GENETIC RECONSTRUCTION OF PARENTAGE AND KINSHIP IN SEMI-FERAL DOMESTIC DOGS, AND ANALYSIS OF EFFECTS OF DOG BREEDING PATTERNS ON AN IMMUNE SYSTEM GENE MARCH7

UNIVERSITY OF
LINCOLN

Masters by Research

Written by Eleanor Karp-Tatham, BSc

09200397

Supervisors: Dr Malgorzata Pilot & Dr Andre Moura

School of Life Sciences – University of Lincoln, United Kingdom

Collaborator: Dr Eugenia Natoli

Dedicated to Stefan Karp

Certificate of Originality

This is to certify that I am responsible for the work submitted in this thesis, that the original work is my own, except as specified in the acknowledgements and in references, and that neither the thesis nor the original work contained therein has been previously submitted to any institution for a degree.

All work carried out and materials used were obtained and completed at the University of Lincoln.

Signature:

_____

Name: Eleanor Karp-Tatham

Date:

Acknowledgments

I would like to dedicate this thesis to my grandfather, Stefan Karp, for his support throughout all my university studies and who made it possible for me to study a Masters. I would like to thank my parents, for continual reassurance and support; Angela Karp and Arthur Tatham.

I wish to thank my supervisors; firstly to Dr Malgorzata Pilot, for her continual support, guidance and belief in my ability throughout this entire project and for years past and secondly Dr Andre Moura for essential training in relation to bioinformatics analysis.

I would like to thank Dr Eugenia Natoli and her collaborators at Azienda USL Roma D, Area Dipartimentale Sanita Pubblica Veterinaria for provision of free-ranging dog samples and behavioural information relating to results. I also wish to thank Prof. Wieslaw Bogdanowicz for samples of free-ranging dogs and Asian wolves.

I would like to thank Chris Rawlings, from Rothamsted research, for his invaluable time, enduring patience and knowledge that made the use of QUARK and Phyre software for protein structural prediction possible.

And last but not least, special thanks to Fernanda Fadel, for provision of pure-breed dog samples, help with bioinformatics analysis, months of support and a valuable friendship.

Table of Contents

List of Tables:

List of Figures:

1. ABSTRACT

Whilst there has been considerable research focusing on the kinship of wolves, data on free-ranging dogs was sparse and there has been a long standing controversial debate over their ability to form packs. One of the aims of this project was to reconstruct kinship relationships in a population of free-ranging dogs, assessing the genetic variability and inbreeding level. For this purpose, I studied a population inhabiting a nature reserve at the outskirts of Rome in Italy. Analysis of twelve microsatellite loci revealed low number of alleles per locus, low levels of heterozygosity and difficulties in assigning parentage, possibly resulting from high levels of inbreeding in the population. Results from parentage analysis suggested multiple breeding individuals to be present in the social groups. One explanation for this is a result of the domestication process as free-ranging dogs no longer follow seasonal reproductive behaviour and have a plentiful supply of human waste to scavenge reducing competition. Although parentage analysis suggested multiple paternity for two litters, results had low statistical support and could be due to low genetic variability in the population.

Recent research has found *MARCH7* as a common candidate gene under diversifying selection between free-breeding dogs and either East Asia or European dog breeds, with a SNP labelled in the intronic region of the gene. *MARCH7* belongs to the membrane-associated RING-CH (MARCH) family, a RING finger protein family of E3 ubiquitin ligases, consisting of 11 members in mammals. The second aim of this study was to test for the possible signals of diversifying selection between free-ranging dogs, pure-breed dogs and wolves in the *MARCH7* gene. This was achieved through three main routes: Sanger sequencing of a targeted region previously identified as being under selection, evolutionary comparison through investigation of nonsynonymous and synonymous patterns and phylogenetic analysis of mammalian species and ab initio prediction of protein structure . Sequence analysis demonstrated the possibility of copy number variation and alternative splicing in *MARCH7* but failed to show polymorphism at the previously identified intronic SNP. Comparative analysis demonstrated *MARCH7* to have highly conserved regions, most notably the RING-CH domain, but also polymorphic regions, where a multitude of both synonymous and nonsynonymous mutations are present across mammalian species studied. Comparison of nonsynonymous and synonymous mutations demonstrated *MARCH7* to be under purifying selection across mammalian species. Ab initio prediction of protein structure indicated a highly

disordered structure across the majority of the gene, with the exception of the RING-CH domain.

2. BACKGROUND

The process of domestication has played a crucial role in human civilization and has been instrumental in allowing the human race to rapidly spread throughout the globe (Wright, 2015). Domestication is dependent on the formation of relationships between humans and animals (Zeder, 2006), with wolves (*Canis lupus)* and dogs (*Canis lupus familiaris*) representing two powerful icons of these types of complex relationships (Lescureux and Linnell, 2014).

The overall aims of this thesis are to investigate the effects domestication on canids and the resulting effects on mating systems, reproductive behaviour and traits under selection, specifically the immune system.

2.1 CANINE DOMESTICATION

Dogs are considered the first human domesticated species (Cagan and Blass, 2016) and broadly speaking the process of domestication can be simplified into two main stages (Figure 2-1). Firstly, dogs were domesticated from their wild ancestor, the gray wolf. Ever since, humans and dogs have lived commensally, utilising the same common food resources and living environment (Wang *et al.,* 2016). Secondly, during the past few hundred year humans have selectively chosen from the gene pool, small sets of dogs, for novel and desired traits resulting in the formation of distinct breeds (Vaysse *et al.,* 2011). New breeds are continually generated through admixture and strong selection for specific physiological, morphological, and behavioural traits (Alvarez and Akey, 2012).

Specific details concerning the process of domestication, for example about geographical location, number of domestication events and time estimates remain highly contentious (Skoglund *et al.,* 2015). A combination of analysis of whole genome sequence data (Freedman *et al.,* 2014), archaeological remains (Ovodov *et al.,* 2011; Germonpré *et al.,* 2012) and mitochondrial genomes of extant and ancient canid lineages (Thalmann *et al.,* 2013) provide evidence for a pre-agricultural origin of dogs, which began through an association with hunter-gatherers (Freedman *et al.,* 2016). Due to the phenomenal diversity witnessed in domestic dogs today the topic of whether a single wild species or multiple species were at the origin had been vastly discussed in the past. Recent evidence from a combination of studies on vocalisations, behaviour, morphology and molecular biology indicate clearly that the wolf is the principal ancestor (for example: Galibert *et al.,* 2011). However, recent results suggest that divergence of modern wolf populations is

contemporaneous with dog domestication, and therefore modern wolf populations cannot be used to determine the location where domestication first occurred (Freedman *et al.*, 2014; Witt *et al.*, 2015).

Pure-breed dogs, free-ranging dogs and wolves remain a single species with evidence for past and ongoing gene flow between them (Alvarez and Akey, 2011; Veradi *et al.*, 2006; vonHoldt *et al.*, 2010). Despite this, each group has distinct characteristics that are outlined briefly in the following sections to provide background context before considering mating systems, selection pressure and the impact of the domestication process on immunity.

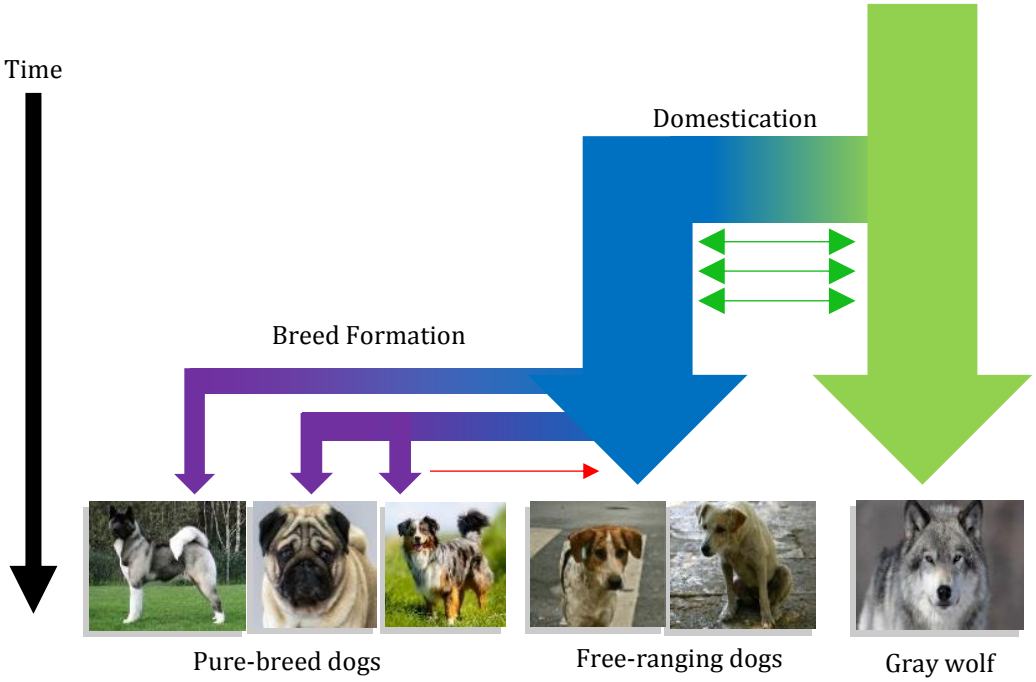


Figure 2-1: Process of canine domestication. Green (wolf) and blue (free-ranging dogs) arrows represent different evolutionary lineages, purple arrow (pure-breeds) represent an ancient (far left) and two modern breeds. Small green arrows depict the possibility of multiple domestication events occurring from wolves, followed by localized dog-wolf introgression. Red arrow represents admixture between pure-breed dogs and free ranging dog populations (Sourced and modified from Boyko, 2011).

2.1.1 *Canis lupus*: The gray wolf

The wolf is a highly adaptable top predator, widely distributed throughout the world (Lucchini *et al.*, 2004; Randi, 2011) and despite multiple differences in phenotypic traits compared to the domestic dog, there is just ~0.047 % difference of nuclear coding-DNA sequence (Cagan and Blass, 2016). Wolves are social carnivores; however pack structure and dynamics are complex and may differ in varying ecological conditions (Randi, 2011). Wolves are known to be territorial, scent marking and defending their territory, which can remain stable for multiple successive breeding pairs (Caniglia *et al.*, 2014).

Wolves are capable of occupying various habitats and have long-distance dispersal capabilities; however gene flow between regional populations is often restricted (Pilot *et al.,* 2006; Stronen *et al.,* 2012; Niskanen *et al.,* 2014). Vilà *et al.,* (1999) analysed mitochondrial DNA (mtDNA) control region sequence data of worldwide wolves, showing that local, fine-scaled population structure exists, likely as a result of recently restricted gene flow.

Topographical barriers are commonly, but not always, associated with differentiation between wolf populations (Aspi *et al.,* 2006). Other influential factors are also seen, such has prey specialisation (Carmichael *et al.,* 2001). In southern parts of Finland, wolves predominately hunt moose, whilst in the eastern and northern regions of the country their diet is mostly made up of semi-domestic and wild reindeer (Kojolo *et al.,* 2004), possibly contributing to population structure.

2.1.2 *Canis lupus familiaris:* The Domestic Dog

The domestic dog (*Canis lupus familaris*) is the result of one of the longest running and largest breeding experiments conducted by humans (Shearman and Wilton, 2011). Dogs are unique in exhibiting the greatest phenotypic diversity among mammalian species (Lequarre *et al.,* 2011), including variation both in conformation and size (Rimbault and Ostrander, 2012). This incredible diversity led Charles Darwin to hypothesize that the domestic dog must have descended from at least two common ancestors (Darwin, 1859). A study by Drake and Klingenberg, (2010), for example, demonstrated that variation in skull shape between dog breeds exceeds that found between species in the Carnivora. When compared to humans, the variation seen within breeds is approximately 100-fold lower however total genetic variation among breeds is similar (Vonholdt *et al.,* 2010).

As previously stated, the formation of dog breeds is considered the second step in the two stage process of domestication. Humans selected small groups of dogs with desirable or novel traits, out of an ancestral domesticated dog gene pool (Vaysse *et al.,* 2011), which together with episodic genetic introgression from local wolf population's caused specific lineages to be produced (Bateson and Sargan, 2012).

2.1.3 Defining a "Free-ranging dog"

Domestic dogs exhibit a wide spectrum of social organisation, spanning the entire range witnessed in canids, e.g. as household pets or as "free-ranging dogs" (Coppinger and

Schneider, 1995; Majumder *et al.*, 2013). Free-ranging dogs are present across the globe displaying diversity in population size as well as social organisation, they can be found as ranging solitary individuals or as part of large social groups (Sparkes *et al.*, 2014).

Various definitions have been used to characterize feral and free-ranging dogs in the literature. Boitani and Ciucci (1995) state that the majority of authors agree that "owned", "stray" and "feral" dogs are not immutable categories, and that dogs are capable of changing status throughout their life (Figure 2-2). A shift in status can result from various processes (Figure 2-2). Changing status may require a significant portion of an individual dog's life to complete and is dependent on local conditions and stimuli present (Boitani and Cuicci, 1995). More recently, Cafazzo *et al.* (2010) defined free-ranging dogs as "domestic dogs that are not under direct human supervision and whose activities and movements are not restricted by human activities" whilst Pilot *et al.*, (2015) state free-ranging dogs can be owned but are not permanently restrained, semi-feral or feral but their common characteristic is lack of artificial restriction concerning individual choice of mate i.e. are free-breeding.

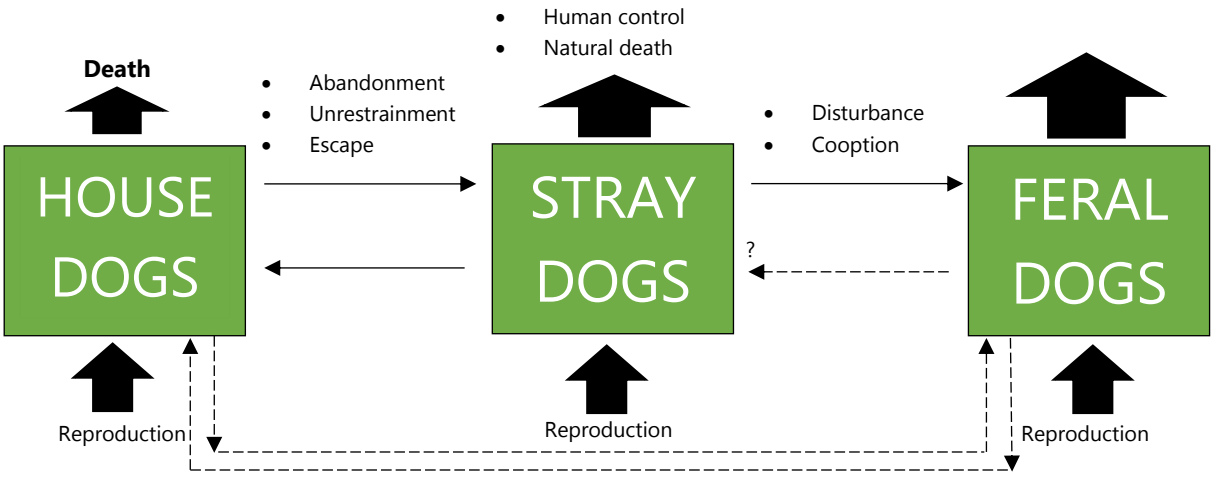


Figure 2-2: Defining a free-ranging dog, edited from Boitani and Cuicci (1995).

Although the genetic history of modern breeds is well researched, it is considerably less understood in free-ranging dogs (Shannon *et al.*, 2015). Current knowledge about free-ranging dog's social behaviour is limited, with the suggestion that they form stable social groups still controversial (Pal, 2015). Some studies state that in areas with abundant resources, whether they are indirectly or directly provided by humans, free-ranging dogs live in packs formed by multiple breeding individuals of both sexes (Cafazzo *et al.*, 2014).

This suggests they may be subject to sexual selection, resulting from a mating system that allows free mate choice.

2.2 MATING SYSTEMS

Mating systems typically form the basis of mammalian social systems and may be defined as 'the association of animals during and the factors contributing to the interaction and identification of partners and eventually fertilisation' (Hennessy *et al.,* 2012). Amongst mammals, monogamy is considered the rarest form (~3-5%), however, it is the most common form of breeding system in canids (Pal, 2011). Some canids are even referred to as "obligate monogamists", which is described as the dependency on cooperation of both parents for success of a litter (Kleiman, 1977). Reinforcement of social monogamy in canids is achieved by behaviours including displayed mating preferences, breeding by a single pair in the social group, continual proximity of the pair during oestrus and a lack of unrelated adult conspecific present in the home range of the breeding pair (Kleiman, 1977).

The ability of free-ranging dogs to form social groups on the basis of dominance relationships, such as those seen in wild relatives has been source of debate (Boitani & Ciucci 1995, Bradshaw *et al.,* 2009). Multiple breeding individuals are found to be present in groups of free-ranging dogs (Pal *et al.,* 1999, Pal 2011), which contrasts with the wolf pack structure characterized by a single breeding pair (Mech & Boitani 2003). Kinship is known to strongly influence the social organisation of group-living animals (Ross, 2001) but there is a lack of specific knowledge concerning the kinship patterns within packs of free-ranging dogs. Therefore, one of the aims of this project is to carry out the analysis of parentage and kin structure in a feral dog population.

Many different factors are known to influence and affect animal mating systems, including; female and male life history, temporal and spatial distribution of mates, resource defence, parental care, use of resources and sexual selection (Klug, 2011). Whilst pure-breed dogs are subject to strong artificial selection, free-ranging dogs and wolves are subject to natural and sexual selection, implying different evolutionary trajectories. Sexual selection is regarded as a powerful evolutionary process as a result of differing reproductive interests of each sex (McKean and Nunney, 2007).

2.3 THE THEORY OF SEXUAL SELECTION

Sexual selection affects genetic variation across a diverse set of traits, influencing both indirectly and directly an individual's ability to compete successfully for fertilisation of gametes (McKean and Nunney, 2007). It is hypothesised to represent important evolutionary pressure acting on variation in the immune system (Sheldon and Verhulst, 1996; Zuk and Stoehr, 2002; Schmid-Hempel, 2003; Lawniczak *et al.*, 2007).

In 1982, Hamilton and Zuk suggested females were capable of continuously basing mate choice on heritable resistance to parasites generated through a process of coadaptational cycles of parasites and their hosts. They hypothesised that this was clearly displayed through individual differences in plumage feathers or displays of the chosen sex. Since their work, three major hypotheses now exist concerning the theory of sexual selection. The first hypothesis states selection in organisms for greater viability is exceeded by the force of sexual selection (Hoekstra *et al.*, 2001). Promotion of a trade-off between secondary sexual characteristics, which are evolutionary exaggerated and energetically expensive and other components of fitness, also occurs (Höglund and Sheldon, 1998). The result of this trade-off is that the genotypes most capable of compromising the demands of reproduction and immunological function are usually the most successful, rather than the genotypes most resistant to disease (Antonovics and Thrall, 1994; Sheldon and Verhulst, 1996; Van Baalen, 1998; Jokela *et al.*, 2000; Zuk and Stoeher, 2002; Schmid-Hempel, 2003). Secondly, promotion of sexual dimorphism in immune function occurs as a result of sexual selection. This is a consequence of the variation produced due to the previously mentioned trade-off, typically involving genes which exhibit sex-specific effects (Zuk, 1990; Zuk and McKean, 1996; Rolff, 2002; McKean and Nunney, 2007). The last hypothesis put forward is the immunocompetence handicap hypothesis (ICHH), which states that individual success in mate competition is a reflection of the genetic variation underlying conditions which are specifically related to pathogen resistance and immune function (Folstad and Karter, 1992; Wedekind and Folstad, 1994). Across mammalian species genes targeted by positive selection are commonly enriched for functions related to immunity and defence, reproduction and chemosensory perception (see review; Kosiol *et al.*, 2008), demonstrating these systems as common targets of both sexual and natural selection.

2.4 DOMESTICATION AND ACCUMULATION OF MALADAPTATION

Whilst many questions surrounding dog domestication remain contentious it is apparent that regardless dogs have evolved many adaptations to reflect their direct interactions with humans (Reiter *et al.,* 2016), which are likely to involve significant genetic changes in response to behavioural and dietary divergence from the gray wolf (Freedman *et al.,* 2016). Accumulation of evidence shows that domestication in both animals and plants results in considerable changes in the genome in comparison to wild ancestors, and selective sweeps at multiple loci as a result of artificial selection are witnessed in many domesticated species (Zeder *et al.,* 2006). In 1930, Fisher suggested increasing segregation of maladaptive mutations in different dog breeds is the result of artificial selection. Decreases in locus-specific effective population size due to selection at linked sites also result in a higher probability of the fixation of deleterious mutations (Hill and Robertson 1966). Combined with this, through the domestication process population bottlenecks may have affected purifying selection by reducing efficacy (Kimura, 1962; Cruz *et al.,* 2008).

As previously stated, free-ranging dogs and wolves are influenced by natural and sexual selection, whilst pure-breed dogs are subject to artificial selection. This is exemplified when comparing pure-breed dogs to free-ranging dogs. Pure-bred dogs exhibit large phenotypic diversity in body size, as a result of selective breeding from humans whilst free-ranging dogs have been shown to be uniformly medium-sized (examples seen in Totton *et al.,* 2010; Ortolani *et al.,* 2009), suggesting body size is influenced by natural selection (Lord *et al.,* 2013). Resulting differences in the variants of some functional genes being favoured in each case may be witnessed between the three groups. Artificial selection results in a relaxation of selection pressures on traits important for independent survival, as well as traits related to mate choice and reproduction. In free-breeding dogs and wolves, such traits are subject to purifying natural selection, as animals carrying detrimental conditions would not survive outside of the domestic environment, or would have limited reproductive success (Pilot *et al.,* 2016). Differences in the strength of natural versus artificial selection in these two groups have important implications for health of individuals representing these groups.

When considering pure-breed dogs it should be noted that both domesticated animals and humans now inhabit environments which are completely different from the

ecological context which their immune systems evolved under. They usually lack, for example, the environmental stresses which are characteristic of a natural population (Turner *et al.,* 2011).

Mating system and social organisation are also known to play a role in influencing parasite prevalence and diversity, thus affecting selection pressure on the immune system. Social organisation is defined by the composition, size of social groups and the intergroup dispersal patterns (Altizer *et al.,* 2003). In general it is expected that monogamous species, with well-defined and defended territories will exhibit fewer parasites as a result of fewer intraspecific contacts when compared to social mammals (with multi-female multi-male groups) will be affected by wider spread of pathogens (Altizer *et al.,* 2003). When you consider this, a differentiation between free-ranging dogs and wolves could be expected due to differences in mating systems utilised and this could represent important implications for immunity.

2.4.1 Influence on immunity

When considering the role domestication has played in shaping the genome of wolves, purebred and free-ranging dogs it is crucial to consider the resulting effects on the immune system. Frequent exposure to a diverse range of pathogens occurs in individuals in natural populations. There is a vast number of genes involved in responding to various types of pathogens and this results in a complicated pathogen-specific selection pressure that acts on the immune system in wild animals (Turner *et al.,* 2012). It is apparent that due to the differences in ecology and habitat that free-ranging dogs, purebred dogs and wolves will encounter different pathogens and thus, be experiencing different pathogen-specific selection pressures.

This is exemplified in a recent study using SNP microarray analysis, which revealed a SNP mutation present in the *MARCH7* gene, where heterozygosity is demonstrated by all free-breeding dogs (GT) and individual wolves (GT, TT) but homozygosity represented in all east Asian dog breeds (TT) and European dog breeds (TT). Interestingly, the Eurasian golden Jackal and black-backed jack also demonstrate homozygosity but GG instead of TT in both cases (Pilot *et al.,* 2016). *MARCH7* is responsible for playing a role in activated T lymphocyte regulation (Metcalfe *et al.,* 2005). Differentiation between pure-breed dogs and free-ranging dogs could be the result of a reduction in selection pressure acting in the immune system of pure-breed dogs, primarily due to living in human households with

access to veterinary care. For these reasons it was considered that *MARCH7* would be an interesting gene to analyse for signatures of diversifying selection between purebred dogs, free-ranging dogs and wolves.

2.5 OBJECTIVES AND THESIS OUTLINE

There are two main objectives to this thesis;

1. To reconstruct kinship relationships in a population of free-ranging dogs, and assess the genetic variability and inbreeding level;
2. To look for possible signals of diversifying selection on immune tolerance between pure-breed dogs, free-ranging dogs and wolves, through analyses of DNA sequence data from the *MARCH7* gene and evolutionary comparison, using phylogenetic analysis of *MARCH7* gene sequences across mammalian species and structural protein prediction

These objectives and the results found are described in the following chapters.

2.5.1 Chapter two; Genetic reconstruction of parentage and kinship in a population of semi-feral domestic dogs

Chapter two will provide new insights into the mating system in free-ranging dogs. This will be achieved studying a sample population of free-ranging dogs from the outskirts of Rome. Microsatellite markers will be utilised to target 14 loci in order to understand parentage and kinship in a population of free-ranging dogs from Rome.

2.5.2 Chapter three; Functional genetic differentiation between pure-breed and free-breeding dogs at *MARCH7* gene

In chapter three, a review of the effect of domestication on the dog genome is provided, with particular focus on immune genes, MARCH family members and the *MARCH7* gene. DNA sequence of *MARCH7* gene is analysed in pure-breed dogs, free-ranging dogs and grey wolves, followed by analysis of patterns of nonsynonymous and synonymous variation to both the carnivore family and a selection of placental mammals and finally analysis of protein structure using *ab initio* modelling.

2.5.3 Chapter four; General discussion

Chapter four, the general discussion, features aspects of both reproductive behaviour and immunity and discusses the main findings of this thesis, and opportunities for further research.

3. GENETIC RECONSTRUCTION OF PARENTAGE AND KINSHIP IN A POPULATION OF SEMI-FERAL DOMESTIC DOGS

3.1 SEASONALITY OF CANINE REPRODUCTIVE BEHAVIOUR

Canids display great diversity in terms of social organisation (Majumder *et al.,* 2014). In response to distribution and quantity of local food resources and strategy to acquire them there is evident inter- and intraspecific variation in social organization witnessed among canids (Cafazzo *et al.,* 2010) and this can also be seen when considering reproductive behaviour.

Many mammalian species display reproductive synchrony to varying degrees to allow conditions to be optimal when breeding occurs (Majumder and Bhadra, 2015). Species that only mate during a specific time of the year are known as seasonal breeders and typically give birth when infant survival is optimal and resources are abundant (Prendergast, 2005). In canids, the grey wolves are seasonal breeders, where from December through to early April they become sexually active, but this activity is known to be dependent on latitude (Hasse, 2000). In contrast, no clear seasonality is observed in domestic dogs, which breed continuously and reproduce aseasonally (Engle, 1946, Lord *et al.,* 2013), although the possibility of seasonality has been suggested through indirect evidence, mostly in regards to free-ranging dogs (Beck, 1973).

Multiple authors suggest that the loss of seasonality witnessed in domestic dogs could be the result of (1) humans directly providing food and shelter in a domestic environment and thus reducing selective pressure or (2) increased fecundity resulting from direct artificial selection (Hasse, 2000; Malm, 1995). Lord *et al.* (2013) argues that genus-typical reproductive behaviours are reduced in domestic dogs as a result of adapting to a new niche. When compared to wild *Canis* species, which experience seasonal fluctuations in food availability, domestic dogs, and more specifically free-ranging dogs, are provided with consistent stationary human refuse, which is generally discarded uniformly and in permanent locations, and is not dependent on seasonal variation such as water or light (Lord *et al.,* 2013).

Despite these notions, evidence for seasonality has been in fact reported in the literature. Throughout India, free-ranging dogs inhabit an array of human habitats as scavengers and provide a ubiquitous presence (Vanak and Gompper, 2009; Vanak *et al.,* 2009). Observations in West Bengal revealed that free-ranging dogs have a distinct mating season, which occurs in conjunction with the wet season or the monsoon (Pal, 2001). As a

result, offspring are typically born during winter, which seems to contradict with the hypothesis of resource abundance. However, as free-ranging dogs are predominately scavengers, resources generally remain constant throughout the year, because the main source is from offerings from humans and human produced waste (Vanak *et al.,* 2009; Bhadra *et al.,* 2015). It has been suggested that the resulting rain from the monsoon triggers reproductive abilities in these free-ranging dogs, even though currently there seems to be an apparent lack of adaptive advantage (Majumder and Bhadra, 2015).

3.2 GRAY WOLF REPRODUCTIVE BEHAVIOUR

Wolves survive in the wild by living in packs, which typically consist of a breeding pair and their offspring, and sometimes also siblings of the breeders and unrelated individuals (Mech 1999; Mech and Nelson 1990, Lehman *et al.,* 1992, Jedrzejewski *et al.,* 2005). In some geographic regions, no cases of unrelated wolves in packs were observed (vonHoldt *et al.,* 2008), whilst in other regions "adoptees" occur in up to 80% of packs studied (Grewel *et al.,* 2004; Rutledge *et al.,* 2010). It has been witnessed that multiple litters born in the same year to separate mothers have occurred in the wild (Meier *et al.,* 1995; Rutledge *et al.,* 2010; Van Ballenberghe 1983; vonHoldt *et al.,* 2008), however the frequency of this occurrence and the mechanisms by which it happens remain unknown (Stenglein *et al.,* 2011). Lastly, inbreeding is generally avoided in wild wolf populations, regardless of the fact that there are numerous opportunities for incestuous mating (Smith *et al.,* 1997; vonHoldt *et al.,* 2008). Inbreeding avoidance is regarded as an important constraint on wolf behavioural ecology (Smith *et al.,* 1997).

Multiple studies have demonstrated mutual mate preferences in captive wolf pack between dominant males and females (Rabb *et al.,* 1967; Zimen, 1976; Jenks, 2011) suggesting similarities between wolves and dogs. In contrast, one study based on a more extensive behaviour data set (Derix and Van Hooff, 1995) found that although highest-ranking members of both sexes were involved in mating, typically it was male-preference of females and control of dominant males reducing the sexual activity of subordinates that influenced reproductive behaviour.

3.3 DOMESTIC DOG REPRODUCTIVE BEHAVIOUR

Modern domestic dog breeds are upheld by using a set of criteria which must be adhered to by breeders. These work by applying a persistent selective pressures on breed defining fixed phenotypes including skull shape, coat colour, leg length, and body size (Rimbault *et al.,* 2013). When considering traits affected by artificial selection, reproduction has been strongly manipulated by humans in order to increase reproductive potential and to produce shorter generation times in dogs (Boitani and Ciucci, 1995). These pressures have resulted in a reduction of phenotypic and genetic heterogeneity within breeds and can be due to additional factors, including repeated use of popular sires, line breeding, and promotion of the breed barrier rule (Farrell *et al.,* 2015). It is also worth noting that surgical sterilisation by orchiectomy or ovariohysterectomy (commonly known as castration, spaying or neutering) is a common procedure undertaken by many pet owners across the world due to perceived behavioural management benefits (Kustritz, 2007; Hoffman *et al.,* 2013). This results in a large section of the domestic dog population being unable to reproduce, restricting the number of breeding individuals. Consequently, the dog genome is characterised by extensive linkage disequilibrium (LD) and low haplotype diversity (Ke *et al.,* 2010; Vaysse *et al.,* 2011).

3.4 FREE-RANGING DOG REPRODUCTIVE BEHAVIOUR

In terms of socio-behavioural ecology wolves and free-ranging dogs demonstrate similar social organisations. They are both able to form and live in packs, exhibiting differential social relationships between members (Marshall-Pescini *et al.,* 2015). In contrast to the monogamous mating system witnessed in wolves, the majority of free-ranging dogs exhibit a polygamous mating system, where both sexes mate with multiple partners (Cafazzo *et al.,* 2014).

Free-ranging dog groups are typically comprised of related individuals, but the proportion of unrelated animals present in these packs is typically higher compared to wolves (Macdonald and Carr, 1995; Bonanni and Cafazzo, 2014). Generally, there is also a greater number of sexually mature individuals of both sexes (Daniels and Beckoff, 1989a; Daniels and Beckoff, 1989b, Macdonald and Carr, 1995; Cafazzo *et al.,* 2010; Bonanni et al, 2010a; Bonanni *et al.,* 2010b; Bonanni *et al.,* 2011; Pal *et al.,* 1999). In contrast to other cooperatively breeding canids, in free-ranging dog populations typically each member of

the group has an equal chance of breeding due to a mating system that is polygamous (Pal, 2011, Paul *et al.*, 2014). Polygamy is a mating system where both sexes are recorded to have variable number of mates and if mating success of male and female is approximately equal (Steyaert *et al.*, 2012).

Cafazzo *et al.* (2014) found an age-graded dominance hierarchy in a pack of free-ranging dogs. This had an effect on multiple aspects of reproductive behaviour, including male copulation rate, reproductive outcome and mate preference. It was observed that both sexes preferentially chose high-ranking partners and that overall their suggested social organisation resembled that of wolves more than previously thought.

In general, unlike wolves, female free-ranging dogs mostly raise their pups alone (Boitaini and Ciucci, 1995; Daniels and Bekoff, 1989) or in rare cases with the help of the male who typically defends the pups but rarely participates in feeding (Pal, 2005; Marshall-Pescini *et al.*, 2015). In a location experiencing harsh weather and limited food availability, only one female produced offspring during a two year period and rearing of the pups was shared amongst several group members (Gipson, 1975). Although information about the social relationship among members is incomplete, it does provide evidence of pack formation in free-ranging dogs and highlights that the species is capable of adapting to harsh conditions.

Cafazzo *et al.* (2014) believe that because dominance hierarchies can be witnessed (Bonanni and Cafazzo, 2014) social regulation of reproduction is likely to operate even in smaller groups of dogs. In fact, in wolf packs a positive relationship between numerous variables can be seen, including dominance, age, reproductive activity and leadership (Mech, 1999; Peterson *et al.*, 2002). Thus this similarity makes it possible to hypothesise that there might be a common mechanism underlying the social organisation of both species. They also believe that the key differentiation between dogs and wolves is linked to the degree of reproductive suppression on subordinates by dominant animals as well as the degree of cooperative breeding, which is typically higher in wolves. Pal *et al.* (1999) described differences in individuals in regard to sexual behaviour both in males and females and they made the presumption that both –intra and intersexual interactions of free-ranging dogs are at least partially dependant on the situation and differing individual personalities.

3.5 STUDY POPULATION

This project focuses on a population inhabiting a nature reserve at the outskirts of Rome in Italy. This population comprised of about 100 adult individuals and their offspring, which were not socialised to humans, even though they relied on the food provided by humans. There are extensive behavioural data on this population, including mating and reproductive patterns (Bonanni *et al.*, 2010a,b, 2011; Cafazzo *et al.*, 2010, 2012, 2014), resulting from a long-term research led by the external collaborator in this project, Dr Eugenia Natoli. This population therefore provides a unique opportunity to study the kinship and parentage patterns in a free-ranging dog population.

3.6 METHODOLOGY

3.6.1 Material

The samples used in this study were collected by Dr Natoli and her collaborators at Azienda USL Roma D, Area Dipartimentale Sanita Pubblica Veterinaria.

Two types of samples were used:

1. Tissue samples obtained from sterilisation of free-ranging dogs from a population living at the outskirts of Rome. This was completed at a veterinary hospital as a result of legislation enforcing sterilisation of non-owned dogs in Rome district. Foetus samples (at an early stage of development) were obtained from the same source.
2. Hair samples from pups from the same free-ranging population were collected through a capture and immediate release; this was also carried out by veterinarians or veterinary technicians.

Pure-breed dog hair samples used for testing DNA extraction methodologies were collected by hand from pet dogs to eliminate the over use of the target samples.

3.6.2 Sample selection

Due to financial constraints of the project it was impossible to genotype all sampled individuals. Three mothers and their known offspring (foetuses) were genotyped alongside selected males. Male selection was completed using previously collected information about social groups and rankings to include males (Table 2) from the same social groups as the mothers (Table 1) or known to frequently visit.

Table 1: Sample ID, number of offspring and name of mother and offspring from free-ranging dogs

Sample ID	No. of offspring	Name
CL_240279	7	Sofia (Petra)
CL_387	9	Snella (Volpe)
CL_922	2	Emma (catt. Vivana)

Table 2: Sample ID and name for all males from free-ranging dogs

Sample ID	Name
CL_8022	Bo
CL_4309	Bernardo
CL_337861	Antonio (Artu)
CL_238379	Fred
CL_931645	Duca
CL_338024	Spider (Duca)
CL_931248	Petto
CL_238158	Angelo

3.6.3 DNA Extraction: Hair

Four different methods for DNA extraction were tested. Hair growth in dogs, as in humans, is not continuous. Instead it occurs in three main stages consisting of three periods of growth; active growth (anagen phase), regression (catagen phase) and a resting period (telogen phase) (Bekaert *et al.*, 2012). During the resting stage the hairs are retained in the hair follicle as dead hair, typically shed hair is made up of telogen hairs which contain insufficient epithelial root cells required for nuclear DNA profiling (Bekaert *et al.*, 2012). In order to avoid this, samples were pulled, as gently as possible, from the dogs to try and ensure sufficient epithelial root cells would be present.

3.6.3.1 DNA Purification Kit

1. Approximately 5-10 strands of hair was cut up into small (approximately 2-3mm) pieces using a sterilised scalpel and Petri dish before being transferred into a 1.5ml Eppendorf tube.
2. 180µl of digestion solution was added, followed by 20µl of Proteinase K solution and the complete solution vortexed until a uniform suspension was achieved.
3. Samples were incubated for a minimum of 8 hours (up to 24), with occasional mixing, until all tissue present was completely lysed and no particles remained.
4. 200µl of Lysis solution was then added and vortexed for approximately 15 seconds until a homogenous mixture was obtained.
5. 400µl of 50% ethanol was added and mixed by vortexing before all the prepared lysate was transferred into a GeneJET Genomic Purification Column inserted in a collection tube.

6. Following centrifugation at 6000 x g for 1 minute the collection tube containing discarded lysate was removed and discarded and the GeneJET Genomic Purification Column inserted into a clean 2ml collection tube.
7. 500µl of Wash buffer I (with ethanol added) was added, followed by centrifugation at 8000 x g for 1 minute and flow-through discarded.
8. The purification column was placed back inside the collection tube and 500µl of Wash Buffer II (with ethanol added) was added to the column.
9. Centrifugation for 3 minutes at maximum speed (>12,000 x g) was followed by an additional 1 minute to ensure thorough removal of ethanol and collection tube including flow-through discarded.
10. The GeneJET Genomic Purification Column was transferred into a sterile 1.5ml Eppendorf tube where 100µl of Elution buffer was added to the column and left to incubate at room temperature for 5 minutes.
11. Centrifugation for 1 minute at 8000 x g was followed by an additional 100µl of Elution buffer, further 5 minutes incubating at room temperature and finally centrifugation for 1 minute at 8000 x g.
12. The purification column was discarded and purified DNA could be immediately used in downstream applications or stored in the freezer.

3.6.3.2 Enzymatic Laundry Power

1. Hair shafts were cut into fragments of about 2 mm.
2. Each sample was digested in 100 ml of extraction reagent (pH 10.3) for 1.5 hours at 50°C.
   - The extraction reagent contained 3 mg enzymatic laundry powder, and 1x PCR buffer (20 mM Tris-HCl (pH 8.4), 20 mM KCl, 10 mM $(NH_4)2SO_4$, 1.5 mM $MgCl_2$.
3. After extraction, extraction solutions were gradually heated up to 95°C to improve extract efficiency, and then subject to 95°C for 10 minutes in order to inactivate enzymes in the extraction reagent.
4. The final DNA extracts were stored at -18°C until use.

3.6.3.3 Sodium Hydroxide

1. 10 hair roots were cut to ~5mm and placed into a 1.5ml microcentrifuge tube.
2. 50 µl of 200mM NaOH solution was added to each tube.

3. All tubes were boiled in a water bath at 94°C for 10 minutes.
4. After boiling, tubes were cooled at room temperature and 50 µl of the following solution was added, containing:
   - 200mM HCl.
   - 100mM Tris-HCl pH 8.5.

3.6.3.4 Chelex

3.6.3.4.1 Sample preparation

Hairs were washed to remove any contaminants or extraneous bodily fluid by fully submerging the hair into 200 µl of sterile deionized water for ~10 minutes.

