker, H. G., Pollinger, J. P., Searle, S. M. J., Sutter, N. B., Thomas, R., Webber, C., Baldwin, J., Abebe, A., Abouelleil, A., Aftuck, L., Ait-zahra Mostafa, Aldredge, T., Allen, N., An, P., Anderson, S., Antoine, C., Arachchi, H., Aslam, A., Ayotte, L., Bachantsang, P., Barry, A., Bayul, T., Benamara, M., Berlin, A., Bessette, D., Blitshteyn, B., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Brown, A., Cahill, P., Calixte, N., Camarata, J., Cheshatsang, Y., Chu, J., Citroen, M., Collymore, A.,

Cooke, P., Dawoe, T., Daza, R., Decktor, K., DeGray, S., Dhargay, N., Dooley, K., Dooley, K., Dorje, P., Dorjee, K., Dorris, L., Duffey, N., Dupes, A., Egbiremolen, O., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Ferreira, P., Fisher, S., FitzGerald, M., Foley, K., Foley, C., Franke, A., Friedrich, D., Gage, D., Garber, M., Gearin, G., Giannoukos, G., Goode, T., Goyette, A., Graham, J., Grandbois, E., Gyaltsen, K., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Healy, C., Hegarty, R., Honan, T., Horn, A., Houde, N., Hughes, L., Hunnicutt, L., Husby, M., Jester, B., Jones, C., Kamat, A., Kanga, B., Kells, C., Khazanovich, D., Kieu, A. C., Kisner, P., Kumar, M., Lance, K., Landers, T., Lara, M., Lee, W., Leger, J.-P., Lennon, N., Leuper, L., LeVine, S., Liu, J., Liu, X., Lokyitsang, Y., Lokyitsang, T., Lui, A., Macdonald, J., Major, J., Marabella, R., Maru, K., Matthews, C., McDonough, S., Mehta, T., Meldrim, J., Melnikov, A., Meneus, L., Mihalev, A., Mihova, T., Miller, K., Mittelman, R., Mlenga, V., Mulrain, L., Munson, G., Navidi, A., Naylor, J., Nguyen, T., Nguyen, N., Nguyen, C., Nguyen, T., Nicol, R., Norbu, N., Norbu, C., Novod, N., Nyima, T., Olandt, P., O'Neill, B., O'Neill, K., Osman, S., Oyono, L., Patti, C., Perrin, D., Phunkhang, P., Pierre, F., Priest, M., Rachupka, A., Raghuraman, S., Rameau, R., Ray, V., Raymond, C., Rege, F., Rise, C., Rogers, J., Rogov, P., Sahalie, J., Settipalli, S., Sharpe, T., Shea, T., Sheehan, M., Sherpa, N., Shi, J., Shih, D., Sloan, J., Smith, C., Sparrow, T., Stalker, J., Stange-Thomann, N., Stavropoulos, S., Stone, C., Stone, S., Sykes, S., Tchuinga, P., Tenzing, P., Tesfaye, S., Thoulutsang, D., Thoulutsang, Y., Topham, K., Topping, I., Tsamla, T., Vassiliev, H., Venkataraman, V., Vo, A., Wangchuk, T., Wangdi, T., Weiand, M., Wilkinson, J., Wilson, A., Yadav, S., Yang, S., Yang, X., Young, G., Yu, Q., Zainoun, J., Zembek, L., Zimmer, A. & Lander, E. S. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438(7069), 803–819.

Liu, G. E., Matukumalli, L. K., Sonstegard, T. S., Shade, L. L. & Tassell, C. P. V. (2006). Genomic divergences among cattle, dog and human estimated from large-scale alignments of genomic sequences. *BMC Genomics* 7(1), 1–13.

Mardis, E. R. (2008). Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics* 9(1), 387–402.

Marklund, L., Moller, M. J., Sandberg, K. & Andersson, L. (1996). A missense mutation in the gene for melanocyte-stimulating hormone receptor (MCIR) is associated with the chestnut coat color in horses. *Mammalian Genome* 7(12), 895–899.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20(9), 1297–1303.

