Detecting Signatures of Selection within the Dog Genome

ABHIRAMI RATNAKUMAR
Abstract


Deciphering the genetic basis of phenotypic diversity is one of the central aims of biological research. Domestic animals provide a unique opportunity for making substantial progress towards this goal. Intense positive selection has lead to a rich reservoir of phenotypes and underlying genotypes that can be interrogated using genetic tools to gain insight into the genetic basis of phenotypic diversity.

The dog is the most phenotypically diverse mammal. It was domesticated from the grey wolf 11-30,000 years ago. After domestication, a period of intense breeding has lead to the massive phenotypic diversity seen amongst dog breeds today. These two phases of strong positive selection at domestication and at breed creation are likely to have left their signature on the genome. In this thesis, we have analysed genome-wide patterns to detect genomic regions involved in selection in both of these phases. We used whole genome sequences from 60 dogs and 12 wolves, to detect dog domestication selective sweeps. We find evidence for genes involved in memory formation, neurotransmission and starch digestion.

To decipher the genetic signals underlying breed diversity, we used genome-wide genotype data from >170,000 SNPs in 509 dogs from 46 different breeds. We find evidence for genes under selection in many breeds, and only a few breeds. In addition, we identify novel sweeps underlying morphology and behavior.

Recombination can influence the configuration of alleles present on a haplotype, and can thus increase or decrease the efficiency of selection. The PRDM9 protein has been shown to be important for determining recombination hotspot locations in humans and other mammals, but of all the mammals studied so far the dog is the only one to have a non-functional PRDM9.

We used the genome-wide genotype data described above to characterise the fine scale recombination map in dogs. We find that recombination hotspots exist in dogs despite the absence of PRDM9. Moreover, we show that these hotspots are enriched for GC rich peaks and that these peaks are getting stronger over time. Our results show that the absence of PRDM9 has lead to the stabilisation of the recombination landscape in dogs.

Keywords: dog, evolution, domestication, PRDM9, recombination, positive selection, selective sweep

Abhirami Ratnakumar, Uppsala University, Department of Medical Biochemistry and Microbiology, Box 582, SE-751 23 Uppsala, Sweden.

© Abhirami Ratnakumar 2013

ISSN 1651-6206
ISBN 978-91-554-8783-6
urn:nbn:se:uu:diva-209335 (http://urn.kb.se/resolve?urn=nbn:se:uu:diva-209335)
List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


II  Ratnakumar, A., Axelsson, E., Webster, M.T., Lindblad-Toh, K., 2013. High Frequency Derived alleles reveal signals of selection within the dog genome *(Manuscript)*


* These authors contributed equally to the work

Reprints were made with permission from the respective publishers.
Additional co-authored publications completed during PhD research

Table of Contents

Introduction .................................................................................................................. 7
  Signatures and tests of Selective Sweeps ............................................................... 8
  Benefits of domestic species ................................................................................. 9
  Benefits of the dog ................................................................................................. 9
  Dog domestication ................................................................................................. 11

Aims of this thesis ..................................................................................................... 13

Papers I & II ............................................................................................................. 14
  Background ........................................................................................................... 14
  Study Design ......................................................................................................... 16
  Results and Discussion for Paper I ..................................................................... 16
  Results and Discussion for Paper II .................................................................... 19

Paper III .................................................................................................................... 22
  Background ........................................................................................................... 22
  Study Design ......................................................................................................... 22
  Results and Discussion ......................................................................................... 23

Paper IV ..................................................................................................................... 27
  Background ........................................................................................................... 27
  Results and Discussion ......................................................................................... 28

General Discussion and Future Perspectives ...................................................... 31

Acknowledgements ................................................................................................. 33

References ............................................................................................................... 35
Abbreviations

bp  base pair
kb  kilobase (1000 bases)
Mb  megabase (10^6 bases)
Gb  gigabase (10^9 bases)
SNP single nucleotide polymorphism
mtDNA mitochondrial DNA
DNA deoxyribonucleic acid
RNA ribonucleic acid
QTL quantitative trait locus
GWA genome-wide association
ENCODE Encyclopedia of DNA Elements
CNV copy number variant
Mya million years ago
Indel insertion or deletion
GERP Genomic Evolutionary Rate Profiling
LD linkage disequilibrium
DAF Derived Allele Frequency
TFBS Transcription Factor Binding Site
GABA Gamma-aminobutyric acid
KEGG Kyoto Encyclopedia of Genes and Genomes
GO Gene Ontology
IBD Identity by Descent
Introduction

Deciphering the genetic basis underlying phenotypic diversity is a central goal in biology. This stems from the human desire to better understand the natural world. In addition, many diseases have a genetic basis so elucidating their genetic underpinnings will help design better treatments.

This central goal seems tractable because of some convenient features of DNA. All species are related to one other through their DNA and DNA can keep record of past events. This is convenient because any detailed investigation into the genes underlying phenotypes in one species can often give insight into the function of those genes and pathways in related species. With the correct statistical tools the genome can be interrogated to reveal the evolutionary past, which can also be of interest from an anthropological perspective.

Genomes are shaped by many forces, one of them is in the form of mutations. The raw material of the genome comes from mutations. Mutations can arise spontaneously at a single base, or as an insertion, deletion or structural variant. Once a new mutation arises its fate is determined by the interplay of the forces of selection and random genetic drift.

Random genetic drift refers to the impact of chance on the fate of mutations [1]. In contrast, selection refers to the force that increases or decreases the frequency of functionally relevant alleles in order to increase the reproductive fitness of the individual.

Mutations can be neutral and have no impact on the reproductive success of the individual or they can be functional. Selection only acts on functional mutations. The new functional mutation can either increase in frequency and become fixed through positive selection or it can decrease in frequency and be removed through purifying selection. The probability of both these outcomes is dependent on the interplay of the initial allele frequency and the effective population size. Variants with low allele frequency are influenced mainly by drift while variants of higher allele frequency are influenced more by selection. In addition, genomes with a small effective population size are more vulnerable to the effects of drift because a new allele needs only to be inherited by a small number of individuals in order to be fixed.

John Maynard Smith and John Haigh [2] suggested in 1974, that as positive selection acts to fix a new mutation it can also impact the allele frequencies of neutral linked loci. They termed this ‘genetic hitchhiking’. This concept forms the basis of a selective sweep, which is when an advantageous
allele rises rapidly in frequency to fixation thereby increasing the frequency of neutral linked loci. ‘Genetic Hitchhiking’ can also occur during the removal of deleterious variants by purifying selection, this is known as background selection [3].