3.6.3.4.2 Extraction

1. Using a sterile scalpel, approximately 1cm of hair from the root end was removed and placed into a sterile 1.5ml microcentrifuge tube.
2. 200µl of 5% Chelex ® 100 was added. If this was insufficient to cover the hair completely then additional µl were added until fully submerged.
3. 2 µl of 10mg/ml of Proteinase K was added for every 200 µl of Chelex ® 100.
4. Samples were vortexed at high speed for 10-30 seconds, ensuring that the samples were fully submerged in the Chelex ® 100 suspension.
5. Samples were incubated at 56°C for 6 hours initially, upon repeat this was extended to 12 hours.
6. Samples were then vortexed at high speed for 5-10 seconds.
7. After vortexing, samples were heated for 8 minutes at 100°C in a heating block, ensuring that the hair was fully submerged in solution.
8. Samples were then vortexed again at high speed for 5-10 seconds
9. This was followed by centrifuging for 3 minutes at approximately 10,000-15,000 x g.
10. The supernatant was transferred to a sterile 1.5 ml microcentrifuge tube for concentration and purification.

3.6.4 DNA Extraction: Tissue

The same method as described in 3.6.3.1 was used for the extraction of DNA from tissue samples, except approximately 5g of tissue sample was utilised in step 1.

3.6.5 DNA precipitation

In order to improve concentrate and purity of DNA from hair extraction, precipitation using 100% and 70% ethanol was conducted, one of the most common methods use for purification and concentration (Fregel *et al.,* 2009). The following methodology was used;

1. DNA sample and 3M Sodium Acetate buffer (pH 5.2) were added into a microcentrifuge tube at a ratio of 1:10 to equalise ion concentrations
2. 2-3 volumes of cold 100% ethanol were added and the samples placed into a -20°C freezer for at least one hour.
3. Samples were then centrifuged for 15 minutes at 12,500 rpm.
4. Using a 1ml pipette as much supernatant as possible was removed, with care exhibited not to disturb the pellet.
5. If not all the supernatant was removed, samples were re-centrifuged briefly and the rest of the supernatant removed using a 200 µL pipette.
6. 250 µL of cold 70% ethanol was then added to each tube.
7. Samples were centrifuged for 5 minutes at 12,500 rpm.
8. Any visible supernatant was removed using a 200 µL pipette and remaining ethanol was evaporated using a 37 °C water bath.
9. The pellet was then suspended in water and stored in the -20°C freezer.

3.6.6 DNA Concentration

DNA concentration and purity was determined by a NanoDrop 1000 spectrophotometer (Thermo Scientific).

3.6.7 Microsatellite primers

Microsatellites, also known as short tandem repeats (STRs), simple sequence repeats (SSRs), or variable number tandem repeats (VNTR) are defined as tandemly repeating units of DNA that can be 1, or 2-6 base pairs in length. Throughout the nuclear genomes of eukaryotes, they are known to be frequently distributed (Bhargava and Fuentes, 2010; Putman and Carbone, 2014). Due to their highly polymorphic nature, microsatellites are used for a multitude of uses including; population genetics, conservation, parentage identification, fingerprinting and genetic mapping (Buschiazzo and Gemmel, 2006; Chistiakov *et al.,* 2006; Guichoux *et al.,* 2011).

Among the most common choices for molecular genetic studies are di-, tri-, and tetranucleotide repeats with dinucleotide repeats accounting for a considerable number of microsatellites across a wide range of species (Li *et al.*, 2002). Trinucleotide and hexanucleotide repeats are the most usual repeat classes to be found in coding regions due to the fact that they do not cause a frameshift (Toth *et al.*, 2000).

For this study, 6 dinucleotide repeats, 7 tetranucleotide repeats and 1 hexanucleotide repeat were chosen (Table 3). Due to mononucleotide repeats being considered less reliable as a result of complications with amplification (Li *et al.*, 2002; Selkoe and Toonen, 2006) they were avoided in this study.

Table 3: PCR primers used for the amplification of microsatellite loci

Multiplex set	Locus	Chromosome	Dye	Length (bp)
ttrAB	FH2088	CFA15	D2	104-136
	FH2010	CFA24	D4	203-235
	FH2017	CFA15	D3	260-272
	FH2054	CFA12	D3	146-178
	C253	CFA20	D3	93-115
ttrC	FH2096	CFA11	D2	88-104
	C213	CFA25	D2	136-172
	FH2079	CFA24	D4	261-285
	VWF	CFA27	D3	129-189
	C250	CFA09	D4	122-144
diC	FH2001	CFA23	D4	129-149
	C466	CFA02	D2	139-163
	C436	CFA27	D4	225-247
	C642	Unknown	D3	178-194
	AHT130	CFA18	D3	108-124

3.6.8 PCR Protocol

1. PCR reaction was prepared as shown in Table 4.

Table 4: PCR Reagents and volumes for microsatellite analysis

Reagent	Volume
QIAGEN Multiplex Mix	4 µL
DNA	1 µL
Primers (Forward and Reverse)	1 µL
BSA	0.1 µL
Water	1.9 µL
Total	8 µL

2. Once prepared samples were loaded onto the thermal cycler
3. An initial activation step of 15 minutes at 95°C was completed.
4. This was followed by 38 cycles of a 3-step process, firstly denaturation for 30 seconds at 94°C, secondly annealing for 90 seconds at 57-63°C and finally extension for 90 seconds at 72°C.
5. This was followed by a final extension period of 10 minutes at 72°C.

3.6.9 Agarose Gel Electrophoresis

Molecular grade agarose (Bioline) gels were used in all cases, using the following methodology;

1. Gels were made to a concentration of 1% e.g. 1 g agarose in 100 mL 1X TAE buffer.
2. Gels were left for 20 minutes to allow setting.
3. Once the gel had set it was completely submerged in 1X TAE buffer in the electrophoresis tank.
4. 2 µL of each PCR product was mixed with 2µL loading buffer (GelRed Dye)
5. All 4 µL of product was then loaded into each lane in the agarose gel.
6. As a size control 0.5µL DNA Ladder was also loaded in one of the empty lanes.
7. Electrophoresis was performed at 100V for approximately 30 minutes.
8. Gels were analysed using a FluorChem™ 5500 Imager (Alpha Innotech).
9. Samples for which a single band of the expected size was visible were considered suitable for purification and sequencing.

3.6.10 GeneMarker

For analysis of electrophoresis traces produced, GeneMarker was used for identification of alleles from chromatograms. GeneMarker is software used for DNA fragment analysis with capillary and gel electrophoresis traces, such as microsatellites. It can also be used for quantification of DNA fragments (SoftGenetics, 2016).

3.6.11 MicroChecker

Microchecker helps identifying genotyping errors due to short allele dominance (aka large allele dropout), scoring errors as the result of stuttering, null alleles and typographic errors. When considering multi-locus genotypes it can also be used for discrimination between inbreeding and Wahlund effect, and deviations from Hardy Weinberg caused by null alleles (Van Oosterhout *et al.,* 2004). For the purpose of this study Microchecker was mainly used to highlight typographic errors and assess the effect of null alleles.

3.6.12 Cervus

Parentage analysis was carried out using Cervus. Cervus is a computer programme designed for the assignment of parents to their offspring through the use of genetic markers. It assumes that species are diploid and markers are autosomal and are inherited independently of one another i.e. they are in linkage equilibrium (Kalinowski *et al.,* 2007).

The simulation of parentage analysis serves two functions:

1) To provide an estimation of the resolving power of a collection of codominant loci considering their allele frequencies.

2) To provide an estimation of the critical values of the log-likelihood statistics Delta or LOD, so that confidence of the parentage assignments made using parentage analysis can be statistically evaluated.

3.6.13 Manual checking of Cervus results

Results that did not fit the expected pattern were manually checked using allele size data. For example, instances where multiple fathers occurred for offspring where checked by comparing allele size of offspring, candidate fathers and mothers to ascertain whether inheritance was correct.

3.6.14 Kinalyzer

To compliment the results from Cervus, Kinalyzer was used to reconstruct sibling groups through the comparison of available individual microsatellite genotypes. This program uses two methods for reconstruction of sibling groups.

For the first option, Kinalyzer uses a combinatorial optimisation approach based on Mendelian inheritance laws to construct the fewest number of sibling groups that contains all the individuals from the population in question, referred to as "2- allele set cover". This "2-allele" property states that assignment within a locus of individual alleles to paternal and maternal parents is such that the number of distinct alleles which are assigned to each parent will never exceed two. Barring genotyping error or mutations all sibling groups must follow this constraint (Ashley *et al.*, 2009).

The second option available is known as a consensus-based approach, sometimes referred to as "greedy consensus" method, which completes reconstruction of sibling groups through the use of subsets of loci and finding consensus of these different solutions. Kinalyzer discards individual loci one by one, reconstructing solutions using the loci remaining with the final solution output consisting of a consensus of the partial solutions (Ashley *et al.*, 2009). Calculation of this consensus is achieved by computing groups in common and then "greedily" (i.e. the quickest solution that means minimum criteria) merging the closest pair of groups iteratively. Kinalyzer computes the distance using costs associated with errors and allelic information shared (Sheikh *et al.*, 2008).

3.7 RESULTS

12 microsatellite loci were amplified for 29 samples. One sample completely failed to amplify and two loci (C436 and C642) provided unreliable results, so they were removed from the data set. Detailed data output can be seen in Appendix 7.4 - 7.14

3.7.1 Gene Marker

Raw microsatellite genotyping results were generally of good quality (Figure 3-1). In two loci, issues with stuttering caused complications with correct interpretation of peaks when three peaks were visible closely positioned together. This was resolved by overlaying of all three dyes to reveal interference between them resulting in one of the peaks being excluded.

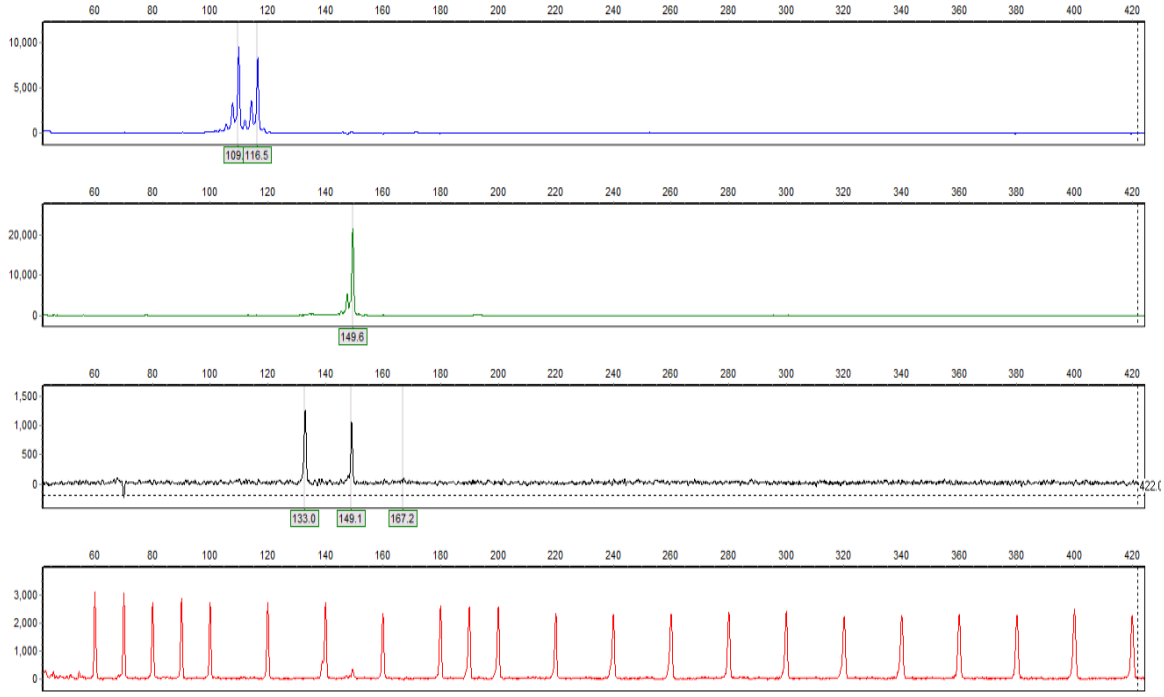


Figure 3-1: Examples of GeneMarker results demonstrating clear, good quality results.

3.7.2 Allele sizes for primer set ttrAB

All five microsatellite primers worked for all individuals apart from Snella and Petto where three loci positions failed to register an interpretable peak (Appendix 7.1, Table 34). Three out of the five loci had four alleles and the remaining two loci had three alleles. The maximum range in allele size varied from 8 to -16 base pair differences across primers, with primer sets 2017 and 253 showing 8 base pair difference, primer sets 2010 and 2088 showing 12 and 2054 showing up to 16 base pair differences, although these maximum ranges were only witnessed for a few individuals in each primer set.

3.7.3 Allele sizes for multiplex set ttRC

This set included four microsatellite loci. DNA sample SN3 failed to amplify at all for any primers and there were multiple instances of failed reaction (Appendix 7.2, Table 35). The maximum range in allele size varied from 8 to 12 base pair differences across primers, with primer 250 showing 6 base pair maximum, primer 2096 showing 8 base pair maximum and primers VwF and 213 both showing 12 base pair maximum difference. In all primers these maximum ranges were seen in only a handful of individuals.

3.7.4 Allele sizes for multiplex set diC

For primer set diC, three primers worked effectively. For individuals SN1 and Emma one locus positions failed to register an interpretable peak. The maximum range in allele size varied from 8 to 12 base pair differences across primers, with primer sets 2001 and 466 showing 8 base pair difference and AHT130 showing up to 8 base pair differences (Appendix 7.3, Table 36), although these maximum ranges were only witnessed for a few individuals in each primer set.

3.7.5 Genetic diversity of the study population

Across all 12 loci there was average of four alleles per locus, which is relatively low. Observed heterozygosity ranged from 0.308 to 0.724 for different loci, averaging 0.570 (Table 5). Expected heterozygosity ranged from 0.440 to 0.705 and was on average 0.5780.

Table 5: Allele frequency analysis; summary of statistics

Locus	k	N	HObs	HExp	PIC	NE-1P	NE-2P	NE-PP	NE-I	NE-SI	HW	F(Null)
1	3	27	0.519	0.469	0.415	0.894	0.757	0.614	0.336	0.604	ND	-0.0463
2	4	27	0.519	0.47	0.418	0.891	0.752	0.604	0.333	0.603	ND	-0.046
3	4	29	0.586	0.619	0.562	0.797	0.631	0.452	0.2	0.496	NS	0.043
4	4	28	0.679	0.662	0.602	0.765	0.595	0.414	0.17	0.468	NS	-0.0234
5	3	28	0.429	0.441	0.375	0.906	0.792	0.669	0.38	0.628	ND	-0.0085
6	3	28	0.679	0.625	0.544	0.811	0.666	0.515	0.219	0.498	NS	-0.0607
7	5	27	0.519	0.518	0.466	0.861	0.71	0.543	0.284	0.567	ND	-0.0407
8	4	18	0.556	0.676	0.599	0.764	0.603	0.431	0.176	0.465	ND	0.0881
9	6	26	0.308	0.481	0.449	0.876	0.712	0.532	0.301	0.589	ND	0.207
10	4	27	0.667	0.584	0.507	0.829	0.688	0.531	0.249	0.526	ND	-0.0973
11	5	29	0.724	0.705	0.641	0.728	0.56	0.383	0.147	0.44	NS	-0.0305
12	4	29	0.655	0.685	0.625	0.745	0.574	0.394	0.155	0.452	NS	0.0358

Where k = number of alleles, n = number of individuals, HObs = observed heterozygosity, HExp = expected heterozygosity, PIC = mean polymorphic content, NE-1P = Average non-exclusion probability (first parent), NE-2P= Average non-exclusion probability (second parent), NE-PP = Average non-exclusion probability (parent pair), NE-I = Average non-exclusion probability (identity), NE-SI = Average non-exclusion probability (sibling identity), HW = Hardy Weinberg and F (Null) = Maximum likelihood estimation of the frequency of null alleles at microsatellite loci.

3.7.1 Genetic variability and heterozygosity in the study population

An examination of heterozygosity, homozygosity and allele frequency revealed a similar pattern for all loci across the population. Generally, one dominant allele was present in half to three quarters of the genotypes present for each locus (Figures 0-2-012). Levels of heterozygosity exceeded those for homozygosity for alleles at most loci, for example see Figure 0-5. There were two exceptions where homozygosity exceeded heterozygosity and in both cases this occurred in the dominant allele (Figure 0-6, Figure 0-10). Locus nine (213) appeared to have the most diversity, having six alleles in the study population (Figure 0-10), however mean polymorphic information content was highest for locus eleven.

The mean expected heterozygosity (He) in the sample population was moderate at 0.578. For the majority of loci only a small difference between observed heterozygosity and expected heterozygosity can be seen (Table 5, Figure 7-13), suggesting adherence to Hardy-Weinberg equilibrium. This was confirmed by carrying out the Hardy-Weinberg equilibrium tests (seen in Appendix 7.57.5), which were non-significant for all loci tested. Mean polymorphic information content (PIC) shows that there is a moderate level of diversity seen in the population and variation can be seen when comparing loci (Figure 7-13). Locus five displayed the lowest PIC value, with locus eleven showing the greatest.

3.7.1 Simulation of parentage

Across all the simulations ran using Cervus low assignment rates were seen for all scenarios; however assignment rates were consistently increased when a genotype of known parent was provided. Small differences can be observed when comparing the simulation of paternity with known mothers to those of no known mother (i.e. Cervus simulated mothers) (Tables 49, 50, 75 and 76). In all cases the assignment rate increased by more than 50% for the strict category and by 6% for the relaxed category when genotypes of known mothers were provided. Information regarding simulation confidence levels, simulation parameters, delta distributions, and breakdown of parentage assignment can be seen in Appendix 7.7 for maternity, 7.9 for maternity, 7.11 for pairs, i.e. maternity and paternity and 7.13 for paternity with known mother given.

3.7.2 Analysis of maternity

Results from parentage analysis in Cervus (Table 6) revealed that the program failed to assign the correct mother to 5 of the offspring and failed to identify any mother to be

present for one offspring (SN5). Correct identification for mother and all offspring occurred in one of three litters studied. For three mother-offspring pairs confidence was below the 80% threshold. In the case of Sofia's offspring No.4 two possible mothers were indicated, one being the actual known mother (Sofia), however pair confidence was higher for incorrectly identified mother (Emma). Identification of mothers was also achieved for two of the adult males sampled, both with 80% confidence. Detailed results in Appendix 7.8.

Table 6: Prediction of candidate mothers (known mothers were not provided to Cervus for this analysis)

Offspring ID	Known mother	Candidate mother ID	Pair loci compared	Pair loci mismatching	Pair confidence
Spider					
Fred		Snella	7	0	+
Angelo					
Sofia					
SO1	Sofia	Sofia	12	0	+
SO2	Sofia	Emma	10	0	+
SO3	Sofia	Sofia	12	0	+
SO4	Sofia	Emma	10	0	+
SO4	Sofia	Sofia	11	0	
SO5	Sofia	Sofia	12	0	+
SO6	Sofia	Sofia	11	0	+
SO7	Sofia	Sofia	12	0	+
Antonio		Snella	7	0	+
Antonio		Sofia	12	0	
Snella					
SN1	Snella	Snella	6	0	+
SN2	Snella	Snella	7	0	+
SN3	Snella	Snella	5	0	+
SN4	Snella	Snella	7	0	+
SN5	Snella				
SN6	Snella	Snella	6	0	+
SN7	Snella	Emma	10	0	+
SN7	Snella	Sofia	12	0	
SN8	Snella	Snella	7	0	+
SN9	Snella	Snella	7	0	+
Bernardo					
Emma					
EM1	Emma	Emma	10	0	*
EM2	Emma	Emma	10	1	+
Petto		Emma	7	0	+
Duca					
Bo					

3.7.3 Analysis of paternity

Results from analysis of paternity revealed multiple candidate fathers for two of the three sets of offspring (Table 7). It was not possible to identify a father for five of the nine offspring from Snella and for one of the two offspring of Emma and Sofia. Pair confidence levels were generally low, with only two offspring being identified with 80% confidence. Antonio was identified as the candidate father for two of the mothers, Sofia and Snella. Although two candidate fathers were identified for Snella, the second male lacked pair confidence, making Antonio a more likely father. It can be noted that the second candidate father for Snella was also identified as the candidate father for three of her offspring. Both mother and offspring No.1 had the same candidate father assigned, although both lacked pair confidence, and offspring No.2 had no candidate father identified. Detailed results are presented in Appendix 7.10.

Table 7: Cervus overview output for analysis of paternity

Offspring ID	Candidate father ID	Pair loci compared	Pair loci mismatching	Pair confidence
Sofia	Antonio	12	0	+
SO1	Bo	12	1	-
SO2	Bo	12	1	-
SO3	Spider	12	1	-
SO4	Petto	8	0	-
SO5				
SO6	Spider	11	0	-
SO7	Bo	12	0	-
Antonio	Angelo	12	0	+
Snella	Antonio	7	0	+
Snella	Fred	7	0	
SN1	Fred	11	1	-
SN2				
SN3	Duca	8	1	-
SN4	Fred	11	0	+
SN5				
SN6				
SN7				
SN8	Fred	11	0	+
SN9				
Bernardo				
Emma	Petto	7	0	-
EM1	Petto	8	0	-
EM2				
Petto	Spider	9	0	-
Duca				
Bo				
Spider	Petto	9	0	-
Fred				
Angelo	Antonio	12	0	+

3.7.4 Assignment of parent pairs

A candidate mother and father were both identified for all offspring from Sofia, however four offspring (out of nine) from Emma and one offspring (out of two) for Snella did not have any candidate parents assigned by Cervus. Candidate parents were also identified for two of the three mothers and three of the eight males in the population (Table 8). Multiple maternal candidates were identified for six offspring and in two of these cases, none of the candidates identified were the true mother. In the remaining four cases the correct mother was identified alongside one other female. All males genotyped were suggested as the candidate father for at least one pup (Table 8).

Pair confidence for mother – offspring pairings showed that 67% of pairs received 80% or higher confidence. Out of 36 pairings, 20 pairs (56%) scored 80% confidence, 4 pairs (11%) scored 95% confidence and the remaining 11 pairs scored a confidence lower than 80%. Pair confidence for father – offspring pairs were lower in comparison to mother – offspring, with just 16% of pairs scoring 80% confidence. Out of 36 pairings, 6 pairs (16%) scored 80% confidence, no pairs scored 95% and the remaining 32 pairs scored a confidence of lower than 80%. Trio confidence scores were lowest of all confidence scores calculated, with just one pair (3%) of 36 pairings scoring 80% confidence.

Table 8: Cervus overview output for assignment of parent pair

Offspring ID	Candidate mother ID	Pair confidence	Candidate father ID	Pair confidence	Trio confidence
Spider					
Fred					
Angelo					
Sofia	Snella		Antonio	+	-
SO1	Sofia	+	Bo	-	+
SO2	Emma	+	Spider		-
SO2	Emma	+	Bernardo		
SO2	Sofia		Bo	-	
SO3	Sofia	+	Bo		-
SO4	Emma	+	Spider		-
SO4	Emma	+	Antonio		
SO4	Emma	+	Petto	-	
SO4	Emma	+	Fred		
SO5	Sofia	+	Bo		-
SO5	Sofia	+	Spider		

SO6	**Emma**		**Spider**	**-**	**-**
SO6	Sofia	+	Spider	-	
SO6	Sofia	+	Bo		
SO7	Sofia	+	Bo	-	-
SO7	Snella		Bo	-	
Antonio	Snella	+	Angelo	+	-
Antonio	Sofia		Angelo	+	
Snella	Sofia		Bernardo		-
Snella	Sofia		Fred		
Snella	Emma		Antonio	+	
SN1					
SN2	Snella	*	Duca		-
SN2	Snella	*	Fred		
SN3	Sofia		Duca	-	-
SN3	Snella	+	Duca	-	
SN4	Snella	+	Fred	+	-
SN5					
SN6					
SN7	Emma	+	Petto		-
SN7	Emma	+	Antonio		
SN7	Sofia		Bo		
SN8	Snella	+	Fred	+	-
SN9					
Bernardo	Emma		Bo		-
Emma					
EM1	Emma	*	Petto	-	-
EM2	Emma	*	Fred		
Petto	Emma	+	Spider	-	-

3.7.5 Paternity analysis with known mothers

Trio confidence for paternal assignment was greatly increased when known mothers were provided, which can be seen when comparing the trio confidence columns (Table 6 and 9). Whereas previously just one pair achieved an 80% confidence score, when known mothers were provided four pairs (22%) achieved a confidence score of 80% and three pairs (17%) achieved a 95% confidence score (Table 9). Eleven pairs, out of eighteen (61%), scored low or failed to achieve a confidence level above 80% for trio confidence and these were generally pairs where the number of trio loci mismatching was higher. Offspring from Sofia are suggested to be from two candidate fathers. The two offspring from mother Emma are suggested to have the same father but trio confidence is low. Five candidate fathers are put forward for the offspring of Snella, making the paternity assignment for this offspring unreliable.

Table 9: Output for paternal parentage with known mothers

Offspring ID	Mother ID	Pair loci mismatching	Candidate father ID	Pair loci mismatching	Pair confidence	Trio loci mismatching	Trio confidence
SO1	Sofia	0	Bo	1	-	1	*
SO2	Sofia	0	Bo	1	-	2	+
SO3	Sofia	0	Bo	1		2	+
SO4	Sofia	0	Spider	0		2	
SO5	Sofia	0	Bo	1		1	+
SO6	Sofia	0	Spider	0	-	1	-
SO7	Sofia	0	Bo	0	-	0	*
SN1	Snella	0	Fred	1	+	3	
SN2	Snella	0	Duca	2		2	-
SN3	Snella	0	Duca	1	-	2	
SN4	Snella	0	Fred	0	+	1	+
SN5	Snella	1	Antonio	1		2	
SN6	Snella	0	Angelo	3		4	
SN7	Snella	0	Bernardo	1		1	
SN8	Snella	0	Fred	0	+	0	*
SN9	Snella	0	Fred	1		3	
EM2	Emma	0	Petto	0	-	0	-
EM2	Emma	1	Petto	1		3	

3.7.6 Paternity assignment for offspring of Sofia

Output from Cervus presents an interesting case for the offspring of Sofia, where it has indicated the possibility of dual paternity (Figure 3-2). Inheritance patterns can be seen in Tables 10-13.

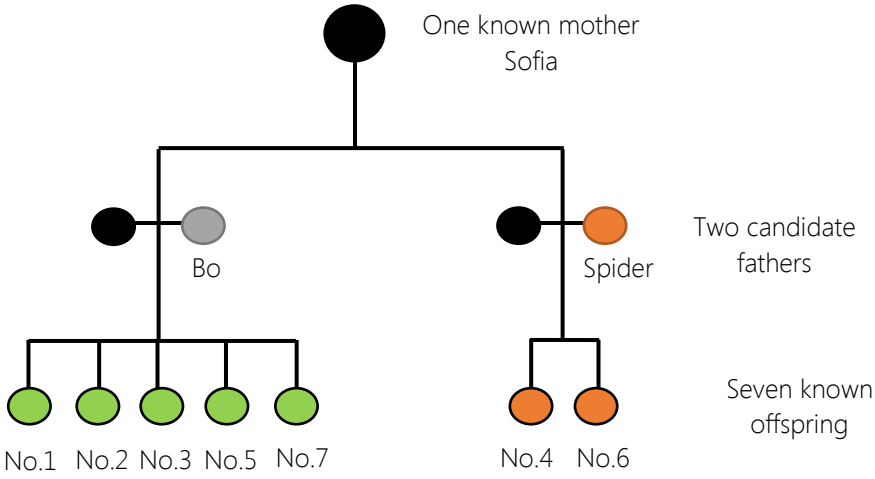


Figure 3-2: Kinship chart showing parentage for all seven offspring, where black = mother, solid orange = candidate father 23024, solid green = Bo, black outline with green middle = offspring of Sofia and Bo and black outline and orange middle = offspring of Sofia and Spider as proposed by Cervus.

Table 10: Paternity assignment for offspring of Sofia. Locus positions one to four (alleles A and B).

Individual	Allele A	Allele B	Allele A	Allele B	Allele A	Allele B	Allele A	Allele B
	Locus 1		Locus 2		Locus 3		Locus 4	
Spider	230	230	266	266	156	172	119	123
Sofia	230	230	266	270	156	168	115	123
SO1	226	230	266	270	156	172	115	123
SO2	226	230	266	266	156	156	119	123
SO3	230	230	266	266	168	172	119	123
SO4	230	230	266	266	156	156	123	123
SO5	230	230	266	270	168	172	123	123
SO6	230	230	266	266	156	156	123	123
SO7	230	230	266	266	168	172	123	123
Bo	226	230	266	266	156	172	119	123

Candidate father one in blue (Spider) and two in red (Bo). Offspring outlined by bold line. Red shading indicates alleles fit with the genotype of the candidate Bo, Blue shading indicates alleles that must have been inherited by candidate 23802, No shading represents allele that may have been inherited from either father or mother.

Table 11: Paternity assignment for offspring of Sofia. Locus positions five to eight (alleles A and B).

Individual	Allele A	Allele B	Allele A	Allele B	Allele A	Allele B	Allele A	Allele B
	Locus 5		Locus 6		Locus 7		Locus 8	
Spider	108	108	98	102	157	163	138	142
Sofia	102	108	102	106	157	157	134	140
SO1	102	108	98	106	157	157	134	134
SO2	108	108	98	102	149	157	134	140
SO3	102	108	98	106	149	157	134	140
SO4	102	108	98	102	149	157	0	0
SO5	108	108	98	102	157	157	134	134
SO6	102	108	98	102	157	157	0	0
SO7	108	108	98	102	157	157	134	134
Bo	108	108	98	98	157	163	134	140

Candidate father one in blue and two in red. Offspring outlined by bold line. Red shading indicates alleles that must have been inherited by Bo, Blue shading indicates alleles that must have been inherited by candidate 23802, No shading represents allele that may have been inherited by either father or from mother, yellow shading indicates alleles not present in any father or mother.

Table 12: Paternity assignment for offspring of Sofia Locus positions nine to twelve (alleles A and B).

Individual	Allele A	Allele B	Allele A	Allele B	Allele A	Allele B	Allele A	Allele B
	Locus 9		Locus 10		Locus 11		Locus 12	
Spider	157	157	**133**	133	116	116	150	160
Sofia	157	157	133	155	116	116	148	152
SO1	157	157	133	145	112	116	148	160
SO2	157	157	133	155	112	116	152	160
SO3	157	157	133	133	112	116	152	160
SO4	157	157	133	155	112	116	150	152
SO5	157	157	133	155	112	116	152	160
SO6	157	157	133	133	112	116	148	160
SO7	155	157	133	155	112	116	148	150
Bo	155	161	133	145	112	112	150	160

Candidate father one in blue and two in red. Offspring outlined by bold line. Red shading indicates alleles that must have been inherited by Bo, Blue shading indicates alleles that must have been inherited by candidate 23802, No shading represents allele that may have been inherited by either father or from mother.

Table 13: Comparison of the candidate fathers from analysis without known mother and with known mother provided, C = confidence

Offspring	Paternity (no known mother)	C	Paternity (known mother)	C
SO1	Bo	-	Bo	*
SO2	Bo	-	Bo	+
SO3	Spider	-	Bo	+
SO4	Petto	-	Spider	
SO5			Bo	+
SO6	Spider	-	Spider	-
SO7	Bo	-	Bo	*

3.7.7 Paternity assignment for offspring of Snella

A similar case of two candidate fathers can be seen for the offspring of Snella (Figure 3-3). Unlike in the case of Sofia where confidence levels were for the most part high (80-95%) for most offspring, only three out of the nine offspring scored an 80% confidence level. When a known mother was provided to Cervus a father was predicted in all individuals compared to analysis ran with no known mother, where a predicated father was allocated for just four offspring (Table 14).

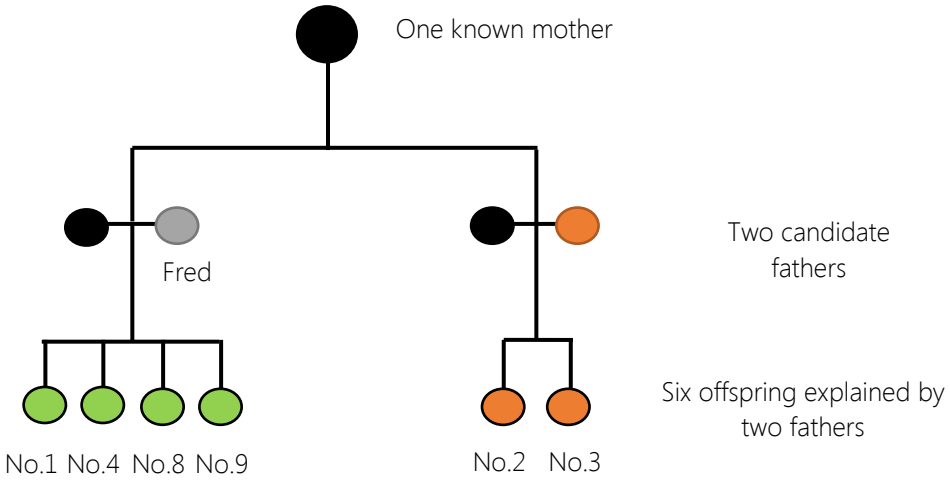


Figure 3-3: Kinship chart showing parentage for six out of nine offspring of Snella, where black = mother, solid orange = candidate father 93145, solid green = candidate father Fred, black outline with green middle = offspring of Snella and Fred and black outline and orange middle = offspring of Snella and Duca

Table 14: Comparison of the candidate fathers from analysis without known mother and with known mother provided, C = confidence

Offspring	Paternity (no known mother)	C	Paternity (known mother)	C
SN1	Fred	-	Fred	+
SN2			Duca	
SN3	Duca	-	Duca	-
SN4	Fred	+	Fred	+
SN5			Antonio	
SN6			Angelo	
SN7			Bernardo	
SN8	Fred	+	Fred	+
SN9			Fred	

Table 15: Paternity assignment for offspring of Sofia Locus positions one to four (alleles A and B).

Individual	Allele A	Allele B	Allele A	Allele B	Allele A	Allele B	Allele A	Allele B
	Locus 1		Locus 2		Locus3		Locus 4	
Duca	230	238	262	270	152	168	123	127
Snella	230	230	266	270	156	168	115	123
SN1	226	230	266	270	156	172	115	123
SN2	226	230	266	266	156	156	119	123
SN3	230	230	266	266	168	172	119	123
SN4	230	230	266	266	156	156	123	123
SN5	230	230	266	270	168	172	123	123
SN6	230	230	266	266	156	156	123	123
SN7	230	230	266	266	168	172	123	123
SN8	230	230	266	270	156	168	115	123
SN9	226	230	266	270	156	172	115	123
Fred	226	230	262	266	156	156	123	127

Candidate father one in blue (Duca) and two in red (Fred). Offspring outlined by bold line. Red shading indicates alleles fit with the genotype of the candidate Fred, Blue shading indicates alleles that must have been inherited by Duca. No shading represents allele that may have been inherited from either father or mother, yellow shading indicates alleles not present in any father or mother.

Table 16: Paternity assignment for offspring of Sofia. Locus positions five to eight (alleles A and B).

Individual	Allele A	Allele B	Allele A	Allele B	Allele A	Allele B	Allele A	Allele B
	Locus 5		Locus 6		Locus 7		Locus 8	
Duca	102	108	98	98	157	157	140	140
Snella	102	108	102	106	157	157	134	140
SN1	102	108	98	106	157	157	134	134
SN2	108	108	98	102	149	157	134	140
SN3	102	108	98	106	149	157	134	140
SN4	102	108	98	102	149	157	0	0
SN5	108	108	98	102	157	157	134	134
SN6	102	108	98	102	157	157	0	0
SN7	108	108	98	102	157	157	134	134
SN8	102	108	102	106	157	157	134	140
SN9	102	108	98	106	157	157	134	134
Fred	108	110	102	102	157	169	134	138

Candidate father one in blue (Duca) and two in red (Fred). Offspring outlined by bold line. Red shading indicates alleles fit with the genotype of the candidate Fred, Blue shading indicates alleles that must have been inherited by Duca. No shading represents allele that may have been inherited from either father or mother, yellow shading indicates alleles not present in any father or mother.

Table 17: Paternity assignment for offspring of Sofia. Locus positions nine to twelve (alleles A and B).

Individual	Allele A	Allele B	Allele A	Allele B	Allele A	Allele B	Allele A	Allele B
	Locus 5		Locus 6		Locus 7		Locus 8	
Duca	157	157	133	133	108	114	150	150
Snella	157	157	133	155	116	116	148	152
SN1	157	157	133	145	112	116	148	160
SN2	157	157	133	155	112	116	152	160
SN3	157	157	133	133	112	116	152	160
SN4	157	157	133	155	112	116	150	152
SN5	157	157	133	155	112	116	152	160
SN6	157	157	133	133	112	116	148	160
SN7	155	157	133	155	112	116	148	150
SN8	157	157	133	155	116	116	148	152
SN9	157	157	133	145	112	116	148	160
Fred	157	157	133	149	110	116	150	150

Candidate father one in blue (Duca) and two in red (Fred). Offspring outlined by bold line. Red shading indicates alleles fit with the genotype of the candidate Fred, Blue shading indicates alleles that must have been inherited by Duca. No shading represents allele that may have been inherited from either father or mother.

3.7.8 Identification of siblings using Kinanalyzer

Kinanalyzer was used to compliment Cervus and produce a prediction of sibling sets. It predicted 6 sets of siblings based on a two allele algorithm. Sibling identification was not accurate for all known sibling groups and in some cases the same individual is identified to be in more than one group (e.g. SO3) is assigned to sibling set 0 and 3.

Figure 3-4 shows a simplified representation of the Kinanalzyer output and demonstrates that the majority of offspring of female 387 (grey circle) were grouped together in sibling set 5, with two offspring grouped in sibling sets 0 and 2. It can be seen in Table 22 that female 387 is grouped with her offspring for sibling set 5. A similar outcome can be seen for Sofia and her seven offspring (blue square), which are mainly clustered in sibling set 3. Female 922 and her two offspring, were each grouped separately in different sibling sets

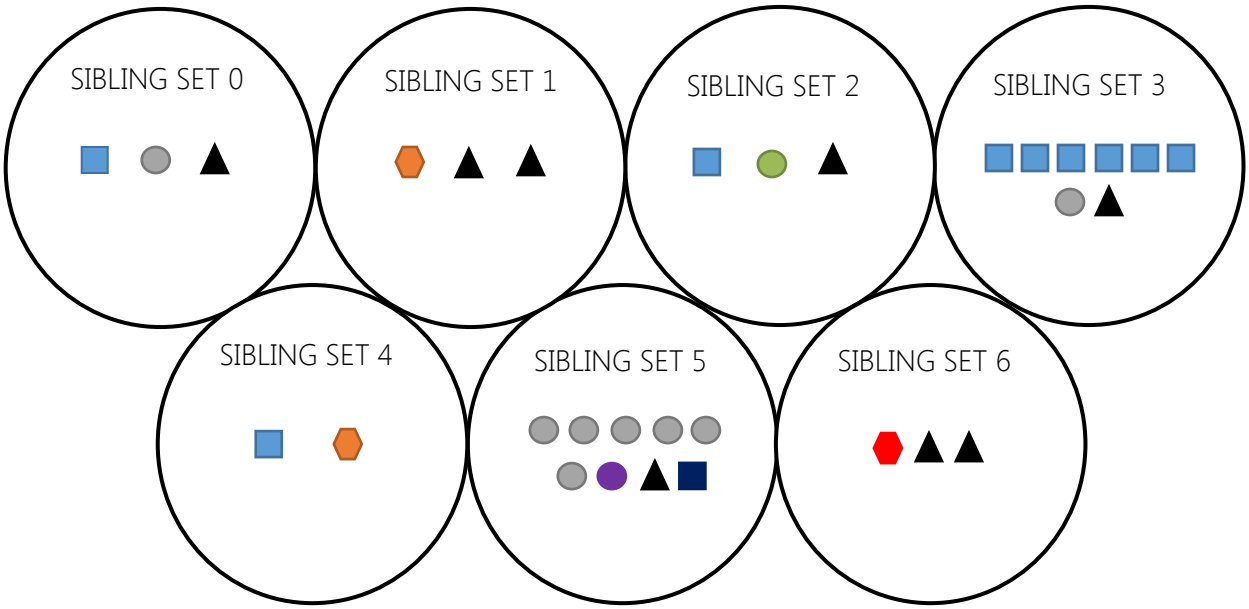


Figure 3-4: Visual representation of Kinalyser groupings, where = Sofia seven offspring, = Sofia, = nine offspring of Snella, = Snella = two offspring of Emma, = Emma, and = all males

3.8 DISCUSSION

Microsatellite analysis of the study population revealed low number of alleles per locus, moderate levels of heterozygosity and difficulties in assigning parentage. Results from parentage analysis revealed multiple breeding individuals to be present in the sample population.