McQueen, H. A., Fantes, J., Cross, S. H., Clark, V. H., Archibald, A. L. & Bird, A. P. (1996). CpG islands of chicken are concentrated on microchromosomes. *Nature Genetics* 12(3), 321–324.

Meredith, R. W., Janečka, J. E., Gatesy, J., Ryder, O. A., Fisher, C. A., Teeling, E. C., Goodbla, A., Eizirik, E., Simão, T. L. L., Stadler, T., Rabosky, D. L., Honeycutt, R. L., Flynn, J. J., Ingram, C. M., Steiner, C., Williams, T. L., Robinson, T. J., Burk-Herrick, A., Westerman, M., Ayoub, N. A., Springer, M. S. & Murphy, W. J. (2011). Impacts of the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification. *Science* 334(6055), 521–524.

Mikawa, S., Morozumi, T., Shimanuki, S.-I., Hayashi, T., Uenishi, H., Domukai, M., Okumura, N. & Awata, T. (2007). Fine mapping of a swine quantitative trait locus for number of vertebrae and analysis of an orphan nuclear receptor, germ cell nuclear factor (NR6A1). *Genome Research* 17(5), 586–593.

Milan, D., Jeon, J.-T., Looft, C., Amarger, V., Robic, A., Thelander, M., Rogel-Gaillard, C., Paul, S., Iannuccelli, N., Rask, L., Ronne, H., Lundström, K., Reinsch, N., Gellin, J., Kalm, E., Roy, P. L., Chardon, P. & Andersson, L. (2000). A Mutation in PRKAG3 Associated with Excess Glycogen Content in Pig Skeletal Muscle. *Science* 288(5469), 1248–1251.

Mochizuki, K., Honma, K., Shimada, M. & Goda, T. (2010). The regulation of jejunal induction of the maltase–glucoamylase gene by a high-starch/low-fat diet in mice. *Molecular Nutrition & Food Research* 54(10), 1445–1451.

Moller, M. J., Chaudhary, R., Hellmén, E., Höyheim, B., Chowdhary, B. & Andersson, L. (1996). Pigs with the dominant white coat color phenotype carry a duplication of the KIT gene encoding the mast/stem cell growth factor receptor. *Mammalian Genome* 7(11), 822–830.

Morris, S. C. & Caron, J.-B. (2014). A primitive fish from the Cambrian of North America. *Nature* 512(7515), 419–422.

Newton, J. M., Wilkie, A. L., He, L., Jordan, S. A., Metallinos, D. L., Holmes, N. G., Jackson, I. J. & Barsh, G. S. (2000). Melanocortin 1 receptor variation in the domestic dog. *Mammalian Genome* 11(1), 24–30.

Nichols, B. L., Avery, S., Sen, P., Swallow, D. M., Hahn, D. & Sterchi, E. (2003). The maltase-glucoamylase gene: Common ancestry to sucrase-isomaltase with complementary starch digestion activities. *Proceedings of the National Academy of Sciences* 100(3), 1432–1437.

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.

Nishizawa, H., Matsuda, M., Yamada, Y., Kawai, K., Suzuki, E., Makishima, M., Kitamura, T. & Shimomura, I. (2004). Musclin, a Novel Skeletal Muscle-derived Secretory Factor. *Journal of Biological Chemistry* 279(19), 19391–19395.

Ostrander, E. A. & Wayne, R. K. (2005). The canine genome. *Genome Research* 15(12), 1706–1716.

Ottoni, C., Flink, L. G., Evin, A., Geörg, C., Cupere, B. D., Neer, W. V., Bartosiewicz, L., Linderholm, A., Barnett, R., Peters, J., Decorte, R., Waelkens, M., Vanderheyden, N., Ricaut, F.-X., Çakırlar, C., Çevik, Ö., Hoelzel, A. R., Mashkour, M., Karimlu, A. F. M., Seno, S. S., Daujat, J., Brock, F., Pinhasi, R., Hongo, H., Perez-Enciso, M., Rasmussen, M., Frantz, L., Megens, H.-J., Crooijmans, R., Groenen, M., Arbuckle, B., Benecke, N., Vidarsdottir, U. S., Burger, J., Cucchi, T., Dobney, K. & Larson, G. (2012). Pig Domestication and Human-Mediated Dispersal in Western Eurasia Revealed through Ancient DNA and Geometric Morphometrics. *Molecular Biology and Evolution* 30(4), 824–832.