Signatures and tests of Selective Sweeps

The rapid increase in frequency to fixation of an advantageous allele characteristic of a selective sweep can leave its signature on the genome at neutral linked loci. Detection of these features can give insight into the genes and pathways under selection as well as the approximate age of the sweep. Some of the features of a selective sweep are: an excess of high frequency derived alleles, an excess of low frequency derived alleles, increased population differentiation, longer than expected haplotype given the derived allele frequency and reduction in heterozygosity.

Before selection the selected haplotype is likely to contain rare private mutations, which are often derived, immediately after the sweep these once rare, derived mutations become high frequency derived mutations. Thus, one feature of a selective sweep is the greater than expected number of high frequency derived alleles. This is detected by Fay and Wu’s H statistic [4]. If some time has passed since the selective event, the selected haplotype will once again accumulate private derived mutations. This will show the signal of an excess of low frequency derived alleles. This is detected by the Tajima’s ‘D’ statistic[5].

Sometimes an advantageous allele can rise in frequency so fast that the background haplotype on which it is located hasn’t had time to be broken down by recombination. This can be detected with the XP-EHH test [6]. Or, if the selected variant is present on a rare or uncommon haplotype and then rapidly increases in frequency due to a sweep, it can lead to increased population differentiation, which can be detected by F_{ST}. The rapid increase in frequency of an advantageous allele can reduce diversity at linked neutral loci, due to genetic hitchhiking. This can be detected by tests that detect reduced heterozygosity.

Immediately after a sweep, the linkage disequilibrium (LD) of the selected locus is high, but over time this LD breaks down due to recombination. The extent to which LD is broken down can be used as an estimate of the age of the sweep. For instance tests of selection that aim to detect very long regions of high LD will be biased for detecting sweeps soon after they have occurred, while tests that are not based on detecting high LD will likely detect older sweeps.

Many of these tests of selection described above detect deviations from what is expected under neutrality. However, many neutral demographic forces can give signals similar to selection. Distinguishing between selection
and drift is one of the biggest challenges in interpreting these signatures. Examples of demographic forces that can mimic selection include population bottlenecks, population expansions and inbreeding. Population bottlenecks can occur when a population rapidly decreases in size, this has the effect of rapidly reducing the number of haplotypes, and can leave the genome vulnerable to random fluctuations in allele frequency due to random genetic drift. If followed by expansion, a population bottleneck can leave its mark on the genome in the form of long runs of homozygosity and high frequency derived alleles. In addition inbreeding can also increase homozygosity within a population. Since these features are similar to what is expected of selective sweeps, interpreting these signatures in populations known to have experienced bottlenecks and inbreeding can be problematic.

Benefits of domestic species

Domestication is the adaptation of non-human species for the benefit of humans. The characteristic feature of domesticated species is the resultant dependence on humans. Jensen and Andersson[7] described some phenotypic changes that can occur as a result of domestication. Broadly categorized these include changes in:

- Morphology (e.g. size, pigmentation patterns),
- Physiology (e.g. altered reproductive cycle),
- Development (e.g. earlier sexual maturity)
- Behaviour (e.g. reduced fear, increased sociability and reduced response to predators)

The domestication process would have entailed the screening of thousands of phenotypes, in thousands of individuals leading to phenotypic diversity that surpasses what is seen in the wild [8, 9]. Elucidating the genetic basis for these phenotypes can give us insight into the biological basis of phenotypic diversity in general.

Benefits of the dog

The dog provides an excellent model for understanding the genotype phenotype relationship. Due to its unique demographic history and extreme phenotypic diversity. It is the most phenotypically diverse mammal [10], and its diversity extends from morphology such as size and coat colour to behavior such as ‘herding’ and ‘pointing’.

In addition to the phenotypic diversity, the genetic structure of dog breeds have proven valuable for mapping genetic traits. Population bottlenecks at domestication and breed creation have lead to long blocks of linkage dise-
equilibrium within breeds and short blocks of linkage disequilibrium across breeds. This enables a two-step approach for mapping traits, the first of which is to detect genomic regions associated with a phenotype within breeds followed by fine-mapping the causal variant across breeds.

![Figure 1. Two population bottlenecks, one at domestication and one at breed creation. (Modified from Lindblad-Toh et al (2005) [11])](image)

Genomic regions important for phenotypic traits can also be detected without knowledge of the phenotype. This is done when the genome is scanned for signatures of selective sweeps. In this regard dogs are also valuable because they have undergone two bouts of strong positive selection, the first at domestication and the second at breed creation, both of which are likely to have left their mark on the genome.
Dog domestication

The dog was the first species to be domesticated [12], and the only large carnivore to be domesticated. Given the extreme phenotypic diversity observed in modern dog breeds, some biologists like Charles Darwin and Konrad Lorenz suggested that the domestic dog may have had multiple canid ancestors [13, 14]. However DNA analysis has shown that the dog was domesticated from the grey wolf [11, 15]. The question remains however, when, where, how and how many times dogs were domesticated.

The earliest evidence of dog-like fossils come from Siberia and Europe and are dated at being more than 30,000 years old [16, 17]. These fossils are of skulls that are smaller in size, with shorter and broader snouts than wolf skulls. DNA analysis of 413 nucleotides in the mitochondrial control region of one of the fossil skulls dated at 33,000 years old from the Altai Mountains in Siberia suggest a closer relatedness to modern dogs and prehistoric canids than to modern wolves[18].

More recent in the fossil record are dog-like fossils buried with humans found in Germany [19] and Switzerland [20] dated at 14,000 years old. More recent than that is again another finding of dog-like fossils in a human burial in the Levant in modern day Israel dated at 11,500 years old [21, 22]. The oldest dog-like fossils in the Americas were found in Utah and are aged at 9,500 years old [23].

DNA analysis has been done to also shed light on this issue. A study based on analysis of mtDNA from both grey wolves and dogs estimated the divergence of dogs and wolves to be at 100,000 years ago [15]. However, this analysis was based on the unrealistic model that all dogs arose from the same matrilineal lineage. In another study of the mtDNA diversity across dog populations suggested domestication occurred 5,400 to 16,000 years ago in South East China near the Yangtze [24]. Coalescent simulations estimated from genotype data from ancient and modern samples, estimate domestication to have occurred between 10,000 to 30,000 years ago[25]. A recent analysis of nuclear DNA haplotypes suggests dog domestication occurred in the Middle East. However this analysis didn’t account for possible admixture between dogs and wolves [26].