A similar study was conducted by Godinho *et al.* (2011), focusing on wolf-dog hybridization in the Iberian Peninsula, utilising 42 autosomal microsatellites. For all genetic diversity measures, Iberian wolves exhibited lower values when compared to the dogs, for example mean expected heterozygosity was Hewolf = 0.617 and Hedog = 0.755. The free-ranging dog population utilised in the present study were 0.177 (He) had lower expected heterozygosity lower than Iberian domestic dogs and 0.039 (He) lower than comparable to Iberian wolves. Garcia-Moreno *et al.* (1996) compared Mexican gray wolves, domestic dogs, northern gray wolves and coyotes using 10 microsatellite loci. They found a mean expected heterozygosity of 0.616 for domestic dogs, 0.675 for coyotes, 0.620 for gray wolf populations (non-hybridizing) and 0.713 for gray wolf populations (hybridizing), and 0.437, 0.103 and 0.253 for 3 captive Mexican wolf populations. In comparison with these values, the free-ranging dog population analysed in the present study had a lower He compared to domestic dogs, coyotes and gray wolves but a higher He in comparison to the Mexican wolf

Adherence to Hardy-Weinberg equilibrium and moderate levels of heterozygosity suggest randomised mating in the population, which fits with the hypothesis that multiple breeding individuals are present in free-ranging dog populations. Adaptation as a result of the domestication process is one functional explanation for the presence of multiple breeding individuals in a pack (Cafazzo *et al.*, 2014). As previously discussed in chapter 3.1, free-ranging dog populations no longer follow seasonal reproductive behaviour and have adapted to scavenging human waste due to its abundance. This may have resulted in allowing reproduction to occur during the first year of life, once full body weight is reached (Lord *et al.*, 2013). A high quantity of food sources available to free-ranging dogs could have resulted in a reduction in the level of competition experienced by the group for food and thus, a reduction in the reproductive suppression of subordinates (Bonanni and Cafazzo, 2014).

3.8.1 Analysis of parentage

Assignment of mothers was completed to assess whether Cervus analysis was capable of correctly predicting the mothers of all offspring. Correct identification was found for fifteen of the eighteen total offspring in total, but for two individuals a second candidate mother was provided (Table 6). Indication of alternative mothers, for example prediction of Emma as a putative mother for the offspring of Sofia could indicate that the mothers analysed in this study are closely related, although missing data at some loci in females could increase the error rate and this should be taken into consideration.

Analysis of fathers was first completed carried out completely blind (i.e. known mothers were not provided to Cervus) and results widely lacked confidence, with just six individuals showing 80% confidence. No coherent evidence was put forward for the paternity of the offspring of females Sofia or Snella, with a multitude of candidate fathers predicted in both cases for all individuals (Table 7). For the two offspring of Emma, only one had a candidate father assigned. Overall, results lacked confidence and reliability.

Simultaneous assignment of both mothers and fathers was also undertaken blind (i.e. no information about mother was provided) and as seen for paternal analysis, confidence levels for trios (offspring, putative mother and putative father) were low, and numerous fathers were provided both for all offspring from the same mother and for individual offspring, with incorrect identification of mothers observed for some individuals. These facts combined suggested a lack of reliability and low confidence in both paternal prediction, and low trio confidence levels corroborated this.

Analysis providing genotypes of known mothers to Cervus increased the confidence levels of results. Notable differences can be seen in the candidate fathers put forward, when a known mother is provided, especially for the offspring of Sofia where only two males were put forward compared to an original five individuals putative fathers proposed in earlier analyses. Results for the offspring of Snella remain unreliable for offspring 5-7, represented which are characterised by low confidence levels and a higher number of mismatched loci. For the remaining offspring (1-4, 8-9) two candidate fathers were identified. One candidate father was identified for both offspring of Emma, the same candidate father which was suggested in the original prediction of paternity. Three mismatched loci were found for the second offspring, representing 25% of all loci used. matched loci were found for the second offspring representing 25% of all loci used.

3.8.2 Indication of multiple paternity

Cervus results indicated the possibility of dual paternity by males Bo and Spider for the offspring of Sofia, by Bo and Spider, but on closer inspection it became apparent that Spider lacked sufficient evidence. Bo's genotype is consisted consistent with the majority of loci used in this study for all the offspring, but still failed to explain all genotypes present across all offspring, although when comparing father Bo with Spider, it is still not possible to explain all genotypes using Spider. At locus seven, three offspring have alleles that match neither their mother nor both candidate fathers. Combining these facts together, it is not likely that Spider is a true father. Taking this into account there are three plausible explanations for the Cervus paternity assignments:; (1) That Bo is the father of all offspring in this litter and the mismatching genotypes are the result of genotyping errors, (2) Bo is the father of four offspring in the litter and the remaining three offspring are fathered by an individual which was not analysed or (3) a different male, which has not been genotyped, is the father of all offspring (i.e. consistent with the notion that all the pups are all full siblings). Behavioural observations from the research group of Dr Eugenia Natoli revealed that male Bo is a likely candidate father as he and the mother of the pups were both seen in the same area, even though they are from neighbouring groups ("Eucalipiti" and "Borgo dei Massimi") (Table 1 and 2). They noted that despite Bo being young he had a high social status and a large body size (Dr Eugenia Natoli, personal communication).

Dual paternity was also suggested for the offspring of Snella, with Duca and Fred being the most likely candidates, however, when a known mother was provided to Cervus, other candidate fathers were also suggested (Table 14). Confidence for both paternity (no known mother) and paternity (known mother provided) are highest for Fred in all cases, with low or no confidence present for all other suggested fathers. For ten loci, genotypes all alleles could have been inherited by derived from either the candidate father or mother. For the remaining two loci, both and Fred (locus one, Table 15) and Duca (Locus 6, Table 16) successfully explain the genotype. Across four loci (three, four, six and seven) a total of 17 genotypes cannot be successfully matched to either Duca or Fred, leading to the assumption that the father is most likely a different individual, not present in among the males sampled genotyped.

3.8.3 Possible sibling groups

Results from Cervus revealed evidence for two pairs of siblings, firstly Sofia and Antonio and secondly Angelo and Antonio. Sofia was predicted as the mother of Antonio (Table 11) and vice versa, Antonio was predicted to be the father of Sofia (Table 14) demonstrating that they are closely related, although the exact relationship cannot be determined based on the data available. In the second case, both Angelo and Antonio were shown to be the candidate father of one another (Table 14) indicating that they are closely related. It should be noted that due to a generation time of ~2 years in dogs but a lifespan of up to 10 years, it is not possible to confirm whether Angelo and Antonio are full siblings, or a father – offspring pair. Kinanalyser failed to pair either of these groups in the same sibling set for a two allele algorithm.

3.8.4 Inbreeding in dog populations

Results from Cervus suggest the possibility of inbreeding (i.e. mating between relatives) occurring in the study population. This is exemplified in the case of Emma and Petto. Results show Emma as the predicted mother for Petto (Table 9, Table 15). Analysis of paternity (no known mother) predicts Petto as the father of Emma, as well as one of her offspring (Table 12), whilst analysis of paternity (with a known mother) indicates Petto as the father of both offspring of Emma. Despite low confidence levels, this would imply that Emma mated with her close relative (father, son, or brother), which resulted the two offspring analysed in this study. Even when there are few or no cases of incest occurring in populations, low levels of heterozygosity in a population provides evidence of inbreeding.

Small populations who accept little or no immigrants into the population will have some level of inbreeding, due to all individuals being likely to have distant relatives present in the population, such as cousins. If inbreeding is occurring in the sample population, there could be detrimental effects in the long run. Inbreeding decreases effective population size (Pollak, 1987), and may result in inbreeding depression or the accumulation of deleterious mutations (Lande and Schemske, 1985; Porcher and Lande, 2005; Porcher and Lande, 2016). Genome-wide homozygosity levels will increase with inbreeding, which can result in fitness reduction in a population (Keller and Waller, 2002; Charlesworth and Willis, 2009).

3.8.5 Sources of error in genotyping of microsatellite loci

In between the extraction of DNA and entering the correct genotype into a database there are numerous steps, at which various errors could occur. Examples of error sources include; misprinting (i.e. incorrect identification of an artefact band/peak as a true allele and including it in the genotype), poor amplification, mislabelling, incorrectly identifying stutter patterns or artefact peaks or data entry errors, and null alleles (Bonin *et al.*, 2004; Selkoe and Toonen, 2006). An error rate of just 1%, i.e. where 1% of alleles into a database are incorrectly identified, can lead to a substantial number of incorrect multilocus genotypes in a big data set and this is an uncommonly small error rate for most studies (Hoffman and Amos, 2005).

3.8.5.1 Levels of heterozygosity

Microsatellites with higher levels of heterozygosity are more powerful at assigning relatedness per locus (Yu *et al.*, 2015). On the other hand, the presence of null alleles (Dakin and Avise, 2004) and a high mutation rate of 10^{-2}-10^{-5} per generation (Agrafioti and Stumpf, 2007) can cause interference when accurately constructing pedigrees (Yu *et al.*, 2015). Identification of parentage based on microsatellites can be problematic for populations with low heterozygosity unless a large number of polymorphic loci can be utilised (Schopen *et al.*, 2008; Tokarska *et al.*, 2009). Additionally, microsatellite discrimination can be significantly weakened when there is a high prevalence of genetic variation and null alleles (Yu *et al.*, 2015). In this study, a total of 14 loci were amplified but two loci failed, leaving a total of 12. If this study were to be replicated then a higher number of loci would be encouraged to increase reliability of results and confidence scores for pairings.

3.8.5.2 Potential causes of null alleles

A null allele is defined as any allele at a given microsatellite locus where amplification consistently fails to reach detectable levels via the polymerase chain reaction (PCR) (Dakin and Avise, 2004). Poor primer annealing is one potential cause of null alleles due to divergence of nucleotide sequence (e.g. from the presence of indels or point mutations) in one or both of the flanking primers. Mutations in the 3' end, where extension begins, of the priming site are thought to be particularly detrimental to PCR amplifications (Kwok *et*

al., 1990). A second source of null alleles involves PCR failure due to inconsistent DNA template quality or low template quantity (Gagneux *et al.*, 1997; Garcia de Leon *et al.*, 1998). When the case occurs that DNA template at a specific locus is poor in selected specimens, the poor samples can sometimes appear homozygous rather than heterozygous for the null allele (Dakin and Avise, 2004). Generation of null alleles via differential amplification of size-variant alleles is another possible source (Wattier *et al.*, 1998). PCR is inherently competitive and alleles of shorter lengths generally amplify more efficiently than larger alleles, such that just the smaller of the two alleles becomes detectable from a heterozygous individual, making them appear homozygous. They are sometimes referred to as 'partial nulls' because this can be easily remedied by the addition of more DNA matrix (Dakin and Avise, 2004). For primer 250 (locus eight), ten individuals failed to amplify at all representing the possibility of null alleles occurring at this location. Further repetition would be required to confirm this.

3.9 CONCLUSION

To conclude, microsatellite analysis for this study population demonstrated moderate heterozygosity, low average number of alleles per locus, and low levels in confidence of parentage assignment rates. Low genetic variability suggested the possibility of inbreeding occurring in this population. Although the parentage analysis suggested the possibility of multiple paternities of two litters, this result had low statistical support, and could have resulted from low genetic variability of the population. Further research using a larger number of loci is required to ascertain a clear and reliable picture.

4. FUNCTIONAL GENETIC DIFFERENTIATION BETWEEN PURE-BREED AND FREE-BREEDING DOGS AT *MARCH7* GENE

4.1 WHAT IS THE IMMUNE SYSTEM?

The immune system is generally divided in two components: innate immunity (inborn components) and adaptive (acquired) immunity (Janeway and Mezhitov, 2002; Palm and Medzhitov, 2009). Innate immunity is comprised of physical barriers (e.g. mucous membranes) and specialised cells (e.g. macrophages and granulocytes) and provides protection requiring no prior exposure to target pathogens (Basset *et al*., 2002; Schley and Field, 2002). In contrast, adaptive immunity relies on previous exposure and consists of B and T lymphocytes that can form immunological memory and provide specific immune responses to targeted antigens (Yabas *et al*., 2016).

4.2 IMMUNE SYSTEM GENES

A broad range of genes are involved in mammalian response to pathogen infections, resulting in a complex pathogen-specific selection pressure acting on the immune system in wild animals (Turner *et al.,* 2012). Due to differences in the ecology and habitat of free-ranging dogs, purebred dogs and wolves, each group will have their own pathogenic environment, and thus potential exposure to differing pathogen-specific selection pressures.

Immune system genes are highly polymorphic in many species due to the evolutionary arms race occurring between pathogens and their hosts, and the fact that maintaining an effective immune response is essential for the survival of small populations (Quintana-Murci *et al.,* 2013; Chae *et al.,* 2014; Niskanen *et al.,* 2014). More specifically, genes associated with immune response are theorised to be under long-term positive selection (Metz *et al.,* 1998, Jansa *et al.,* 2003).

One factor contributing to the extinction and population decline of wild mammalian carnivores across the world is infectious disease-driven mortality (Gompper, 2014; Knobel *et al.* 2014). In combination with other endangerment factors this can have a particular impact on small or declining populations that are experiencing habitat loss or fragmentation. This can also be the case for canids where disease transmission is influenced by humans inhabiting the same geographic region (Knobel *et al.,* 2014), which has been demonstrated in North America where grey wolf populations have suffered long-term pup mortality due to parvovirus infection (Mech *et al.,* 2008).

Immune function is known to be highly heritable (da Craen *et al.,* 2005; Sorci *et al.,* 1997; Cooke and Hill, 2001) and control of this is linked to a combination of alleles, which encode functionally relevant immune molecules (Bulher and Sanchez-Mazas, 2011; Reche and Reinherz, 2003; Sanchez-Mazas and Meyer, 2014). Alleles belonging to immune genes are hypothesised to coevolve in direct interaction with pathogens (Dodds and Thrall, 2009). The Red Queen hypothesises states that pathogens create a constant pressure for the introduction of new alleles in populations, resulting in high variability within immune genes (Woolhouse *et al.,* 2002; Těšický and Vinkler, 2015).Immune genes are in fact among the most polymorphic protein coding genes within the genome (Morris *et al.,* 2015), and studies in several vertebrate species have demonstrated that the rate of adaptive evolution is higher in immune genes in comparison to other gene classes (Huang *et al.,* 2004; Tonteri *et al.* 2010). Resistance in both wild and laboratory animals for a diverse range of diseases has been witnessed as a result of polymorphism in immune genes (for examples see; Paterson *et al.,* 1998; Piertney and Oliver, 2006), implying that they can provide adaptive potential to wild populations (Morris *et al.,* 2015).

. Homozygosity in major histocompatibility complex (MHC) have been shown to increase the risk of parasite infection and autoimmune diseases (e.g. Meyer-Lucht and Sommer, 2005; Kennedy *et al.,* 2006). Niskanen *et al.* (2014) found that MHC-heterozygous wolves or carriers of a specific DLA-DRB1 allele exhibited fewer infections when compared to homozygotes or carriers of other DLA-DRB1 alleles. This confirms that pathogen load, in this case parasites, can be an important source of selection in wolves (Niskanen *et al.,* 2014).

4.3 SELECTION PRESSURES ON THE DOMESTIC DOG

Today, there is an estimated population of ~1 billion dogs worldwide (Gompper, 2014). It is theorised that the first domesticated population was the result of just a few founder individuals and that this small population size could have resulted in an accumulation of deleterious mutations (Vilà *et al.,* 1997; 2005; Savolainen *et al.,* 2002). Initially humans selected strongly for beneficial behavioural traits, such as tameness (Saetre *et al.,* 2004) but the artificial selection pressures imposed by humans resulted in a relaxation of pressures on other traits. As a result of positive selection in the dog, an accumulation of nonsynonymous mutations in the entire dog genome was more feasible (Björnerfeldt *et al.,*

2006). Differences between the gray wolf and dog genome are hypothesised to be largest when considering regions which were influenced by selection during the early stages of the domestication process. Similarly, genetic differentiation between dog breeds is expected to be clearly seen in regions that experienced selection as the result of breed formation (Ramirez *et al.,* 2014). Apart from nonsynonymous changes in proteins, other types of genetic variation are witnessed when comparing dogs with wolves. Although there is a lack of direct evidence to suggest that frequency of these changes is higher in dogs compared to wolves, it has been suggested that variation in tandem repeats (Fondon and Garnder, 2004), and presence of short interspersed elements (SINEs) (Wang and Kirkness, 2005) could contribute to phenotypic diversity. As well as phenotypic diversity, it is possible that a reduction in selective constraints acting on the domestic dog is linked to a large number of diseases and conditions affecting multiple dog breeds (Ostrander and Krugzak, 2000).

There is evidence of an increased accumulation of nonsynonymous mutations in the dog's genome since domestication (Cruz *et al.,* 2008). This has been attributed to two key factors, (1) relaxation of selective pressures and (2) the effect of positive selection on linked sites as a result of Hill-Robertson interference (Hill and Robertson 1966), which can reduce the probability that these deleterious mutations will be eradicated from the population. A comparative study of mtDNA lineages demonstrated that there has been a greater accumulation of nonsynonymous mutations in dogs compared to wolves (Björnerfeldt *et al.,* 2006). Axelsson *et al.* (2013) identified 36 genomic regions that are likely to represent targets for selection between wolves and dogs. Ten of these genes were found to have key roles in fat metabolism and starch digestion. Axelsson *et al*. (2013) hypothesised that domestic dogs have acquired a greater ability to digest starch when compared to wolves as a direct result of human-dog interaction. In a similar study, nine genes relating to high-altitude adaption were found to display signatures of positive selection between Tibetan Mastiffs and native Chinese dogs (Li *et al.,* 2014). Evidence for positive selection in genes involved in metabolism (particularly lipid metabolism), pigmentation and those influencing behaviour, neuropsychiatric disorders and brain function in pure-breed dogs has also been demonstrated (Freedman *et al.,* 2016). A study by Marsden *et al.* (2016) found that pure-breed dogs had higher levels of deleterious genetic variation than gray wolves at a genome–wide level, whilst free-ranging dogs displayed intermediate values.

As previously mentioned in chapter 2.4.1 a study assessing genome-wide differentiation between free-ranging dogs and pure-breed dogs found strong differentiation in an immune system gene *MARCH7* (Pilot *et al.,* 2016). This leads to the intriguing question whether *MARCH7*may show diversifying selection between pure-breed and free-ranging dogs, as well as grey wolves.

4.4 MARCH GENES

The membrane-associated RINGCH-type finger (MARCH) family is a RING finger protein family of E3 ubiquitin ligases, consisting of 11 members in mammals (Zhao *et al.,* 2013, Szigyarto *et al.,* 2010). The RING (Really Interesting New Gene) family is the largest type of E3 ubiquitin ligases (Chasapis and Spyroulias, 2009). E3 ligases play a role in providing specificity to ubiquitination by recognizing target substrates and mediating the transfer of ubiquitin from an E2 ubiquitin-conjugating enzyme to substrate (Deshaies and Joazerio, 2009; Iyenger *et al.,* 2011).

Ubiquitination is a post-translational modification in which the 76-amino acid polypeptide ubiquitin (Ub) is covalently attached to lysine residues in target proteins (Iyenger *et al.,* 2011). Protein modification by ubiquitin serves a critical signalling function across a diverse range of cellular processes. The combinatorial diversity within the ubiquitin pathway suggests that it is the most complex regulatory system of the eukaryotic cell (Nathan *et al.,* 2008). With only three amino acid differences between mammals, yeast and plants, ubiquitin displays a remarkable evolutionary conservation (Shaid *et al.,* 2013). Ubiquitination is catalysed by the sequential actions of E1 Ub-activating, E2 Ub-conjugating, and E3 Ub ligase enzymes (Iyenger *et al.,* 2011). It has been shown that ubiquitination also regulates key cellular processes including gene transcription, cell cycle progression, DNA repair, apoptosis, virus budding and receptor endocytosis (Shaid *et al.,* 2013).

Out of the 11 MARCH members found in mammals, 9 members possess hydrophobic transmembrane spaces that are known to be localised to the intracellular organelle membrane and plasma membrane (Iyengar *et al.,* 2011). The remaining 2 members, *MARCH7* and *MARCH10* have no transmembrane domain (Iyengar *et al.,* 2011; Nakamura, 2011). Generally, MARCH proteins are known for having multiple cellular

functions, such as immune regulation, protein quality control and membrane trafficking (Zhao *et al.*, 2013; Hu *et al.*, 2015).

4.5 MARCH7

The *MARCH7* gene, also known as axotrophin, codes a protein of 693 amino acids with a single recognized functional motif, the RING-CH domain, close to the C-terminus (Figure 4-1) (Nathan *et al.*, 2008).

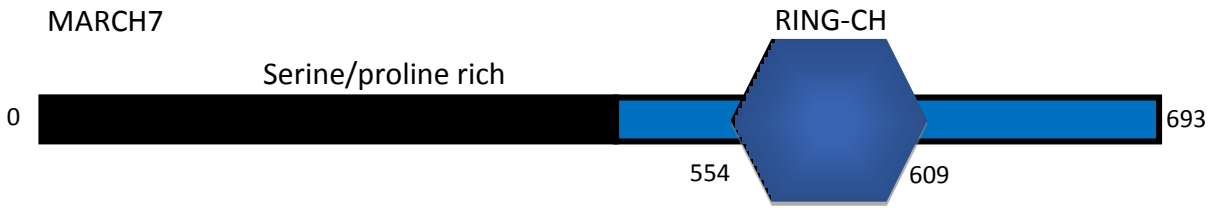


Figure 4-1: Schematic of MARCH7 structure adapted from Nathan et al. 2008

Expression of *MARCH7* has been demonstrated to be high in neurones, stem cells and lymphocytes, thus representing a possible involvement in both development and the immune system (Su *et al.*, 2002). Conservation of *MARCH7* appears to be strong amongst vertebrates, especially in mammals where homology between mouse and human is 85% and an identical RING-CH domain is witnessed (Nathan *et al.*, 2008). Localisation of *MARCH7* has been demonstrated to occur in the nucleus and cytosol of transfected cultured cells (Nathan *et al.*, 2008) and studies involving *MARCH7*-null mice provided evidence that it could play a fundamental role in immune tolerance and T-cell proliferation (Metcalfe *et al.*, 2005; Metcalfe and Muthukumarana 2005). Metcalfe *et al.* (2005) detected early axonal degradation of the dorsal root ganglia and agenesis of the corpus callosum in *MARCH7*-null mice and concluded that *MARCH7* was only mildly important in normal development. A subtractive gene array study by Metcalfe and Muthukumarana (2005) indicated that *MARCH7* may play a role in immunity when they demonstrated a specific link to immune tolerance for eight genes, one of which being *MARCH7.* They discovered feedback regulation of T lymphocytes fails to regulate when *MARCH7* is absent and T-cell mediated immunity becomes activated and noted both a five-fold overproduction and eight-fold hyperprolifertation of leukaemia inhibitory factor (LIF). LIF is a member of the Interleukin 6 family of cytokines and demonstrates pleiotropic effects on numerous organs and cell types (Graf *et al.*, 2011; Mathieu *et al.*, 2012). LIF has many functions, including involvement in the systematic inflammatory

response, suppressing differentiation of embryonic stem cells, facilitating endometrial implantation of embryos conversion of sympathetic neurons to the cholinergic phenotype from adrenergic and enhancing proliferation of myoblasts (Blanchard *et al.,* 2000). In conjunction with a finding that B lymphocytes are unaffected in *MARCH7*-null mice (Metcalfe *et al.,* 2005) it became evident that *MARCH7* is specifically linked to active T lymphocytes and provides negative regulation. Gao *et al.* (2009) demonstrated that *MARCH7* is fundamental for the degradation of LIF receptor gp190 subunit, a hetero-oligomeric receptor complex that binds to LIF to allow exertion of biological activities (Hisaka *et al.,* 2004).

MARCH7 is involved in the ubiquitination reaction (Szigyarto *et al.,* 2010) a key mechanism linked to regulation of the stability, activity and location of the Hedgehog (HH) signaling components (Hsia *et al.,* 2015). Regulation of *MARCH7* is achieved through degradation (auto-ubiquitination) and preservation (deubiquitination) through specialised deubiquitination enzymes, USP-7 and USP-93 (Nathan *et al.,* 2008). Flierman *et al. (*2006) demonstrated an association of *MARCH7* with E2-25K protein, which is typically known as the huntingtin-interacting protein due involvement in the ubiquitination of the gene product for Huntington's disease, huntingtin (Szigyarto *et al.,* 2010).

4.5.1 Hedgehog signalling pathway

In their study, Pilot *et al.* (2016) suggested that regulatory functions of three candidate genes under diversifying selection between pure-breed and free-breeding dogs (*MARCH7*, *PKD1L1* and *CALCB*) are linked through the Hedgehog (HH) signaling pathway. This led them to hypothesise that diversifying selection between free-ranging and pure-bred dogs did not occur independently on individual genes but through a common developmental and genetic mechanism.

The HH pathway is one of major signalling pathways that control key steps of embryonic development (Yao and Chuang, 2015). HH signalling controls numerous processes during insect and vertebrate embryonic development and adult homeostasis including tissue/organ pattering (more specifically of the neural tube, lung, skin, axial skeleton, and gastrointestinal tract) (Saqui-Salces and Merchant, 2010), cellular proliferation and differentiation, pathfinding, left/right asymmetry and stem cell maintenance (Yao and Chuang, 2015).

4.6 OBJECTIVES OF THIS STUDY

This study aims to expand on the findings by Pilot *et al.* (2016), investigating the level of genetic variation found in the *MARCH7* gene.

Objective 1: Study patterns of genetic variation in free-ranging dogs, pure-breed dogs and wolves, at a targeted site covering a SNP site identified to be under diversifying selection in the three canid groups by Pilot *et al.* (2016).

- **Hypothesis**: Differences in the patterns of genetic variation will be present across free-ranging dogs, pure-breed dogs and wolves, resulting from differing selection pressures.

Objective 2: Study patterns of non-synonymous versus synonymous mutations across a range of mammalian species to identify any signatures of positive and/or purifying selection.

- **Hypothesis**: Differences in patterns of nonsynonymous versus synonymous mutation will vary across mammalian species and could result in changes to protein function

Objective 3: Utilise structural protein prediction software to study protein conformation patterns for non-synonymous mutations found to be uniquely present in the dog when compared to other mammalian species.

- **Hypothesis**: non-synonymous mutations will result in changes to protein confirmation and tertiary structure.

4.7 METHODOLOGY
4.8 PATTERNS OF GENETIC VARIATION OF *MARCH7* IN CANIDS

4.8.1 Sample collection

All samples used for this project have been obtained from existing collections and/or databases and none were obtained specifically for this study.

DNA samples were obtained from four sources:

1) Wolf DNA samples were obtained from the collection of the Museum and Institute of Zoology, Polish Academy of Sciences. The Museum and Institute of Zoology is a CITES institute and has obtained all necessary permits to import the samples.
2) DNA samples from most free-ranging dogs were obtained from the collection of the Museum and Institute of Zoology, Polish Academy of Sciences.
3) The remaining free-ranging dog samples were provided by Dr Eugenia Natoli.
4) DNA from pure-breed dogs was collected at the University of Lincoln by Fernanda Fadel for her PhD study, which underwent all relevant ethical approvals.

4.8.2 Free-ranging dog samples

Free-ranging dog samples obtained from the Museum and Institute of Zoology came from across the world (Table 18).

Table 18: Sampling site and region for free-ranging dog samples used

ID	Sampling site	Region
3SL	Portoroz, Slovenia	Europe
8SL	Skofije, Slovenia	Europe
387	Rome, Italy	Europe
9916010	Rome, Italy	Europe
981538	Rome, Italy	Europe
3PL	Zduny, Poland	Europe
19AS	Riyadh, Saudi Arabia	Middle East
20AS	Riyadh, Saudi Arabia	Middle East
6CH	Zibo, Shandong Province, China	East Asia
5TAJ	Mueang Khon Kaen District, Thailand	East Asia
14TDZ	Dushanbe, Tajikistan	Central/West Asia
7KZ	Almaty, Kazakhstan	Central/West Asia

4.8.3 Pure-breed domestic dog samples

DNA samples from pure-breed dogs were assessed for concentration and purity using a NanoDrop 1000 spectrophotometer (Thermo Scientific). A total of 30 dogs were selected from breeds chosen to represent differences in body size, shape and other morphological characteristics to reflect the diversity witnessed present in pure-breed dogs (Table 19).

Table 19: Pure-breed domestic dog samples with associated NanoDrop scores

Sample ID	Breed	260/280	260/230
2786	Springer Spaniel	1.89	1.35
3364	Smooth Dachshund	1.882	1.81
1698	Labrador	1.71	0.82
1827	Flat coated retriever	1.82	0.84
1914	Chow Chow	2.03	1.45
1725	Shar Pei	1.93	1.14
1669	Hovawart	1.82	0.92
2727	Beagle	1.94	1.34
2285	German Shepard	1.87	1.11
540	Greyhound	1.96	1.54
1818	Staffordshire bull terrier	2.01	1.52
296	Tibetan terrier	1.92	1.91
2883	Border Collie	1.98	1.61
4444	Bearded Collie	1.84	1.26
1357	Shetland Sheepdog	1.79	1.04
1363	Lakeland terrier	1.79	1.21
588	Jack Russel terrier	1.83	1.01
4274	Rottweiler	1.9	1.42
1310	Dalmatian	1.73	1.03
594	Keeshond	1.83	1.14
726	German spitz mittel	1.66	0.9
1800	Golden retriever	1.88	1.19
1933	Akita	1.92	1.29
1664	Weineramer	1.88	1.1
1964	French bulldog	1.7	0.8
1832	Cocker spaniel	1.83	1.11
2956	Australian cattle dog	1.88	1.19
1720	Basenji	1.93	1.31
1673	Japanese Shikoku	1.89	1.1
4710	Rhodesian ridgeback	1.73	0.9

4.8.4 DNA extraction

The same method as described in 3.6.4 was followed for the free-ranging dog samples provided by Eugenia Natoli only.

4.8.5 DNA concentration

The same method as described in 3.6.6 was followed.

4.8.6 Primer design

Primers were designed using the built in primer 3 algorithm (Untergasser *et al.*, 2012) in Geneious (Biomatters Ltd, 2016) to encompass a previously identified SNP mutation present in the *MARCH7* gene, which is located in chromosome 36 (5,499,129-5,531,823 Canfam 3.1). Specific location of the mutation was established in the intronic region at chromosome position 43,900 on Canfam 3.1.4 based on results from Lindblad-Toh *et al.* (2005). The forward primer sequence selected was; CTGCTTAGTGGGGAGTCTGC, the reverse primer sequence selected was AAGGTGGAAGCAGAATGGGG and product size 714 base pairs. Primers had a melting temperature of around 60°C and Hairpin, Self-Dimer and Pair Dimer structures non-existent.

4.8.7 PCR Protocol

1. PCR reaction was prepared in the following proportions (Table 20):

Table 20: PCR reagents and volumes

Reagent	Volume
PCR Master Mix (Thermo Scientific Fisher)	8 µL
DNA	1 µL
Primers (2mM, forward and reverse)	1 µL
Water	6 µL
Total	16 µL

Once prepared, samples were loaded onto the thermal cycler with an initial activation step of 3 minutes at 95°C. This was followed by a 3-step cycling process, firstly denaturation for 30 seconds at 95°C, secondly annealing for 45 seconds at 62°C and finally extension for 45 seconds at 72°C. 38 cycles were used and this was followed by a final extension period of 10 minutes at 72°C.

4.8.8 Agarose Gel Electrophoresis

The same method as described in 3.6.9 was followed.

4.8.9 Purification of PCR Products

To clean PCR products for sequencing Exonuclease 1 (Exo) and Thermosensitive Alkaline Phosphatase (FastAP) (Thermo Fisher Scientific) were utilised.

1. One volume of Exo was mixed with two volume of FastAP to make EXOSAP in 50ul aliquots to use for several sequencing reactions.
2. The EXOSAP mixture was directly added to PCR products at the rate of 1.5ul EXOSAP into a 5ul PCR reaction.
3. PCR products mixed with EXOSAP were incubated at 37°C for 45 minutes, followed by 80°C for 15 minutes then held at 8°C.

4.8.10 DNA Sequencing

After purification PCR products were sent to a sequencing service DBS Genomics at the School of Biological and Biomedical Sciences, Durham University, where Sanger sequencing was completed and sequences returned via email.

4.8.11Analysis of DNA Sequences

Analysis of sequences was completed using the Geneious 7.1.9 package (http://www.geneious.com, Keasrse *et al.* 2012).

4.8.12 Analysis of data from the Dog Genome SNP Database (DoGSD):

DoGSD is an online database (http://dogsd.big.ac.cn/snp/pages/search/search_snp.jsp) containing information on variation in genomes of dogs of different breeds, free-ranging dogs and wolves (Table 20).

A 1000 base-pair region (between positions 5,525,000-5,526,000) in chromosome 36 surrounding the SNP identified in microarray study, positioned at 5,525,355, was searched to assess if that mutation or other mutations are present in these individuals.

4.8.13 Analysis of *MARCH7* polymorphism data present in Ensembl

Ensembl is a genome browser containing vertebrate genomes, found at http://www.ensembl.org/index.html. Data about polymorphisms present in CanFam 3.1 were recorded for comparison to both the *MARCH7* region targeted and data from DoGSD.

4.9 PATTERNS OF NON-SYNONYMOUS VERSUS SYNONYMOUS MUTATIONS IN MAMMALIAN SPECIES

4.9.1 Analysis of signatures of selection

Comparison of the dog *MARCH7* sequence (based on CanFam 3.1 genome assembly) with the orthologous sequences of other members of the Carnivora was completed for all species with sequenced *MARCH7* available on the NCBI database (Table 21).

Table 21: Scientific name. Common name and Family of all Carnivores used in selection analysis

Scientific name	Common name	Family
Canis familiaris	Domestic dog	Canidae
Felis catus	Domestic cat	Felidae
Panthera tigris altaica	Amur tiger	Felidae
Acinonyx jubatus	Cheetah	Felidae
Mustela putorius furo	Domestic ferret	Mustelidae
Odobenus rosmarus divergens	Pacific walrus	Odobenidae
Leptonychotes weddellii	Weddell seal	Phocidae
Ailuropoda melanoleuca	Giant Panda	Ursidae
Ursus maritimus	Polar bear	Ursidae

For comparative analysis and studying signatures of selection when comparing the dog to representatives of other placental mammals, exons of *MARCH7* for the dog were extracted using Geneious and aligned with the orthologous exons of a representative for each family of placental mammals obtained from the NCBI database (Table 22). A total of 37 placental mammals were used.

Table 22: Scientific name, Common name, Family and Order for all placental mammal representatives used

Scientific name	Common name	Family	Order
Aotus nancymaae	Ma's night monkey	Aotidae	Primates
Bison Bison	Bison	Bovidae	Artiodactyla
Callithrix jacchus	White-tufted ear marmoset	Callitrichidae	Primates
Canis familiaris	Domestic dog	Canidae	Carnivora
Saimiri boliviensis boliviensis	Bolivian squirrel monkey	Cebidae	Primates
Rhinopithecus roxellana	Snub nosed monkey	Cercopithecidae	Primates
Microcebus murinus	Gray mouse lemur	Cheirogaleidae	Gray mouse lemur
Chrysochloris asiatica	Cape golden mole	Chrysochlorinae	Afrosoricida
Cricetulus griseus	Chinese hamster	Cricetidae	Rodentia
Galeopterus variegatus	Sunda flying lemur	Cynocephalidae	Dermoptera
Dasypus novemcinctus	Armadillo	Dasypodidae	Cingulata
Tursiops truncates	Bottlenose dolphin	Delphinidae	Artiodactyla
Loxodonta africana	African savanna elephant	Elephantidae	Proboscidea
Equus ferus caballus	Horse	Equidae	Perissodactyla
Felis catus	Domestic cat	Felidae	Carnivora
Panthera tigris altaica	Amur tiger	Felidae	Carnivora
Acinonyx jubatus	Cheetah	Felidae	Carnivora
Otolemur garnettii	Small-eared galago	Galagidae	Primates
Homo sapiens	Human	Hominidae	Primates
Gorilla Gorilla Gorilla	Gorilla	Hominidae	Primates
Nomascus leucogenys	Northern white-cheeked gibbon	Hylobatidae	Primates
Propithecus coquereli	Coquerel's sifaka	Indriidae	Primates

Elephantulus edwardii	Cape elephant shrew	Macroscelididae	Macroscelidea
Mustela putorius furo	Domestic ferret	Mustelidae	Carnivora
Ochotona princeps	American pika	Ochotonidae	Lagomorpha
Odobenus rosmarus divergens	Pacific walrus	Odobenidae	Carnivora
Orycteropus afer afer	Aardvark	Orycteropodidae	Tubulidentata
Leptonychotes weddellii	Weddell seal	Phocidae	Carnivora
Pteropus alecto	Black flying fox	Pteropodidae	Chiroptera
Ceratotherium simum simum	Rhino	Rhinocerotidae	Perissodactyla
Condylura cristata	Star nosed mole	Talpidae	Eulipotyphla
Tarsius syrichta	Tarsier	Tarsiidae	Primates
Echinops telfairi	Small Madagascar hedgehog	Tenrecinae	Afrosoricida
Trichechus manatus latirostris	Florida manatee	Trichechidae	Sirenia
Tupaia belangeri chinensis	Chinese tree shrew	Tupaiidae	Scandentia
Ailuropoda melanoleuca	Giant Panda	Ursidae	Carnivora
Ursus maritimus	Polar bear	Ursidae	Carnivora

4.9.1.1 Pairwise d_N/d_S analysis

Pairwise dN/dS analysis was carried out with the software package DNAsp (DNA Sequence Polymorphism) version 5 (Librado and Rozas, 2009). The ratio of d_N (nonsynonymous) and d_S (synonymous) mutations demonstrates the selective pressure at a protein level (Kryazhimskiy *et al.* 2008). As a result of natural selection acting predominantly on a protein level, the fixation rates of nonsynonymous and synonymous mutations are different and this comparison can reveal information about the direction and strength of natural selection on a protein (Yang, 2007). When ω is less than one there is a fitness advantage to the protein (and thus the individual) through nonsynonymous mutations and as a result their fixation rate is higher than that of synonymous mutations (Yang *et al.* 2000; Kryazhimskiy *et al.* 2008). In the case of genes experiencing diversifying selection, ω would be expected to significantly exceed 1, and ω to be below 1 if the gene is under purifying selection. 1 is the expected value for a gene under neutral selection, thus, deviation from 1 provides information regarding potential selection.

4.9.1.2 Testing for signatures of selection using TOPALi

TOPALi (tree TOPology-related analysis of Alignments Interface) version 2 (Milne et al, 2009) was utilised to run maximum-likelihood analyses of selection, using the the PAML package (Phylogenetic Analysis by Maximum Likelihood; Ziheng Yang 2006) as implemented in the software TOPALi. As part of the analysis, PAML models dN/dS onto a phylogenetic tree, so production of an accurate phylogenetic tree is necessary.

4.9.1.2.1 MrBayes phylogenetic analysis

Topali completes Bayesian tree estimation using MrBayes tree estimation method. In this study a one model approach was used, where MrBayes relies on a single substitution model for the whole alignment (Huelsenbeck and Ronquist, 2001: Ronquist and Huelsenbeck, 2003).

Geneious (7.1.9 package) was also utilised to run MrBayes analysis (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) to construct a phylogenetic tree of *MARCH7* for the placental mammals and compared with previously published phylogenetic mammalian trees for use in conjunction with *ab initio* modelling of protein structure.

4.9.1.2.2 Model selection

Model selection was completed prior to analysis. Tree generation was completed using PhyML, with gamma and invariable sites selected. Model selection was completed assuming protein-coding DNA, resulting in three analyses. AIC2/BIC calculations were completed using sequence length for sample size. All output from model selection can be found in Appendix 7.23.1 and 7.23.2.

4.9.1.2.3 Maximum-likelihood d_N/d_S site model analysis

The site model treats the ɯ ratio for any site (codon) in the gene as a random variable from a statistical distribution, thus allowing ɯ to vary among codons (Nielsen and Yang, 1998; Yang *et al.,* 2000). Positive selection is defined as a presence of some codons at which $ɯ > 1$. A likelihood ratio test (LRT) is constructed to compare a null model that does not allow for any codons with $ɯ > 1$ against a more general model that does allow it.

The testing for positive selection involves comparing a null hypothesis of no selection with the alternative hypothesis of positive selection. In total, three independent

comparison of model pairs was completed (Table 23), based on well described robustness to detect selection (Wong *et al.*, 2004; Yang 2007, 2009).