Ovodov, N. D., Crockford, S. J., Kuzmin, Y. V., Higham, T. F. G., Hodgins, G. W. L. & van der Plicht, J. (2011). A 33,000-year-old incipient dog from the Altai Mountains of Siberia: evidence of the earliest domestication disrupted by the Last Glacial Maximum. *PloS One* 6(7), e22821.

Pang, J.-F., Kluetsch, C., Zou, X.-J., Zhang, A., Luo, L.-Y., Angleby, H., Ardalan, A., Ekström, C., Sköllermo, 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* 26(12), 2849–2864.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T. & Reich, D. (2012). Ancient Admixture in Human History. *Genetics* 192(3), 1065–1093.

Patterson, N., Price, A. L. & Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genet* 2(12), e190.

Pennisi, E. (2007). Human Genetic Variation. *Science* 318(5858), 1842–1843.

Primmer, C. R., Raudsepp, T., Chowdhary, B. P., Møller, A. P. & Ellegren, H. (1997). Low Frequency of Microsatellites in the Avian Genome. *Genome Research* 7(5), 471–482.

Pritchard, J. K., Stephens, M. & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155(2), 945–959.

Pryce, J. E., Hayes, B. J., Bolormaa, S. & Goddard, M. E. (2011). Polymorphic Regions Affecting Human Height Also Control Stature in Cattle. *Genetics* 187(3), 981–984.

Przeworski, M. (2002). The Signature of Positive Selection at Randomly Chosen Loci. *Genetics* 160(3), 1179–1189.

Pushkarev, D., Neff, N. F. & Quake, S. R. (2009). Single-molecule sequencing of an individual human genome. *Nature Biotechnology* 27(9), 847–850.

R Development Core Team (2011). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.* [online],. Available from: http://www.R-project.org.

Racine, J. S. (2012). RStudio: A Platform-Independent IDE for R and Sweave. *Journal of Applied Econometrics* 27(1), 167–172.

Rands, C. M., Darling, A., Fujita, M., Kong, L., Webster, M. T., Clabaut, C., Emes, R. D., Heger, A., Meader, S., Hawkins, M. B., Eisen, M. B., Teiling, C., Affourtit, J., Boese, B., Grant, P. R., Grant, B. R., Eisen, J. A., Abzhanov, A. & Ponting, C. P. (2013). Insights into the evolution of Darwin's finches from comparative analysis of the Geospiza magnirostris genome sequence. *BMC Genomics* 14(1), 95.

Reisz, R. R. & Müller, J. (2004). Molecular timescales and the fossil record: a paleontological perspective. *Trends in Genetics* 20(5), 237–241 (Pagination error in this issue, see Publisher's note in Vol. 21 issue 1 p. 36).

Rubin, C.-J., Megens, H.-J., Barrio, A. M., Maqbool, K., Sayyab, S., Schwochow, D., Wang, C., Carlborg, Ö., Jern, P., Jørgensen, C. B., Archibald, A. L., Fredholm, M., Groenen, M. A. M. & Andersson, L. (2012). Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences, USA* 201217149.

Rubin, C.-J., Zody, M. C., Eriksson, J., Meadows, J. R. S., Sherwood, E., Webster, M. T., Jiang, L., Ingman, M., Sharpe, T., Ka, S., Hallböök, F., Besnier, F., Carlborg, Ö., Bed'hom, B., Tixier-Boichard, M., Jensen, P., Siegel, P., Lindblad-Toh, K. & Andersson, L. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464(7288), 587–591.