Wolves and early human hunter gatherers had much in common, they were both social, lived in large groups and hunted large game in groups during the day unlike most other predators which are most active at night. Given the very large wolf population worldwide at the time it’s not hard to imagine that a mutually beneficial relationship formed leading to the eventual domestication of the dog.

There are two predominant theories about how dog domestication occurred. The first is one where early humans adopted wolf puppies as pets and over generations they became domesticated. The second is one where the wolves domesticated themselves[27]. In this scenario bold wolves would
have approached settlements to scavenge food and early humans would have removed aggressive individuals and over time they would have become less fearful and less aggressive. After this initial stage it is plausible that early humans realized the benefits of the domestic dog, and the process may have become more intentional.

The dog is a promising model for deciphering the genetic basis underlying phenotypic diversity. In this thesis we describe the detection of both dog domestication selective sweeps, and breed specific selective sweeps, and the use of genome-wide genotype data to characterize the recombination landscape within the dog genome.
Aims of this thesis

• To detect the genomic targets of positive selection during dog domestication

• To detect the genomic targets of artificial selection during breed creation in the domestic dog

• To characterise the recombination landscape in the domestic dog
Background

Benefits of domestic animals

Domestic animals present a unique opportunity to investigate the relationship between genotype and phenotype. Intense scrutiny of thousands of individuals at domestication has lead to a rich reservoir of phenotypes and associated genotypes[9].

A previous report suggests that whole genome sequencing should be useful for detecting variants that distinguish domesticates from their wild ancestors[28]. Thus far, a number of genes have been identified as being functionally different between domestic species and their wild ancestors. These include the coat colour gene \textit{MC1R} [29] and the muscle growth gene \textit{IGF2} [30] in pigs, and the yellow skin gene \textit{BCDO2} [31] and thyroid pathway gene \textit{TSHR} [32] in chickens.

The domestic dog displays striking phenotypic differences from its ancestor the grey wolf. The most striking differences are behavioural [33-37]. This potential for uncovering genes and pathways underlying behaviour makes the detection of domestication genes in dogs especially appealing.

An evolution experiment started in Russia in 1959 has become very insightful for understanding how domestication may have lead to this rapid behavioural transformation. In this experiment foxes were obtained from fur farms and selected for tameness [38-40]. Foxes became “tame” after only a few generations and interestingly displayed phenotypes seen in dogs today.

These traits include changes in coat colour, ear morphology, tail wagging and barking [38, 39]. This suggests that the genes involved have pleiotropic effects [38, 39] and gives us crucial insight into how dog domestication may have happened. It makes plausible the notion that domestication may have happened multiple times in multiple locations.

Using patterns of selection to infer function

There have been major efforts by the international research community to annotate the human and mouse genomes. The main purpose of this is to highlight functional regions. These annotations include ultra-conserved ele-
ments\cite{41,42}, conserved bases\cite{43,44}, chromatin marks\cite{44}, DNase hypersensitivity sites\cite{44}, gene annotations\cite{44}, Chip-seq binding sites\cite{44}, non-coding RNA annotations\cite{44}, and gene expression annotations\cite{44}.

Another enormous effort made by the genomics research community is the 1000 genomes project. The purpose of which was to uncover most of the variation present in the population\cite{45,46}. Particularly rare variants as these were deemed most likely to be functional. The resolution of this variant detection meant they had the power to estimate the strength of purifying selection on different classes of variants.

One method for doing this is to detect an excess of low frequency variants in the derived allele frequency spectrum. An analysis performed by the ‘1000 genomes’ \cite{46} (Figure 2) project illustrates that the strength of purifying selection is correlated with base conservation and shows that the extent of purifying selection is the strongest at stop codons and at highly conserved splice variants and highly conserved non-synonymous mutations \cite{46}. The variants experiencing the strongest purifying selection are the ones that are likely to be very important.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{plot.png}
\caption{The 1000 genomes project analysis shows purifying selection at conserved bases (Figure taken from 1000 genomes consortium (2012)\cite{46})}
\end{figure}

The reduction of sequencing costs has made whole genome sequencing more common. In order to better prioritise the variants obtained from this data it is useful to identify those that are functionally relevant. One way to do that is to detect those that have signals of selection.
Study Design

Papers I and II are based on the whole genome resequencing of a pool of 12 wolves and a pool of 60 dogs. The pool of wolves that was used represents a worldwide distribution (Sweden (n=1), Spain (n=3), Russia (n=3), Belarus (n=2), Bulgaria (n=1), USA (n=1) and Canada (n=1)). The dog samples were obtained from 14 different breeds; with 2 breeds (Drever and Belgian Ter-vuren) containing 12 dogs each and the remaining 12 breeds containing 3 dogs each.

After sequencing, the dog and wolf reads were aligned to the canfam2 dog genome assembly, the dog reads aligned to 94.6% of the genome, while the wolf reads aligned to 91.6%. The average coverage for the dog pools was 29.8x while the average coverage of the wolf pool was 6.2x.

After the reads were aligned, dog and wolf reads were combined for SNP calling. This was done in order to increase sensitivity to rare variants. For a SNP to be called, the variant allele needed to be supported by at least 3 reads. This resulted in the detection of 3.78 million SNPs, 46.8% of which were only segregating in dogs whereas 3.7% were segregating only in wolves. In addition, the aligned reads were used to detect 506,148 short indels and 26, 619 CNVs.

Results and Discussion for Paper I

The aim of the paper was to detect genomic targets of positive selection at dog domestication. To do this, we implemented statistics aimed at detecting signatures of reduced heterozygosity and increased population differentiation in 200kb sliding across the genome.

The statistic for determining reduced heterozygosity ($H_P$) is described in Rubin et al [32]. The number of reads that support the major allele $n_{MAJ}$ and minor alleles $n_{MIN}$ were summed across each SNP in each window, such that the more homozygous the window the lower the value for $H_P$. The resulting $H_P$ values were then z-transformed and windows at the extreme of the distribution with a cut-off of $Z(H_P) < -5$ and where $Z(H_P)_WOLF < Z(H_P)_DOG$ were detected as sites of putative dog domestication selective sweeps.

In addition to detecting regions of reduced heterozygosity we also aimed to detect regions of high population differentiation. We did this by implementing the $F_{ST}$ statistic. The method that was implemented was one that adjusts for sample size differences. We calculated $F_{ST}$ for each SNP and then averaged the scores across the window. Windows with $Z(F_{ST}) > 5$ were deemed as being putative dog domestication sweeps.