Table 23: Description and assumption of all three model pairs used

Model	Description	Reference
M0 vs. M3	Simplest model M0 assumes single dN/dS value (ω) across all sites M3 assumes three different ω categories across sites, which can have any value	Goldman and Yang (1994) Yang and Nielsen (2000)
M1a vs. M2a	M1a is a nearly neutral model M1a assumes two classes (ω_0 varying from 0-1 and ω_1 fixed at 1) M2a adds an additional class to M1a, ω_1 which may have any value above 1, and therefore represents positive selection	Wong *et al.* (2004) Yang *et al.* (2005)
M7 vs. M8	M7 assumes a β distribution of ω across sites, where ω can vary between 0-1. M8 allows increased proportion of sites to have a value of ω greater than 1 (ω_2)	Yang *et al.* (2000)

4.9.2 Alignment conservation annotation

Jalview Version 2 (Waterhouse *et al.*, 2009; Troshin *et al.*, 2011) was used for the visualisation of alignment conservation across the placental mammals. Protein structure conservation is automatically calculated and measures the number of physico-chemical properties that are conserved for each column of the alignment. The calculations required for this annotation are based on those used in the AMAS method of multiple sequence alignment analysis in the Livingstone and Barton (1993) study.

4.10 PROTEIN CONFORMATION PATTERNS

4.10.1*Ab Initio* Protein Structure Prediction

Comparison of the *MARCH7* sequence between the dog and other placental mammals revealed a number of non-synonymous mutations in exons. In order to fully assess whether these resulted in changes in protein structure, and if so how they altered ab *initio* protein structure, prediction software was implemented. Protein tertiary structures reveal crucial information for the understanding of the relationship between protein amino acid sequences and their biological functions (Baker and Sali, 2001).

4.10.1.1QUARK software

QUARK provides a computer algorithm for protein structure prediction and *ab initio* protein folding, constructing a protein 3D model from amino acid sequence. Models are built from small fragments (up to 20 residues long) using replica-exchange Monte Carlo simulation through the guidance of an atomic-level knowledge-based force field (Xu and Zhang, 2012). Replica exchange Monte Carlo algorithms are capable of maintaining multiple independent replicas of possible solutions i.e. protein conformations. Every replica is set using a different temperature and locally runs a Markov process sampling from the Boltzmann distribution in energy space (Thachuk *et al.,* 2007). A force field is defined as a mathematical expression that describes the dependence of the energy a system possesses on the coordinates of its particles. This expression consists of an analytical form of U (the interatomic potential energy) combined with a set of defined parameters entering into this form (González, 2011). Exons from selected mammalian species, alongside the dog, were translated into amino acid sequence and submitted to QUARK for processing and model formation.

QUARK was chosen for analysis over other *ab initio* protein structure prediction software programmes due to research demonstrating it producing superior results. In one experiment, a specific score (total Z-score) was 18% higher for QUARK than the second best programme and 47% better than the third best program (Xu and Zhang, 2012). When it was directly compared to Rosetta, another algorithm for ab inito protein structure prediction, it was shown for 145 benchmark proteins (i.e. proteins used on both algorithms) that the template modelling (TM) score by QUARK was 10% higher than Rosetta (Xu and Zhang, 2012).

QUARK is only capable of predicting proteins shorter than 200 amino acids in length. This meant that complete analysis of the entire coding region for *MARCH7* was not possible, as it is known to code for 693 amino acids (Nathan *et al.*, 2008). In order to counter this, individual exons were submitted one at a time for individual species.

QUARK produces a submitted primary sequence, predicted secondary structure, predicted 3-state secondary structure types, predicted starting Beta-turn position, predicted Real-value Phi-angle, predicted Real-value Psi-angle, distance profile from fragments and the clustered torsion angle pairs from fragment, predicted solvent accessibility and predicted tertiary structure. The majority of the analysis ran included the use of predicted tertiary structure and use of the predicted secondary structure.

Initially exon 1 for the dog and cat were submitted and compared, which provided a considerably higher difference in protein structure than expected (Appendix 7.24). Six differences are present in the nucleotide sequence (Figure 7-13), but just one resulting change in amino acid is seen (Figure 7-14). Whilst a single amino acid is capable of resulting in changes to protein structure, a lack of consistently between the models for individual species (i.e. between the ten models produced for the dog) made the reliability of comparison questionable. After this, using Bayesian Inference of Phylogeny, a phylogenetic tree for *MARCH7* was produced and compared to the phylogenetic trees of mammals produced on Ensembl to infer two closely related species (Figure 4-13, Figure 4-14). Comparison of the human and gorilla tertiary structure also revealed differences beyond that expected (Figure 4-17), which led to the additional use of Phyre^2 software.

4.10.2 Phyre^2 software

Despite QUARK'S highly regarded status for ab initio protein structure prediction, complications with analysis resulted in the use of alternative protein structure prediction software. Phyre^2 (Protein Homology/analogy Recognition Engine V 2.0) is a suite of tools accessible online used for the prediction and analysis of protein structure, function and mutations (Kelley *et al.*, 2015).

Phyre also provides the following options: sequence analysis, secondary structure and disorder prediction (example seen in Appendix 7.26), domain analysis (Figure 23), detailed template information, binding site prediction and transmembrane helix prediction.

4.10.3 PDBeFold software

For comparison of QUARK model outputs, of different or the same species, PDBeFold (Krissinel and Henrick, 2004) was used for pairwise comparison and 3D alignment of protein structures. Protein structure comparison service PDBeFold is supplied by the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm).

Two or more models can be layered on top of each other, providing a similarity percentage score, indicating regions of similarity in contrast to those regions which do not align. A similar approach was utilised when comparing models produced from Phyre for different species.

4.11 RESULTS

4.12 PATTERNS OF GENETIC VARIATION OF *MARCH7* IN CANIDS

A targeted region of *MARCH7* gene was successfully sequenced for 35 canids (10 pure-bred dogs, 15 free-ranging dogs and 10 wolves). Sequence quality varied, with the lowest seen in wolf samples. Due to poor quality of some sequencing results and time constraints it was not possible to acquire sequence for all samples available. Polymorphism data from Dog Genome SNP database and Ensembl are also presented for the entire length of the *MARCH7* gene.

4.12.1Sequence analysis for targeted *MARCH7* region

Three polymorphic sites were found in *March7* DNA sequences of pure-breed dogs and wolves (Table 24). No mutations were found for free-ranging dogs.

Among the pure-breed dogs, the Tibetan terrier and Border collie were the most variable, with mutations seen at all three chromosome positions. There was no evidence of the original mutation (at chromosome position 43,900) found in a study designing the canine SNP microarray (Vaysse *et al.*, 2011) and by Pilot *et al*. (2016) based on the SNP microarray data. A table of frequency for all mutations present in *MARCH7* (Table 27) can be seen included in chapter 4.12.4.

Table 24: Sequencing results from Geneious

Chromosome Position	Original	Mutation	Intron/Exon	Samples
43,371	T	T/A	Intron	Tibetan terrier, Bearded collie, Border Collie, Beagle One wolf
43,510	C	C/A	Intron	Tibetan terrier, Border collie Two wolves
43,516	T	T/A	Intron	Tibetan terrier, Border collie Three wolves

4.12.2 Sequence variation for *MARCH7* gene from online resources

4.12.2.1 Dog Genome SNP Database

Out of all the dogs available on the Dog Genome SNP database, 12 dogs showed no variation as compared with the reference dog genome sequence, whilst the remaining dogs had just one SNP mutation present (Table 25). Wolves had two SNPs located in *MARCH7* compared to just one seen in both pure-bred dogs and free-ranging dogs, although the second SNP at chromosome position 44,347 was only seen in a few individuals.

Table 25: Single nucleotide polymorphisms discovered on Dog Genome SNP database

Chromosome Position	Original	Mutation	Intron/Exon	Samples
43,674	CC	CT/TT	Intron	Wolf, Free-ranging dogs, Pure-breed dogs
44,347	CC	CT/TT	Intron	Wolf

Neither of the mutations found were present in any of the samples analysed in the present study via Sanger sequencing, however the SNP found at position 43,674 was outside of the region sequenced. No clear pattern can be witnessed to differentiate free-ranging dogs, wolves and pure-breed dogs. German shepherd dogs all presented with a CT genotype, but otherwise all other groups presented with CT and TT in substitute of CC.

4.12.2.2 Ensembl Variation Database

Three polymorphisms have been labelled on the Ensembl variation database occurring in the MARCH7 gene in CanFam 3.1 (Table 26). All three described occur in exons but are synonymous and occur in a much earlier chromosome position compared to DogSD.

Table 26: Single nucleotide polymorphisms present in MARCH7 from Ensembl database, O = original, M = mutation.

Chromosome Position	O	M	Intron/Exon	Ambiguity code
30,485	A	G	Exon	R
34,926	A	G	Exon	R
35,274	C	T	Exon	Y

4.12.3 Frequency of SNPs

Frequency of polymorphisms from both sequencing and the DogGSD were higher in pure-breed dogs and wolves when compared to free-ranging dogs. The highest frequency for all SNPs was found for chromosome position 43,674 (Table 27).

Table 27: Frequency of single nucleotide polymorphism from sequencing analysis and online databases

SNP	Position	Wolves	Free-ranging dogs	Pure-breed dogs	Number of individuals	Source
C/T						DogGSD
CC	43,674	2.5%	10%	2.5%	79	DogGSD
TT	43,674	2.5%	28%	21.5%	79	DogGSD
CT	43,674	9%	18%	6%	79	DogGSD
C/T						DogGSD
CC	44,347	10%	56%	30%	79	DogGSD
TT	44,347	0%	0%	0%	79	DogGSD
CT	44,347	4%	0%	0%	79	DogGSD
T/A						Sequencing
TT	43,371	26%	43%	17%	35	Sequencing
AA	43,371	0%	0%	0%	35	Sequencing
TA	43,371	3%	0%	11%	35	Sequencing
C/A						Sequencing
CC	43,510	23%	43%	23%	35	Sequencing
AA	43,510	0%	0%	0%	35	Sequencing
CA	43,510	5.5%	0%	5.5%	35	Sequencing
T/A						Sequencing
TT	43,516	20% 7	43%10	23%	35	Sequencing
AA	43,516	0%	0%	0%	35	Sequencing
TA	43,516	8.5% 3	0%	5.5%	35	Sequencing
Original SNP identified						
TT		X				
GT		X	X			
GG				X		

4.12.4 Variation in DNA sequence quality

As previously mentioned, quality of DNA sequence varied amongst samples, with sequences of wolves generally having the poorest quality. Examples of high quality sequence used for can be seen below, representing two wolf and six domestic dog samples.

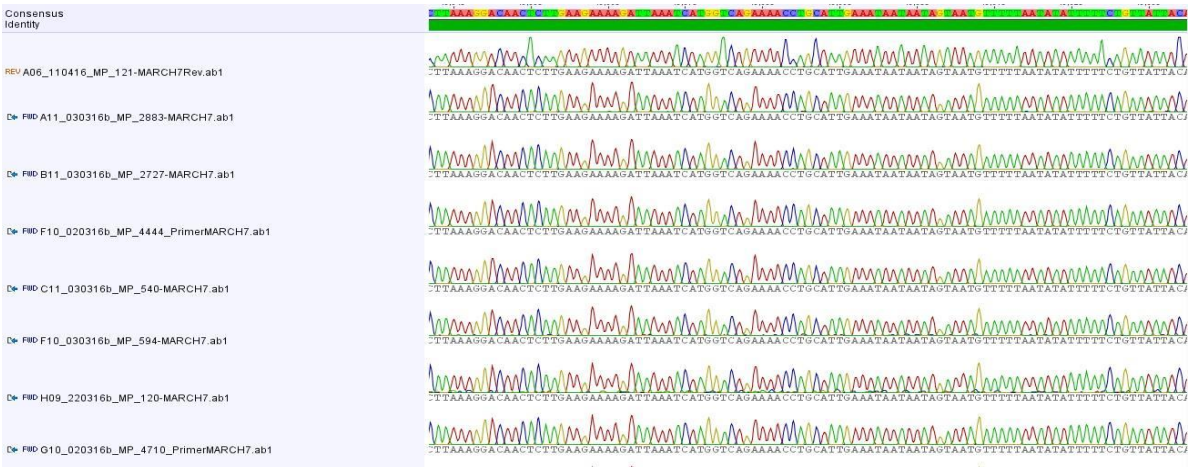


Figure 4-2: Example of high quality sequencing results on Geneious

4.12.4.1 Poor quality sequencing results

Poor sequence quality presented as an issue in a considerable number of wolf and a number of free-ranging dog samples. It should be noted that all samples sent for sequencing had produced clearly distinguishable bands of expected size when ran on a gel, with examples seen in Figure 4-3 for wolves and Figure 4-4 for free-ranging dogs. All samples where no band was seen or only a weak or poorly visible band produced were re-run or discarded from the dataset.

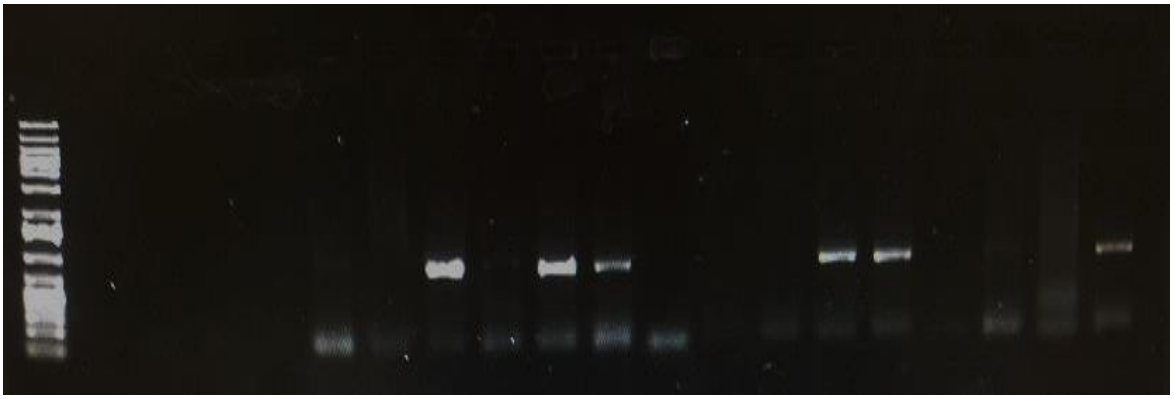


Figure 4-3: Example of gel produced from agarose gel electrophoresis for wolf samples

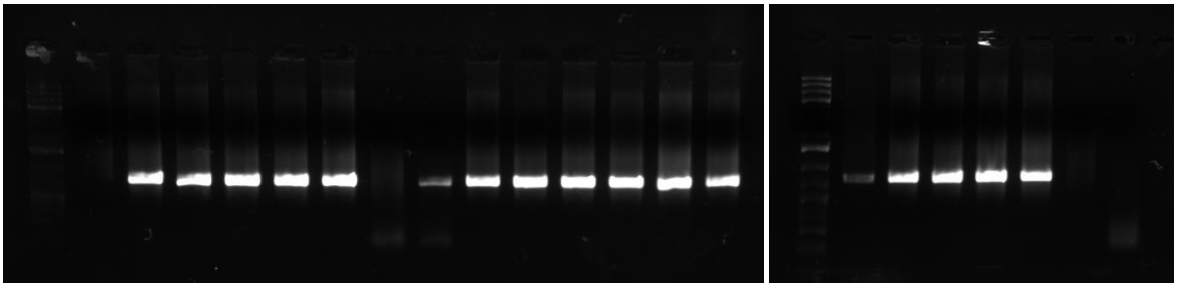


Figure 4-4: Example of gel produced from agarose gel electrophoresis for free-ranging dog samples. The ladder used on the left is not clearly visible but still provides required information about length of product)

There are multiple reasons why DNA sequencing reactions might fail, including: poor quality DNA, loss of sequencing reaction products during purification, bad water, dead sequencing chemistry, too much template DNA, degraded or failed synthesis primer or blocked capillary (Nucleics, 2016). Successful amplification and sequencing of other samples, using the same methodology and carried out simultaneously using the same reagents and instrument make it unlikely to be due to technical failure or human error.

Failed reactions are indicated by N's or by a noisy baseline (Figure 4-6). In the case of most failed reactions, analysed data is not present due to insufficient signal strength and failure to reach the threshold required for analysis (Iowa State University, 2016). Dye terminator peaks (such as those seen at the beginning of the following sequences) can result in the threshold being met by artificially raising the signal strength high enough. This produces analysed data, but in a very unreliable form that is low quality because the base caller is essentially analysing background noise (Iowa State University, 2016).

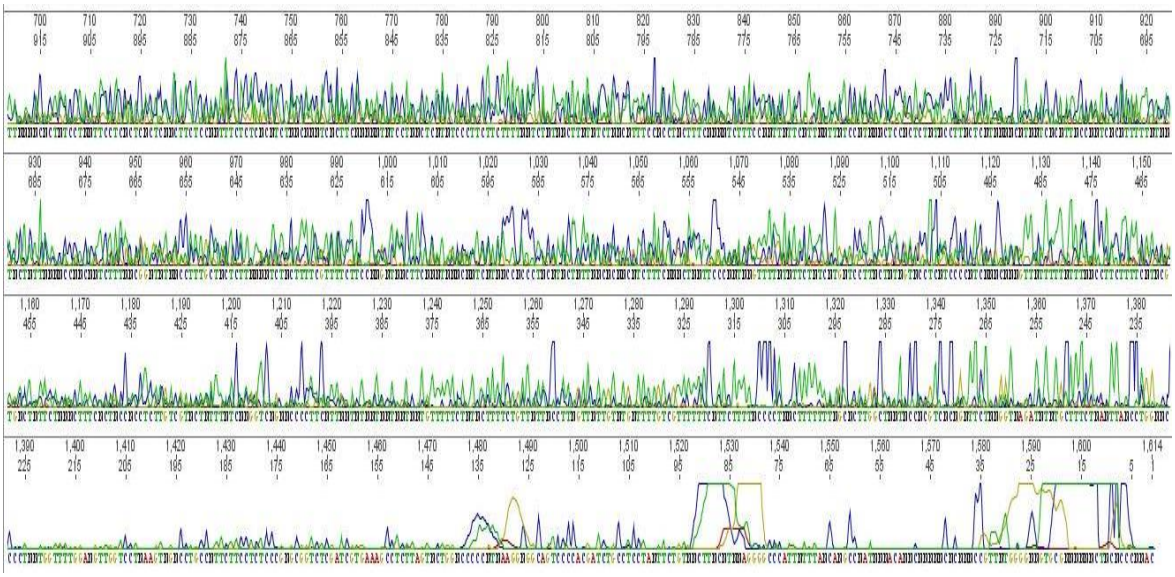


Figure 4-5 Poor sequencing reaction resulting in unreadable sequence due to excessive noise

4.13 PATTERNS OF NON-SYNONYMOUS VERSUS SYNONYMOUS MUTATIONS IN MAMMALIAN SPECIES

Comparative evolutionary analysis of the dog to both the carnivores and placental mammals demonstrates considerable variation across species at the nucleotide level but a high level of conservation at both the amino acid level and protein structure level.

4.13.1 Comparison of the *MARCH7* DNA sequence in the domestic dog to other representatives of the order Carnivora

Alignment of the *MARCH7* DNA sequence for 9 species of Carnivora revealed a total of 70 SNPs, 40 of which being present only in the dog (See Appendix 7.19, Table 83).

4.13.1.1 DNAsp: Pairwise d_N/d_S analysis of dog vs other carnivores

Evaluation of d_N/d_S ratios provided evidence of purifying selection, where all values were consistently below 1 (Figure 4-8). Data output from DNAsp can be found in Appendix7.22.1.1.

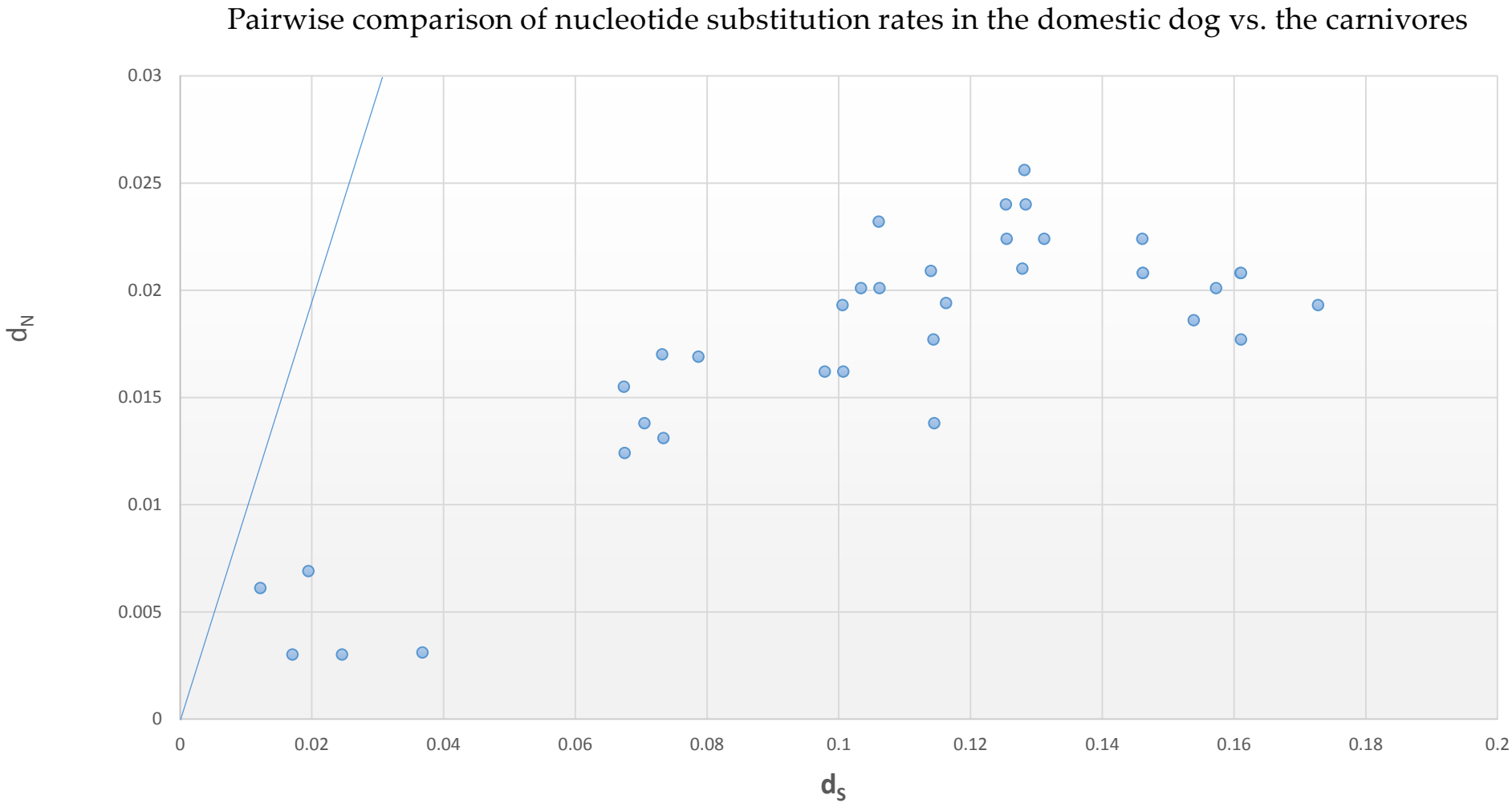


Figure 4-6: Graph showing d_S (synonymous) and d_N (nonsynonymous) values for comparing the dog to other carnivores. The diagonal line indicates dN = dS, meaning neutral selection. Points above the line represent diversifying selection with w = dN / dS > 1. Points below the line indicate purifying selection with w = dN / dS < 1.

4.13.1.2 Topali: PAML results

To detect the particular amino acid sites under diversifying election in the *MARCH7 gene*, three pairs of ML models of codon substitution were applied: M3/M0, M8/M7 and M2/M1 (Nielsen and Yang 1998; Yang *et al.*, 2000).

4.13.1.2.1 Site model analysis

Site model analysis did not support positive selection in *MARCH7*, as none of the three model comparisons were significant. The **w** ratios for M0 and M3 ranged from 0.123 to 0.469, for M1a and M2a from 0.057 to 0.92, which instead suggests purifying selection (Table 28).

Table 28: Site Model Analysis Output for comparison of the dog to other carnivores

Model	ℓ	P_0	P_1	P_2	w_0	w_1	w_2	p	q	df	-2ΔL	Sig	PSS
M0 (one-ratio)	-4163.7				0.123								--
M3 (discrete with 3 categories)	-4155.78	0.73	0.28	0	0	0.461	0.469			4	15.839		
M1a (Nearly Neutral)	-4156.4	0.92	0.08		0.057	1							--
M2a (Positive Selection)	-4156.4	0.92	0.04		0.057	1	1			2	0	NS	
M7 (beta (10 categories)	-4155.94							0.154	1.027				--
M8 (beta&w>1 (11 categories)	-4155.94	1	0				1	0.154	1.027	2	0	NS	

*Where ℓ = log likelihood, P_0, P_1, P_2 = the proportion of sites that are included in the different dN/dS classes included in each model, **ω** = d_N/d_S ratio, B(p, q) is the beta function, df= degrees of freedom, 2 L = Likelihood ratio, Sig = Significance and PSS = Positive selected sites (Nucleic Position, Amino acid).*

4.13.2 Comparison of the domestic dog to representatives of the placental mammals

Amino-acid alignment revealed regions of high conservation. Conservation can be considered as a numerical index measuring the conservation of physico-chemical properties seen in the alignment. Identities are scored the highest then the next most conserved group will contain substitutions to the amino acids, which lie in the same physico-chemical class (Livingstone and Barton, 1993).

Conservation seen in figure 4-9, 4-10 and 4-11 is represented as a histogram giving a score for each column. Conserved columns are represented by a * (a score of 11 with default amino acid property grouping). Columns with mutation where all properties remain conserved are represented by a + (score of 10, indicating property conservation). Figure 4-8 shows there is variation in the levels of conservation, which is a representative view of the majority of *MARCH7* alignment. Figures 4-9 and 4-10 show the region of *MARCH7* which covers the functional motif element, a RING-CH domain located at amino acid positions 554-609 (Nathan *et al.*, 2008). Apart from one amino acid, position 589, all other amino acids conservation of function is 100%.

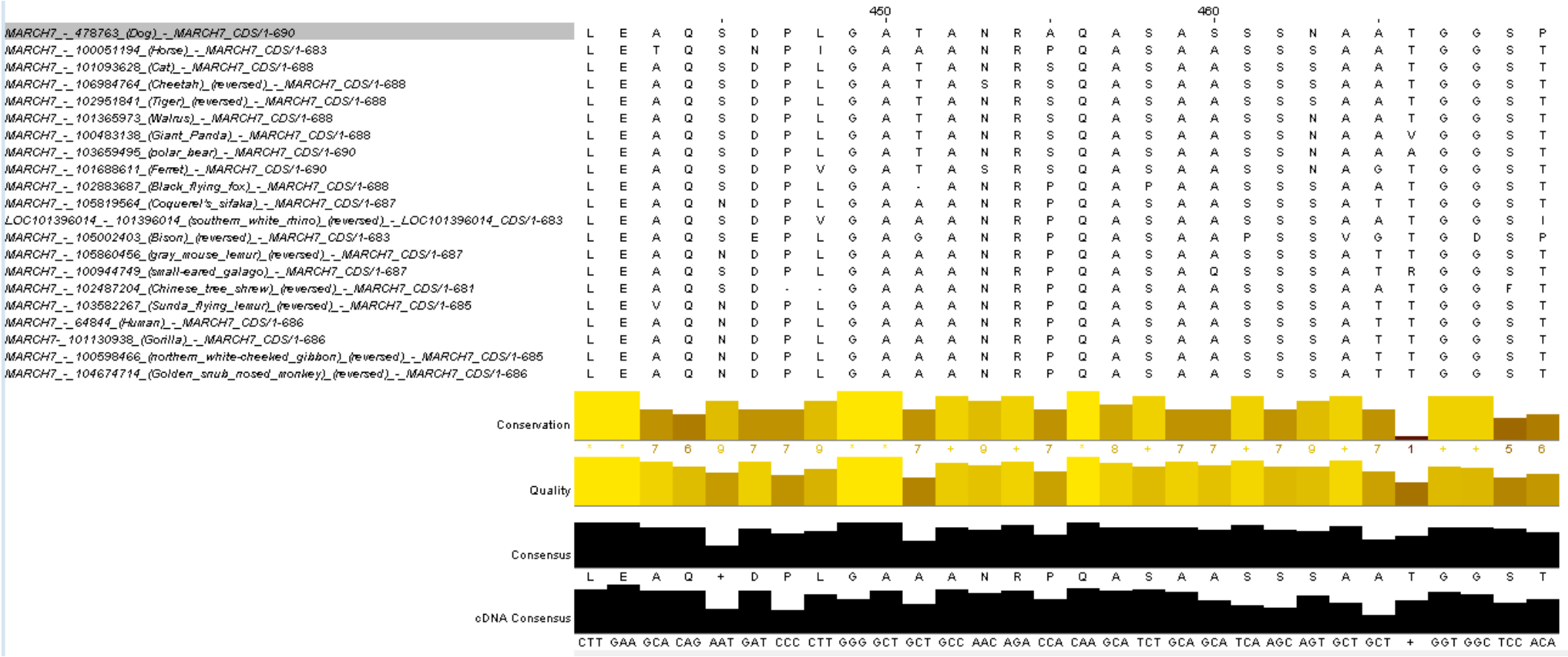


Figure 4-7: Conservation of MARCH7 gene across placental mammal representatives. It is important to note that all species utilised were used to form the histogram represented below but they are not all presented in the side panel of names.

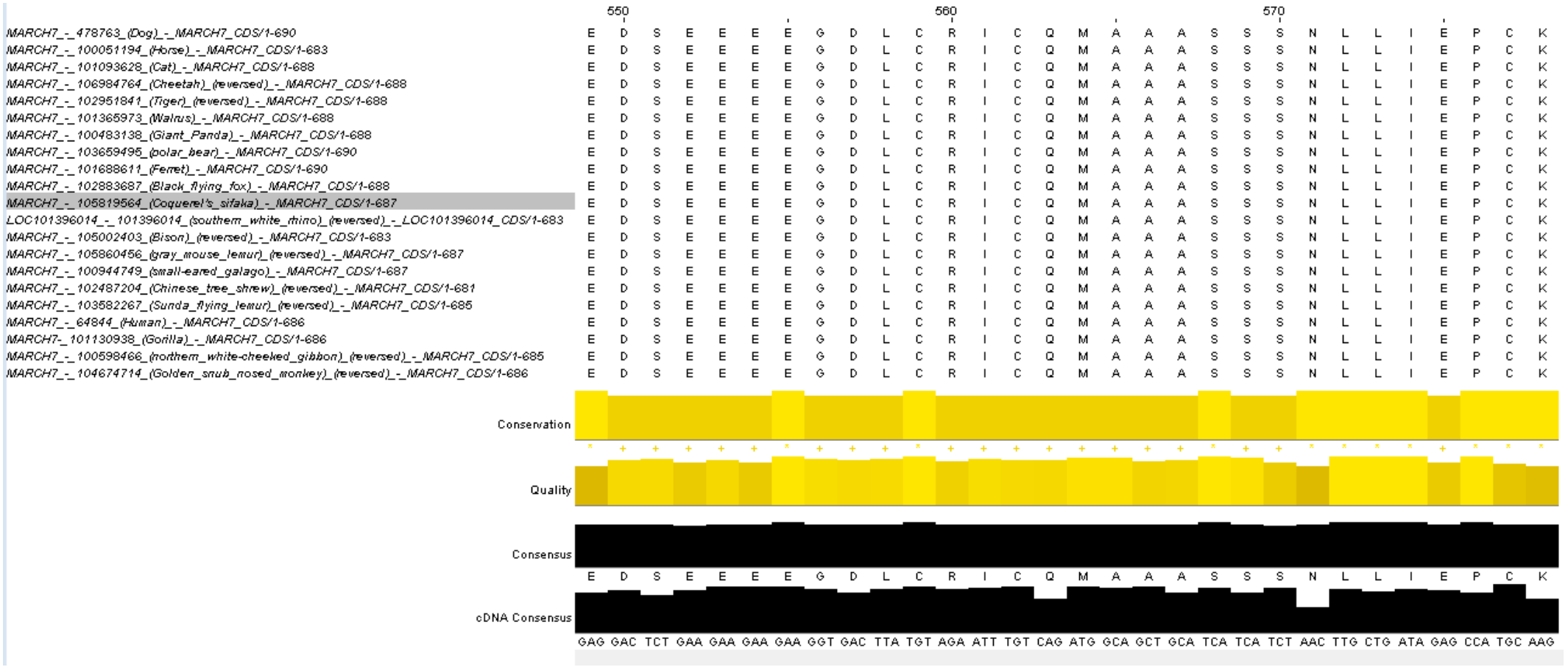


Figure 4-8: Conservation of MARCH7 gene across placental mammal representatives. It is important to note that all species utilised were used to form the histogram represented below but they are not all presented in the side panel of names.

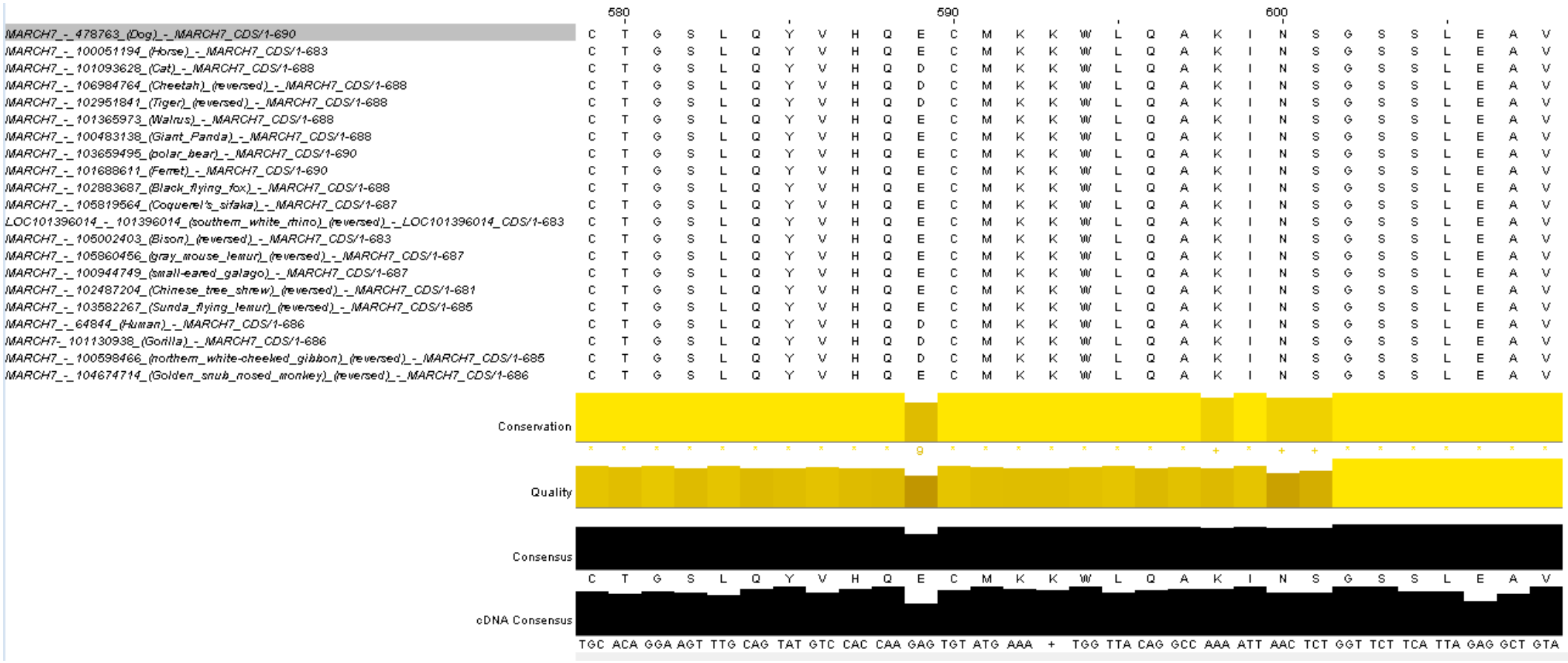


Figure 4-9: Conservation of MARCH7 gene across placental mammal representatives. It is important to note that all species utilised were used to form the histogram represented below but they are not all presented in the side panel of names.

4.13.2.1 Single nucleotide polymorphisms

Comparison of the dog with placental mammals revealed 123 polymorphic sites, ranging from single nucleotide polymorphisms to changes of multiple adjacent base pairs. A total of 14 polymorphic sites were present only in the dog. Of these 14, 11 were SNPs, 1 consisted of a two base pair change and 2 consisted of a three base pair change. Of particular interest were the two cases where a three base pair change occurred, primarily of GTG (dogs) compared to AGA (all other mammals) and TTT (dog) compared to GCC (all other mammals). This prompted analysis at an amino acid level to be completed.

4.13.2.2 Amino acid substitution table

For the 123 SNP positions identified, amino acid translation was completed (Appendix 7.21, Table 85). All amino acids present in other mammalian species were noted in one column with the amino acid present in the dog in the adjacent column, with a further column showing whether this amino acid substitution was only witnessed in the dog when compared to the other mammals. Position of the SNP within the codon (1st, 2nd or 3rd) was recorded (X) to observe differences in synonymous and nonsynonymous rates.

Out of the 123 positions, 29 (24%) were the result of changes to codon position one (CP1), 16 (13%) were the result of changes to codon position two (CP2) 72 (59%) were the result of changes to codon position three (CP3). There were two cases which resulted from changes to CP2 and CP3, and a further three cases which resulted from changes to all three codon positions. (4%)

A total of 29 changes were present just in the dog. Twelve (41%) were the result of changes to CP1, seven (24%) were the result of changes to CP2 and 8 (28%) were the result of changes to CP3. In one case (3.5%) a change to amino acid resulted from all three coding positions and in the last case (3.5%) a change to CP2 and CP3 resulted in a change.

4.13.2.3 DNAsp: MARCH7 d_N and d_S comparison of dog vs mammals

To test for deviation in the substitution pattern of different regions of *MARCH7*, d_N and d_S were calculated for the domestic dog and representatives of the placental mammals. In all cases, d_N was less than d_S ($w = dN / dS < 1$) indicating purifying selection (Figure 4-12) . All data output from DNAsp can be seen in Appendix 7.22.2.1.

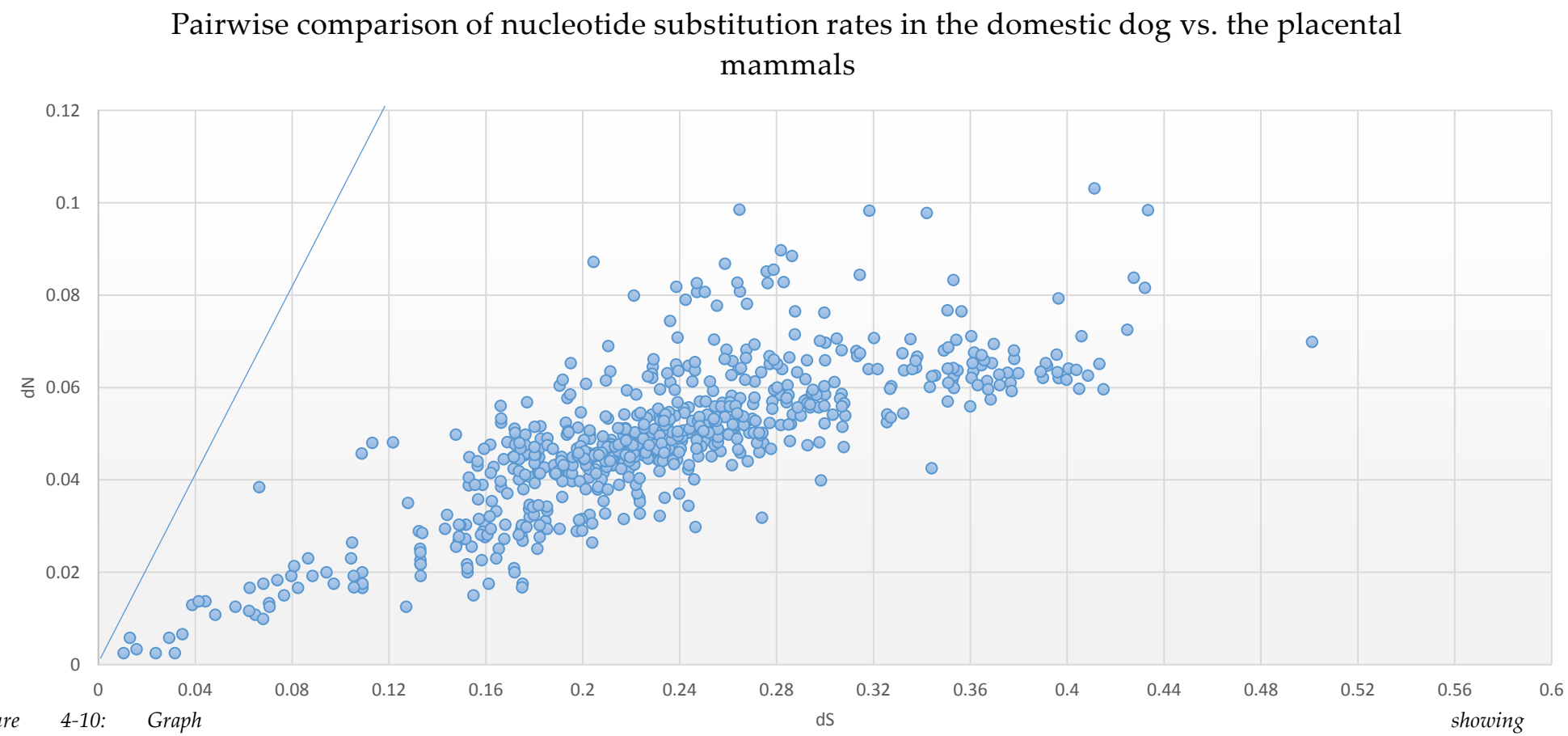


Figure 4-10: Graph showing dS (synonymous and dN (nonsynonymous) values for when comparing the dog to placental mammal representatives. The diagonal line indicates dN = dS, meaning neutral selection. Points above the line represent diversifying selection with w = dN / dS > 1. Points below the line indicate purifying selection with w = dN / dS < 1.