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., McCarroll, S. A., Gaudet, R., Schaffner, S. F., Lander, E. S., Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S. B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R. C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M. M. Y., Tsui, S. K. W., Xue, H., Wong, J. T.-F., Galver, L. M., Fan, J.-B., Gunderson, K., Murray, S. S., Oliphant, A. R., Chee, M. S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.-F., Phillips, M. S., Roumy, S., Sallée, C., Verner, A., Hudson, T. J., Kwok, P.-Y., Cai, D., Koboldt, D. C., Miller, R. D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.-C., Mak, W., Song, Y. Q., Tam, P. K. H., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C. P., Delgado, M., Dermitzakis, E. T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B. E., Whittaker, P., Bentley, D. R., Daly, M. J., Bakker, P. I. W. de, Barrett, J., Chretien, Y. R., Maller, J., McCarroll, S., Patterson,

N., Pe'er, I., Price, A., Purcell, S., Richter, D. J., Sabeti, P., Saxena, R., Sham, P. C., Stein, L. D., Krishnan, L., Smith, A. V., Tello-Ruiz, M. K., Thorisson, G. A., Chakravarti, A., Chen, P. E., Cutler, D. J., Kashuk, C. S., Lin, S., Abecasis, G. R., Guan, W., Li, Y., Munro, H. M., Qin, Z. S., Thomas, D. J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L. R., Clarke, G., Evans, D. M., Morris, A. P., Weir, B. S., Johnson, T. A., Mullikin, J. C., Sherry, S. T., Feolo, M., Skol, A., Zhang, H., Matsuda, I., Fukushima, Y., Macer, D. R., Suda, E., Rotimi, C. N., Adebamowo, C. A., Ajayi, I., Aniagwu, T., Marshall, P. A., Nkwodimmah, C., Royal, C. D. M., Leppert, M. F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I. F., Knoppers, B. M., Foster, M. W., Clayton, E. W., Watkin, J., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G. M., Yakub, I., Birren, B. W., Wilson, R. K., Fulton, L. L., Rogers, J., Burton, J., Carter, N. P., Clee, C. M., Griffiths, M., Jones, M. C., McLay, K., Plumb, R. W., Ross, M. T., Sims, S. K., Willey, D. L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J. C., L'Archevêque, P., Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A. L., Brooks, L. D., McEwen, J. E., Guyer, M. S., Wang, V. O., Peterson, J. L., Shi, M., Spiegel, J., Sung, L. M., Zacharia, L. F., Collins, F. S., Kennedy, K., Jamieson, R. & Stewart, J. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature* 449(7164), 913–918.

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.

Schmid, M., Nanda, I. & Burt, D. W. (2005). Second report on chicken genes and chromosomes 2005. *Cytogenetic and Genome Research* 109(4), 415–479.

Seilacher, A., Bose, P. K. & Pflüger, F. (1998). Triploblastic Animals More Than 1 Billion Years Ago: Trace Fossil Evidence from India. *Science* 282(5386), 80–83.

Sharma, D. K., Maldonado, J. E., Jhala, Y. V. & Fleischer, R. C. (2004). Ancient wolf lineages in India. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271(Suppl 3), S1–S4.

Shendure, J. & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology* 26(10), 1135–1145.

Signer-Hasler, H., Flury, C., Haase, B., Burger, D., Simianer, H., Leeb, T. & Rieder, S. (2012). A Genome-Wide Association Study Reveals Loci Influencing Height and Other Conformation Traits in Horses. *PLoS ONE* 7(5), e37282.

Skoglund, P., Götherström, A. & Jakobsson, M. (2011). Estimation of Population Divergence Times from Non-Overlapping Genomic

Sequences: Examples from Dogs and Wolves. *Molecular Biology and Evolution* 28(4), 1505–1517.

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

Smith, J. J. & Voss, S. R. (2007). Bird and Mammal Sex-Chromosome Orthologs Map to the Same Autosomal Region in a Salamander (Ambystoma). *Genetics* 177(1), 607–613.