Overlapping windows at the extremes of the distribution were collapsed into single regions. In order to minimize the detection of regions that have been fixed by drift, we required that the signal be at least 200kb long. This
identified 14 regions of low heterozygosity with the Hp statistic, with an average length of 400kb, and 35 regions of elevated FST, with an average length of 340kb. Comparison of the two lists of regions revealed 36 unique autosomal regions which together contain 122 genes.

Genes that were identified to be within the boundaries 100kb either side of each sweep were retrieved, the human orthologues of these genes were obtained and subjected to Gene Ontology analysis using the GOStat program [47]. This analysis revealed enrichment for the term ‘nervous system development’ (P= 0.013) and ‘regulation of neuron differentiation’ (P=0.005). Genes in sweeps that were included in this category include MBP, VWC2, SMO, TLX3, CYFIP1 and SH3GL2.

This analysis also revealed enrichment for the terms ‘starch catabolic process’ (P=0.039) and ‘fatty acid metabolism’ (P= 0.031) the genes included in these categories include MGAM which encodes the maltase enzyme and SGLT1 which encodes a glucose transporter.

Analysis of predicted CNVs within the putative sweep regions revealed the strongest signal of a CNV fixed in all dogs on chr6. EST sequences within this region were obtained and aligned across the whole genome, which detected a match to the amylase gene on chromosome unknown. The detection of the amylase CNV in combination with maltase and SGLT1 genes strongly supported the notion that modulation of the starch digestion pathway was important in dog domestication. In order to consolidate this hypothesis, functional studies we undertaken.

Confirmation of the amylase CNV was undertaken by assessing the number of copies present in dogs and wolves using quantitative-PCR. Analysis of 136 dogs and 35 wolves revealed all tested wolves to be carrying 2 copies and all tested dogs to be carrying between 4 and 30 copies. In order to determine whether a difference in copy numbers translated to differences in expression, quantitative-PCR was performed for the amylase gene using pancreatic tissue from 9 dogs and 12 wolves. This revealed a 28-fold increase in expression of amylase in dogs compared to wolves. Serum samples were also collected from dogs and wolves and amylase activity assays were performed to determine whether differences in expression also translated to differences in enzyme activity. Activity assays performed on frozen serum samples from 12 dogs and 13 wolves revealed a 4.7 fold higher activity in dogs.

The starch digestion pathway is such that starch is first cleaved into maltose by amylase in the intestine, this is then digested by maltase, sucrose, and isomaltase to form glucose, and then this glucose is transported across the plasma membrane by SGLT1.

Functional analysis suggests differential activity of amylase in dogs versus wolves. In order to determine if this was also the case for maltase, expression analysis and enzyme activity assays were also performed for maltase.
Gene expression analysis was performed on pancreas tissue from 9 dogs and 8 wolves and revealed a 12-fold increase in expression in dogs. In order to determine whether this difference in expression translated to a difference in enzyme activity, an enzyme activity assay was performed on 7 dogs and 8 wolves and revealed a 2-fold increase in activity in dogs compared to the wolves.

In addition to performing functional studies to estimate any functional differences in maltase and amylase, rigorous analysis was also performed to detect putative causal variants within all of the sweeps. This proved to be enormously challenging due to a number of reasons. The first is the inherent nature of a sweep, which based on the detection methods employed here, is by definition a region where most if not all variants are high frequency or fixed. Another factor is based on the experimental design. The pooled design of the experiment means that allele frequency estimates have limitations. All reads covering an allele could in theory come from just one or small number of individuals, meaning that most individuals in the pool are not sampled which can lead to incorrect estimates of allele frequency. This is further exacerbated by the low average coverage of wolf reads at 6.2x. Related to this is the possibility that the causal variant may not have had enough coverage of reads to have been called a SNP. The final complicating factor is uncertainty regarding the number of putative causal variants likely to be present in the sweep region.

Traditional annotations of non-synonymous, stop codon, UTRs etc. were compiled. Patterns in the haplotype structure were also investigated. Although putative sweeps show increased frequency overall there often still remain regions where consecutive SNPs show higher allele frequency than the sweep overall. The largest of these fixation blocks were used as likely locations of putative causal variants. In addition to haplotype patterns, extensive use was made of available human annotations. These include conservation estimates in terms of conserved elements [42], and base conservation scores. In addition, locations of DNase hypersensitivity sites and chromatin marks that reflect enhancer, promoter and insulator activity [44, 48] were also used. The TRANSFAC [49] database was also used to predict transcription factor binding site motifs, and vertebrate whole genome alignments [50] were used to detect the distribution of species that contain dog and wolf alleles such that inferences could be drawn about whether the species distribution in alleles is consistent with the perceived phenotype underlying the sweep.

Several candidate mutations were identified in this way for the maltase sweep. These mutations include 2 non-synonymous mutations and 1 indel. A non-synonymous SNP was also detected in the SGLT1 sweep.

In addition, genotyping was performed on SNPs detected in the maltase and SGLT1 sweeps in 71 dogs from 38 different breeds and 19 wolves from
a worldwide distribution. For the maltase sweep 68/71 dogs were found to contain at least 1 copy of the selected haplotype whereas no wolves carried this haplotype. Of these 68 dogs, 55 were found to be homozygous while 13 were found to be heterozygous.

Haplotype analysis of 48 random SNPs in the SGLT1 sweep revealed a 50.5kb haplotype that was present in all dogs tested, for which 63 were homozygous and 8 were heterozygous. In contrast only 1 of the 19 wolves was found to carry a single copy of this haplotype.

In conclusion, the study described in paper I above indicates that genes involved in behavior and starch digestion were important for dog domestication. It suggests that the advent of the agricultural revolution is likely to have triggered positive selection that have left its mark on the dog genome.

**Results and Discussion for Paper II**

The aim of this study was to use diversity patterns and the derived allele frequency spectrum to gain insight into how selection has shaped the dog genome. In particular we wanted to firstly evaluate the extent of selection genome-wide, and then detect evidence of positive selection and purifying selection, and finally we wanted to detect selective sweeps that have been targets of artificial selection during dog domestication.

In order to estimate the extent of selection we compared genome-wide recombination rates obtained from Axelsson et al [51] to genome-wide diversity estimates. In doing so, we found a strong positive correlation (Pearson's, $r = 0.47$, $P < 2.2 \times 10^{-16}$). This strong correlation is consistent with either pervasive positive or background selection or increased mutation rate.

The selective arguments for this pattern are based on the idea that diversity at linked neutral loci is reduced due to genetic hitchhiking as advantageous alleles are fixed by positive selection or deleterious alleles are removed by purifying selection. This diversity reducing effect is countered in regions of high recombination due to the increased rate of shuffling of background haplotypes.