4.13.2.4Topali results

In the same way as used previously for comparison to carnivores, to detect the particular amino acid sites under diversifying selection in the *MARCH7* gene, three pairs of ML models of codon substitution were used: M3/M0, M8/M7 and M2/M1 (Nielsen and Yang 1998; Yang *et al.*, 2000).

4.13.2.4.1Site model analysis

Site model analysis comparing the dog to placental mammal's demonstrated purifying selection. Models M3 and M8 were found to be significant, but M2a was non-significant.

Table 29: Site model analysis comparing dog to placental mammals

Model	ℓ	P_0	P_1	P_2	w_0	w_1	w_2	p	q	df	-2ΔL	Sig	PSS
M0	-13770				0.1								--
M3	-13513.72	0.710	0.290	0.000	0.038	0.495	54.033			4	512.981	<0.001	
M1a	-13553.78	0.854	0.146		0.085	1.000							--
M2a	-13553.41	0.854	0.001		0.085	1.000	3.206			2	0.734	NS	
M7	-13497.25							0.290	1.342				
M8	-13491.57	0.984	0.016				1.591	0.335	1.787	2	11.346	<0.01	

Where ℓ = log likelihood, P_0, P_1, P_2 = the proportion of sites that are included in the different dN/dS classes included in each model, ω = d_N/d_S ratio, B(p, q) is the beta function, df= degrees of freedom, 2 L = Likelihood ratio, Sig = Significance and PSS = Positive selected sites (Nucleic Position, Amino acid).

4.14 PROTEIN CONFORMATION PATTERNS

4.14.1 Protein structure prediction

To ascertain whether or not mutations present in *MARCH7* exons resulted in considerable changes to protein structure, structural prediction of protein structure was completed using a combination of QUARK, Phyre2 and PDBe Fold software packages. Complications arising from a highly disordered structure of *MARCH7* (Nathan *et al.*, 2008) meant that producing reliable models was difficult. Although initially QUARK was chosen due to researching showing it be superior to other prediction software packages (Xu and Zhang, 2012), Phyre provided output relating to percentage of disorder, revealing the extent of disorder present in the *MARCH7* gene. A general overview of the output of both programmes is provided below, which are further discussed in 4.15.3.

4.14.1.1 QUARK results

A large number of results were produced from QUARK, with key outputs shown below for one example (Dog Exon 1).

4.14.1.1.1 Predicted secondary structure

QUARKs prediction of secondary structure is displayed in the following format;

```
>C-coil;H-helix;E-sheet;T-beta turn
MESKPSRIPRRISVQPSSSVSARMMSGSRGNSLNDTYHSRDSSFRLDSEYQ
CCCCTTTTTTTTEECCCCCCHHHHTTTTTTTTTTTTTTTTTTCTTTTC
12345678901234567890123456789012345678901234567890
--------10--------20--------30--------40--------50
```

Figure 4-11: Predicted secondary sequence of Exon 1 Dog

4.14.1.1.2 Predicted tertiary structure

QUARK produces 10 models for each sequence provided, with Model 1 representing the most robust model. For all further analysis and comparisons model 1 was used for all species. On the QUARK server is it possible to interact with each model using Jsmol (JavaScript-Based Molecular Viewer).

As can be seen in Figure 0-212 there is clear differentiation between the models predicted below for exon 1 in the dog which bought in to question the reliability of QUARK for accurate model prediction as models are expected to be similar.

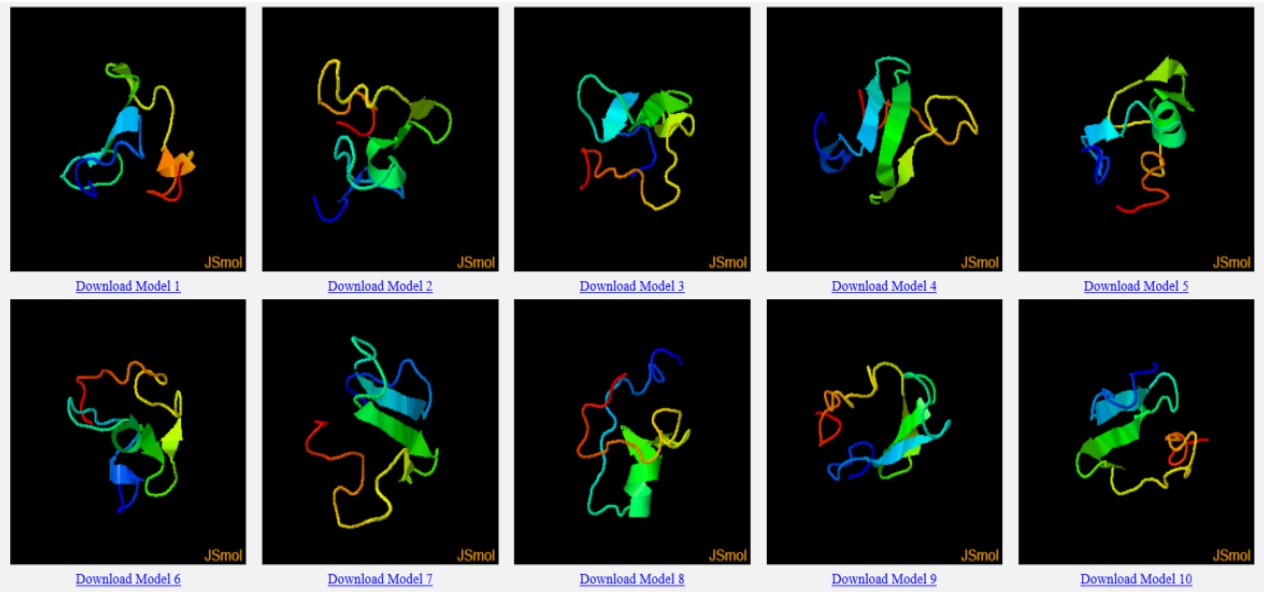


Figure 4-12: Predicted models of tertiary structure for Dog Exon 1with estimated TM scores for Model 1 and best top 10 model predictions. Estimated TM-score of Model 1: 0.3096 ±0.0833. Estimated TM-score of the Best of Top 10 Model: 0.3548 ±0.0764

4.14.1.2 Phyre results

Due to complications with QUARK, Phyre was used to predict the entire structure of *MARCH7* rather than a single exon.

Analysis using Phyre revealed a highly disordered structure of MARCH7. 84% of *MARCH7* gene in the human and 85% of *MARCH7* in the dog were predicted to be disordered and just 9% of amino acids could be modelled with confidence. An example output can be seen in Appendix 7.26.

4.14.1.3 PDBe Fold

PDBe fold was used to superimpose two predicted tertiary models from QUARK or Phyre of different species to provide a comparison of structural prediction effectively.

4.14.1.3.1 Comparison of human and gorilla

Human and gorilla were chosen for comparison due to their close phylogenetic proximity; as such predicted structures were expected to be highly homologous. Initial results from comparison of exon 1 revealed a low similarity between the two species (Table 32), regardless of just one change in amino acid seen at sequence level (Figure 4-15), raising questions about the reliability of QUARKs model prediction for use in modelling *MARCH7*.

Table 30: PDBe fold output for comparison of Human and Gorilla Exon 1

Human				Alignment (1 of 1)				Gorilla			
N_{res}	$\%_{res}$	N_{SSE}	$\%_{SSE}$	Q	P	RMSD	N_{align}	N_{res}	$\%_{res}$	N_{SSE}	$\%_{SSE}$
51	41	3	33	0.0922	0.82	2.749	21	51	41	2	50

$\%_{seq}$	Z	N_{SSE}	N_{gaps}
19.0	3.58	1	5

Where Q= quality function of C_a-alignment, P = minus logarithm of P-value (probability of achieving the same of better quality of match at a chance), Z = statistical significance of a match in terms of Gaussian statistics, RMSD = Root Mean Square Deviation, calculated between C_a- atoms of matches residues at best 3D superposition of the query and target structures, N_{algn} = length of N_{algn}, or number of matched residues, $\%_{seq}$ = sequence identity is a quality of characteristics of C_a-alignment, N_{res} = the size of target chain (expressed in number of residues), $\%_{SSE}$ = the perfect of matched SSEs (what fraction of Secondary structures of target chain was identified in the query).

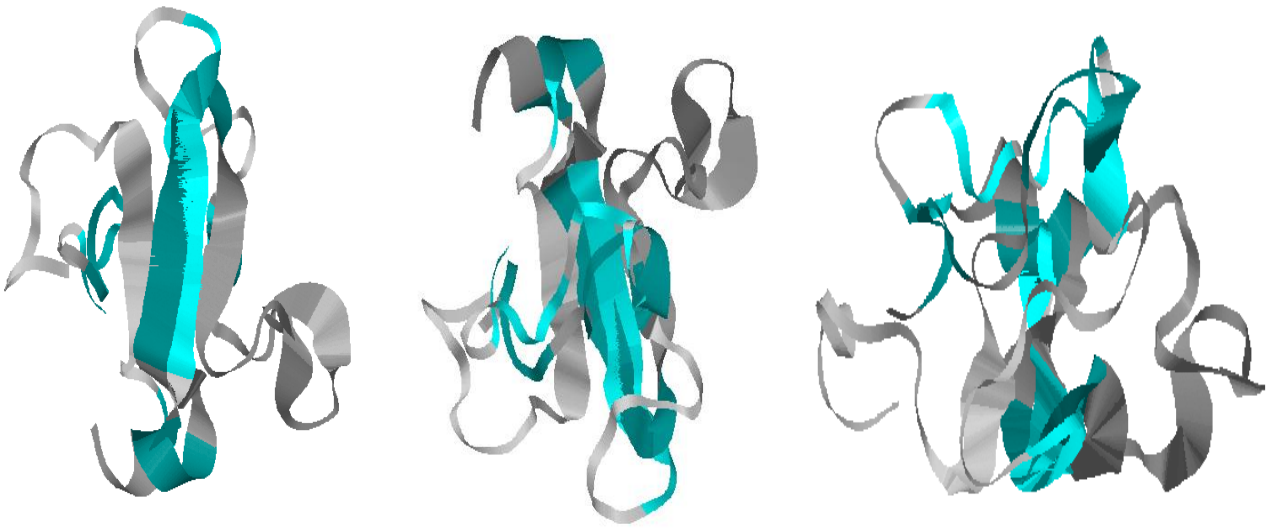


Figure 4-13: Three alternative angles of superimposed 3D predictions of Human (Grey) and Gorilla (Blue) Exon. Regions in blue demonstrate similarity between human and gorilla predicted structures, which in this case is low.

4.14.1.3.2Comparison of human and dog

Information from Phyre regarding protein disorder (Appendix 7.26) led to the analysis and comparison of the RING-CH domain in human and dog, which revealed considerably more accurate tertiary structure prediction from QUARK and a much higher similarity of structure (as predicted from high conservation of RING-CH domain witnessed when studying *MARCH7* sequence). A 9.7% difference in sequence identity is observed (Table 33) between species (indicated by the grey regions in Figure 4-16).

Table 31: PDBe fold output for comparison of Human and Dog RING-CH domain

Human				Alignment (1 of 1)				Dog			
N_{res}	$\%_{res}$	N_{SSE}	$\%_{SSE}$	Q	P	RMSD	N_{align}	N_{res}	$\%_{res}$	N_{SSE}	$\%_{SSE}$
73	99	1	100	0.918	3.37	0.736	72	73	99	1	100

$\%_{seq}$	Z	N_{SSE}	N_{gaps}
90.3	5.24	1	1

Where Q= quality function of C_a-alignment, P = minus logarithm of P-value (probability of achieving the same of better quality of match at a chance), Z = statistical significance of a match in terms of Gaussian statistics, RMSD = Root Mean Square Deviation, calculated between C_a- atoms of matches residues at best 3D superposition of the query and target structures, N_{algn} = length of N_{algn}, or number of matched residues, $\%_{seq}$ = sequence identity is a quality of characteristics of C_a-alignment, N_{res} = the size of target chain (expressed in number of residues), $\%_{SSE}$ = the perfect of matched SSEs (what fraction of Secondary structures of target chain was identified in the query).

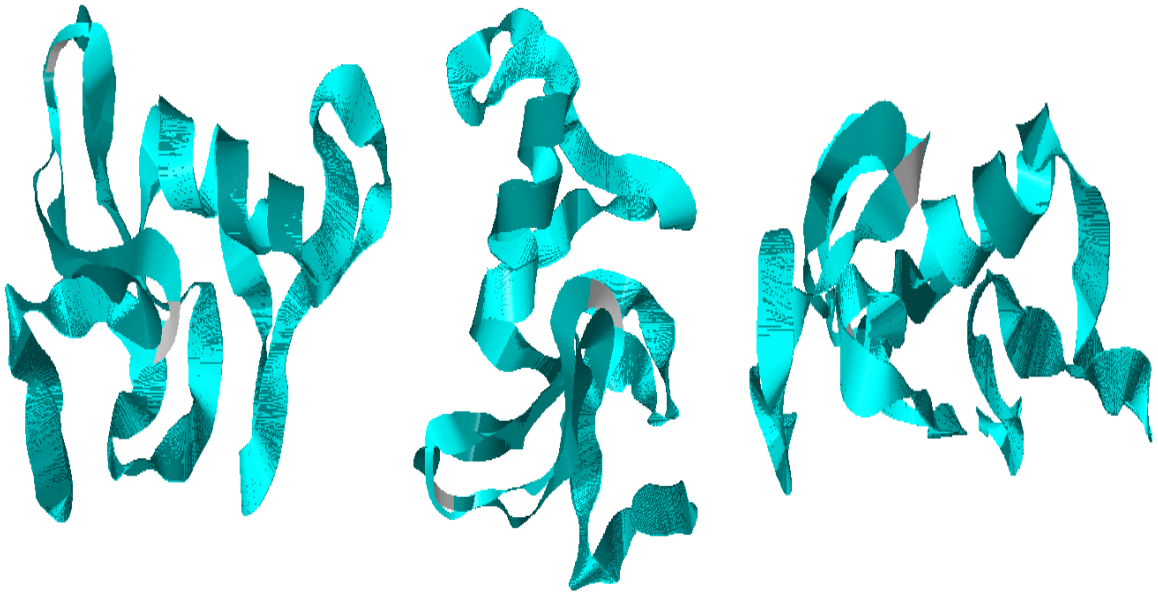


Figure 4-14: Three alternative angles of superimposed 3D predictions of Human (Grey) and Dog (Blue) RING-CH region. Regions in blue demonstrate similarity between human and dog predicted structures, which in this case is high.

4.15 DISCUSSION

Overall the sequencing results from the targeted region showed patterns of genetic variation in *MARCH7* to be highest in wolves and pure-breed dogs and lowest in free-ranging dogs. Comparison of the domestic dog to other carnivores demonstrated 57.14% of SNPs to be present in just the dog, and further comparison to a range of placental mammals revealed 14% of SNPS present just in the domestic dog. Studying patterns of nonsynonymous and synonymous mutations across mammalian species demonstrated evidence for purifying selection in *MARCH7* and high conservation of function in the RING-CH domain across specie (Figure 4-9, 4-10). Patterns of protein conformation indicated a highly disordered structure of *MARCH7* but high conservation particularly across the RING-CH domain.

4.15.1 Patterns of genetic variation in *MARCH7* in canids

Sequencing results combined with data from the Dog Genome SNP database show that pure-breed dogs and wolves have a higher number of mutations present in *MARCH7* than free-ranging dogs. Out of five polymorphic sites discovered, variation in free-ranging dogs was only exhibited at one site, whilst pure-breed dogs showed variation at four and wolves at all five positions. A further three polymorphisms were present in pure breed dogs on the Ensembl database, suggesting that variation in *MARCH7* is highest in pure-breed dogs, but it should be noted that there is currently no sequenced or annotated *MARCH7* gene available on Ensembl for the gray wolf or free-ranging dogs so a direct comparison is not possible.

Sequence analysis of the targeted region of *MARCH7* gene revealed single nucleotide polymorphisms (SNPs) present in the intronic regions in domestic dogs and wolves, but no polymorphisms in any of the free-ranging dogs. Out of the thirty pure-bred dogs selected for this study, mutations were present in four breeds (40%) for one mutation and two breeds (20%) for further two mutations. The Tibetan terrier and Border collie represented the most divergent of all breeds used; having mutations present at all three chromosomal positions (Table 24).

In contrast to results found by Pilot *et al.* (2016) (described in chapter 2.4.1) the sequencing data in this study revealed no evidence of genetic variation at the same base pair position. Although they noted a TT genotype for both East Asian dog breeds and European dog

breeds, which was corroborated by results for pure-breed dogs in this study, Pilot *et al.* (2016) also observed differences in wolves and free-ranging dogs whereas, here, individuals from both these groups all had a TT genotype. Issues with sequencing caused complications concerning the sequence analysis for free-breeding dog and wolf samples. Quality at the beginning of the sequence was particularly poor. Poor sequence quality combined with the lack of evidence supporting the SNP microarray data indicate a possibility that variation in the *MARCH7* gene could result from segmental duplication or copy number variation (CNV). Segmental duplications can result in issues with sequencing, and sequence quality (Treangen and Salzberg, 2011).

4.15.1.1Copy number variation in MARCH7

CNVs are defined as DNA segments of variable length, up to several megabases (Mb) that vary in copy number in comparison to a reference genome (Molin *et al.*, 2014). Differing types of CNVs include deletions and duplications. Phenotypic effects of CNVs result from altered gene dosage and regulation, changes in gene structure, changes in gene expression, unmasking of recessive alleles, indirectly through position effect or downstream pathways and regulatory networks (Li *et al.*, 2014; Molin *et al.*, 2014). Deletions and duplications can result in significant effects on various phenotypic traits, including some breed-defining traits (Nicholas *et al.*, 2011). For example, duplication of a three fibroblast growth factor (FGF) genes is associated with the dorsal hair ridge in Rhodesian and Thai Ridgebacks dogs (Salmon Hillbertz *et al.*,. 2007).

Gene duplication provides raw genetic materials for functional and structural modifications whilst still conserving parental function (Acharya *et al.*, 2015). It has long been recognised that gene duplication is a major driving force behind shaping organism and genome evolution (Ohno *et al.*, 1968, Stephens, 1951; Ba *et al.*, 2014). Duplication events can involve whole genomes, individual genes or genomic segments (Berglund *et al.*, 2012; Pasek and Górecki, 2016). Two or multiple gene copies in a genome can provide a "back-up" mechanism to allow organisms to remain phenotypically stable under varying environmental, genetic or stochastic perturbations (Espinosa-Cantú *et al.*, 2015; Gu *et al.*, 2003; Wagner, 2005). Nicholas *et al.* (2009) estimated that segmental duplications comprise ~4.21% of the canine genome. Genomic rearrangements of this kind can be representative of polymorphisms that are functionally neutral or convey phenotypes through diverse mechanisms, including deletions, insertions, variation in copy number or

dosage-sensitive genes, the production of fusion genes and other mechanisms (Figure 4-17) (Lupsi and Stankiewicz, 2005; Gu *et al.*, 2008).

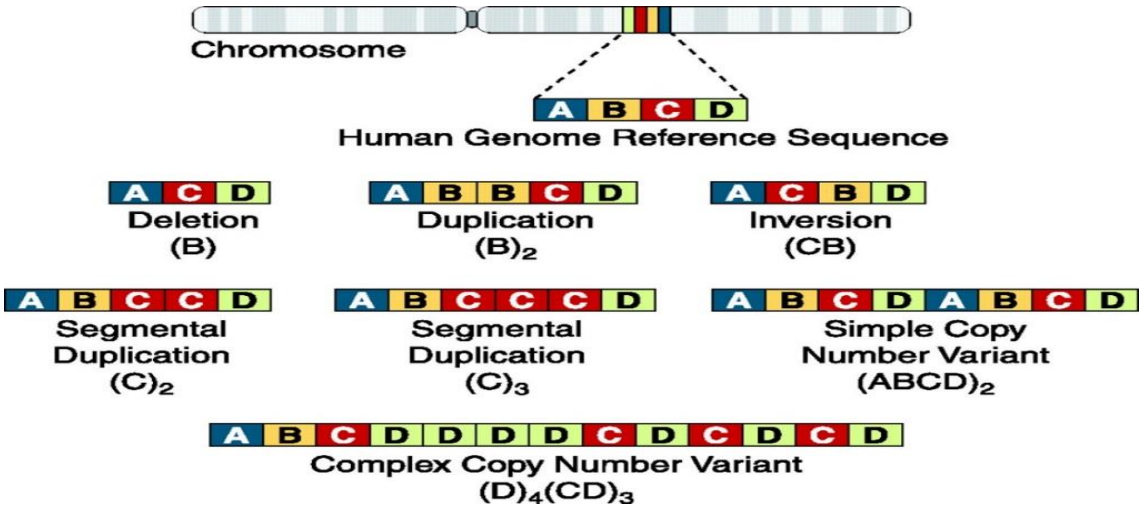


Figure 4-15: ***Main mechanisms that lead to CNV changes****._Non-homologous recombination between sequences with a high level of identity (segmental duplications, low copy repeats, or duplicons) (over 90%) might cause the duplication or deletion of genetic material. Depending on the peculiarities of genes involved in the rearrangements, a clinical phenotype could be observed. Inverted duplicons might cause an inversion of the genetic material, but other types of changes could occur depending on the complexity of the duplicons, which often contain sequences that are in a parallel orientation, in which case deletions or duplications can result. Sourced from Dierssen et al, 2009.*

CNV dispersion is affected by numerous genetic mechanisms (Hastings *et al.*, 2009). Typically it is thought that there are three main classes of mutational mechanisms (Alvarez and Akey, 2012), with the most common known to be non-allelic homologous recombination (NAHR), where during meiosis and mitosis misalignment and crossover occurs between regions of extended homology (Ramirez *et al.*, 2014). The remaining two mutational mechanisms include the fork-stalling and template-switching (FoSTeS) model proposed by Slack *et al.* (2006), which was extended to a more general replicative template-switching model that is referred to as microhomology-mediated break-induced replication (MMBIR) by Hastings *et al.* (2009). Finally, retrotransposition through an RNA-mediated process can lead to the development of CNVs.

4.15.1.1.1 Copy number variation in relation to immunity

Genes involved in immune response (adaptive and innate), chemosensation, fertility and reproduction are commonly affected by CNVs and this phenomenon is seen across many

mammalian genomes (Wolfe *et al.,* 2003; Emes *et al.,* 2003; Lindblad-Toh *et al.,* 2005). CNVs are known to be important in allowing adaptation to novel environments (Nguyen *et al.,* 2006; de Smith *et al.,* 2009). As a result, CNVs may provide a substantial proportion of the genetic variability, which is the substrate of natural selection, increasing genetic plasticity to allow organisms to evolve rapidly to external pressures and thus, playing a fundamental role in their adaptability and fitness (Feuk *et al.,* 2006; de Smith *et al.,* 2009).

CNV in *MARCH7* could allow for an increased level of genetic plasticity, in wolves and free-ranging dogs. In pure-breed dogs the pressure on the immune system to continually adapt may be relaxed in comparison to free-ranging dogs and wolves. Further investigation would be required to ascertain whether CNV is occurring in *MARCH7* and this is discussed in 5.3.2.2.

4.15.1.2Alternative splicing

The polymorphisms in the *MARCH7* gene discovered here in multiple individuals, whilst in the intronic region, were close to the intron/exon border and as such could affect crucial regulatory functions or splicing. Splicing of precursor mRNA (pre-mRNA) is a crucial regulatory stage involved in the pathways related to gene expression, which involves the removal of introns and ligation of exons to produce mRNA (Keren *et al.,* 2010). Alternative splicing (AS) relies on alternative use of exons, promoters, introns and polyadenylation sites to produce variation within mRNA's from individual genes that vastly increases the diversity of transcripts that these genes can express (Ward and Cooper, 2010). Three sites, known as the core splicing signals, are involved in every splicing reaction and can be found in all introns. These are the branch point sequence, the 3' splice site and the 5' splice site (Wang and Burge, 2008). Alternative splicing of pre-mRNA's is known to influence the control of gene expression levels and proteomic diversity (Wang and Burge, 2008) affecting nearly 95% of mammalian genes (Kornblihtt *et al.,* 2013). AS is also known for playing a crucial role in differentiation, development and disease (Luco *et al.,* 2011) and is tightly regulated in different tissues and developmental stages, as disruption to AS can lead to a wide range of diseases (Wang and Burge, 2008).

Alternatively spliced exons have unusually low rates of evolution at synonymous sites (Lida and Akashi, 2000; Xing and Lee, 2006). Combining this with evidence that synonymous rates of evolution can be especially low in exonic domains associated with

splice control (Hurst and Pal, 2001; Obran and Olah, 2001), has led to the understanding that most selection on synonymous mutations in mammals is associated with perturbation of splicing (Parmley and Hurst, 2007).

Generation of antigen receptor diversity expressed by T and B cells is one advantageous use of alternative splicing in the immune system (Hozumi and Tonegawa, 1976; Tonegawa *et al.,* 1974). An essential function of the immune system is ability to differentiate between non-self and self in order to provide protection against autoimmunity as well as mounting an effective immune response to protect from disease. A balance is required between removal of autoreactive B and T cells (during their development) that express self-reactive receptors and the ability to maintain an adequately diverse repertoire of lymphocytes to enable response to any pathogen (Yabas *et al.,* 2016). *MARCH7* is known to be a regulator of T lymphocytes (Gao *et al.,* 2009) so alternative splicing may allow activation of T lymphocytes in response to a greater range of pathogens.

4.15.2 Patterns of non-synonymous and synonymous mutations in mammalian species

Animals inhabit a variable and diverse environment and as a result their immune systems must be able to successfully interact with and respond to equally diverse immunobiome (the specific set of components that generate evolutionary and ecological selective forces on the immune system) (Horrocks *et al.,* 2011). Using sequence alignments from diverse mammalian taxa can be beneficial in identifying conserved regions of proteins with low rates of amino acid substitution, which are subject to strong purifying selection. Such regions can be interpreted to have important function (Springer and Murphy, 2007).

When considering overall genetic variation in *MARCH7* across all species, differences can be seen at the nucleotide level in differing positions. This is likely to be linked to different evolutionary routes and selection pressures on individual species. Hwang and Green (2004) provided evidence that cytosine deamination, biased gene conversion and context-dependant DNA replication errors are a large explanation for naturally occurring SNPs. They explain that their relative contribution throughout mammalian evolution has varied due to different generation times, recombination rates and effective population sizes. They also note that C-G transitions have accumulated in a clock-like fashion when compared to other context-dependant substitution types. There is evidence to support the

widespread selection pressure on the nucleotide level in eukaryotic genomes and demonstration of the importance of synonymous positions for regulation of translation and alternative splicing (Chamary *et al.,* 2006; Cartegni *et al.,* 2002; Fairbrother *et al.,* 2004). These observations support the theory that synonymous positions may be under selection and codon bias is maintained by a balance between selection, mutation and genetic drift (Bulmer, 1991; Duret, 2002).

Carnivores represent the most ecologically diverse order inside the mammalian class, which is exemplified by their body size spanning more than three orders of magnitude (Christiansen and Wroe, 2007). Within the Caniformia, the Canidae are known to be the most ancient lineage amongst all living families (Wang *et al.,* 2004). As a result a certain level of naturally occuring variation would be expected to occur in *MARCH7*. A total of 70 mutations were detected in *MARCH7* in the dog when aligned with the carnivores, with over half (57.14%) occurring exclusively in the dog (Table 83, Appendix 7.19). Synonymous and non-synonymous mutations both have implications for the functionality of *MARCH7*. SNPs can result in significant changes to the function and structure of mRNA, through processes such as mutagenesis and splicing errors, and these provide vast possibilities when considering gene expression regulation (Shabalina *et al.,* 2013). Translational selection is responsible for the unequal usage of synonymous codons in protein coding genes in a wide variety of organisms (dos Reis *et al.,* 2004).

Non-synonymous SNPs that lead to an amino acid change in the protein product are of major interest, because amino acid substitutions are known to result in numerous inherited diseases (Ng and Henikoff, 2003; Krawczak *et al.,* 2000). Out of total 123 amino acid substitutions present in the dog when compared to placental mammal species, 29 (24%) nonsynonymous substitutions were exclusively present in the dog. Due to time constraints it was not possible to calculate the same percentage for all species utilised, so direct comparison is not possible. Differentiation in evolutionary patterns may be caused by variation in selection and mutation across a sequence. When considering codons, different coding positions (CPs) are constrained evolutionarily to varying degrees due to the functional constraints imposed on them by the genetic code and the physicochemical properties of amino acids (Bofkin and Goldman, 2007). Out of the three CPs, the most functionally constrained is the second, i.e. any change to this position results in a nonsynonymous change in coding sequence. Comparatively, the least constrained

position is the third (Bokfin and Goldman, 2007). This is generally explained by assuming nucleotides substitutions at the second CP are nonsynonymous and influenced by strong purifying selection, whilst substitutions are the third CP and a proportion of those at the first CP are synonymous and receive weaker purifying selection and thus will evolve faster (Xia, 1998).

4.15.2.1 MARCH7 under purifying selection in mammalian species

Comparison of d_N and d_S was completed to test for deviation in the substitution pattern of different regions of *MARCH7*. Comparison of the dog to the carnivores, and to the placental mammals both showed w to be less than one, indicating that *MARCH7* is undergoing purifying selection. Purifying, sometimes termed negative, selection is one of the major reasons that orthologous sequences remain well conserved between species (Hardison *et al.,* 2003; Ellegren *et al.,* 2003). Thus, these sequences that are considerably more similar than expected under the model of neutral evolution are likely to be serving crucial functional roles (Siepel *et al.,* 2005).

Site model analysis for both comparisons to the carnivores and to placental mammal's representatives provided further support for *MARCH7* being under purifying selection. For example, no differentiation between model 2a (positive selection) and model 1a (nearly neutral) was found when comparing the dog to the carnivores (**Error! Reference source not found.**).

4.15.2.2 Conservation of MARCH7 sequence

MARCH7 gene sequence has been shown to have 85% homology between mouse and human, with an identical RING-CH domain (Nathan *et al.,* 2008). The results from sequence alignment in the present study confer with this, as the RING-CH domain is also witnessed to be the most conversed region of the gene. No functional amino acid changes in the RING-CH domain occur throughout placental mammals and all the carnivores, even if nucleotide substitutions are present (Figure 4-9, 4-10, 4-11). This demonstrates similarities to the research of *BRCA1* in primates. *BRCA1* is a tumour suppressor gene implicated in transcription, DNA damage control and cell cycle regulations (Rosen *et al.* 2013) and analysis indicates the BRCT domains and terminal RING as being the most conversed regions of the protein (Pavlicek *et al.,* 2004). The RING-CH domain of RING-type ubiquitin ligases is known to be required for ubiquitination (Fujita *et al.,* 2013),

explaining the strong purifying selection pressure acting on it to keep it identical amongst species.

4.15.3 Protein conformation patterns for non-synonymous mutations

Ab initio modelling of protein structure using QUARK presented a multitude of complications. Comparison of the Gorilla with the Human revealed stark differences in the models predicted through QUARK for these two species, even though there is just one amino acid difference in sequence and no changes in the predicted secondary structure (Appendix 7.25, Figure 7-18, 7-19). These results prompted questions about the reliability of results from the models predicted in QUARK, so analysis was completed to ascertain the level of protein disorder present in *MARCH7*. A change of just one amino acid in a sequence can result in direct changes to protein structure, and here we see a difference between cysteine (Gorilla) and phenylalanine (human). Cysteine (Cys) residues are known for their involvement in intra- and intermolecular disulphide bonds and for particular enzymes can form part of the catalytic activity site. Presence of disulphide bonds in proteins and peptides has been demonstrated to impose conformational rigidity and during the folding process formatting of non-native intramolecular disulphide bones can result in protein misfolding, leading to precipitation and aggregation (Trivedit *et al.*, 2009). In contrast, phenylalanine side chains are non-reactive and rarely involved directly in protein function, although it can play a role in substrate recognition (Betts and Russell, 2003). Clear differences in the properties of cysteine and phenylalanine are described and it can be noted that whilst cysteine is hydrophilic, phenylalanine is strongly hydrophobic (Betts and Russell, 2003). These could attribute to differences in *MARCH7* structure and function between species however as overall homology and conservation is generally very high it is still unexpected for comparison of tertiary structure to reveal such a low percentage similarity.

The schematic structure of *MARCH7* in Nathan *et al.* (2008), revealed a RING-CH close to the C-terminus and serine/proline-rich region at the N-terminus, and they have noted a disordered structure. Proline is known in many respects for being anomalous and its unique features help to contribute to the roles it plays in protein function and structure (Morgan and Rubenstein, 2013). The nitrogen atom in proline is covalently bound within a five-membered ring, which causes marked restriction of the phi (Φ) angular range in

peptide bond formation in a protein or peptide at this locus. Furthermore, it is known that in response to subtle influences, such as changes in local charge distribution, proline can adopt both a *cis* and *trans* configuration. This results in a tendency of prolyls (i.e. multiple prolines) to produce bending in the regional amino acid alignment and thus the fold of the protein (Morgan and Rubenstein, 2013). Prolyls tend to be excluded from the alpha helices and beta sheets, however, sometimes they can be found situated at the ends of these motifs. Proline causes disruption in the secondary structure of a protein through inhibition of the backbone, preventing conformation to an alpha-helix or beta-sheet (Morgan and Rubenstein, 2013). A more complex alternative is that proline can impose a secondary structure that poses a confined phi angle which is chosen over secondary structure forms. Because of their hydrophobicity they tend to adopt positions within the interior of a protein (Morgan and Rubenstein, 2013). The disruptive effects of proline could help to explain some of the difficulties faced by QUARK in producing reliable models, due to an inability to accurately predict the correct conformation to an alpha-helix or beta-sheet in these proline rich regions at secondary structure. If secondary structure cannot be produced clearly then evidently there will be further complications in production of a reliable and accurate tertiary structure. Complications with QUARK appeared to be universal across all mammalian species, with differences seen in all 10 models for all species submitted. Similarity between models produced increased in the exons coding for the RING-CH region of *MARCH7*, due to an increased order in the gene but homology between models still remained low.

4.15.3.1 Prediction of disorder

Analysis using Phyre revealed that 84% of the human *MARCH7* gene and 85% of *MARCH7* in the dog are disordered and just 9% of amino acids could be modelled with confidence (Appendix 7.26). Generally speaking proteins adopt stable, localized structures but in certain cases regions of the protein chain fail to do so. These are regions whose coordinates are hard to determine by experimental techniques, or that simply do not fold into stable structures (Tompa, 2002; Receveur-Bréchot *et al.*, 2006). Such regions are known as disordered regions or intrinsically disordered proteins (IDPs) (Forman-kay and Mittag, 2013).

IDP's allow binding of either of a single protein to numerous proteins at different times or binding of a number of proteins to a common partner (Huart and Hupp, 2013). The lack of

an "intrinsic" structure gives IDP's evolutionary advantages, including their capability of binding to multiple partners, participating in various reactions and pathways and allowing changes in – or fine tuning of molecular interaction networks (Dunker and Obradović, 2001; Dunker *et al.*, 2002; Dyson and Wright, 2005; Deng *et al.*, 2009). As *MARCH7* is involved in the immune response, including immune tolerance and regulation of T lymphocytes (Metcalfe and Muthukumarana, 2005), a disordered protein structure may be explained through the evolutionary advantages this kind of structure provides.

4.16 CONCLUSION

Results of this study indicated that genetic variation in *MARCH7* within canids was greatest in wolves and pure-breed dogs, and lowest in free-ranging dogs, but provided a suggestion of copy number variation affecting this gene. Studying patterns of nonsynonymous and synonymous mutations revealed *MARCH7* to be under purifying selection across mammalian species, with sequence analysis demonstrating high conservation across the gene and identical functional conservation in the RING-CH domain. Patterns of protein conformation indicated a highly disordered structure of *MARCH7* outside of RING-CH domain, which could offer an evolutionary advantage, given the involvement of *MARCH7* in the immune response.

5. GENERAL DISCUSSION

5.1 RECONSTRUCTION OF KINSHIP RELATIONSHIPS IN A FREE-RANGING DOG POPULATION

The aim of this project was to assess the genetic variability, inbreeding levels and to reconstruct the kinship relationships for a population of free-ranging dogs. Whilst there has been considerable research focusing on the kinship of wolves, data on free-ranging dogs was sparse and there has been a long standing debate over their ability to form social groups with a clear dominance structure, similar to wolf packs. In this study, microsatellite analysis of genetic variability in a free-ranging dog population revealed moderate heterozygosity, a low average number of alleles per locus, deviation from HWE and difficulties in assigning correct parentage in the sample population. Low genetic variability pointed towards inbreeding in the population studied. Inbreeding is generally avoided in wild wolf populations (Smith *et al.,* 1997) but is affecting purebred dogs, due to artificial selection by humans.

When compared to wolves, there is an apparent difference in the number of breeding individuals present in a pack. Paternity results for the offspring of Snella and Sofia also suggested the possibility of multiple fathers but lacked statistical support. Typical wolf packs consist of a single breeding pair, whilst evidence from microsatellite results shows multiple breeding individuals to be present in this free-ranging dog population. One explanation for this is that it is a result of the domestication process (Cafazzo *et al.* 2014), as free-ranging dogs are no longer influenced by seasonal reproductive behaviour, have an abundance of human waste (and food provided by humans in the case of the study population) to scavenge reducing competition between conspecifics and facilitating reproduction to occur during the first year of life (Lord *et al.,* 2013; Bonanni and Cafazzo, 2014).

This study shows evidence for considerable differences between the mating systems in free-ranging dogs, wolves, and pure-breed dogs. Mating systems are known to directly influence sexual selection which impact on functional diversification between the three canid groups, particularly at coding genes influenced by sexual selection. Immune system genes are one set of genes influenced through this mechanism and in order to study this analysis of the *MARCH7* gene was undertaken. *MARCH7* was chosen due to previous research providing an indication of diversifying selection occurring in this gene, with differences witnessed between the three canid groups.

5.2 FUNCTIONAL GENETIC DIFFERENTIATION BETWEEN PURE-BREED AND FREE-RANGING DOGS AT *MARCH7* GENE

In order to investigate evolutionary patterns in the canine *MARCH7* gene three main methods were used; studying the patterns of genetic variation in *MARCH7* in canids, comparing the patterns of nonsynonymous and synonymous variation between canids and other mammalian species, and studying *MARCH7* protein conformation patterns.

Sequencing results and data from online sources demonstrate a higher variation in pure-breed dogs and wolves than in free-ranging dogs, but not evidence for the polymorphism at the SNP site described by Pilot *et al.* (2016). Poor sequencing quality and lack of the previously identified SNP might suggest gene duplication or CNV in the canine *March7* gene. CNVs in immune related genes are known to increase the organisms' ability to adapt to novel environments (Nguyen *et al.,* 2006; de Smith *et al.,* 2007), increasing genetic variability and fitness (Feuk *et al.,* 2006). Mutations present on the intron-exon boundary could indicate alternative splicing in the *MARCH7* gene, with one explanation being that the immune system must be able to produce a diverse repertoire of antigen receptors in T and B cells (Yabas *et al.,* 2016) and *MARCH7* plays an important roles in T lymphocyte production (Gao *et al.,* 2009).