Sundström, E., Komisarczuk, A. Z., Jiang, L., Golovko, A., Navratilova, P., Rinkwitz, S., Becker, T. S. & Andersson, L. (2012). Identification of a melanocyte-specific, microphthalmia-associated transcription factor-dependent regulatory element in the intronic duplication causing hair greying and melanoma in horses. *Pigment Cell & Melanoma Research* 25(1), 28–36.

Tang, K., Thornton, K. R. & Stoneking, M. (2007). A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS biology* [online], 5(7), e171.

Teo, S. M., Pawitan, Y., Ku, C. S., Chia, K. S. & Salim, A. (2012). Statistical challenges associated with detecting copy number variations with next-generation sequencing. *Bioinformatics* 28(21), 2711–2718.

Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V. J., Sawyer, S. K., Greenfield, D. L., Germonpré, M. B., Sablin, M. V., López-Giráldez, 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., Pääbo, S., Green, R. E. & Wayne, R. K. (2013). Complete Mitochondrial Genomes of Ancient Canids Suggest a European Origin of Domestic Dogs. *Science* 342(6160), 871–874.

Thomas, G., Moffatt, P., Salois, P., Gaumond, M.-H., Gingras, R., Godin, É., Miao, D., Goltzman, D. & Lanctôt, C. (2003). Osteocrin, a Novel Bone-specific Secreted Protein That Modulates the Osteoblast Phenotype. *Journal of Biological Chemistry* 278(50), 50563–50571.

Troshina, A., Gustavsson, I. & Tikhonov, V. N. (1985). Investigation of two centric fusion translocations of wild pigs by different banding techniques. *Hereditas* 102(1), 155–158.

Trut, L. (1999). Early Canid Domestication: The Farm-Fox Experiment. *American Scientist* 87(2), 160.

Trut, L., Oskina, I. & Kharlamova, A. (2009). Animal evolution during domestication: the domesticated fox as a model. *BioEssays : news and reviews in molecular, cellular and developmental biology* 31(3), 349–360.

Vaysse, A., Ratnakumar, A., Derrien, T., Axelsson, E., Rosengren Pielberg, G., Sigurdsson, S., Fall, T., Seppälä, E. H., Hansen, M. S. T., Lawley, C. T., Karlsson, E. K., Bannasch, D., Vilà, C., Lohi, H., Galibert, F., Fredholm, M., Häggström, J., Hedhammar, Å., André, C., Lindblad-

Toh, K., Hitte, C., Webster, M. T. & The LUPA Consortium (2011). Identification of Genomic Regions Associated with Phenotypic Variation between Dog Breeds using Selection Mapping. *PLoS Genet* 7(10), e1002316.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Miklos, G. L. G., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Francesco, V. D., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R.-R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z. Y., Wang, A., Wang, X., Wang, J., Wei, M.-H., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S. C., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.-H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigó, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.-H., Coyne, M., Dahlke, C., Mays, A. D., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K.,

Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. & Zhu, X. (2001). The Sequence of the Human Genome. *Science* 291(5507), 1304–1351.

Verginelli, F., Capelli, C., Coia, V., Musiani, M., Falchetti, M., Ottini, L., Palmirotta, R., Tagliacozzo, A., Mazzorin, I. D. G. & Mariani-Costantini, R. (2005). Mitochondrial DNA from Prehistoric Canids Highlights Relationships Between Dogs and South-East European Wolves. *Molecular Biology and Evolution* 22(12), 2541–2551.

Vilà, C., Maldonado, J. E. & Wayne, R. K. (1999). Phylogenetic relationships, evolution, and genetic diversity of the domestic dog. *Journal of Heredity* 90(1), 71–77.

Vilà, C., Savolainen, P., Maldonado, J. E., Amorim, I. R., Rice, J. E., Honeycutt, R. L., Crandall, K. A., Lundeberg, J. & Wayne, R. K. (1997). Multiple and Ancient Origins of the Domestic Dog. *Science* 276(5319), 1687–1689.

vonHoldt, B. M., Pollinger, J. P., Lohmueller, K. E., Han, E., Parker, H. G., Quignon, P., Degenhardt, J. D., Boyko, A. R., Earl, D. A., Auton, A., Reynolds, A., Bryc, K., Brisbin, A., Knowles, J. C., Mosher, D. S., Spady, T. C., Elkahloun, A., Geffen, E., Pilot, M., Jedrzejewski, W., Greco, C., Randi, E., Bannasch, D., Wilton, A., Shearman, J., Musiani, M., Cargill, M., Jones, P. G., Qian, Z., Huang, W., Ding, Z.-L., Zhang, Y., Bustamante, C. D., Ostrander, E. A., Novembre, J. & Wayne, R. K. (2010). Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* 464(7290), 898–902.