One interesting idea is that this result may be due to pervasive positive selection at dog domestication, as seems to be the case in chicken [52]. This can be tested in the future by comparing the strength of correlation between wolf diversity and recombination rate to dog diversity and recombination rate.

We were then interested in detecting evidence of purifying selection and whether this can give insight into the functional relevance of variants. To do this we plotted the derived allele frequency spectrum. Purifying selection aims to remove deleterious variants from the population, this can leave a signature in the form of skew in the derived allele frequency spectrum towards an excess of low allele frequency variants.
This signature was observed for SNPs with high base conservation scores and SNPs in conserved elements. The mean derived allele frequency for all alleles was found to be 0.383 while the mean DAF in conserved elements was 0.373 ($P$ < $2.2 \times 10^{-16}$, Wilcoxon) and the mean DAF for alleles with GERP score $\geq$ 6 was 0.278 ($P$ < $2.2 \times 10^{-16}$, Wilcoxon). This method of using genomic signatures of purifying selection is useful for identifying classes of annotations that are functionally relevant in dogs. This can help in the prioritisation of functional variants in dog re-sequencing studies.

In addition to detecting SNPs subject to purifying selection we also aimed to detect SNPs shaped by positive selection, to do this we obtained SNPs that have evidence of being strongly differentiated between dogs and wolves. These were determined by filtering SNPs with an estimated allele frequency greater than 80% in dogs and less than 20% in wolves. This revealed 88,208 SNPs. The cat and panda genomes were then used as outgroups, to detect SNPs that were likely to be dog derived, these were defined to be those SNPs where the wolf allele matches that of cat or panda ($n$=18,832). In order to consolidate the hypothesis that these SNPs are under positive selection, we determined the mean diversity around them.

We found the mean diversity at all differentiated SNPs to be 2.935 ($P$ < $2.2 \times 10^{-16}$, Wilcoxon), while mean diversity at all SNPs was 3.687. This is consistent with the notion that the set of highly differentiated SNPs are under positive selection. Then we retrieved only the SNPs with dog derived alleles and sub-categorized them based on variant class.

Mean diversity was found to be reduced from 3.436 in conserved elements to 2.902 if also overlapping a chromatin mark classified as being ‘Active Promoter’ [48] ($P$ = $7.25 \times 10^{-10}$, Wilcoxon), and to 3.204 if also overlapping a ‘Strong Enhancer’ chromatin mark[48] ($P$ = $2.64 \times 10^{-15}$, Wilcoxon). A statistically significant reduction in mean diversity was found for the chromatin marks ‘Enhancer’ ($P$ = $1.39 \times 10^{-8}$) and ‘Promoter’, ($P$ = $1.45 \times 10^{-8}$) but was not found for the chromatin mark ‘Insulator’ ($P$ = 0.998).

This suggests that there is stronger selection at conserved elements that also overlap chromatin mark classifications. These chromatin marks were generated in humans by the ENCODE project[44]. This analysis suggests that adding this information can inform function in dogs, and suggests that the use of human chromatin mark annotations are useful for prioritising variants in functional regions in dogs.

We then wanted to determine if the dog derived highly differentiated alleles were enriched in particular parts of the genome, we found that this set of alleles were enriched for being in intergenic regions and at sites overlapping perfect match transcription factor binding site motifs. The proportion of high frequency dog derived alleles that overlap TFBS motifs is 0.46 whereas the proportion of all derived alleles overlapping TFBS motifs is 0.44 ($P$ = $1.42 \times 10^{-6}$, Fishers Exact Test).
Our final aim in this study was to detect genes and pathways under selection due to dog domestication. To do this we detected clusters of high frequency derived alleles that were also associated with depletion in intermediate frequency derived alleles. We detected 823 such regions, containing 4,810 SNPs. We then performed randomisation analyses to determine how likely these clusters are to occur by chance, the mean number of SNPs forming clusters in the 1000 randomisations was 825.

The density of the SNPs within each cluster was used to enrich for putative selective sweep regions, after applying a filter of 0.05, 531 regions containing a total of 451 genes were defined. We then performed KEGG pathway analysis and found an enrichment for “long term depression” ($P=0.0054$) and “axon guidance”($P=0.0039$). “Long term depression” refers to the molecular basis for memory formation at the synapses in the brain, “axon guidance” refers to the process by which neurons send out axons to find targets in the brain. In addition, the sweep regions contained genes and receptors for the main neurotransmitters in the brain.

Our conclusion from this study is that the dog genome may have been subjected to pervasive positive selection at domestication, and that genes involved in memory formation and neurotransmission were likely to be important in the behavioural transformation of the ancestral grey wolf to the domestic dog.
Paper III

Background

After domestication dogs became successful probably due to the agricultural revolution and spread around the world [53] and then a few hundred years ago during the Victorian era were subject to intense breeding [54, 55]. This breeding was aimed at creating and maintaining breed standards and involved intense artificial selection on many traits. This period was important for the extreme phenotypic diversity observed in modern dogs today. This extreme diversity encompasses morphology such as ear type, coat type, coat colour, tail type and size. It also includes diversity in behaviour, as there are breed differences in; aggression, temperament, intelligence, pointing, herding and retrieving behaviour. Since much of the extreme diversity observed in dogs occurred over a very short period of time it is likely that at least some of the variation underlying these extremes in phenotype are controlled by mutations of large effect[56].

Studies based on genome-wide association have already uncovered genes involved in morphology. For example *IGF1* contributes to breed size, *FGF4* contributes to short legs and *HAS2* contributes to the skin wrinkling phenotype seen in the Shar Pei breed.

In addition to associating genotype to phenotype, another strategy is to search for signatures of selective sweeps, and present a map of selection. So that further studies can associate them to phenotypes.

Study Design

The aim of this study was to apply multiple tests of selection to genotyping data from ~170k SNPs from 509 dogs from 30 breeds in order to identify regions under intense artificial selection within these breeds. We divided our search into two components. We wanted to identify common variants under selection that are shared by many breeds with a shared phenotype and also rare, breed-defining variants under selection in one or a few breeds.
Results and Discussion

Common variants
In order to identify common variants shared by many breeds, we applied single SNP $F_{ST}$ and across breed genome-wide association (GWA) studies for both morphological and behavioural traits. A genome wide association study looks for differences in alleles or allele frequency across the genome that associate with a phenotype. They are often applied in a case-control manner within a single breed of interest, but we decided to apply it to identify variants shared across groups of breeds. We did this by assigning breeds to two groups based on a phenotypic value and then searching for associations to the phenotype across the two groups.