Analysis focusing on the patterns of non-synonymous and synonymous mutations in mammalian species revealed *MARCH7* to be under purifying selection. Comparative analysis demonstrates that whilst *MARCH7* has highly conserved regions, most notably the RING-CH domain, it is polymorphic and a multitude of both synonymous and nonsynonymous mutations are present in all mammals studied. As already discussed, nonsynonymous mutations are of interest due to the resulting change in amino acid (Ng and Henikoff, 2003; Krawczak *et al.,* 2000), but synonymous mutations are now recognised to be essential for the function and maintenance of diverse regulatory signals located in protein coding regions (Shabalina *et al.,* 2013). It is clear that protein-coding sequences in higher eukaryotes require diversification for functional integrity, and this is achieved by the use of different codons in their variable and constitutive regions through different selection mechanisms (Resch *et al.,* 2007).

Ab initio modelling of *MARCH7* using QUARK and Phyre software revealed a highly disordered structure, making accurate prediction of tertiary structures difficult. Analysis demonstrated high conservation across the RING-CH domain in particular, which is

similar to findings by Pavlicek *et al.* (2004) who focused on primates. Regions in proteins which fail to fold into structures are referred to as disordered regions or intrinsically disordered proteins. One of the evolutionary advantages of proteins which have disordered regions is their capability to bind to multiple partners and participle in various reactions and pathways (Dunker and Obradović, 2001; Dunker *et al.*, 2002; Dyson and Wright, 2005; Deng *et al.*, 2009). *MARCH7* is involved in immune tolerance and regulation of T lymphocytes (Metcalfe and Muthukumarana, 2005) so capability to bind to multiple partners may be expected due to involvement in signalling pathways (i.e. hedgehog signalling pathway) which involves multiple partners.

5.3 FUTURE CONSIDERATIONS

When considering the results obtained from this study, there are a number of future considerations which could improve results or provide additional support, including: changes to primer design and amplification, using genome-wide SNP data for parentage analysis and analysis of alternative splicing, codon usage bias and copy number variation.

5.3.1 Studying kinship patterns

Failure of two loci to work in this study affected the confidence levels and validity of results, an increased number of target loci for future studies would be beneficial, as well as consideration of primer design and amplification and the utilisation of genome-wide SNPs in replacement of microsatellites for parentage analysis to increase reliability of paternity assignment.

5.3.1.1 Primer design and amplification

Redesigning primers to bind to a different region of the flanking sequence, or adjusting PCR conditions can often ameliorate null allele problems (Callen *et al.*, 1993; Pemberton *et al.*, 1995). Re-amplifying individuals homozygous for shorter-length alleles and increasing the sample concentration in the DNA sequencer run is one way to combat this source of genotyping error (Selkoe and Toonen, 2006)

5.3.1.2 Utilising SNPs vs microsatellites

An alternative to using microsatellites for parentage analysis is the use of SNPS, which could offer benefits over microsatellites. A study by Yu *et al.* (2015) studying the effectiveness of microsatellite and single nucleotide polymorphism markers for parentage

analysis in European domestic pigs found that SNPs offer several advantages over microsatellites in parentage analysis. SNPs can be used for the investigation of both non-coding and coding regions meaning they provide wider genome coverage than microsatellites (Yu *et al.,* 2015). They are located throughout the genome, have a lower genotyping error rate and low mutation rate and can be easily genotyped through high throughput microarray analysis (Werner *et al.,* 2004; Honda *et al.,* 2009).

5.3.2 Studying genetic variation in *MARCH7*

To gain more in-depth understanding of the resulting changes to protein function caused by the CNV and alternative splicing, as indicated by results, further analysis would be required.

5.3.2.1 Alternative splicing analysis

One common method used to detect alternative splicing is the use of RNA sequencing (RNA-seq), including the use multivariate analysis of transcript splicing (MATS) which can be used to detect differential alternative splicing events (Park *et al.,* 2013). RNA sequencing of the intronic regions where SNPs were found to be present in *MARCH7* would enable confirmation of whether alternative splicing is occurring.

5.3.2.2 Copy number variation analysis

To ascertain whether copy number variation is occurring, further analysis would be required and there are multiple methodologies available that can be applied to genotype. These methods are based on either ultra-dense genotyping with SNP chips, high-throughput sequencing or the hybridization of DNA in BAC/PAC/oligonucleotide arrays (Clop *et al.,* 2012; Foong *et al.,* 2015). Other methods include fluorescence in situ hybridisation (FISH), multiple-ligation-dependant probe amplification (MLPA) and array comparative genomic hybridisation (Olsson *et al.,* 2016).

6. REFERENCES

Acharya, D., Mukherjee, D., Podder, S. and Ghosh, T. (2015) Investigation of different duplication patterns of essential genes in mouse and human, *PLoS One,* 10: e0120784.

Agrafioti, I. and Stumpf, M. (2007) SNPSTR: a database of compound microsatellite SNP markers, *Nucleic Acids Research,* 35: 71-75.

Almeida, M., Betancor, E., Fregel, R., Suárez, N. and Pestano, J. (2011) Efficient DNA extraction from hair shafts, *Forensic Science International: Genetics Supplement Series,* 3: e319-320.

Altizer, S., Nunn, C., Thrall, P., Gittleman, J., Antonovics, J., Cunnigham, A., Dobson, A., Ezenwa, V., Jones, K., Pedersen, A., Poss, M. and Pulliam, J. (2003) Social organisation and parasite risk in mammals: integrating theory and empirical studies, *Annual Review of Ecology, Evolution and Systematics,* 34: 517-547.

Alvarez, C. and Akey, J. (2012) Copy number variation in the domestic dog, *Mammalian Genome,* 23: 14-163.

Antonovics, J. and Thrall, P. (1994) The cost of resistance and the maintenance of genetic polymorphism in host-pathogen systems, *Proceedings: Biological Sciences*, 257: 105-110.

Ashley, M., Caballero, C., Chaovalitwongse, W., Dasgupta, D., Govindan, P., Sheikh, S. and Berger-Wolf, T. (2009) KINALYZER, a computer program for reconstructing sibling groups, *Molecular Ecology Resources,* 9: 1127-1131.

Aspi, J., Roininen, E., Ruokonen, M., Kojola, I. and Vila, C. (2006) Genetic diversity, population structure, effective population size and demographic history of the Finnish wolf population, *Molecular Ecology,* 15: 1561-1576.

Ba, A., Strome, B., Hua, J., Desmond, J., Gagnon-Arsenault, I., Weiss, E., Landry, C. and Moses, A. (2014) Detecting functional divergence after gene duplication through evolutionary changes in posttranslational regulatory sequences, *PLoS Computational Biology,* 10: e1003977.

Baker, D. and Sali, A. (2001) Protein structure prediction and structural genomics, *Science,* 294: 93-96.

Basset, C., Holton, J., O'Mahony, R. and Roitt, I. (2003) Innate immunity and pathogen-host interaction, *Vaccine,* 21: 12-23.

Bateson, P. and Sargan, D. (2012) Analysis of the canine genome and canine health: A commentary. *The Veterinary Journal,* 194: 265-269.

Beck, A. (1973). *The ecology of stray dogs: a study of free-ranging urban animals*. Baltimore: York Press. 98-99.

Bekaert, B., Larmuseau, M., Vanhove, M., Opdekamp, A. and Decorte, R. (2012) Automated DNA extraction of single dog hairs without roots for mitochondrial DNA analysis, *Forensic Science International: Genetics,* 6: 277-281.

Berglund, J., Nevalainen, E., Molin, A., Perloski, M., The LUPA Consortium, André, C., Zody, M., Sharpe, T., Hitte, C., Lindblad-Toh, K., Lohi, H. and Webster, M. (2012) Novel origins of copy number variation in the dog genome, *Genome Biology*, 13: R73.

Betts, M. J. and Russell, R. B. (2003) Amino Acid Properties and Consequences of Substitutions, in Bioinformatics for Geneticists (eds M. R. Barnes and I. C. Gray), John Wiley & Sons, Ltd, Chichester.

Bhadra, A., Bhattacharjee. D., Paul, M., Singh, A., Gade, P., Shrestha, P. and Bhadra, A. (2015) The meat of the matter: a rule of thumb for scavenging dogs? *Ethology, Ecology and Evolution*, 28: 427-440.

Bhargava, A. and Fuentes, F. (2010) Mutational dynamics of microsatellites, *Molecular Biotechnology,* 44: 250-266.

Björnerfeldt, S., Webster, M. and Vilà, C. (2006) Relaxation of selective constraint on dog mitochondrial DNA following domestication, *Genome Research,* 16: 990-994.

Blanchard, F., Duplomb, L., Wang, Y., Robledo, O., Kinzie, E., Pitard, V., Godard, A., Jacques, Y. and Baumann, H. (2000) Stimulation of Leukemia Inhibitory Factor Receptor degradation by extracellular-regulated kinase, *The Journal of Biological Chemistry,* 275: 28793-28801.

Bofkin, L. and Goldman, N. (2007) Variation in evolutionary processes at different codon positions, *Molecular Biology and Evolution*, 24: 513-521.

Boitani, L. and Ciucci, P. (1995) Comparative social ecology of feral dogs and wolves, *Ethology, Ecology and Evolution*, 7: 49-72.

Bonanni, R. and Cafazzo, S. (2014**)** The social organization of a population of free-ranging dogs in a suburban area of Rome: a reassessment of the effects of domestication on dog behaviour. In Kaminski J, Marshall-Pescini S editors. The social dog: behaviour and cognition. Elsevier Inc. 65–104.

Bonanni, R., Cafazzo, S., Valsecchi, P. and Natoli, E. (2010) Effect of affiliative and agonistic relationships on leadership behaviour in free-ranging dogs, *Animal Behaviour*, 79: 981-991.

Bonanni, R., Natoli, E., Cafazzo, S. and Valsecchi, P. (2011) Free-ranging dogs assess the quantity of opponents in intergroup conflicts, *Animal Cognition*, 14: 103-115.

Bonanni, R., Valsecchi, P. and Natoli, E. (2010) Pattern of individual participation and cheating conflicts between groups of free-ranging dogs, *Animal Behaviour*, 79: 957-968.

Bonin, A., Bellemain, E., Bronkon Eidesen, P., Pompanon, F., Borchmann, C. and Taberlet, P. (2004) How to track and assess genotyping errors in population genetics studies, *Molecular Ecology*, 13: 3261-3273.

Boyko, A. (2011) The domestic dog: man's best friend in the genomic era, *Genome Biology*, 12: 216.

Buhler, S. and Sanchez-Mazas, A. (2011) HLA DNA sequence variation among human populations: molecular signatures of demographic and selective events, *PLoS One*, 6: e14643.

Bulmer, M. (1991) The selection-mutation-drift theory of synonymous codon usage, *Genetics*, 129: 897-907.

Buschiazzo, E. and Gemmell, N. (2006) The rise, fall and renaissance of microsatellites in eukaryotic genomes, *Bioessays*, 28: 1040-1050.

Cafazzo, S., Bonanni, R., Valsecchi, P. and Natoli, E. (2014) Social variables affecting mate preferences, copulation and reproduction outcome in a pack of free-ranging dogs, *PLoS One*, 9: e98594.

Cafazzo, S., Natoli, E. and Valsecchi, P. (2012) Scent-marking behaviour in a pack of free-ranging domestic dogs, *Ethology*, 118: 955-966.

Cafazzo, S., Valsecchi, P., Bonanni, R. and Natoli, E. (2010) Dominance in relation to age, sex, and competitive contexts in a group of free-ranging domestic dogs. *Behavioural Ecology*, 21; 443-455.

Cagan, A. and Blass, T. (2016) Identification of genomic variants putatively targeted by selection during dog domestication, *BMC Evolutionary Biology*, 16: 10.

Callen, D., Thompson, A., Shen, Y., Phillips, H., Richards, R., Mulley, J. and Sutherland, G. (1993) Incidence and origin of "null" alleles in the (AC)n microsatellite markers, *American Journal of Human Genetics*, 52: 922-927.

Caniglia, R., Fabbri, E., Galaverni, M., Milanesi, P. and Randi, E. (2014) Non-invasive sampling and genetic variability, pack structure, and dynamics in an expanding wolf population, *Journal of Mammalogy*, 95: 41-59.

Carmichael, L., Nagy, J., Larter, N. and Strobeck, C. (2001) Prey specialization may influence patterns of gene flow in wolves of the Canadian Northwest, *Molecular Ecology*, 10: 2787-2798.

Cartegni, L., Chew, S. and Krainer, A. (2002) Listening to silence and understanding nonsense: exonic mutations affect splicing, *Nature Review Genetics*, 3: 285-298.

Casapis, C. and Spyroulias, G. (2009) RING Finger E3 ubiquitin ligases: structure and drug discovery, *Current Pharmaceutical Design*, 15 (31): 3716-3731.

Chae, E., Bomblies, K., Kim, S., Karelina, D., Zaidem, M., Ossowski, S., Martín-Pizarro, C., Laitinen, R., Rowan, B., Tenenboim, H., Lechner, S., Demar, M., Habring-Müller. A., Lanz, C., Rätsch, G. and Weigel, D. (2014) Species-wide genetic incompatibility analysis identifies immune genes as hot spots of deleterious epistasis, *Cell*, 159: 1341-1351.

Chamary, J., Parmley, J. and Hurst, L. (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals, *Nature Review Genetics*, 7: 98-108. .

Charlesworth, D. and Willis, J. (2009) The genetics of inbreeding depression, *Nature Review Genetics*, 10: 783-796.

Chasapis, C. and Spyroulias, G. (2009) Ring finger E(3) ubiquitin ligases: structure and drug discovery, *Current Pharmaceutical design*, 15 (31): 3716-3731.

Chistiakov, D., Hellemans, B. and Volckaert, F. (2006) Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics, *Agriculture,* 255 (1-4): 1-29.

Christiansen, P. and Wroe, S. (2007) Bite forces and evolutionary adaptions to feeding ecology in carnivores, *Ecology*, 88: 437-358.

Clop, A., Vidal, O. and Amills, M. (2012) Copy number variation in the genomes of domestic animals, *Animal Genetics,* 43: 503-517.

Cookes, G. and Hill, A. (2001) Genetics of susceptibility to human infectious disease, *Nature Reviews Genetics*, 2: 967-977.

Coppinger R, Schneider R. Evolution of working dogs. In: Serpell J, editor. The domestic dog: its evolution, behaviour and interactions with people. Cambridge: Cambridge University Press; 1995. 21–47

Cruz, F., Vilà. C. and Webster, M. (2008) The legacy of domestication: accumulation of deleterious mutations in the dog genome, *Molecular Biology and Evolution,* 25: 2331-2336.

da Craen, A., Posthuma, D., Remarque, E., van den Biggelaar, A., Westendorp, R., Boomsma, D. (2005) Heritability estimates of innate immunity: an extended twin study, *Genes and Immunity*, 6: 167-170.

Dakin, E. and Avise, J. (2004) Microsatellite null alleles in parentage analysis, *Heredity*, 93: 504-509.

Daniels, T. and Bekoff, M. (1989a) Population and social biology of free-ranging dogs, *Canis familaris*, *Journal of Mammalogy*, 70: 754-762.

Daniels, T. and Bekoff, M. (1989b) Spatial and temporal resource use by feral and abandoned dogs, *Ethology*, 81: 300-312.

Darwin, C. (1859). *Origin of the species.* London: J. Murray. 14-25

de Smith, A., Walters, R., Froguel, P. and Blakemore, A. (2009) Human genes involved in copy number variation: mechanisms of origin, functional effects and implications for disease, *Cytogenetic and Genome Research,* 123: 17-26.

Deng, S., Eickholt, J. and Cheng, J. (2009) PreDisorder: ab initio sequence-based prediction of protein disordered regions, *BMC Bioinformatics*, 10: 436-442.

Derix, R. and Van Hooff, J. (1995) Male and female partner preferences in a captive wolf pack (*Canis lupus*): specificity versus spread of sexual attention, *Behaviour,* 132: 127-149.

Deshaies, R. and Joazeiro, C. (2009) RING Domain E3 Ubiquitin Ligases, *Annual Review of Biochemistry*, 78: 399-434.

Dierssen, M., Herault, Y. and Estivill, X. (2009) Aneuploidy: From a physiological mechanism of variance to down syndrome, *Physiological Reviews,* 89: 887-920.

Dodds, P. and Thrall, P. (2009) Recognition events and host-pathogen co-evolution in gene-for-gene resistance to flax rust, *Functional Plant Biology,* 36: 395-408.

Dos Reis, M., Savva, R. and Wernisch, L. (2004) Solving the riddle of codon usage preferences: a test for translational selection, *Nucleic Acids Research,* 32: 5056-5044.

Drake, A. and Klingenberg, C. (2010) Large-scale diversification of skull shape in domestic dogs: disparity and modularity, *The American Naturalist,* 175: 289-301.

Dunker, A. and Obradović, Z. (2001) The protein trinity—linking function and disorder, *Nature Biotechnology,* 19: 805-806.

Dunker, A., Brown, C., Lawson, J., Iakoucheva, L. and Obradović, Z. (2002) Intrinsic disorder and protein function, *Biochemistry,* 41: 6573-6582.

Duret, L. (2002) Evolution of synonymous codon usage in metazoans, *Current Opinion in Genetics& Development,* 12: 640-649.

Dyson, H. and Wright, P. (2005) Intrinsically unstructured proteins and their functions, *Nature Reviews; Molecular Cell Biology,* 6: 197-208.

Ellegren, H., Smith, N. and Webster, M. (2003) Mutation rate variation in the mammalian genome, *Current Opinions on Genetic Developments,* 13: 562-568.

Emes, R., Goodstadt, L., Winter, E. and Ponting, C. (2003) Comparison of the genomes of human and mouse lays the foundation of genome zoology, *Human Molecular Genetics,* 12: 701-709.

Engle, E. (1946) No seasonal breeding cycle in dogs, *Journal of Mammalogy,* 27: 79-81.

Espinosa-Cantú, A., Ascencio, D., Barona-Gómez, F. and DeLuna, A. (2015) Gene duplication and the evolution of moonlighting proteins, *Frontiers in Genetics,* 6: 1-7.

Fairbrother, W., Holste, D., Burge, C. and Sharp P. (2004) Single nucleotide polymorphism-based validation of exonic splicing enhancers, *PLoS Biology*, 2: E268.

Farrell, L., Schoenebeck, J., Weiner, P., Clements, D. and Summers, K. (2015) The challenges of pedigree dog health: approaches to combating inherited disease. *Canine, Genetics and Epidemiology*, 2; 3-17.

Farrell, L., Schoenebeck, J., Weiner, P., Clements, D. and Summers, K. (2015) The challenges of pedigree dog health: approaches to combating inherited disease, *Canine Genetics and Epidemiology*, 2: 3.

Feuk, L., Carson, A. and Scherer, S. (2006) Structural variation in the human genome, *Nature Reviews Genetics*, 7: 85-87.

Flierman, D., Coleman, C., Pickart, C., Rapoprt, T. and Chau, V. (2006) E2-25K mediates US11-triggered retro-translocation of MHC class I heavy chains in a permeabiziled cell system. *PNAS*, 103: 11589-11594.

Fondon III, J. and Garner, H. (2004) Molecular origins of rapid and continuous morphological evolution, *Proceedings of the National Academy of Science of the United States of America*, 101: 18058-18063.

Foong, J., Girdea, M., Stavropoulos, J. and Brudno, M. (2015) Prioritizing clinically relevant copy number variation from genetic interactions and gene function data, *PLoS One*, 10: 1-15.

Forman-Kay, J. and Mittag, T. (2013) From sequences and forces to structure, function, and evolution of intrinsically disordered proteins, *Structure*, 21: 1492-1429.

Freedman, A., Schweizer. R, Vecchyo, D., Hang, E., Davis, D., Gronau, I., Silva, P., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., Ramierz, O., Hormozdiari, F., Alkan, C., Parker, G., Lee, C., Tadigotla, V., Siepel, A., Bustamnate, C., Harkins, T., Nelson, S., Marques-Bonet, T., Ostrander, E., Wayne, R. and Novembre, J. (2016) Demographically-Based Evaluation of Genomic Regions under Selection in Domestic Dogs, *PLoS Genetics*, 12 : e1005851.

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., Via, 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. and Novembre, J. (2014). Genome sequencing highlights the dynamic early history of dogs. *PLoS Genetics*. 10: e1004016.

Fujita, H., Iwabu, Y., Tokunaga, K. and Tanaka, Y. (2013) Membrane-associated RING-CH (MARCH) 8 mediates the ubiquitination and lysosomal degradation of the transferrin receptor, *Journal of Cell Science,* 126: 2798-2809.

Gagneux, P., Boesch, C. and Woodruff, D. (1997) Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair, *Molecular Ecology,* 6: 861-868.

Galibert, F., Quignon, P., Hitte, C. and André. C. (2011) Toward understanding dog evolutionary and domestication history, *Comptes Rendus Biologies*, 334: 190-196.

Gao, W., Thompson, L., Zhou, Q., Putheti, P., Fahmy, T., Strom, T. and Metcalfe, S. (2009) Treg versus Th17 lymphocyte lineages are cross-regulated by LIF versus IL-6. *Cell Cycle,* 8: 1444-1450.

Garcia de Leon, F., Canonne, M., Quillet, E., Bonhomme, F. and Chatain, B. (1998) The application of microsatellite markers to breeding programmes in the sea bass, *Dicentrarchus labrax*, 159: 303-316.

Garcia-Monreno, J., Matocq, M., Roy, M., Geffen, E. and Wayne, R. (1996) Relationships and genetic purity of the endangered Mexican wolf based on analysis of microsatellite loci, *Conservation Biology,* 10: 376-389.

Germonpré. M., Lázničková. M. and Sablin, M. (2012) Palaeolthic dog skulls at the Gravettian Předmostí site, the Czech Republic, *Journal of Archaeological Science,* 39: 184-202.

Ghatak, S., Muthukumaran, R. and Nachimuthu, S. (2003) A simple method of genomic DNA extraction from human samples for PCR-RFLP analysis, *Journal of Biomolecular Techniques*, 24: 224-231.

Gipson, P., Gipson, I. and Sealander, J. (1975) Reproductive biology of wild *Canis* (Canidae) in Arkansas, *Journal of Mammalogy*, 56: 606-612.

Godinho, R., Llaneza, L., Blanco, J., Lopes, S., Alvares, F., Garcia, E., Palacios, V., Cortes, Y., Talegon, J., Ferrand, N. (2011) Genetic evidence for multiple events of hybridization

between wolves and domestic dogs in the Iberian Peninusla, *Molecular Ecology,* 20: 5154-5166.

Goldman, N. and Yang, Z. (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences, *Molecular Biology and Evolution,* 11: 725-736.

Gompper, M. (2014). *Free-Ranging Dogs and Wildlife Conservation*. Oxford: Oxford University Press. 15.

González, M. (2011) Force fields and molecular dynamics simulations, *Collection SFN,* 12: 169-200.

Graf, U., Casanova, E. and Cinelli, P. (2011) The role of Leukemia Inhibitory Factor (LIF) – pathway in derivation and maintenance of murine pluripotent stem cells, *Genes,* 2: 280-297.

Graffy, E. and Foran, D. (2005) A simplified method for mitochondrial DNA extraction from head hair shafts, *Journal of Forensic Science,* 50: 1119-1122.

Grewal, S., Wilson, P., King, T., Shami, K., Theberge, M., Theberge, J. and White, B. (2004) A Genetic Assessment of the Eastern Wolf (*Canis lycaon*) in Algonquin Provincial Park, *Journal of Mammalogy,* 85: 625-632.

Gu, W., Zhang, F. and Lupski, J. (2008) Mechanisms for human genomic rearrangements, *Pathogenetics,* 1:4-21.

Gu, Z., Steinmetz, L., Gu, X., Scharfe, C., Davis, R. and Li, W. (2003) Role of duplicate genes in genetic robustness against null mutations, *Nature,* 421: 63-66.

Guichoix, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F. and Petit, R. (2011) Current trends in microsatellite genotyping, *Molecular Ecology Resources,* 11: 591-611.

Haase, E. (2000) Comparison of reproductive biological parameters in male wolves and domestic dogs, *Zeitschrift für Säugetierkunde*, 65: 257-270.

Hardison, R., Roskin, K., Yang, S., Diekhans, M., Kent, W., Weber, R., Elnitski, L., Li, J., O'Conner, M., Kolbe, D., Schwartz, S., Furey, T., Whelan, S., Goldman, N., Smit, A., Miller, W., Chairomonte, F. and Haussler, D. (2003) Covariation in frequencies of

substitution, deletion, transposition, and recombination during eutherian evolution, *Genome Research,* 13: 13-26.

Hastings, P., Lupski, J., Rosenburg, S. and Ira. G. (2009) Mechanisms of change in gene copy number, *Nature Review Genetics,* 10: 55-564.

Hennessy, C., Dubach, J. and Gehrt, S. (2012) Long-term pair bonding and genetic evidence for monogamy among urban coyotes (Canis Latrans), *Journal of Mammalogy,* 93: 732-742.

Hill, W. and Robertson, A. (1966) The effect of linkage on limits to artificial selection, *Genetic Research,* 8: 269-294.

Hisaka, T., Desmouliére, A., Taupin, J., Daburon, S., Neaud, V., Senant, N., Blanc, J., Moreau, J. and Rosenbaum, J. (2004) Expression of leukemia inhibitory factor (LIF) and its receptor gp190 in human liver and in cultured liver myofibroblasts. Cloning of new isoforms of LIF mRNA, *Complete Hepatology,* 3: 10.

Hoekstra, H., Hoekstra, J., Berrigan, D., Vignieri, S., Hoang, A., Hill, C., Beerli, P. and Kingsolver, J. (2001) Strength and tempo of directional selection in the wild, *Proceedings of the National Academy of Sciences of the United States of America,* 98: 9157-9160.

Hoffman, J. and Amos, W. (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion, *Molecular Ecology,* 14: 599-612,

Hoffman, J., Creevy, K. and Promislow, D. (2013) Reproductive capability is associated with lifespan and cause of death in companion dogs, *PLoS One,* 8 (4): e61082.

Höglund, J. and Sheldon, B. (1998) The cost of reproduction and sexual Selection, *Oikos,* 83: 478-483.

Holmgren, A., McConkey, C. and Shin. S. (2017) Outrunning the Red Queen: bystander activation as a means of outpacing innate immune subversion by intracellular pathogens, *Cellular & Molecular Immunology,* 14: 14-21.

Honda, T., Kasuta, T. and Mukai, F. (2009) Simulation study of parentage analysis with SNPs in Japanese Black Cattle population, *Asian-Australian Journal of Animal Sciences,* 22: 1351-1358.

Horrocks, N., Matson, K. and Tieleman, B. (2011) Pathogen pressure puts immune defense into perspective, *Integrative and Comparative Biology*, 51: 563-576.

Hozumi, N. and Tonegawa, S. (1976) Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions, *Proceedings of the National Academy of Sciences of the United States of America*, 73: 3628-3632.

Hsia, E., Gui, Y. and Zheng, X. (2015) Regulation of hedgehog signalling by ubiquitnation, *Frontiers in Biology*, 10: 203-220.

Hu, J., Meng, Y., Yu, T., Hu, L. and Mao, M. (2015) Ubiquitin E3 ligase *MARCH7* promotes ovarian tumour growth, *Oncotarget*, 6: 12174-12187

Huang, H., Winter, E., Wang, H., Weinstock, K., Xing, H., Goodstadt, L., Stenson, P., Cooper, D., Smith, D. Albà, M., Ponting, C. and Fechtel, K. (2004) Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes, *Genome Biology*, 5: R47.

Huart, A. and Hupp, T. (2013) Evolution of conformational disorder and diversity of the P53 interactome, *BioDiscovery*, 8: 1-10.

Huelsenbeck, J. and Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogeny, *Bioinformatics*, 17: 754-755.

Hurst, L. and Pal, C. (2001) Evidence for purifying selection acting on silent sites in BRAC1, *Trends in Genetics*, 17: 62-65.

Hwang, D., and Green, P. (2004) Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substation patterns in mammalian evolution, *Proceedings of the National Academy of Sciences of the United States of America*, 1010: 12994-14001.

Iowa State University. (2017). Troubleshooting Sanger Sequencing Results. Available: http://www.dna.iastate.edu/troubleshooting-ss.html. Last accessed 10/10/2017.

Irion, D., Schaffer, A., Famula, T., Egglesotn, M., Hughes, S. and Pedersen, N. (2003) Analysis of Genetic variation in 28 dog breed populations with 100 microsatellite markers, *Journal of Heredity*, 94: 81-87.

Iyengar, P., Hirota, T., Hirose, S. and Nakamura, N. (2011) Membrane-associated RING-CH 10 (MARCH10 Protein) is a microtubule-associated E3 ubiquitin ligase of the spermatid flagella. *The Journal of Biological Chemistry,* 286: 39082-39090.

Janeway, C. and Medzhitov, R. (2002) Innate immune recognition, *Annual Reviews Immunology,* 20: 197-216.

Jansa, S., Lundrigan, B. And Tucker, P. (2003) Tests for positive selection on immune and reproductive genes in closely related species of the murine genus *Mus, Journal of Molecular Evolution,* 56: 294-307.

Jędrzejewski, W., Branicki, W., Veit, C., Međugorac, I, Pilot, M., Bunevich, A., Jędrzejewski, B., Schmidt, K., Theuerkauf, J., Okarma, H., Gula, R., Szymura, L. and Förster, M. (2005) Genetic diversity and relatedness within packs in an intensely hunted population of wolves *Canis lupus, Acta Theriologica*, 50: 3-22.

Jenks, S. (2011) A longitudinal study of the sociosexual dynamics in a captive family group of wolves: the University of Connecticut wolf project, *Behavioural Genetics,* 41: 810-829.

Jokela, J., Schmid-Hempel, P. and Rigby, M. (2000) Dr. Pangloss restrained by the Red Queen – steps towards a unified defence theory, *Oikos,* 89: 267-274.

Kalinoswki, S., Taper, M. and Marshall, T. (2007) Revising how the computer program Cervus accommodates genotyping error increases success in paternity assignment, *Molecular Ecology,* 16: 1099-1106.

Ke, X., Kennedy, L., Short, A., Seppala , E., Barnes, A., Clements, D., Wood, S., Carter, S., Happ, G., Lohi, H. and Ollier, W. (2010) Assessment of the functionality of genome-wide canine SNP arrays and implications for canine disease association studies. *Animal Genetics,* 42: 181-190.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P. and Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, *Bioinformatics*, 28: 1647-1649.

Keller, L. and Waller, D. (2002) Inbreeding effects in wild populations, *Trends in Ecology & Evolution,* 17: 230-241.

Kelley, L., Mezulis, S., Yates, C., Wass, M. and Sternberg, M. (2015) The Phyre2 web portal for protein modelling, prediction and analysis, *Nature Protocols*¸ 10: 845-858.

Kennedy, L., Davidson, L., Barnes, A., Short, A., Fretwell, N., Jones, C., Lee, A., Ollier, W. and Catchpole, B. (2006) Identification of susceptibility and protection major histocompatibility complex haplotypes in canine diabetes mellitus, *Tissue Antigens,* 68: 467-476.

Keren, H., Lev-Maor, G., Ast, G. (2010) Alternative splicing and evolution: diversification, exon definition and function, *Nature Reviews: Genetics,* 11: 345-355.

Kimura, M. (1962) On the probability of fixation of mutant genes in a population, *Genetics,* 47: 713-719.

Kleiman, D. (1977) Monogamy in mammals. *The Quarterly review of Biology,* 52: 39-69.

Klug, H. (2011) Animal mating systems. eLS.

Knobel, D., Butler, J., Lembo, T., Critchlow, R., Gompper, M. (2013). Dogs, disease and wildlife. In: Gompper, M. *Free-Ranging Dogs and Wildlife Conservation*. Oxford: Oxford University Press. 144-169.

Kojola, P., Ronkainen, S., Hakala, A., Kokko, S. (2004) Interactions between wolves *Canis lupus* and dogs *C. familiaris* in Finland, *Wildlife Biology,* 10: 101-105.

Kornblihtt, A., Schor, I., Allo, M., Dujardin, G., Petrillo, E., Munoz, M. (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation, *Nature Reviews: Molecular Cell Biology,* 14: 152-165.

Kosiol, C., Vinar, T., da Fonseca, R., Hubisz, M., Bustamante, C., Nielsen, R. and Siepel, A. (2008) Patterns of positive selection in six mammalian genomes, *PLoS Genetics,* 4: e1000144.

Krawczak, M., Ball, E., Fenton, I., Stenson, P., Abeysinghe, S., Thomas, N. and Cooper, D. (2000) Human gene mutation database – A biomedical information and research resource, *Human Mutation,* 15: 45-51.

Krissinel, E. and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta Crystallographica Section D;* 60: 2256-2268.

Kryazhimskiy, S. and Plotkin, J. (2008) The population genetics of dN/dS, *PLoS Genetics,* 4: e1000304.

Kustritz, M. (2007) Determining the optimal ages for gonadectomy of dogs and cats, *Vet Med Today: Reference Point*, 231: 1665-1675.

Kwok, S., Kellogg, D., McKinney, N., Spasic, D., Goda, L., Levenson, C. and Sninsky, J. (1990) Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies, *Nucleic Acids Research,* 18: 999-1005.

Lande, R. and Schemske, D. (1985) The evolution of self-fertilization and inbreeding depression in plants, *Evolution,* 39 (1): 24-40.

Lawniczak, M., Barnes, A., Linklater, J., Boone, J., Wigby, S. and Chapman, T. (2007) Mating and immunity in vertebrates, *Trends in Ecology & Evolution,* 22: 48-55.

Lehman, N., Clarkson, P., Mech, D., Meier, T. and Wayne, R. (1992) A study of the genetic relationships within and among wolf packs using DNA fingerprinting and mitochondrial DNA, *Behavioural Ecology and Sociobiology,* 30: 83-94.

Lequarré, A., Andersson, L., André, C., Fredholm, M., Hitte, C., Leeb, T., Lohi, H., Lindblad-Toh, K. and Georges, M. (2011) LUPA: A European initiative taking advantage of the canine genome architecture for unravelling complex disorders in both human and dogs. *The Veterinary Journal*, 189: 155-159.

Lescureux, N. and Linnell, J. (2014) Warring brothers: The complex interactions between wolves (*Canis lupus*) and dogs (*Canis familiaris*) in a conservation context, *Biological Conservation*, 171: 232-245.

Li, Y. Korol, A., Fahima, T., Beiles, A. and Nevo, E. (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review, *Molecular Ecology,* 11: 2453-2465.

Li, Y., Wu, D., Boyko, A., Wang, G., Wu, S., Irwin, D. and Zhang, Y. (2014) Population variation revealed high-altitude adaption of Tibetan Mastiffs, *Molecular Biology and Evolution,* 31: 1200-1205.7

Li, Y., Wu, D., Boyko, A., Wang, G., Wu, S., Irwin, D. and Zhang, Y. (2014) Population variation revealed high-altitude adaption of Tibetan Mastiffs, *Molecular Biology and Evolution,* 31: 1200-1205.

Librado, P. and Rozas, J. (2009) DnaSp V5: a software for comprehensive analysis of DNA polymorphism data, *Bioinformatics,* 25: 1451-1452.

Lida, K. and Akashi, H. (2000) A test of translational selection at 'silent' sites in the human genome: base composition comparisons in alternatively spliced genes, *Gene¸*261: 93-105.

Lindblad-Toh, K., Wade, C., Mikkelsen, T., Karlsson, E., Jaffe, D., Kamal, M., Champ, M., Chang, J., Kulbokas III, E., Zody, M., Mauceli, E., Xie, X., Breen, M., Wayne, R., Ostrander, E., Ponting, C., Galibert, F., Smith, D., deJong, P., Kirkness, E., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C., Cook, A., Cuff, J., Daly, M., DeCaprio, D., Gnrerre, S., Grabherr, M., Kellis, M., Kleber, M., Bardeleben, C., Goodstadt, L., Heger, A., Hitte, C., Kim, L., Koepfli, K., Parker, H., Pollinger, J., Searle, S., Sutter, N., Thomas, R., Webber, C., Broad Institute Genome Sequencing Platform and Lander, E., (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature,* 438: 803-819.

Livingstone, C. and Barton, G. (1993) Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation, *Computer Applications in the Biosciences,* 9: 745-756.

Lord, K. Feinstein, M., Smith, B. and Coppinger, R. (2013) Variation in reproductive traits of members of the genus *Canis* with special attention to the domestic dog (*Canis familaris*), *Behavioural Processes,* 92: 131-142.

Lucchini, V., Galov, A., and Randi, E. (2004) Evidence of genetic distinction and long-term population decline in wolves (*Canis lupus*) in the Italian Apennines, *Molecular Ecology,* 13: 523-536.

Luco, R., Allo, M., Schor, I., Kornblihtt, A. and Misteli, T. (2011) Epigenetics in alternative pre-mRNA splicing, *Cell,* 144: 16-26.

Lupski, J. and Stankiewcz, P. (2005) Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes, *PLoS Genetics,* 1: e49.

Macdonald DW, Carr GM. Variation in dog society: between resource dispersion and social flux. In: Serpell J, editor. The domestic dog: its evolution, behaviour and interactions with people. Cambridge: Cambridge University Press; 1995. 199–216.

Majumder, A., Bhadra, A., Ghosh, A., Mitra, S., Bhattacharjee, D., Chatterjee, J., Nandi, A. and Bhadra, A. (2014) To be or not to be social: foraging associations of free-ranging dogs in an urban ecosystem, *Acta Ethologica*, 17: 1-8.

Majumder, S. and Bhadra, A. (2013) To be or not to be social: foraging associations of free-ranging dogs in an urban ecosystem, *PLoS One*, 10: 1-15.

Malm, K. (1995) Regurgitation in relation to weaning in the domestic dog: a questionnaire study, *Applied Animal Behaviour Science*, 43: 111-122.

Marsden, C., Vecchyo, D., O'Brien, D., Taylor, J., Ramierz, O., Vilà, C., Marques-Bonet, T., Schnabel, R., Wayne, R. and Lohmueller, K. (2016) Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs, *Proceedings of the National Academy of Sciences of the United States of America*, 113: 152-157.

Marshall-Pescini, S., Virányi, Z. and Range, F. (2015) The effect of domestication on inhibitory control: wolves and dogs compared, *PLOS One*, 10: 1-16.

Mathieu, M., Saucuort, C., Mournetas, V., Gauthereau, X., Thézé, N., Praloran, V., Thiébaud, P., and Boeuf, H. (2012) LIF-Dependant signaling: new pieces in the lego, *Stem Cell Reviews and Reports*, 8:1-15.

McKean, K. and Nunney, L. (2007) Sexual selection and immune function in Drosophila melanogaster, *Evolution*, 62: 386-400.

Mech, D. (1999) Alpha status, dominance, and division of labor in wolf packs, *Canadian Journal of Zoology*, 77: 1196-1203.

Mech, D. and Boitani, L. (2003). Wolf social ecology. In: Mech, D. and Boitaini, L. *Wolves: Behaviour, Ecology and Conservation*. Chicago: University of Chicago Press. 15-30.

Mech, D. and Nelson, M. (1990) Non-family wolf, *Canis lupus*, packs, *Canadian Field-Naturalist*, 104: 482-483.

Mech, D., Goyal, S., Paul, W. and Newton, W. (2008) Demographic effects of canine parvovirus on a free-ranging wolf population over 30 years, *Journal of Wildlife Diseases*, 44: 824-836.

Meier, T., Burch, J., Mech, L. and Adams, L. (1995) Pack structure and genetic relatedness among wolf packs in a naturally regulated population, Edmonton, Canada, *Proceedings of the 2nd North American Symposium on Wolves*, 293-309.

Metcalfe, S. (2005) Axotrophin and leukaemia inhibitory factor (LIF) in transplantation tolerance. *Philosophical Transactions of the Royal Society B*, 360; 1687-1694.

Metcalfe, S. and Muthukumarana, P. (2005) Transplantation tolerance: gene expression profiles comparing allotolerance vs. allorejection. *International Immunopharmacology*, 5: 33-39.

Metz, E., Robles-Sikisaka, R. and Vacquier, V. (1998) Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA, *Proceedings of the National Academy of Science of the United States of America*, 95: 10676-10681.

Meyer-Lucht, Y. and Sommer, S. (2005) MHC diversity and the association to nematode parasitism in the yellow-necked mouse (*Apodemus flavicollis*), *Molecular Ecology*, 14: 2233-2243.