Våge, D. I., Klungland, H., Lu, D. & Cone, R. D. (1999). Molecular and pharmacological characterization of dominant black coat color in sheep. *Mammalian Genome* 10(1), 39–43.

Våge, D. I., Lu, D., Klungland, H., Lien, S., Adalsteinsson, S. & Cone, R. D. (1997). A non-epistatic interaction of agouti and extension in the fox, Vulpes vulpes. *Nature Genetics* 15(3), 311–315.

Völker, M., Backström, N., Skinner, B. M., Langley, E. J., Bunzey, S. K., Ellegren, H. & Griffin, D. K. (2010). Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. *Genome Research* 20(4), 503–511.

Wade, C. M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Lear, T. L., Adelson, D. L., Bailey, E., Bellone, R. R., Blöcker, H., Distl, O., Edgar, R. C., Garber, M., Leeb, T., Mauceli, E., MacLeod, J. N., Penedo, M. C. T., Raison, J. M., Sharpe, T., Vogel, J., Andersson, L., Antczak, D. F., Biagi, T., Binns, M. M., Chowdhary, B. P.,

Coleman, S. J., Valle, G. D., Fryc, S., Guérin, G., Hasegawa, T., Hill, E. W., Jurka, J., Kiialainen, A., Lindgren, G., Liu, J., Magnani, E., Mickelson, J. R., Murray, J., Nergadze, S. G., Onofrio, R., Pedroni, S., Piras, M. F., Raudsepp, T., Rocchi, M., Røed, K. H., Ryder, O. A., Searle, S., Skow, L., Swinburne, J. E., Syvänen, A. C., Tozaki, T., Valberg, S. J., Vaudin, M., White, J. R., Zody, M. C., Lander, E. S. & Lindblad-Toh, K. (2009). Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse. *Science* 326(5954), 865–867.

Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., Fan, W., Zhang, J., Li, J., Zhang, J., Guo, Y., Feng, B., Li, H., Lu, Y., Fang, X., Liang, H., Du, Z., Li, D., Zhao, Y., Hu, Y., Yang, Z., Zheng, H., Hellmann, I., Inouye, M., Pool, J., Yi, X., Zhao, J., Duan, J., Zhou, Y., Qin, J., Ma, L., Li, G., Yang, Z., Zhang, G., Yang, B., Yu, C., Liang, F., Li, W., Li, S., Li, D., Ni, P., Ruan, J., Li, Q., Zhu, H., Liu, D., Lu, Z., Li, N., Guo, G., Zhang, J., Ye, J., Fang, L., Hao, Q., Chen, Q., Liang, Y., Su, Y., San, A., Ping, C., Yang, S., Chen, F., Li, L., Zhou, K., Zheng, H., Ren, Y., Yang, L., Gao, Y., Yang, G., Li, Z., Feng, X., Kristiansen, K., Wong, G. K.-S., Nielsen, R., Durbin, R., Bolund, L., Zhang, X., Li, S., Yang, H. & Wang, J. (2008). The diploid genome sequence of an Asian individual. *Nature* 456(7218), 60–65.