Using this method we were able to identify associations for the previously mapped traits of furnishings\cite{57}, size\cite{58} and ear morphology\cite{56} (dropped ear vs pricked ear) and we were also able to map associations to the curly tail phenotype which hasn’t been mapped before.

In addition to across breed GWAs performed on morphological traits we also performed across breed GWAs on behavioural traits. To do this we made use of the “boldness” phenotypic classifications made by expert breeders, in addition to the behavioural phenotypic data available from the Swedish Kennel Club. The Swedish Kennel Club classifies behavioural characteristics into the five categories of sociability, curiosity, playfulness, chase-proneness and aggressiveness.

By making use of these resources we were able to map associations to the boldness and sociability behavioural traits. Interestingly we found that the boldness and dropped ear phenotypes mapped to adjacent/overlapping regions in the genome.

GWA studies require that a phenotype is known in advance so that genotypic associations to the phenotype can be mapped, however it is likely that there are many regions of the genome that have been shaped by selection that associate with a phenotype which is as yet unknown. In order to capture these regions we applied the single SNP $F_{ST}$ statistic to the data.

We did this by applying this statistic to every SNP in the dataset, and found that 240 SNPs had an $F_{ST} > 0.55$ and a minor allele frequency $> 0.15$. We then merged SNPs within 500kb of each other to identify a final set of 44 regions. Each of these 44 regions contains between 1-94 SNPs with elevated $F_{ST}$ and 8 of the 9 regions comprising of more than 3 high $F_{ST}$ SNPs overlap known trait-associated regions, including body size, skull shape, dropped ear, limb length and tail length \cite{56, 58}. These regions also identify the boldness and sociability phenotypes identified by the across breed GWAs.
We also applied a related statistic known as the pairwise fixation index. We did this by multiplying the number of breeds with an allele frequency greater than 0.95 to the number of breeds with a different allele at the same SNP also with an allele frequency greater than 0.95. This statistic highlights the previously mapped regions of dropped ear, furnishings[57], size, long coat and it also picks up regions found in the across breed GWAs for curly tail. In addition to identifying regions that have already been mapped the single SNP $F_{ST}$ and pairwise fixation index highlight many novel regions, and provide a rich resource which can be used to map causal variants.

**Breed-specific variants**

In addition to detecting regions containing common variants shared by multiple breeds, we also applied statistics to detect regions under selection found in 1 or just a few breeds. To do this we applied the $S_i$ and $D_i$ statistics. The $S_i$ test identifies blocks of the genome where one breed has little or no variation, consistent with fixation of a long haplotype by a selective sweep. This statistic was implemented by comparing the number of segregating sites in each breed to all other breeds within 150kb sliding windows. The $D_i$ statistic was calculated in a similar manner, by averaging $F_{ST}$ calculated for each pairwise breed combination across the same 150kb sliding windows used for the $S_i$ analysis. The top $S_i$ and $D_i$ regions were obtained by retrieving the top 1% of all windows and collapsing overlapping windows, this gave a total of 7618 $S_i$ regions and 6404 $D_i$ regions.

Once these statistics were applied, we performed coalescent simulations based on a neutral demographic scenario to create simulated genotypes to which the $S_i$ and $D_i$ statistics were then applied. This was done to estimate the neutral expectation. Comparison of the real and simulated data indicated that the real data contained much longer regions. So we decided to use the length of the regions as a way of enriching for true selective sweeps, we did this by ranking the regions by length and then applying a 5% False Discovery Rate (FDR) to the data, this gave us 524 high confidence $S_i$ regions and 724 high confidence $D_i$ regions.

The $S_i$ statistic was designed to detect drops in heterozygosity while the $D_i$ statistic was designed to detect regions of high differentiation. Using the $S_i$ statistic we were able to identify previously mapped regions including a 590kb region in Dachshunds overlapping the $FGF4$ gene which is associated with the achondroplasia [59], a 1.4Mb region overlapping the $HAS2$ gene associated with the Shar Pei skin wrinkling phenotype [60], homozygous regions were also identified in Yorkshire Terriers and Standard Poodles overlapping the $RSPO2$ gene which is associated with the furnishings phenotype[57]. After ranking the 524 high confidence $S_i$ regions we found that the top 20 regions are all greater than 1 Mb in length, the longest of these re-
regions was 3Mb long and was found in Beagles and was found to overlaps regions from 4 other breeds within the top 20 $S_i$ regions.

**Figure 3.** *FGF4* sweep in Dachshunds and *HAS2* sweep in Shar Peis. The plot on the left shows the drop in relative heterozygosity in Dachshunds associated with the achondroplasia phenotype. The plot on the right shows the drop in heterozygosity for the sweep associated with the Shar Pei skin wrinkling phenotype.

**Figure 4.** Two novel sweeps. The plots represent novel sweeps for phenotypes that are yet to be mapped. The plot on the left corresponds to the longest sweep in Beagles on chr22 and the plot on the right corresponds to the second longest sweep in English Bulldogs on chr26.

After performing the $S_i$ analysis we wanted to devise a strategy that would take advantage of shared sweeps to narrow down the sweep to smaller regions likely to contain the causal variant. To do this we performed IBD analyses by identifying $S_i$ regions shared across multiple breeds and then finding the minimum region of shared fixation across all of the breeds. Applying this technique allowed us to successfully narrow down the homozy-
gous regions overlapping the RSPO2 gene to a 180kb region that contains the previously mapped furnishings[57] mutation. In addition to narrowing down this known mutation we were able to narrow down the largest region of homozygosity which was 3Mb long and which has sub sections shared by 8 breeds to a 484 kb region containing only 2 genes. We were also able to narrow down another region shared by 7 breeds to a 376kb region. This region contains 6 genes one of which is the myostatin gene, mutations in which have been mapped to muscle phenotypes in the Belgian blue cattle breed [61] and the whippet dog breed [62].

The Di statistic revealed highly differentiated segments within large homozygous blocks, three of the top 20 regions identified by the Di statistic overlapped the longest region identified by Si which was also shared among four other breeds within the top 20 Si regions.

In addition to the Si and Di statistics we also applied the XP-EHH statistic. This statistic was developed to detect recent selection across human populations, and is aimed at identifying long haplotypes that are fixed or at high frequency in single breeds compared to others.

However, unlike the human population for which the statistic was designed, the dog genome is very homozygous. This meant that the statistic was too noisy when applied to the dog genome in the standard way. So we decided to use it to validate the regions we had already found using Si and Di. When we did this we found that the XP- EHH provided support for both the Si and Di regions (binomial test: \( P < 10^{-9} \)).