Milne, I., Linder, D., Bayer, M., Husmeier, D., McGuire, G., Marshall, D. and Wright, F. (2009) TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments of HPC clusters and multi-core desktops, *Bioinformatics*, 25: 126-127.

Molin, A., Berglund, J., Webster, M. and Lindblad-Toh, K. (2014) Genome-wide copy number variant discovery in dogs using the CanineHD genotyping array, *BMC Genomics*, 15: 210.

Morgan, A. and Rubenstein, E. (2013) Proline: The Distribution, Frequency, Postioning, and Common Functional Roles of Proline and Polyproline Sequences in the Human Proteome, *PLoS One*, 8: e53785.

Morris, K., Wright, B., Grueber, C., Hogg, C. and Belov, K. (2015) Lack of genetic diversity across diverse immune genes in an endangered animal, the Tasmanian devil (*Sacrophilus harrisii*), *Molecular Ecology*, 24: 3860-3872.

Nakamura, N. (2011) The Role of the Transmembrane RING Finger Proteins in Cellular and Organelle Function. *Membranes*, 1: 354-393.

Nathan, J., Sengupta, S., Wood, S., Admon, A., Markson, G., Sanderson, C., Lehner, P. (2009) The ubiquitin E3 ligase *MARCH7* is differentially regulated by the deubiquitylating enzymes USP7 and USP9X. *Traffic*, 9: 1130-1145.

Ng, P. and Henikoff, S. (2003) SIFT: predicting amino acid changes that affect protein function, *Nucleic Acids Research*, 31: 3812-3814.

Nguyen, D., Webber, C. and Ponting, C. (2006) Bias of selection on human copy-number variants, *PLoS Genetics*, 2: e20.

Nicholas, T., Baker, C., Eichler, E. and Akey, J. (2011) A high-resolution integrated map of copy number polymorphisms within and between breeds of the modern domesticated dog, *BMC Genomics*, 12: 414.

Nicholas, T., Cheng, Z., Ventura, M., Mealey, K., Eichler, E. and Akey, J. (2009) The genomic architecture of segmental duplications and associated copy number variants in dogs, *Genome Research*, 19: 491-499.

Nielsen, R. and Yang, R. (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene, *Genetics*, 148: 929-936.

Niskanen, A., Kennedy, L., Ruokonen, M., Kojola, I., Lohi, H., Isomursu, M., Jansson, E., Pyhäjärvi, T. and Aspi, J. (2014) Balancing selection and heterozygote advantage in major histocompatibility complex loci of the bottlenecked Finnish wolf population, *Molecular Ecology*, 23: 875-889.

Nucleics. (2017). *DNA Sequencing Troubleshooting.* Available: https://www.nucleics.com/DNA_sequencing_support/DNA-sequencing-troubleshooting.html. Last accessed 10/01/2017.

Ohno, S., Wolf, U. and Atkin, N. (1968) Evolution from fish to mammals by gene duplication, *Hereditas*, 59: 169-187.

Olsson, M., Kierczak, M., Karlsson, Å., Jabłońska, J., Leegwater, P., Koltookian, M., Abadie, J., Citres, D., Thomas, Å., Hedhammar, A.,Tintle, L, Lindblad-Toh, K. and

Meadows, J. (2016) Absolute quantification reveals the stable transmission of high copy number variant linked to autoinflammatory disease, *BMC Genomics*, 12: 299.

Orban, T. and Olah, E. (2001) Purifying selection on silent sites – a constraint from splicing regulation? *Trends in Genetics*, 17: 252-253.

Ortolani, A., Vernooij, H. and Coppinger, R. (2009) Ethiopian village dogs: Behavioural responses to a stranger's approach, *Applied Animal Behaviour Science*, 119: 210-218.

Ostrander, E. and Kruglyak, L. (2000) Unleashing the canine genome, *Genome Research*, 10: 1271-1274.

Ovodov, N., Crockford, S., Kuzmin, Y., Higham, T., Hodgins, G. and 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: e22821.

Pal, S. (2001b) Population ecology of free-ranging urban dogs in West Bengal, India, *Acta Theriologica*, 46: 69-78.

Pal, S. (2005) Parental care in free-ranging dogs, *Canis familiaris Applied Animal Behaviour Science*, 90: 31-47.

Pal, S. (2011) Mating system off free-ranging dogs (Canis familiaris), *International Journal of Zoology*, 2011: 1-11.

Pal, S. (2015) Factors influencing intergroup agonistic behaviour in free-ranging domestic dogs (Canis familiaris), *Acta ethol*, 18: 209-220.

Pal, S., Ghosh, B. and Roy, S. (1999) Inter- and intra-sexual behaviour of free-ranging dogs (*Canis familaris*), *Applied Animal Behaviour Science*, 62: 267-278.

Palm, N. and Medzhitov, R. (2009) Pattern recognition receptors and control of adaptive immunity, *Immunology Reviews*, 227: 221-233.

Park, J., Tokeim, C., Shen, S. and Xing, Y. (2013) Identifying differential alternative splicing events from RNA sequencing data using RNASeq-MATS, *Methods in Molecular Biology*, 1038: 171-179.

Parmley, J. and Hurst, L. (2007) How do synonymous mutations affect fitness? *BioEssays*, 29: 515-519.

Parmley, J. and Hurst, L. (2007) How do synonymous mutations affect fitness? *Bioessays*, 29: 515-519.

Paszek, J. and Górecki, P. (2016) Genomic duplication problems for unrooted gene trees, *BMC Genomics*, 17: 165-175.

Paterson, S., Wilson, K. and Pemberton, J. (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population, *Proceedings of the National Academy of Science*, 95: 3714-3719.

Paul, M., Majumder, S. and Bhadra, A. (2014) Grandmotherly care: a case study in Indian free-ranging dogs, *Journal of Ethology*, 32: 75-82.

Pavlicek, A., Noskov, N.,Kouprina, N., Barrett, J., Jurka, J., Larionov, V. (2004) Evolution of the tumour suppressor *BRAC1* locus in primates: implications for cancer predisposition, *Human Molecular Genetics*, 13: 2737-2751.

Pemberton, J., Slate, J., Bancroft, D. and Barrett, J. (1995) Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies, *Molecular Ecology*, 4: 249-252.

Peterson, R., Jacobs, A., Drummer, T., Mech, D. and Smith, D. (2002) Leadership behaviour in relation to dominance and reproductive status in gray wolves, *Canis lupus*, *Canadian Journal of Zoology*, 80: 1405-1412.

Pfeiffer, I., Völkel, I., Täubert, H. and Brenig, B. (2004) Forensic DNA-typing of dog hair: DNA extraction and PCR amplification, *Forensic Science International*, 141: 149-151.

Piertney, S. and Oliver, M. (2006) The evolutionary ecology of the major histocompatibility complex, *Heredity*, 96: 7-21.

Pilot, M., Malewski, T., Moura, A., Grzybowski, T., Olenski, K., Kaminski, S., Fadel, F., Alagaili, A., Mohammed, O. and Bogdanowicz, W. (2016) Diversifying selection between pure-breed and free-breeding dogs inferred from genome-wide SNP analysis, *G*, 171: 232-245.

Pollak, E. (1987) On the theory of partially inbreeding finite populations. I. Partial selfing, *Genetics*, 117: 353-360.

Porcher, E. and Lande, R. (2005) The evolution of self-fertilization and inbreeding depression under pollen discounting and pollen limitation, *Journal of Evolutionary Biology,* 18: 497-508.

Porcher, E. and Lande, R. (2016) Inbreeding depression under mixed outcrossing, self-fertilisation and sib-mating, *BMC Evolutionary Biology,* 16:105.

Prendergast, B. (2005) Internalization of seasonal time, *Hormones and Behaviour*, 48: 503-511.

Putman, A. and Carbone, I. (2014) Challenges in analysis and interpretation of microsatellite data for population genetic studies, *Ecology and Evolution,* 4: 4399-4428.

Quintana-Murci, L. and Clark, A. (2013) Population genetic tools to dissect innate immunity in humans, *Nature Reviews Immunology,* 13: 280-293.

Rabb, G., Woolpy, J. and Ginsburg, B. (1967) Social Relationships in a Group of Captive Wolves, *American Zoologist*¸ 7: 305-311.

Ramierz, O., Olade, I., Berglund, J., Lorente-Galdos, B., Hernandez-Rodriguez, J., Quilez, J., Webster, M., Wayne., R., Lalueza-Fox, C., Vilà, C. and Marques-Bonet, M. (2014) Analysis of structural diversity in wolf-like canids reveals post-domestication variants, *BMC Genomics,* 15: 465.

Randi, E. (2011) Genetics and conservation of wolves *Canis lupus* in Europe, *Mammal Review*, 41 (2): 99-111.

Receveur-Bréchot, V., Bourhis, J., Uversky, V., Canard, B. and Longhi, S. (2006) Assessing protein disorder and induced folding, *Proteins,* 62: 24-45.

Reche, P. and Reinherz, E. (2003) Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms, *Journal of Molecular Biology,* 331: 623-641.

Reiter, T., Jagoda, E. and Capellini, T. (2016) Dietary Variation and Evolution of Gene Copy Number among Dog Breeds, *PLoS One,* 11: e0148899.

Resch, A., Carmel, L., Marino-Ramierz, L., Oqurtsov, A., Shabalina, S., Rogozin, I., Koonin, E. (2007) Widespread positive selection in synonymous sites of mammalian genes, *Molecular Biology and Evolution,* 24: 1821-1831.

Rimbault, M, Beale, H., Schoenebeck, H., Hoopes, B., Allen, J., Kilroy-Glynn, P., Wayne, R., Sutter, N. and Ostrander, E. (2013) Derived variants at six genes explain nearly half of size reduction in dog breeds, *Genome Research,* 23: 1985-1995.

Rimbault, M. and Ostrander, E. (2012) So many doggone traits: mapping genetics of multiple phenotypes in the domestic dogs, *Human Molecular Genetics,* 15: 52-57.

Rolff, J. (2002) Bateman's principle and immunity, *Proceedings: Biological Sciences,* 269: 867-872.

Ronquist, F. and Huelsenbeck, J. (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models, *Bioinformatics,* 19: 1572-1574.

Rosen, E. (2013) *BRCA1* in the DNA damage response and at telomeres, *Frontiers in Genetics,* 4: 85.

Ross KG (2001) Molecular ecology of social behaviour: analyses of breeding systems and genetic structure. *Molecular Ecology* 10: 265–284.

Rutledge, L., Patterson, B., Mills, K., Loveless, K., Murray, D. and White, D. (2010) Protection from harvesting restores the natural social structure of eastern wolf packs, *Biological Conservation,* 143: 332-339.

Saetre, P., Lindberg, J., Leonard, J., Olsson, K., Pettersson, U., Ellegren, H., Bergström. T., Vilà. C. and Jazin, E. (2004) From wild wolf to domestic dog: gene expression changes in the brain, *Molecular Brain Research,* 126: 198-206.

Salmon Hillbertz, N., Isaksson, M., Karlsson, E., Hellmén, E., Peilberg., G., Savolainen, P., Wade, C., von Euler, H., Gustafson, U., Hedhammar, A., Nilsson, M., Lindblad-Toh., K., Andersson, L. and Andersson, G. (2007) Duplication of *FGF3*, *FGF4*, *FGF19* and *ORAOV1* causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs, *Nature Genetics,* 39: 1318-1320.

Sanchez- Mazas, A. and Meyer, D. (2014) The Relevance of *HLA* Sequencing in Population Genetic Studies, *Journal of Immunology Research,* 2014: 1-12.

Saqui-Salces, M. and Merchant, J. (2010) Hedgehog signaling and gastrointestinal cancer, *Biochimica et Biophysica Acta,* 1803: 786-795.

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

Schley, P. and Field, C. (2002) The immune-enhancing effects of dietary fibres and prebiotics, *British Journal of Nutrition,* 87: 221-230.

Schmid-Hempel, P. (2003) Variation in immune defence as a question of evolutionary ecology, *Proceedings of the Royal Society B: Biological Sciences,* 270: 357-366.

Schopen, G., Bovenhuis, H., Visker, M. and van Arendonk, J. (2008) comparison of information content for microsatellites and SNPs in poultry and cattle, *Animal Genetics,* 39: 451-453.

Selkoe, K. and Toonen, R. (2006) Microsatellites for ecologist: a practical guide for using and evaluating microsatellite markers, *Ecology Letters,* 9: 615-629.

Shabalina, S., Spiridonov, N. and Kashina, A. (2013) Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity, *Nucleic Acids Research,* 41: 2073-2094.

Shaid, S., Brandts, C., Serve., H and Dikic, I. (2013) Ubiquitination and selective autophagy. *Cell Death and Differentiation,* 20; 21-30.

Shannon, L., Boyko, R., Castelhano, M., Corey, E., Hayward, J., McLean, C., White, M., Said, M., Anita, B., Bondjengo, N., Calero, J., Galov, A., Hedimbi, M., Imam, B., Khalap, R., Lally, D., Masta, A., Oliveira, K., Pérez, L., Randall, J., Tam, N., Trujillo-Cornejo, F., Valeriano, C., Sutter, N., Todhunter, R., Bustamante, C. and Boyko, A. (2015) Genetic structure in village dogs reveal a Central Asian domestication origin, *Proceedings of the National Academy of Sciences of the United States of America,* 112: 13639-13644.

Shearman, J. and Wilton, A. (2011) Origins of the Domestic Dog and the Rich Potential for Gene Mapping, *Genetic Research International,* 2011: 1-6.

Sheikh, S., Berger-Wolf, T., Ashley, M., Caballero, I., Chaovalitwongse, W. and DasGupta, B. (2008) Error tolerant sibship reconstruction in wild populations, *Computational Systems Bioinformatics Conference,* 7: 273-284.

Sheldon, B. and Verhulst, S. (1996) Ecological immunology: costly parasite defences and trade-offs in evolutionary ecology, *Trends in Ecology & Evolution,* 11: 317-321.

Sipel, A., Bejerano, G., Pedersen, J., Hinrichs, A., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L., Richards, S., Weinstock, G., Wilson, R., Gibbs, R., Kent, W., Miller, W. and Haussler, D. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes, *Genome Research,* 15: 1034-1050.

Skoglund, P., Ersmark, E., Palkopoulou, E., Dalen, L. (2015) Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds, *Current Biology,* 25: 1515-1519.

Slack, A., Thorton, P., Magner, D., Rosenberg, S. and Hastings, P. (2006) On the mechanism of gene amplification induced under stress in *Escherichia coli, PLoS Genetics*, 2: e48.

Smith, D., Meier, T., Geffen, E., Mech, L., Burch, J., Adams, L. and Wayne, R. (1997) Is incest common in gray wolf packs? *Behavioural Ecology,* 8: 384-391.

SoftGenetics. (2017) *Gene Marker*. Available: http://www.softgenetics.com/GeneMarker.php. Last accessed 10/10/2017.

Sorci, G., MØller, A. and Boulinier, T. (1997) Genetics of host-parasite interactions, *Trends in Ecology & Evolution,* 12: 196-200.

Sparkes, J., Körtner, G., Ballard, G., Fleming, P. and Brown, W. (2014) Effects of sex and reproductive state on interactions between free-roaming domestic dogs, *PLoS One,* 9: 1-13.

Stenglein, J., Waits, L., Ausband, D., Zager, P. and Mack, C. (2011) Estimating gray wolf pack size and family relationships using noninvasive genetic sampling at rendezvous sites, *Journal of Mammalogy,* 92: 784-795.

Stephens, S. (1951) Possible significance of duplication in evolution, *Advanced Genetics,* 4: 247-265.

Steyaert, S., EndrestØl, A., Hackländer, K., Swenson, J. and Zedrosser, A. (2011) The mating system of the brown bear *Ursus arctos*, *Mammal Review,* 42: 12-34.

Stronen, A., Forbes, G., Paquet, P., Goulet, G., Sallows, T. and Musiani, M. (2012) Dispersal in a plain landscape: short distance genetic differentiation in southwestern Manitoba wolves, Canada, *Conservation Genetics,* 13: 359-371.

Su, A., Cooke, M., Ching, K., Hakak, Y., Walker, J., Wiltshire, T., Orth, A., Vega, R., Sapinoso, L., Moqrich, A., Patapoutlan, A., Hampton, G., Schultz, P. and Hogenesch, J. (2002) Large-scale analysis of the human and mouse transcriptomes, *Proceedings of National Academy of Sciences of the United States of America,* 99: 4465-4470.

Suenaga, E. and Nakamura, H. (2005) Evaluation of three methods for effective extraction of DNA from human hair, *Journal of Chromatography B,* 820: 137-141.

Szabo, C., Masiello, A., Ryan, J. and Brody, L. (2000) The breast cancer information core: database design, structure, and scope. *Human Mutations,* (2): 123-131.

Szigyarto, C., Sibbons, P., Williams, G., Uhlen, M. and Metcalfe, S. (2010) The E3 Ligase Axotrophin/*MARCH-7*: protein expression profiling of human tissues reveals links to adult stem cells. *Journal of Histochemistry and Cytochemistry,* 58: 301-308.

Takayanagi, K., Asamura, H., Tsukada, K., Ota, M., Saito, S. and Fukushima, H. (2003) Investigation of DNA extraction from hair shafts, *International Congress Series,* 1239: 759-764.

Těšický, M. and Vinkler, M. (2015) Trans-species polymorphism in immune genes: general pattern of MHC-Restricted phenomenon, *Journal of Immunology Research,* 2015: 1-10.

Thachuk, C., Shmygelska, A. and Hoos, H. (2007) A replica exchange Monte Carlo algorithm for protein folding in the HP model, *BMC Bioinformatics,* 8: 342.

Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V., Sawyer, S., Greenfield, D., Germonpré, M., Sablin, M., Lóperz-Girálex, F., Domingo-Roura, X., Napierala, H., Uerpmann, H., Loponte, M., Acosta, A., Giemsch, L., Schmitz, R., Worthington, B., Buikstra, J., Druzhkova, A., Graphodatsky, A., Ovodov, N., Wahlberg, N., Freedman, A., Schweizer, R., Koepfli, K., Leonard, J., Meyer, M., Krause, J., Pääbo, S., Green, R. and Wayne, R. (2013) Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs, *Science,* 342: 871-874.

Tokarska, M., Marshall, T., Kowalczyk, R., Wójcik, J., Pertoldi, C., Kristensen, T., Loeschcke, V., Gregersen, V. and Bendixen, C. (2009) Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison, *Heredity,* 103: 326-332.

Tompa, P. (2002) Intrinsically unstructured proteins, *Trends in Biochemical Science*, 27: 527-533.

Tonegawa, S., Steinberg, C., Dube, S. and Bernardini, A. (1974) Evidence for somatic generation of antibody diversity, *Proceedings of the National Academy of Sciences of the United States of America*, 71: 4027-4031.

Tonteri, A., Vasemägi, A., Lumme, J. and Primmer, C. (2010) Beyond MHC: signals of elevated selection pressure on Atlantic salmon (*Salmo salar*) immune-relevant loci, *Molecular Ecology*, 19: 1273-1282.

Totton, S., Wandeler, A., Zinsstag, J., Bauch, C., Ribble, C., Rosatte, R. and McEwen, S. (2010) Stray dog population demographics in Jodhpur, India following a population control/rabies vaccination program, *Preventive Veterinary Medicine*, 97: 51-57.

Treangen, T. and Salzberg, S. (2011) Repetitive DNA and next-generation sequencing: computational challenges and solutions, *Nature Review Genetics*, 13: 36-46.

Trivedi, M., Laurence, J. and Siahaan, T. (2009) The role of thiols and disulfides in protein chemical and physical stability, *Current protein & peptide science*, 10: 614-625.

Troshin, P., Procter, J. and Barton, G. (2011) Java bioinformatics analysis web services for multiple sequence alignment – JABAWS: MSA. *Bioinformatics*, 27: 2001-2002.

Turner, A., Begon, Jackson, J., Bradley, H. and Paterson, S. (2011) Genetic diversity in cytokines associated with immune variation and resistance to multiple pathogens in a natural rodent population, *PLoS Genetics*, 7: e1002343.

Turner, A., Begon, M., Jackson, J. and Paterson, S. (2012) Evidence for selection at cytokine loci in a natural population of field voles (*Microtus agrestis*). *Molecular Ecology*, 21: 1632-1646.

Tóth, G., Gáspári, Z. and Jurka, J. (2000) Microsatellites in different eukaryotic genomes: survey and analysis, *Genome Research*, 10: 967-981.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B., Remm, M. and Rozen, S. (2012) Primer3 – new capabilities and interfaces, *Nucleic Acids Research*, 40: e115.

Van Baalen, M. (1998) Coevolution of recovery ability and virulence, *Proceedings: Biological Sciences*, 265: 317-325.

Van Ballenberghe, V. (1983) Two litters raised in one year by a wolf pack, *Journal of Mammalogy,* 64: 171-172.

Van Oosterhout, C., Hutchinson, W., Wills, D. and Shipley, P. (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data, *Molecular Ecology Notes*, 4: 535-538.

Vanak, A. and Gompper, M. (2009) Dietary niche separation between sympatric free-ranging domestic dogs and Indian foxes in central India, *Journal of Mammalogy,* 90: 1058-1065.

Vanak, A., Thaker, M and Gompper, M. (2009) Experiemntal examination of behavioural interactions between free-ranging wild and domestic canids, *Behavioural Ecology and Sociobiology,* 64: 279-287.

Vaysse, A., Ratnakumar, A., Derrien, T., Axelsson, E., Pielberg, G., Sigurdsson, S., Fall, T., Seppala, E., Hansen, M., Lawley, C., Karlsson, E., The LUPA Consortium, Bannasch, D., Vila, C., Lohi, H., Galibert, F., Fredholm, M., Haggstrom, J., Hedhammar, A., Andre, C., Lindblad-Toh, K., Hitte, C. and Webster, M. (2011) Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping, *PLoS Genetics,* 7: e1002316.

Veradri, A., Lucchini, V. and Randi, E. (2006) Detecting introgressive hybridization between free-ranging domestic dogs and wild wolves (Canis lupus) by admixture linkage disequilibrium analysis, *Molecular Ecology,* 15: 2845-2855.

Vila, C., Maldonado, J., and Wayne, R. (1999) Phylogenetic relationships, evolution, and genetic diversity of the domestic dog, *The American Genetic Association*, 90: 71-77.

Vilà, C., Savolainen, P., Maldonado, J., Amorim, I., Rice, J., Honeycutt, R., Crandall, K., Lundeberg, J. and Wayne, R. (1997) Multiple and ancient origins of the domestic dog, *Science,* 276: 1687-1689.

Vilà, C., Seddon, J. and Ellegren, H. (2005) Genes of domestic mammals augmented by backcrossing with wild ancestors, *Trends in Genetics,* 21: 214-218.

vonHoldt, B., Pollinger, J., Lohmueller, K., Han, E., Parker, H., Quignon, P., Degenhardt, J., Boyko, A., Earl, D., Auton, A., Reynolds, A., Bryc, K., Brisbin, A., Knowles, J., Mosher, D., Spady, T., Elkahloun, A., Geffen, E, Pilot, M., Jedrzejewski, W., Greco, C., Randi, E.,

Bannasch, D., Wilton, A., Shearman, J., Musiani, M., Cargill, M., Jones, P., Qian, Z., Huang, W., Ding, Z., Zhang, Y., Bustamante, C., Ostrander, E, Novembre, J. and Wayne, R. (2010) Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication, *Nature,* 464: 898-902.

Vonholdt, B., Stahler, D, Smith, D., Earl, D., Pollinger, J. and Wayne, R. (2008) The genealogy and genetic viability of reintroduced Yellowstone grey wolves, *Molecular Ecology,* 17: 252-274.

Wagner, A. (2005) Distributed robustness versus redundancy as causes of mutational robustness, *Bioessays*, 27: 176-188.

Wang, G., Zhai, W., Yang, H., Wang, L., Zhong, L., Liu, Y., Fan, R., Yin, T., Zhu, C., Poyarkov, A., Irwin, D., Hytönen, M., Lohi, H., Wu, C., Savlainen, P. and Zhang, Y. (2016) Out of southern East Asia: the natural history of domestic dogs across the world, *Cell Research,* 26: 21-33.

Wang, W. and Kirkness, E. (2005) Short interspersed elements (SINEs) are a major source of canine genomic diversity, *Genome Research,* 15: 1798-1808.

Wang, X., Tedford, R., Van Valkenburgh, B. and Wayne, R. (2007). Evolutionary history, molecular systematics, and evolutionary ecology of Canidae. In: MacDonald, D. and Sillero-Zubiri, C. *Biology and conservation of wild canids.* Oxford: OUP Oxford. 39-42.

Wang, Z. and Burge, C. (2008) Splicing regulation: From a parts list of regulatory elements to an integrated splicing code, *RNA,* 14: 802-813.

Ward, A. and Cooper, T. (2010) The Pathobiology of Splicing, *Journal of Pathology,* 220: 152-163.

Waterhouse, A., Procter, J., Martin, D., Clamp, M. and Barton, G. (2009) Jalview Version 2 – a multiple sequence alignment editor and analysis workbench, *Bioinformatics,* 25: 1189-1191.

Wattier, R., Engel, C., Saumitou-Laprade, P. and Valero, M. (1998) Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta), *Molecular Ecology*, 7: 1569-1573.

Wedekind, C. and Folstad, I. (1994) Adaptive or nonadaptive immunosuppression by sex hormones. *American Naturalist*, 143: 936–938.

Werner, F., Durstewitz, G., Habermann, F., Thaller, G., Kramer, W., Kollers, S., Buitkamp, J., Georges, M., Brem, G., Mosner, J. and Fries, R. (2004) Detection and characterisation of SNPs useful for identity control and parentage testing in major European dairy breeds, *Animal Genetics,* 35: 44-49.

Wetton, J., Higgs, J., Spriggs, A., Roney, C., Tsang, C. and Foster, A. (2003) Mitochondrial profiling of dog hairs, *Forensic Science International,* 133: 235-241.

Wilson, M., Polanskey, D., Butler, J., DiZinno, J., Replogle, J. and Budowle, B. (1995) Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shafts, *BioTechniques* 18: 662-664.

Witt, K., Judd. K., Kitchen, A., Grier, C., Kohler, T., Ortman, S., Kemp, B. and Malhi, R. (2015) DNA analysis of ancient dogs of the Americas: Identifying possible founding haplotypes and reconstructing population histories, *Journal of Human Evoltuon,* 79: 106-118.

Wolfe, K. and Li, W. (2003) Molecular evolution meets the genomics revolution, *Nature Genetics,* 33: 255-265.

Wong, W., Yang, Z., Goldman, N. and Nielsen, R. (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites, *Genetics,* 168: 1041-1051.

Woolhouse, M., Webster, J., Domingo, E., Charlesworth, B. and Levin, B. (2002) Biological and biomedical implications of the co-evolution of pathogens and their hosts, *Nature Genetics,* 32: 569-577.

Wright, D. (2015) The genetic architecture of domestication in animals, *Bioinformatics and Biology Insights,* 9: 11-20.

Xia, X. (1998) The rate heterogeneity of nonsynonymous substitutions in mammalian mitochondrial genes, *Molecular Biology and Evolution,* 15: 336-344.

Xing, Y. and Lee, C. (2006) Alternative splicing and RNA selection pressure – evolutionary consequences for eukaryotic genomes, *Nature Review Genetics,* 7: 499-509.

Xu, D. and Zhang, Y. (2012) *Ab initio* protein structure assembly using continuous structure fragments and optimised knowledge-based force field, *Proteins,* 80: 1715-1735.

Yabas, M., Elliott, H. and Hoyne, G. (2016) The Role of Alternative Splicing in the Control of Immune Homeostasis and Cellular Differentiation, *International Journal of Molecular Sciences*, 17: 3.

Yang, Z. (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution, *Molecular Biology and Evolution,* 15: 568-573.

Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood, *Molecular Biology and Evolution,* 24: 1586-1591.

Yang, Z. and Nielsen, R. (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models, *Molecular Biology and Evolution,* 17: 32-43.

Yang, Z. and Nielsen, R. (2002) Codon-substitution models for detecting molecular adaption at individual sites along specific lineages, *Molecular Biology and Evolution,* 19: 908-917.

Yang, Z., Nielsen, R., Goldman, N., Pedersen, A. (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites, *Genetics,* 155: 431-439.

Yao, E. and Chuang, P. (2015) Hedgehog signaling: From basic research to clinical applications, *Journal of the Formosan Medical Association,* 114: 569-576.

Yoshii, T., Tamura, K. and Ishiyama, I. (1992) Presence of a PCR-inhibitor in hairs, *Japanese Journal of Legal Medicine,* 46: 313-316.

Yoshii, T., Tamura, K., Taniguchi, T., Akiyama, K. and Ishiyama, I. (1993) Water-soluble eumelanin as a PCR-inhibitor and a simple method for its removal, *Japanese Journal of Legal Medicine,* 47: 323-329.

Yu, G., Tang, Q., Long, K., Chex, T., Li, M. and Shaui, S. (2015) Effectiveness of microsatellite and single nucleotide polymorphism markers for parentage analysis in European domestic pigs, *Genetic Molecular Research,* 14: 1362-1370.

Zhao, B., Ito, K., Iyengar, P., Hirose, S., and Nakamura, N. (2013) *MARCH7* E3 ubiquitin ligase is highly expressed in developing spermatids of rats and its possible involvement in head and tail formation. *Histochemistry and Cell Biology,* 139: 447-460.

Zimen, E. (1976) On the regulation of pack size in wolves, *Zeitschrift für Tierpsychologie*, 40: 300-341.

Zuk, M. (1990) Reproductive strategies and disease susceptibility: an evolutionary viewpoint. *Parasitology Today*, 6: 231-233.

Zuk, M. and McKean, K. (1996) Sex differences in parasite infections: Patterns and processes, *International Journal for Parasitology*, 26: 1009-1024.

Zuk, M. and Stoehr, A. (2002) Immune defence and host life history, *The American Naturalist*, 160: 9-22.

7. APPENDIX

7.1 ALLELE SIZES FOR PRIMER SET TTRAB

Table 32: Allele sizes for ttRAB primer set (2010, 107, 2054, 2088 and 253). Expected primer range seen in row 3, directly below primer name. Mothers are indicated by blue shading with all offspring listed directly below. Males clearly separated by thick black line.

Individual		2010		2017		2054		2088		253	
		203-235		260-272		146-178		104-136		93-115	
Sofia	Mother	230	230	266	270	156	168	115	123	102	108
SO1	Offspring	226	230	266	270	156	172	115	123	102	108
SO2	Offspring	226	230	266	266	156	156	119	123	108	108
SO3	Offspring	230	230	266	266	168	172	119	123	102	108
SO4	Offspring	230	230	266	266	156	156	123	123	102	108
SO5	Offspring	230	230	266	270	168	172	123	123	108	108
SO6	Offspring	230	230	266	266	156	156	123	123	102	108
SO7	Offspring	230	230	266	266	168	172	123	123	108	108
Snella	Mother	0	0	0	0	156	168	115	123	0	0
SN1	Offspring	226	230	262	266	152	156	115	115	108	110
SN2	Offspring	230	238	262	270	152	168	115	127	108	108
SN3	Offspring	230	238	266	270	152	168	115	127	108	108
SN4	Offspring	226	230	262	266	152	156	123	127	108	108
SN5	Offspring	230	230	258	266	156	156	115	115	108	108
SN6	Offspring	230	230	266	270	168	168	115	119	108	108
SN7	Offspring	230	230	266	266	156	156	123	123	102	108
SN8	Offspring	230	238	266	270	156	156	123	127	108	110
SN9	Offspring	230	238	266	270	152	156	115	127	108	108
Emma	Mother	226	230	266	266	156	156	123	123	102	108
EM1	Offspring	226	238	266	266	156	156	123	123	102	102
EM2	Offspring	226	238	266	266	156	156	123	127	102	102
Petto	Male	0	0	0	0	156	172	0	0	102	108
Duca	Male	230	238	262	270	152	168	123	127	102	108
Bo	Male	226	230	266	266	156	172	119	123	108	108
Antonio	Male	230	230	266	266	156	168	115	123	108	108
Spider	Male	230	230	266	266	156	172	119	123	108	108
Fred	Male	226	230	262	266	156	156	123	127	108	110
Angelo	Male	230	230	266	270	156	168	119	123	108	108
Bernardo	Male	230	230	266	266	156	156	119	123	108	108

7.2 ALLELE SIZES FOR PRIMER SET TTRC

Table 33: Allele positions for ttRC primer set (2096, VwF, 250 and 213). Expected primer range seen in row 3, directly below primer name. Mothers are indicated by blue shading with all offspring listed directly below. Males clearly separated by thick black line.

Individual		Ttrc							
		2096		VwF		250		213	
		88-104		129-189		122-144		136-172	
Sofia	Mother	102	106	157	157	134	140	157	157
SO1	Offspring	98	106	157	157	134	134	157	157
SO2	Offspring	98	102	149	157	134	140	157	157
SO3	Offspring	98	106	149	157	134	140	157	157
SO4	Offspring	98	102	149	157	0	0	157	157
SO5	Offspring	98	102	157	157	134	134	157	157
SO6	Offspring	98	102	157	157	0	0	157	157
SO7	Offspring	98	102	157	157	134	134	155	157
Snella	Mother	102	106	157	169	0	0	0	0
SN1	Offspring	98	102	157	157	134	134	157	157
SN2	Offspring	102	106	157	169	0	0	0	0
SN3	Offspring	0	0	0	0	0	0	0	0
SN4	Offspring	102	106	157	169	0	0	157	157
SN5	Offspring	102	102	157	157	140	140	155	157
SN6	Offspring	102	106	0	0	134	134	163	163
SN7	Offspring	102	102	157	157	134	140	155	157
SN8	Offspring	102	102	169	169	0	0	157	157
SN9	Offspring	102	106	157	169	0	0	161	161
Emma	Mother	98	102	147	157	0	0	157	157
EM1	Offspring	102	102	157	157	0	0	157	157
EM2	Offspring	102	102	147	157	0	0	159	159
Petto	Male	98	102	157	157	138	138	155	157
Duca	Male	98	98	157	157	140	140	157	157
Bo	Male	98	98	157	163	134	140	155	161
Antonio	Male	102	106	157	169	140	142	157	161
Spider	Male	98	102	157	163	138	142	157	157
Fred	Male	102		102	157	169	134	138	157	157
Angelo	Male	106		106	157	169	138	142	157	169
Bernardo	Male	102		106	169	169	134	140	157	161

7.3 ALLELE SIZES FOR PRIMER SET TTRC

Table 34: Allele positions for diC primer set (2001, AHT130 and 466). Expected primer range seen in row 3, directly below primer name. Mothers are indicated by blue shading with all offspring listed directly below. Males clearly separated by thick black line.

Individual		dic					
		2001		AHT130		466	
		129-149		108-124		139-163	
Sofia	Mother	133	155	116	116	148	152
SO1	Offspring	133	145	112	116	148	160
SO2	Offspring	133	155	112	116	152	160
SO3	Offspring	133	133	112	116	152	160
SO4	Offspring	133	155	112	116	150	152
SO5	Offspring	133	155	112	116	152	160
SO6	Offspring	133	133	112	116	148	160
SO7	Offspring	133	155	112	116	148	150
Snella	Mother	133	155	110	116	150	150
SN1	Offspring	0	0	110	116	150	150
SN2	Offspring	133	133	110	110	150	150
SN3	Offspring	133	155	110	110	148	150
SN4	Offspring	133	155	110	116	150	150
SN5	Offspring	145	145	114	116	150	150
SN6	Offspring	133	155	108	116	150	150
SN7	Offspring	133	155	112	116	150	152
SN8	Offspring	133	155	110	116	150	150
SN9	Offspring	133	149	110	110	148	150
Emma	Mother	0	0	110	112	152	152
EM1	Offspring	133	133	112	116	150	152
EM2	Offspring	133	155	108	110	152	160
Petto	Male	133	155	110	116	150	152
Duca	Male	133	133	108	114	150	150
Bo	Male	133	145	112	112	150	160
Antonio	Male	155	155	116	116	148	150
Spider	Male	**133**	133	116	116	150	160
Fred	Male	133	149	110	116	150	150
Angelo	Male	145	155	116	116	148	152
Bernardo	Male	133	133	110	112	152	160

7.4 FREQUENCY AND DISTRIBUTION OF ALLELES

Allele frequency and the number of heterozygotes and homozygotes for individual alleles are provided below for multiplex set ttRAB.

7.4.1 Locus one (Primer 2010)

Three alleles were amplified for locus one, but one allele (230) predominated in most genotypes (Figure 7-1). Allele 230 was the only one to be present in both heterozygote and homozygote form, with the other two alleles only showing heterozygosity (Figure 7-1).

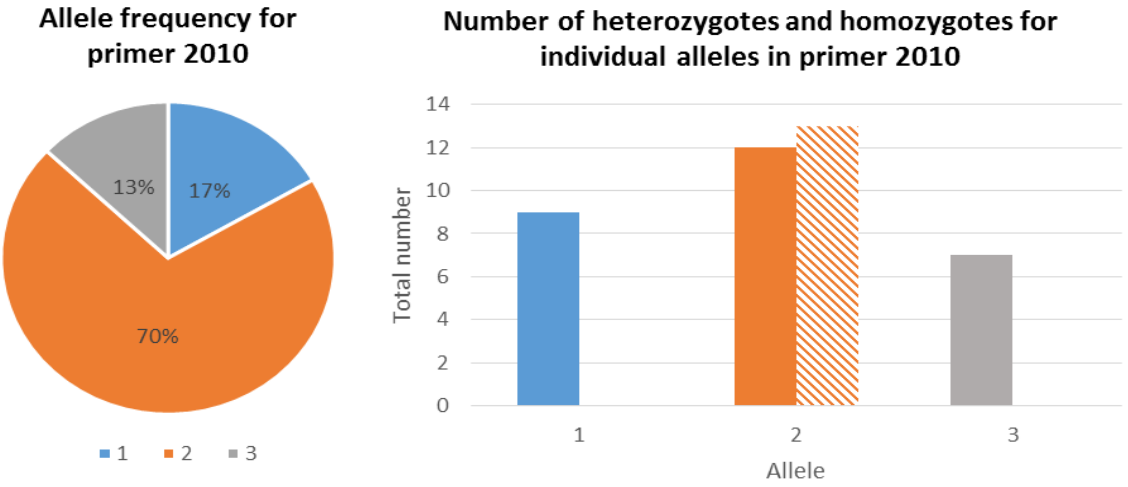


Figure 7-1: Allele frequency and number of heterozygotes and homozygotes for Locus one (Primer 2010)

7.4.2 Locus two (Primer 2017)

Four alleles were amplified for locus two, but one allele (266) predominated in most genotypes and allele 258 was seen in just one genotype (Figure 7-2). Allele 266 was the only allele to be present in both heterozygote and homozygote form with all other alleles showing just heterozygosity (Figure 7-2).

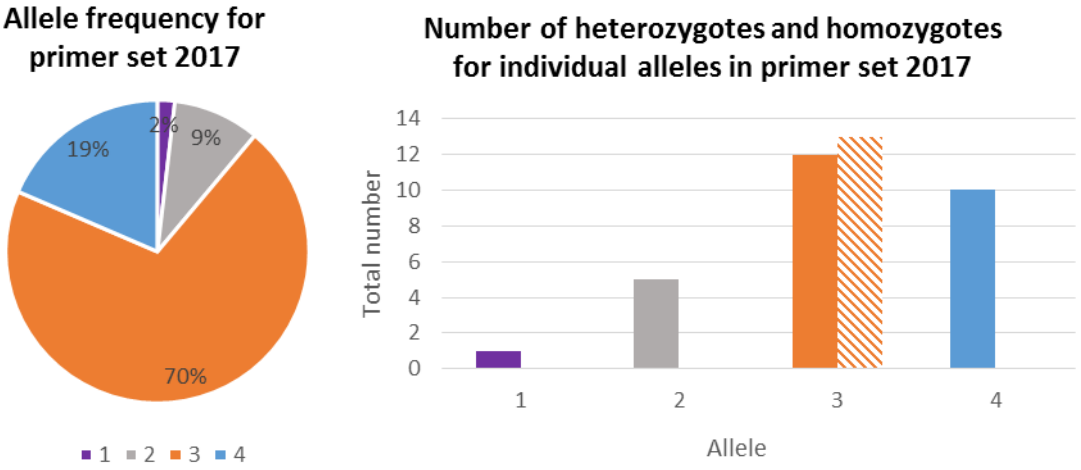


Figure 7-2: Allele frequency and number of heterozygotes and homozygotes for Locus two (Primer 2017)

7.4.3 Locus three (Primer 2054)

Four alleles were amplified for locus three, with allele 156 representing slightly more than half of the genotypes (Figure 7-3). Two out of four alleles presented homozygote and heterozygote forms, allele 156 demonstrating equal distribution whilst allele 168 was homozygous in just one genotype. The remaining two alleles were found just in heterozygote form (Figure 7-3).