Warren, W. C., Clayton, D. F., Ellegren, H., Arnold, A. P., Hillier, L. W., Künstner, A., Searle, S., White, S., Vilella, A. J., Fairley, S., Heger, A., Kong, L., Ponting, C. P., Jarvis, E. D., Mello, C. V., Minx, P., Lovell, P., Velho, T. A. F., Ferris, M., Balakrishnan, C. N., Sinha, S., Blatti, C., London, S. E., Li, Y., Lin, Y.-C., George, J., Sweedler, J., Southey, B., Gunaratne, P., Watson, M., Nam, K., Backström, N., Smeds, L., Nabholz, B., Itoh, Y., Whitney, O., Pfenning, A. R., Howard, J., Völker, M., Skinner, B. M., Griffin, D. K., Ye, L., McLaren, W. M., Flicek, P., Quesada, V., Velasco, G., Lopez-Otin, C., Puente, X. S., Olender, T., Lancet, D., Smit, A. F. A., Hubley, R., Konkel, M. K., Walker, J. A., Batzer, M. A., Gu, W., Pollock, D. D., Chen, L., Cheng, Z., Eichler, E. E., Stapley, J., Slate, J., Ekblom, R., Birkhead, T., Burke, T., Burt, D., Scharff, C., Adam, I., Richard, H., Sultan, M., Soldatov, A., Lehrach, H., Edwards, S. V., Yang, S.-P., Li, X., Graves, T., Fulton, L., Nelson, J., Chinwalla, A., Hou, S., Mardis, E. R. & Wilson, R. K. (2010). The genome of a songbird. *Nature* 464(7289), 757–762.

Weir, B. S. & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 38(6), 1358–1370.

Welch, J. J., Fontanillas, E. & Bromham, L. (2005). Molecular Dates for the "Cambrian Explosion": The Influence of Prior Assumptions. *Systematic Biology* 54(4), 672–678.

Wernersson, R., Schierup, M. H., Jørgensen, F. G., Gorodkin, J., Panitz, F., Stærfeldt, H.-H., Christensen, O. F., Mailund, T., Hornshøj, H., Klein, A., Wang, J., Liu, B., Hu, S., Dong, W., Li, W., Wong, G. K., Yu, J.,

Wang, J., Bendixen, C., Fredholm, M., Brunak, S., Yang, H. & Bolund, L. (2005). Pigs in sequence space: A 0.66X coverage pig genome survey based on shotgun sequencing. *BMC Genomics* 6(1), 70.

Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.-J., Makhijani, V., Roth, G. T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C. L., Irzyk, G. P., Lupski, J. R., Chinault, C., Song, X., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D. M., Margulies, M., Weinstock, G. M., Gibbs, R. A. & Rothberg, J. M. (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452(7189), 872–876.

Wright, E. M., Loo, D. D. F. & Hirayama, B. A. (2011). Biology of Human Sodium Glucose Transporters. *Physiological Reviews* 91(2), 733–794.

Acknowledgements

I would like to thank all colleagues in IMBIM, the department of Animal Breeding and Genetics, friends and family and all who have supported, helped and contributed towards my PhD and this thesis. Special thanks go to:

My supervisor Leif Andersson who accepted me as a PhD student. I am grateful to you for your consistent encouragement, polite behavior, guardian and sympathetic attitude and an overall dynamic supervision throughout the whole process. Without your immense experience, scientific ingenuity and valuable advice this work never have been done. My co-supervisors, Lars Rönnegård, Erik Bongcam-Rudloff, Alvaro Martinez Barrio and Carl-Johan Rubin have been an invaluable source of knowledge and I want to thank them for always having time for questions. Finally, I would like to thank all who gave me valuable time for nice discussions and providing me, directly and indirectly, the opportunity to work with them.

I would like to thank Higher Education Commission (HEC) of Pakistan for providing scholarship during the period of research.

Parents are, truly, a dense of my humblest obligations and my words will fail to give them credit especially to my father, Maqbool-ul-Haq. His love, patience, sacrifice, encouragement and spiritual motivation have always been a great source of inspiration for me. I am greatly indebted and submit my earnest thank to him as he has potentially tolerated agony and provided me all the support to complete my studies. I wish he could live to see his efforts made me successful. Moreover, I must acknowledge my family and relatives for their prayers, patience, sacrifice, understanding and support during all the years of my stay here.

Finally, I again thank all those whom I have mentioned and whom I did not, for helping me in completing this work. The errors that remain are entirely mine.