We also performed a GO analysis on the genes contained within the regions identified by single SNP FST, Si and Di. GO analysis performed on the genes closest to singleton SNPs with very high FST values revealed an enrichment for developmental processes, GO analysis performed on regions identified by Si and Di containing single genes revealed an enrichment for genes in developmental processes, central nervous system and pigmentation pathways.

The main outcome from this study is that we defined a highly comprehensive map of selection within the dog genome, we were able to map association to previously unmapped traits such as curly tail, and we were able to map association to behavioural phenotypes which haven’t been done before. In addition we provide a strategy for refining sweep regions shared across multiple breeds.
Background

Recombination is the process by which haplotypes are shuffled. It is important for shaping the combination and configuration of alleles present within a population consequently recombination can have important implications for selection. Linkage between sites under selection can reduce the efficiency of selection in what is known as the 'Hills-Robertson effect'[63]. In addition, the absence of recombination can lead to the accumulation of deleterious known as “Muller’s ratchet”[64]. In order to really understand the relative influence of the various forces shaping a genome, it is important to have knowledge of the recombination landscape.

The predominant method for creating recombination maps in the past has been based on linkage mapping using pedigrees, this method has also been employed in dogs[65] and is in general effective at the broad scale but is limited for estimating recombination at the fine scale.

Software tools based on coalescent theory are now available for inferring recombination rates and recombination hotspots based on linkage disequilibrium structure [66]. These tools work by relating the rate of decay of linkage disequilibrium to infer recombination rates. These tools have been used to construct recombination maps in humans[67-70] and mice[71] and yeast[72, 73]. In humans this analysis has shown that up to 80% of recombination occurs in mostly narrow regions comprising 20 % of the genome known as recombination ‘hotspots’ [69]. In addition as much as 41% of these recombination hotspots contain a 13bp sequence motif [74] which is bound by the protein PRDM9 (PR domain zinc finger protein 9)[75-77].

The PRDM9 protein contains a zinc finger DNA binding domain and a methyltransferase domain. The proposed mechanism of action is that it recognises a sequence motif via its zinc finger binding array and then on binding DNA methylates nearby histones which then attract the recombination machinery and initiates recombination.

Curiously, the locations of recombination hotspots are not shared even among closely related species like humans and chimpanzees[77]. It can vary even within the human population[67, 68], and across different mouse strains[75, 76, 78]. Consistent with this observation is the finding that the PRDM9 protein is under strong positive selection[79]. Since PRDM9 binds DNA via its zinc-finger binding array, variation in this binding array leads to
variation in the sequences recognized thus leading to variation in the specified hotspot locations[79, 80].

GC-biased gene conversion is the neutral recombination associated process, which leads to a bias in the transmission of ‘G’ and ‘C’ alleles during mismatch repair in meiotic recombination. In regions of high recombination this biased transmission of alleles can disrupt sequence motifs originally recognized by PRDM9. Positive selection on PRDM9 and the extinction of hotspots due to biased gene conversion is collectively known as the “hotspot conversion paradox” and underlies the constantly shifting recombination landscape.

Loss of PRDM9 in dogs
Analysis of the PRDM9 protein sequence in various metazoans including 13 rodent species, chimpanzee, orangutan, rhesus macaque, marmoset, African savannah elephant, sloth, European shrew, domestic cat, pig, 4 cattle species and rabbits revealed the dog to be the only mammal studied to have a non-functional PRDM9[79]. This unique feature of dogs leads to fundamental questions about how recombination is initiated in dogs.

The construction of the 172K dog SNP array described in paper III [81] and the initial genotyping of 509 samples on this array provided a valuable opportunity to survey the genome for selective sweeps (paper III). In addition, it also provided the valuable opportunity to survey the recombination landscape within the dog genome. It was also at this time that the non-functionalisation of PRDM9 was reported setting the stage for the detailed investigation of the consequences of this in dogs.

Results and Discussion
We first wanted to determine when PRDM9 become non-functional, did it occur after domestication, or did it occur further back in evolutionary time. To do this, we sequenced exon 7 from 5 canids representing the Canidae family, these included the Bush dog, African wild dog, Red fox, Black backed jackal and Golden jackal. PRDM9 was found to be non-functional in all of these canids. Since Canidae diversified ~7.8 million years ago, the non-functionalisation of PRDM9 must have occurred prior to that. To go further back in evolutionary time, we inspected the PRDM9 sequences in the cat and panda reference assemblies, and found that PRDM9 appeared to be functional in those species. This sets the non-functionalisation of PRDM9 to have occurred at the earliest 49 million years ago, after the lineage leading to the dog split from the lineage leading to the panda.

We then wanted to estimate genome-wide fine scale recombination rates, we did this by obtaining genome-wide genotype data from 173, 622 SNPs
typed in 471 dogs from 30 different breeds. We then randomly sampled 100 haplotypes and then used the ‘Interval’ program from the ‘LDhat’[66] package to estimate recombination rates from this data. ‘Interval’ uses the coalescent framework to estimate linkage disequilibrium boundaries, which are then used to estimate recombination rates. We found that the difference between the lowest and highest rates on all autosomes and X chromosome spans 5 orders of magnitude, which is similar to humans. We also found that in dogs 80% of recombination takes place in 46% of the genome; this is in contrast to humans where 80% of recombination takes place in 20% of the genome.

‘Interval’ assumes a Wright-Fisher population model, but we know that the dog has a unique demographic history with strong selection and extensive bottlenecks, so we wanted to estimate the accuracy of the recombination estimates and also predict our power to detect recombination events. To do this we performed simulations. We used the MACs[82] program to simulate genotype data using the previously published dog demography parameters[83]. We wanted to estimate the strength of breed specific bottlenecks. We did this by simulating genotype data with different bottleneck sizes, we then compared the simulated data to the real data and used the least squares method to estimate the optimal bottleneck parameters for each breed.

We then simulated recombination hotspots of different lengths and intensities and used the program ‘SequenceLDhot’ [84] to detect hotspots in both the real and simulated data. This revealed we had the power to detect approximately 10% of hotspots, and our power to detect increased with the strength and intensity of the hotspots. Overall we detected 4,373 hotspots genome-wide, and given our power to detect was estimated to be 10%, this means that there are likely to be approximately 40,000 recombination hotspots in the dog genome.

We then wanted to determine, if these hotspots were enriched for sequence motifs as is observed in humans. To do this we retained only the 1,683 narrow hotspots (18kb) and then compiled a list of equivalent cold spots. Matched for hotspot regions such as whether it is located within or outside of genes. We then used the Repeatmasker program to search the hot and cold spots for enriched motifs. We found that compared to cold spots, hotspots were enriched for GC rich repeats $rr=2.57$, $p<1\times10^{-16}$.

We then performed analyses on base composition, and to do this we defined GC peaks to be regions with a $>50\%$ increase in GC content. Analysis of GC peaks in hotspots and cold spots revealed 29% of hotspots contained a GC peak compared to 14% of cold spots.

We then wanted to determine whether the GC peaks were a cause or consequence of recombination. We did this through comparisons of alignments to cat and panda. These revealed evidence of weak GC peaks already present in panda. The GC content of GC peaks in panda was 59% whereas in dogs it was found to be 67%. We then analysed the substitution patterns at the GC
peaks from alignments to cat and panda, and found substitution bias from ‘G’ and ‘C’ alleles in the lineage leading to dog. This substitution bias is consistent with the action of biased gene conversion. We then did a comparison of human hotspots to an equivalent list of cold spots defined in the same way as in dog and showed in humans 10% of hotspots have GC peaks versus 13% of cold spots.

In conclusion, we find that on the broad scale recombination rate varies in dog to the same extent as humans, but we find that the recombination landscape in dogs appears to be more stable. Evidence of this stability are the GC biased patterns at recombination hotspots.
General Discussion and Future Perspectives

In Papers I and II we identify targets of selection during dog domestication. We show genes involved in starch digestion and behavior to be important. In Paper III we identify targets of selection at breed creation, including genes involved in morphology and behavior. In Paper IV we show that \textit{PRDM9} has been lost in canids resulting in the stabilization of the recombination landscape, and that dog recombination hotspots are associated with GC peaks.

Papers I-III detect both breed-defining selective sweeps and domestication selective sweeps. In the future it would be fruitful to find the causal variants underlying these sweeps. The most important challenge in this regard is the identification of the associated phenotype. For at least some of the sweeps, insight can be gained by performing literature searches on the genes and by reviewing known expression patterns. In addition, overlaying the sweeps with results from QTL studies can also be beneficial. Once a putative phenotype is associated with the sweep, it would be beneficial to genotype more dogs by sequencing. This would help refine the sweep boundaries, and make causal variant detection easier.

In the case of the domestication sweeps, one approach might be to perform phenotypic characterisation of dog-wolf hybrids and then association analysis to assess whether segregating phenotypes associate with segregating genotypes.

Paper I revealed starch digestion to be important in dog domestication. This suggests that these sweeps occurred at the time of the agricultural revolution. Papers I and II also detected signatures of selection at behaviour genes. An interesting comparison would be to quantify the decay in linkage disequilibrium at the starch digestion sweeps and compare it to the decay in linkage disequilibrium in the behavior associated sweeps. This comparison would need correction for recombination rate but it could prove insightful for dating sweeps and determining which phenotype accompanied the very first transition from wild wolf to the domestic dog.

Another insightful comparison would be to compare the genes and pathways identified in the studies in Papers I and II to any genes and pathways that are identified in the future as distinguishing the tame from aggressive foxes from the Belyaev “farm fox” experiment. This can be extended further to similar studies in other domestic species and can elucidate whether domestication involved convergent evolution.
Paper II also described patterns of both positive and purifying selection at variant classes. In the future it would be useful to perform sequencing of more individuals and to a greater depth of coverage in order to detect rare variants. This can increase power to detect evidence of purifying selection using the derived allele frequency spectrum. This can be useful for giving insight into the types of annotations that are relevant for prioritising putative causal variants in dogs.

In Paper III we detected targets of artificial selection at breed creation. One of the challenges faced was the difficulty in distinguishing selection from drift. This challenge can be reduced if the number of dogs is increased. Repeating the analysis on more dogs from more breeds would result in a higher resolution map with less false positives.

The finding of an association between GC peaks and recombination in Paper IV, raises the question about what extent this is a cause and to what extent it is a consequence of recombination. One way to answer this would be to characterize further the exact mechanism of action of PRDM9 in dogs and whether there are features associated with the GC peaks that attract recombination. In addition, PRDM9 has been dubbed a ‘speciation gene’ [85]because of its role in hybrid sterility[85, 86], however this is not the case in canids. This suggests the existence of compensatory mutations or pathways in canids that account for their fertility. In the future it might be useful to perform sequencing to uncover these pathways.
Acknowledgements

Kerstin and Matt for your advice, support and supervision throughout my PhD. Kerstin, for giving me the opportunity to work on really exciting projects, I think its great that you are bold in your approach to science! Matt, for hiring me as a PhD student and for giving me useful advice and suggestions for analyses. When I first started, I had many ideas and wanted to go in many different directions, but working with you has shown me the importance of being focused.

Erik and Maja for the collaboration on the dog domestication project. Erik you are always relaxed and smiling! Thanks for your work on detecting sweeps. Maja, for collecting samples and showing me how to do qPCR.

Cecilia, for being really helpful whenever I had questions

Leif, for interesting discussions, I really enjoy talking to you about my work because you are so passionate about Science!

Birgitta, for helping me design the maltase enzyme activity assay.

Sergey, for helping me with lab work.

Eva, for your work on the sequence capture experiment.

Ulla for being helpful whenever I had questions.

Doreen, Yulna and Alex for waiting with me at the hospital when I tried to implement sophisticated manoeuvres and my horse didn’t want to cooperate.

Hava, for entertaining conversations during teaching. Next time a student says ‘trevlig helg’ I will say ‘detsamma’ instead of ‘tillsammans’….

I want to thank everyone in D11:3 for being helpful and for the interesting scientific discussions and for making my time in Uppsala fun!
I’d like to thank all the amazing friends I’ve made in Uppsala for the many fun times! Especially Sonja, Fabiana, Camille and Vera for being extra supportive in the last year, when I’ve had little time for socialising.

I would also like to thank my friends in Sydney, Ang, Dhiv, Cindy, Nicole and Rachael, for being supportive and still keeping in touch even though it’s been nearly 10 years since I left Sydney.

I’d like to thank my two sisters Karthiga and Jarn, for being the best support network anyone could ask for, even though all three of us live in three different time zones, I look forward to the daily emails and regular Skype conversations.

Last but not least, I would like to thank my parents for always supporting me in whatever I want to do, even if sometimes you think I’m a little bit nuts.
References


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 938

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine.

Distribution: publications.uu.se
urn:nbn:se:uu:diva-209335