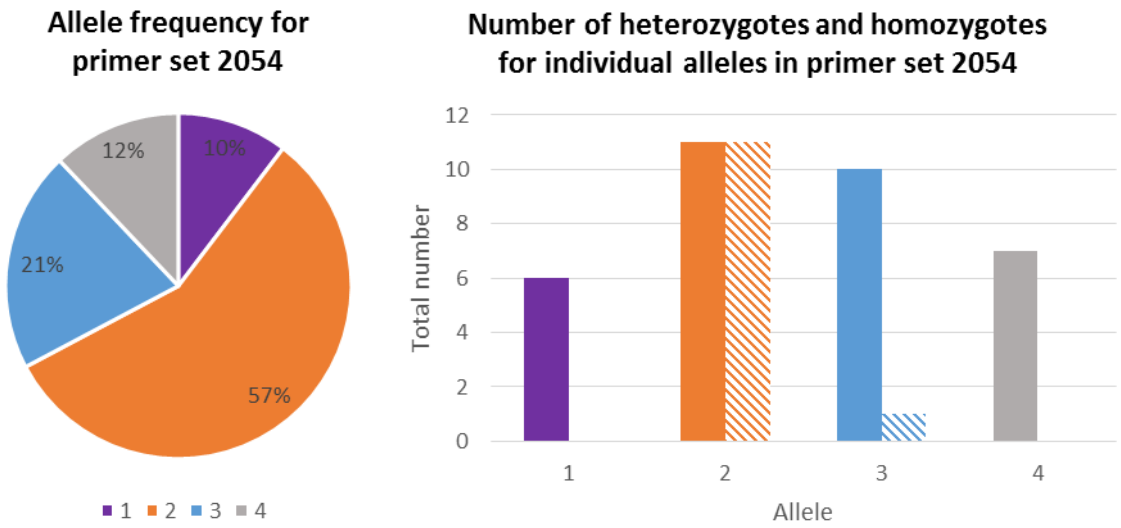


Figure 7-3: Allele frequency and number of heterozygotes and homozygotes for Locus three (Primer 2054)

7.4.4 Locus four (Primer 2088)

Four alleles were amplified for locus three, with allele 123 representing slightly more than half of the genotypes (Figure 7-4). Two out of four alleles presented both homozygote and heterozygote forms but heterozygosity was more frequent in both alleles. The remaining two alleles were found just in heterozygote form (Figure 7-4).

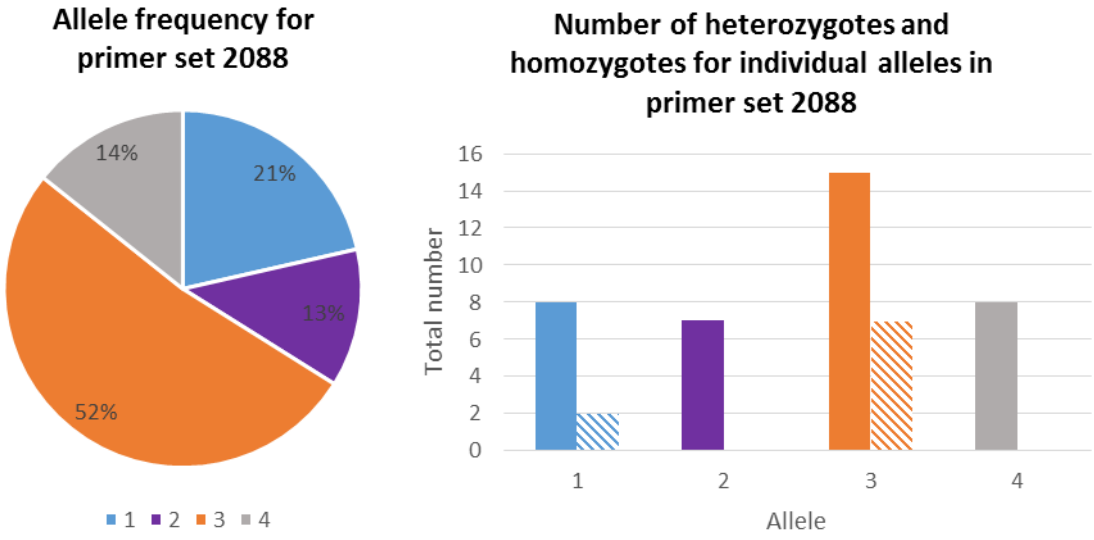


Figure 7-4: Allele frequency and number of heterozygotes and homozygotes for locus four (Primer 2088)

7.4.5 Locus five (Primer 253)

Three alleles were amplified for locus five, with allele 108 predominating almost three quarters of all genotypes (Figure 7-5). Allele 108 was more commonly found in homozygote form, whilst allele 102 was more commonly present in heterozygote form and allele 110 was infrequent across genotypes and only found in homozygote form (Figure 7-5).

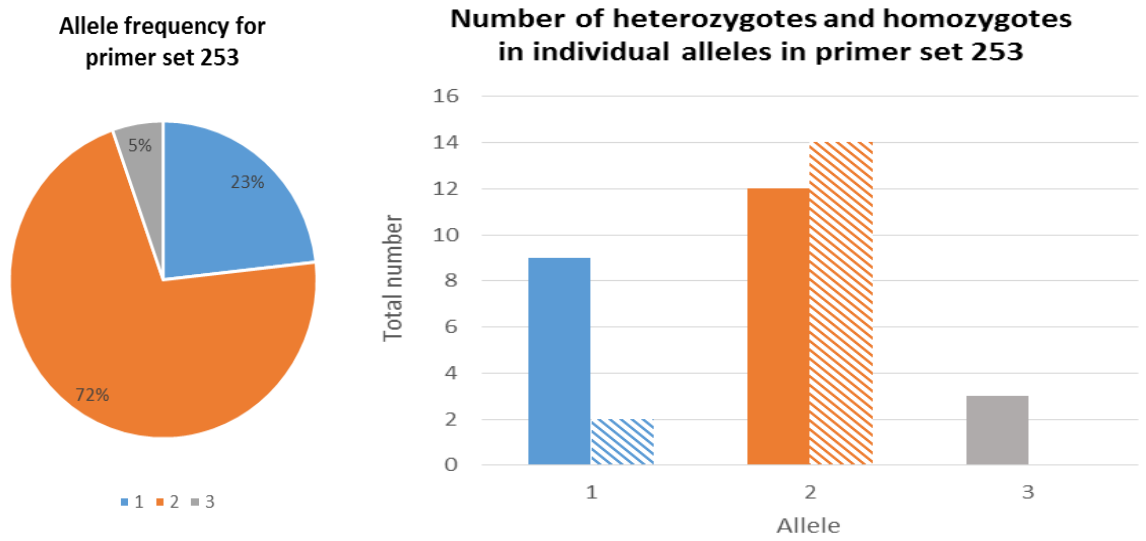


Figure 7-5: Allele frequency and number of heterozygotes and homozygotes at locus five (Primer 253)

7.4.6 Locus six (Primer 2096)

Three alleles were amplified for locus five, with approximately half of all genotypes represented by allele 102 (Figure 7-6). Allele 106 was only found in heterozygote form whilst the other two alleles present were found in both heterozygote and homozygote form (Figure 7-6).

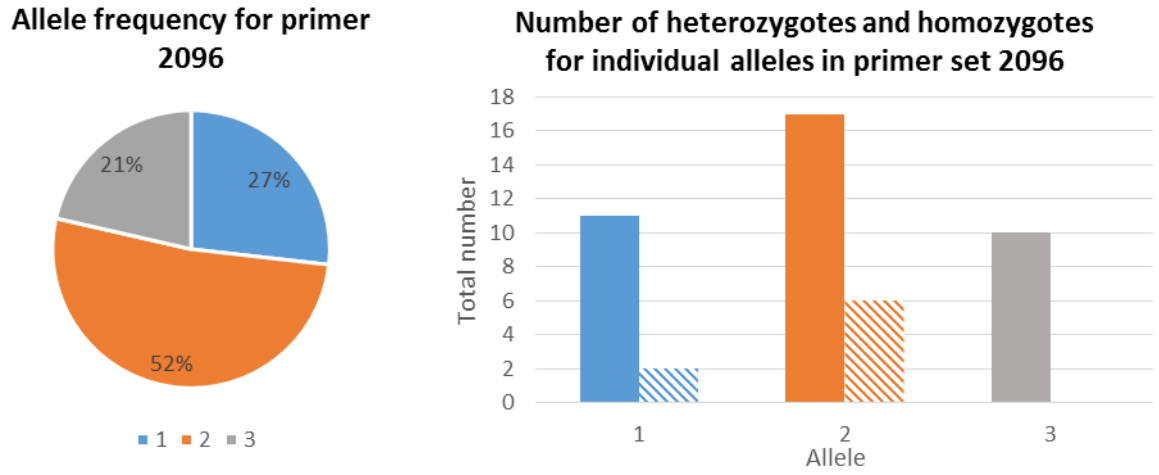


Figure 7-6: Allele frequency and number of heterozygotes and homozygotes for locus 7 (Primer 2096)

7.4.7 Locus seven (Primer VwF)

Five alleles were amplified for locus seven, with allele 157 predominating most genotypes, with three alleles present in low frequency (Figure 7-7). Allele 157 represents the only allele to be found in both heterozygote and homozygote form (Figure 7-7).

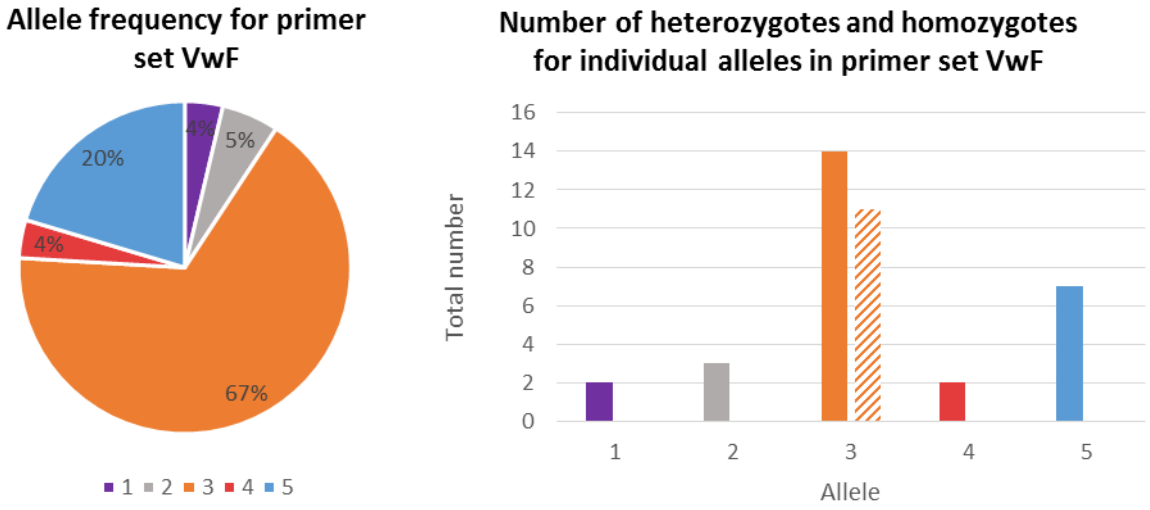


Figure 7-7: Allele frequency and number of heterozygotes and homozygotes for locus 7 (Primer VwF)

7.4.8 Locus eight (Primer 250)

Five alleles were amplifed for locus eight, with three quarters of genotypes represented by two alleles (Figure 7-8). Allele 142 was the only allele to be soley found in heterzoygote form. All other alleles were found in both heterzygote and homozygote but heterozygosity was consistenytly found to be more frequent (Figure 7-8).

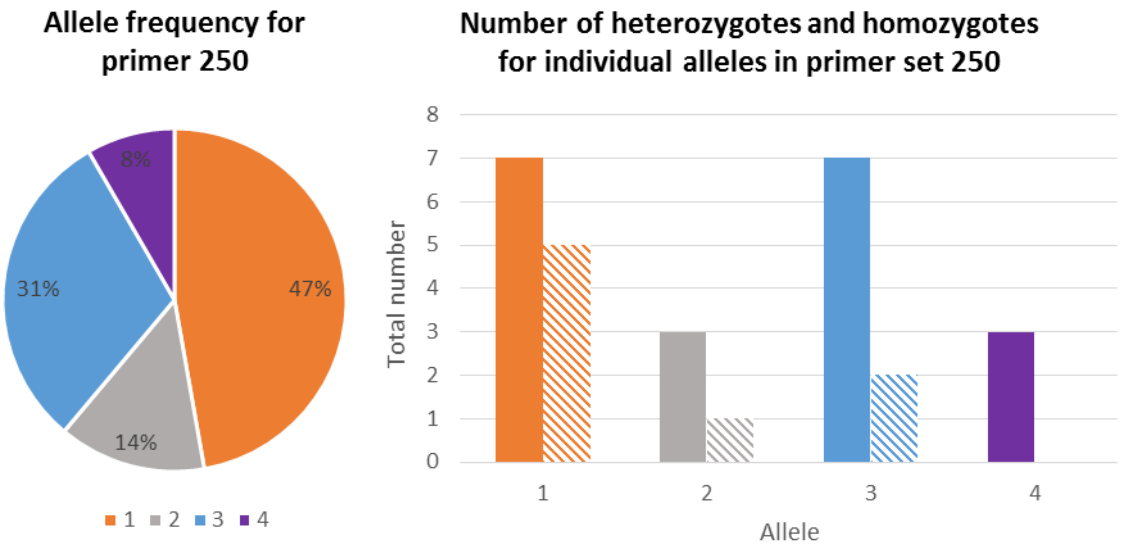


Figure 7-8: Allele frequency and number of heterozygotes and homozygotes for Locus eight (Primer 250)

7.4.9 Locus nine (Primer 213)

Locus nine was the most diverse with six alleles amplified, almost three quarters of genotypes presented allele 157 (Figure 7-9), which was found to be homozygote more than double the amount of times it was heterozygote. Locus nine was the only locus found to have alleles present in only homozygote form, seen in alleles 159 and 163 (Figure 7-9).

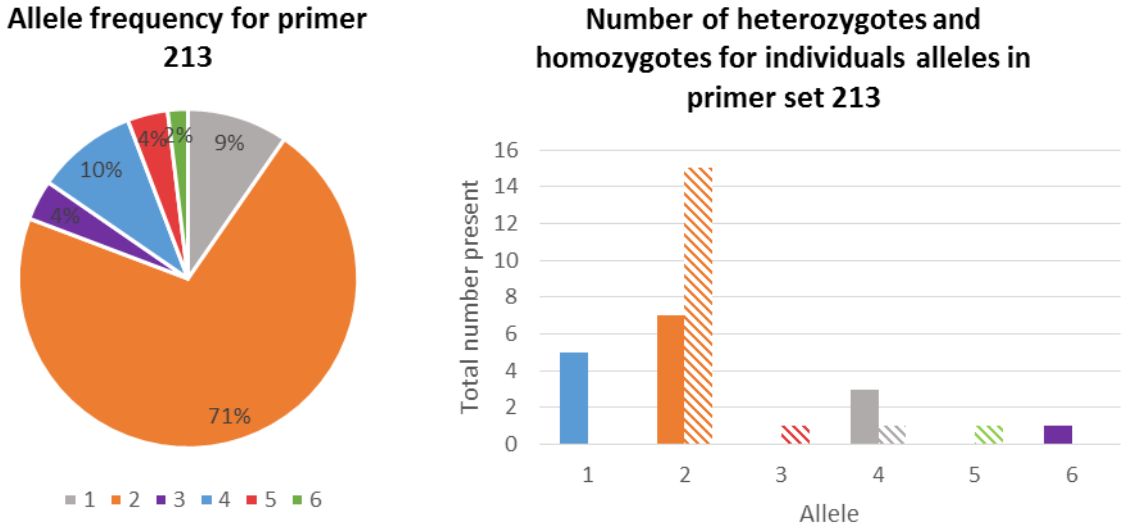


Figure 7-9: Allele frequency and number of heterozygotes and homozygotes for Locus nine (Primer 213)

7.4.10 Locus ten (Primer 2001)

Four alleles were amplified for locus ten with over three quarters of all genotypes represented by two alleles (Figure 7-10). Allele 133 was the most predominant and was more frequently found in heterozygote form. Across all alleles frequency of homozygotes was found to be considerably lower, with allele showing no homozygosity (Figure 7-10).

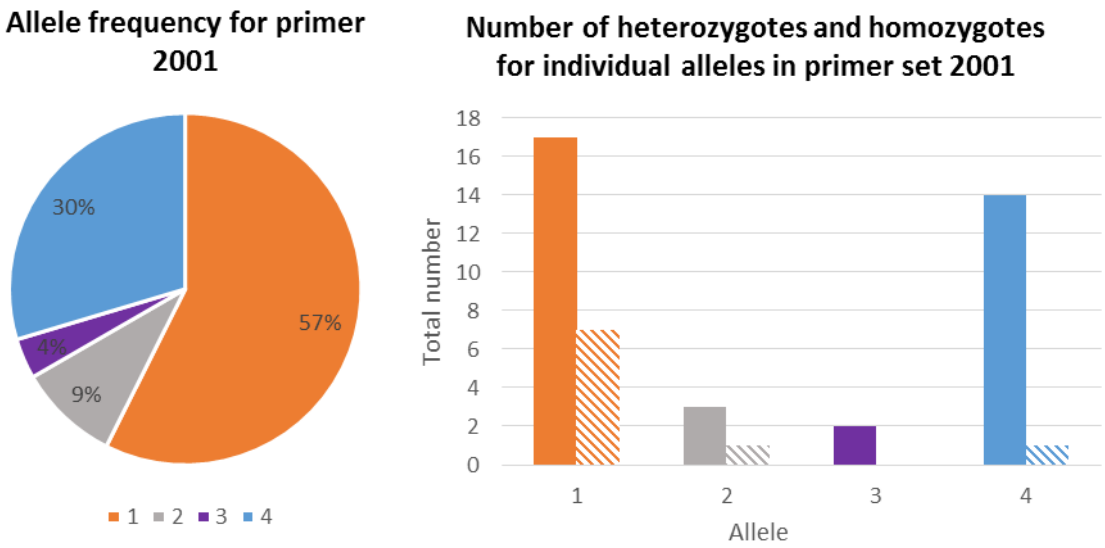


Figure 7-10: Allele frequency and number of heterozygotes and homozygotes for locus ten (Primer 2001)

7.4.11 Locus eleven (Primer AHT130)

Five alleles were amplified for locus eleven with three common alleles and two rare alleles, found in less than 10% of all genotypes (Figure 7-11). Heterozygosity was more frequent in all alleles, with two alleles showing no homozygosity in any genotype (Figure 7-11).

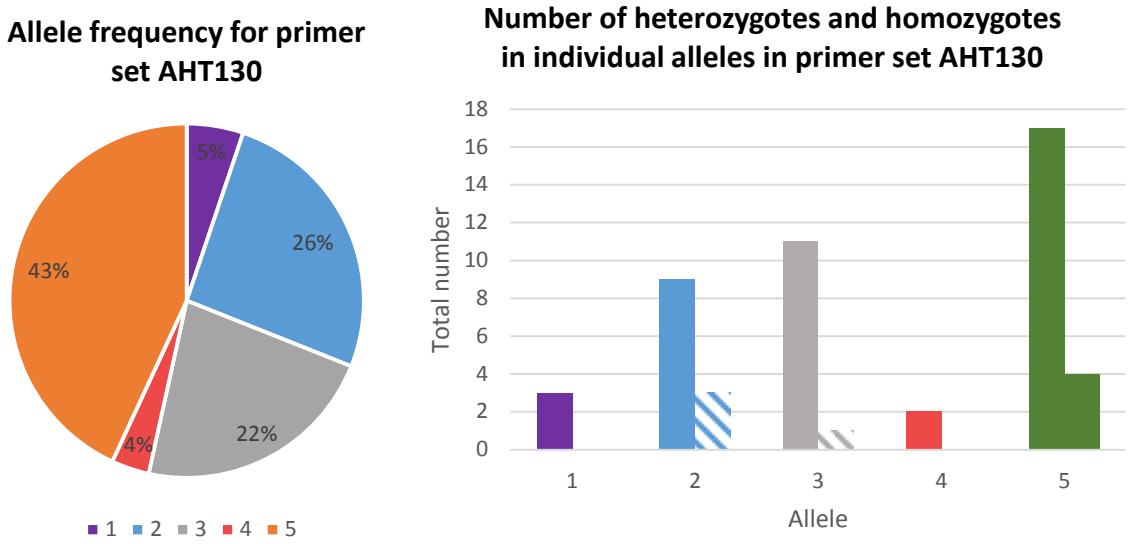


Figure 7-11: Allele frequency and number of heterozygotes and homozygotes for locus eleven (Primer AHT130)

7.4.12 Locus twelve (Primer set 466)

Four alleles were amplified for locus twelve, where almost half of all genotypes is predominated by one allele (Figure 7-12). Frequency of heterozygotes and homozygotes is almost equal for allele 150 but for all others frequency of homozygotes is lower or not witnessed at all (Figure 7-12).

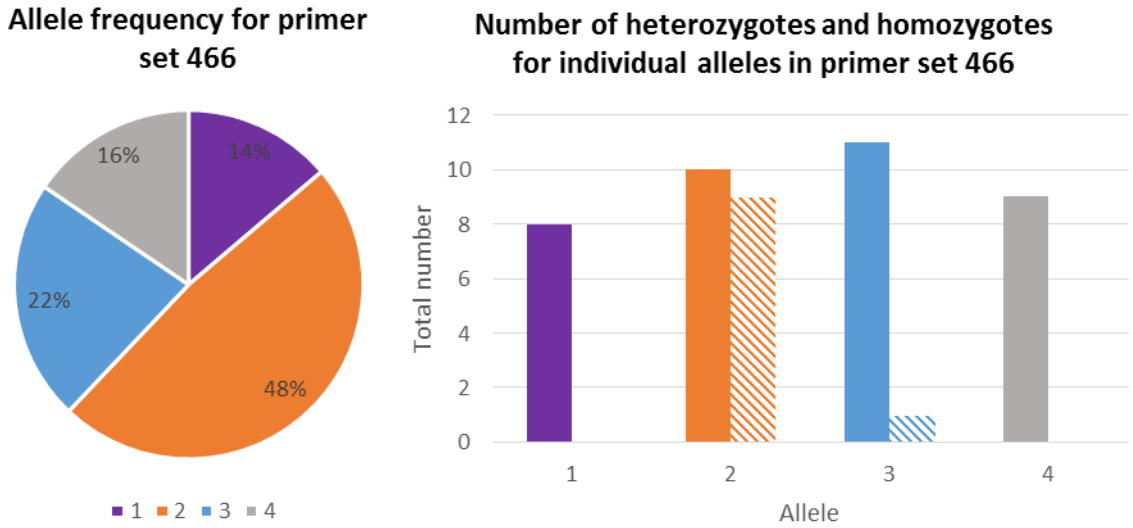


Figure 7-12: Allele frequency and number of heterozygotes and homozygotes for locus twelve (Primer 466)

7.5 ALLELE FREQUENCY TABLES

7.5.1 Locus one (Primer 2010)

Table 35: Allele frequency for locus one (Primer 2010)

Allele	Count	Hets	Homs	Freq	Freq with null
226	9	9	0	0.1667	0.1831
230	38	12	13	0.7037	0.7242
238	7	7	0	0.01296	0.139

Number of individuals typed:	27
Heterozygotes:	14
Homozygotes:	13
Number of alleles:	3
Observed heterozygosity:	0.5185
Expected heterozygosity	0.4689
Polymorphic information content (PIC):	0.4151
Average non-exclusion probability (first parent):	0.8941
Average non-exclusion probability (second parent):	0.7573
Average non-exclusion probability (parent pair):	0.6145
Average non-exclusion probability (identity):	0.3365
Average non-exclusion probability (sib identity):	0.6040
Hardy-Weinberg equilibrium test:	Not completed
Null allele frequency estimate:	-0.0463

7.5.2 Locus two (Primer 2017)

Table 36: Allele frequency for locus two (Primer 2017)

Allele	Count	Hets	Homs	Freq	Freq with null
258	1	1	0	0.0185	0.0187
262	5	5	0	0.0926	0.0971
266	38	12	13	0.7037	0.7243
270	10	10	0	0.1852	0.2060

Number of individuals typed:	27
Heterozygotes:	14
Homozygotes:	13
Number of alleles:	4
Observed heterozygosity:	0.5185
Expected heterozygosity	0.4703
Polymorphic information content (PIC):	0.4182
Average non-exclusion probability (first parent):	0.8908
Average non-exclusion probability (second parent):	0.7520
Average non-exclusion probability (parent pair):	0.6041
Average non-exclusion probability (identity):	0.3333
Average non-exclusion probability (sib identity):	0.6025
Hardy-Weinberg equilibrium test:	Not completed
Null allele frequency estimate:	-0.0460

7.5.3 Locus three (Primer 2054)

Table 37: Allele frequency for locus three (Primer 2054)

Allele	Count	Hets	Homs	Freq	Freq with null
152	6	6	0	0.1034	0.1092
156	33	11	11	0.5690	0.5073
168	12	10	1	0.2069	0.2117
172	7	7	0	0.1207	0.1288

Number of individuals typed:	29
Heterozygotes:	17
Homozygotes:	12
Number of alleles:	4
Observed heterozygosity:	0.5862
Expected heterozygosity	0.6189
Polymorphic information content (PIC):	0.5617
Average non-exclusion probability (first parent):	0.7974
Average non-exclusion probability (second parent):	0.6306
Average non-exclusion probability (parent pair):	0.4516
Average non-exclusion probability (identity):	0.2001
Average non-exclusion probability (sib identity):	0.4959
Hardy-Weinberg equilibrium test:	
Minimum expected frequency:	5.0
Chi-square values (using Yates' correction):	1.0668
Degrees of freedom:	1
P-value	0.3017
Significance (with Bonferroni correction):	NS
Null allele frequency estimate:	0.0430

7.5.4 Locus four (Primer 2088)

Table 38: Allele frequency for locus four (Primer 2088)

Allele	Count	Hets	Homs	Freq	Freq with null
115	12	8	2	0.2143	0.1981
119	7	7	0	0.1250	0.1339
123	29	15	7	0.5179	0.5366
127	8	8	0	0.1429	0.1548

Number of individuals typed:	28
Heterozygotes:	19
Homozygotes:	9
Number of alleles:	4
Observed heterozygosity:	0.6786
Expected heterozygosity	0.6617
Polymorphic information content (PIC):	0.6020
Average non-exclusion probability (first parent):	0.7651
Average non-exclusion probability (second parent):	0.5945
Average non-exclusion probability (parent pair):	0.4139
Average non-exclusion probability (identity):	0.1705
Average non-exclusion probability (sib identity):	0.4677

Hardy-Weinberg equilibrium test:	
Minimum expected frequency:	5.0
Chi-square values (using Yates' correction):	0.0090
Degrees of freedom:	1
P-value	0.9246
Significance (with Bonferroni correction):	NS
Null allele frequency estimate:	-0.0234

7.5.5 Locus five (Primer 253)

Table 39: Allele frequency for locus five (Primer 253)

Allele	Count	Hets	Homs	Freq	Freq with null
102	13	9	2	0.2321	0.2208
108	40	12	14	0.7143	07326
110	3	3	0	0.0536	0.0551

Number of individuals typed:	28
Heterozygotes:	12
Homozygotes:	16
Number of alleles:	3
Observed heterozygosity:	0.4268
Expected heterozygosity	0.4409
Polymorphic information content (PIC):	0.3748
Average non-exclusion probability (first parent):	0.9062
Average non-exclusion probability (second parent):	0.7917
Average non-exclusion probability (parent pair):	0.6692
Average non-exclusion probability (identity):	0.3797
Average non-exclusion probability (sib identity):	0.6284
Hardy-Weinberg equilibrium test:	Not completed
Null allele frequency estimate:	-0.0085

7.5.6 Locus six (Primer 2096)

Table 40: Allele frequency for locus six (Primer 2096)

Allele	Count	Hets	Homs	Freq	Freq with null
98	15	11	2	0.2679	0.2669
102	29	17	6	0.5179	0.5739
106	12	10	0	0.2143	0.2199

Number of individuals typed:	28
Heterozygotes:	19
Homozygotes:	9
Number of alleles:	3
Observed heterozygosity:	0.6786
Expected heterozygosity	0.6253
Polymorphic information content (PIC):	0.5445
Average non-exclusion probability (first parent):	0.8114
Average non-exclusion probability (second parent):	0.6660
Average non-exclusion probability (parent pair):	0.5151
Average non-exclusion probability (identity):	0.2186

Average non-exclusion probability (sib identity):	0.4976
Hardy-Weinberg equilibrium test:	
Minimum expected frequency:	5.0
Chi-square values (using Yates' correction):	0.5871
Degrees of freedom:	1
P-value	0.4435
Significance (with Bonferroni correction):	NS
Null allele frequency estimate:	-0.0607

7.5.7 Locus seven (Primer VwF)

Table 41: Allele frequency for locus seven (Primer VwF)

Allele	Count	Hets	Homs	Freq	Freq with null
147	2	2	0	0.0370	0.0377
149	3	3	0	0.0556	0.0571
157	36	14	11	0.6667	0.7250
163	2	2	0	0.0370	0.0377
169	11	7	0	0.2037	0.1832

Number of individuals typed:	27
Heterozygotes:	14
Homozygotes:	13
Number of alleles:	5
Observed heterozygosity:	0.5185
Expected heterozygosity	0.5178
Polymorphic information content (PIC):	0.4657
Average non-exclusion probability (first parent):	0.8611
Average non-exclusion probability (second parent):	0.7098
Average non-exclusion probability (parent pair):	0.5433
Average non-exclusion probability (identity):	0.2844
Average non-exclusion probability (sib identity):	0.5670
Hardy-Weinberg equilibrium test:	Not completed
Null allele frequency estimate:	-0.0407

7.5.8 Locus eight (Primer 250)

Table 42: Allele frequency for locus eight (Primer 250)

Allele	Count	Hets	Homs	Freq	Freq with null
134	17	7	5	0.4722	0.4182
138	5	3	1	0.1389	0.1171
140	11	7	2	0.3056	0.2902
142	3	3	0	0.0833	0.0864

Number of individuals typed:	18
Heterozygotes:	10
Homozygotes:	8
Number of alleles:	4
Observed heterozygosity:	0.556
Expected heterozygosity	0.6762
Polymorphic information content (PIC):	0.5989
Average non-exclusion probability (first parent):	0.7639
Average non-exclusion probability (second parent):	0.6028

Average non-exclusion probability (parent pair):	0.4307
Average non-exclusion probability (identity):	0.1759
Average non-exclusion probability (sib identity):	0.4653
Hardy-Weinberg equilibrium test:	Not completed
Null allele frequency estimate:	0.0881

7.5.9 Locus nine (Primer 213)

Table 43: Allele frequency for locus nine (Primer 213)

Allele	Count	Hets	Homs	Freq	Freq with null
155	5	5	0	0.0962	0.0967
157	37	7	15	0.7115	0.5640
159	2	0	1	0.0385	0.0186
161	5	3	1	0.0962	0.0766
163	2	0	1	0.0385	0.0186
169	1	1	0	0.0192	0.0186

Number of individuals typed:	26
Heterozygotes:	8
Homozygotes:	18
Number of alleles:	6
Observed heterozygosity:	0.3077
Expected heterozygosity	0.4811
Polymorphic information content (PIC):	0.4495
Average non-exclusion probability (first parent):	0.8756
Average non-exclusion probability (second parent):	0.7123
Average non-exclusion probability (parent pair):	0.5318
Average non-exclusion probability (identity):	0.3013
Average non-exclusion probability (sib identity):	0.5894
Hardy-Weinberg equilibrium test:	Not completed
Null allele frequency estimate:	0.2070

7.5.10 Locus ten (Primer 2001)

Table 44: Allele frequency for locus ten (Primer 2001)

Allele	Count	Hets	Homs	Freq	Freq with null
133	31	17	7	0.5741	0.6543
145	5	3	1	0.0926	0.0763
149	2	2	0	0.0370	0.0374
155	16	14	1	0.2963	0.3294

Number of individuals typed:	27
Heterozygotes:	18
Homozygotes:	9
Number of alleles:	4
Observed heterozygosity:	0.6667
Expected heterozygosity	0.5835
Polymorphic information content (PIC):	05065
Average non-exclusion probability (first parent):	0.8290
Average non-exclusion probability (second parent):	0.6879

Average non-exclusion probability (parent pair):	0.5312
Average non-exclusion probability (identity):	0.2488
Average non-exclusion probability (sib identity):	0.5258
Hardy-Weinberg equilibrium test:	Not completed
Null allele frequency estimate:	-0.0973

7.5.11 Locus eleven (Primer AHT130)

Table 45: Allele frequency for locus eleven (Primer AHT130)

Allele	Count	Hets	Homs	Freq	Freq with null
108	3	3	0	0.0517	0.0531
110	15	9	3	0.2586	0.2341
112	13	11	1	0.2241	0.2341
114	2	2	0	0.0345	0.0351
116	25	17	4	0.4310	0.4741

Number of individuals typed:	29
Heterozygotes:	21
Homozygotes:	8
Number of alleles:	5
Observed heterozygosity:	0.7241
Expected heterozygosity	0.7054
Polymorphic information content (PIC):	0.6406
Average non-exclusion probability (first parent):	0.7282
Average non-exclusion probability (second parent):	0.5596
Average non-exclusion probability (parent pair):	0.3830
Average non-exclusion probability (identity):	0.1467
Average non-exclusion probability (sib identity):	0.4401
Hardy-Weinberg equilibrium test:	
Minimum expected frequency:	5.0
Chi-square values (using Yates' correction):	0.4572
Degrees of freedom:	1
P-value	0.4989
Significance (with Bonferroni correction):	NS
Null allele frequency estimate:	-0.0305

7.5.12 Locus twelve (Primer set 466)

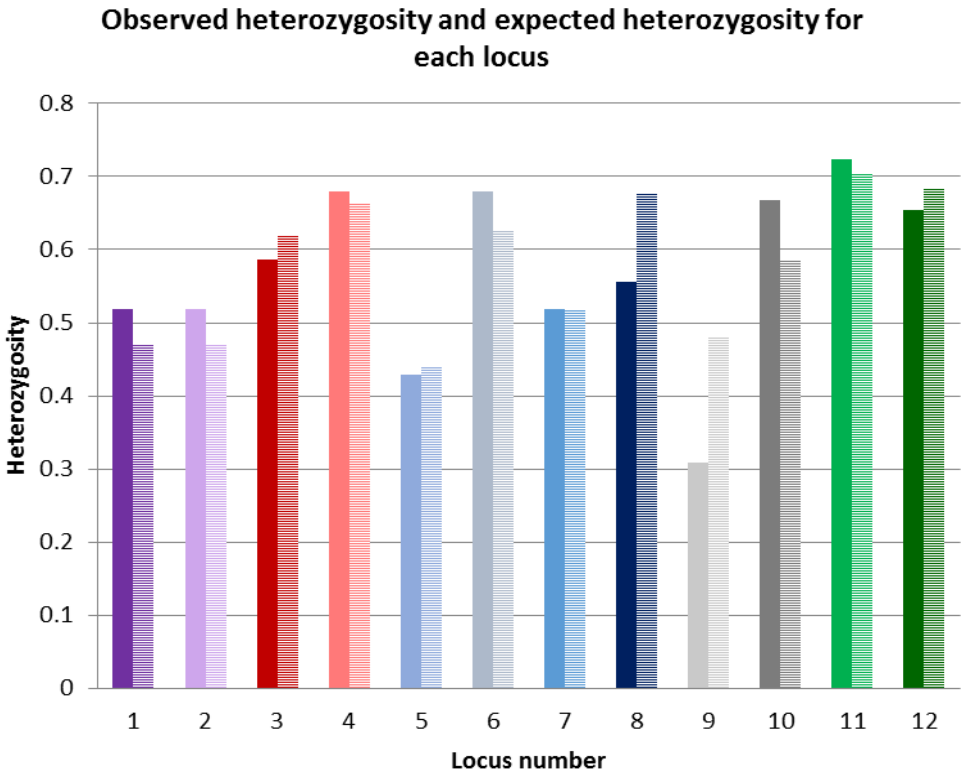
Table 46: Allele frequency for locus twelve (Primer 466)

Allele	Count	Hets	Homs	Freq	Freq with null
148	8	8	0	0.1379	0.1488
150	28	10	9	0.4828	0.4121
152	13	11	1	0.2241	0.2340
160	9	9	0	0.1552	0.1693

Number of individuals typed:	29
Heterozygotes:	19
Homozygotes:	10
Number of alleles:	4
Observed heterozygosity:	0.6552

Expected heterozygosity	0.6854
Polymorphic information content (PIC):	0.6248
Average non-exclusion probability (first parent):	0.7453
Average non-exclusion probability (second parent):	0.5735
Average non-exclusion probability (parent pair):	0.3936
Average non-exclusion probability (identity):	0.1553
Average non-exclusion probability (sib identity):	0.4520
Hardy-Weinberg equilibrium test:	
Minimum expected frequency:	5.0
Chi-square values (using Yates' correction):	2.2020
Degrees of freedom:	1
P-value	0.1378
Significance (with Bonferroni correction):	NS
Null allele frequency estimate:	0.0358

7.6 OBSERVED AND EXPECTED HETEROZYGOSITY AND PIC ACROSS ALL LOCI



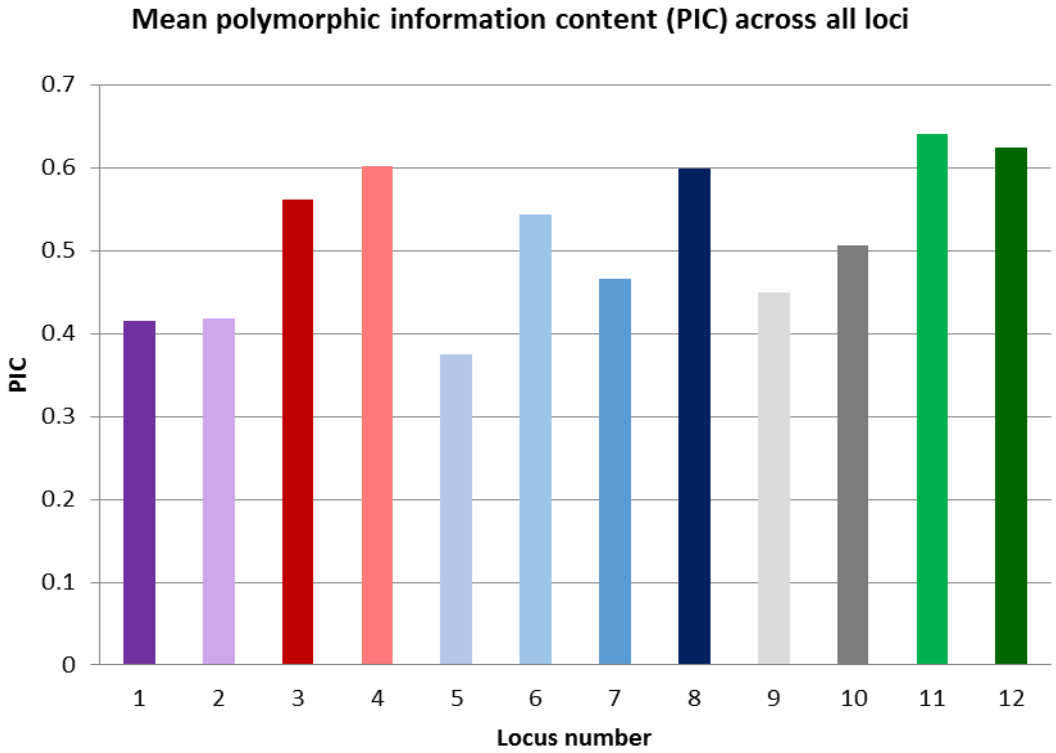


Figure 7-13: Observed heterozygosity and expected heterozygosity levels for all loci (left) and mean polymorphic content across all loci (right)

7.7 SIMULATION OF MATERNITY

7.7.1 Confidence level analysis for maternal parentage assignment

Simulation of maternity yielded relatively low assignment rates at a strict or relaxed level rate, improving by 6% for strict and decreasing by 1% if fathers are known (Table 7, Table 8).

Table 47: Cervus overview output for mother alone simulation

Level	Confidence (%)	Critical Delta	Assignments	Assignments rate
Strict	95.00	2.47	1130	11%
Relaxed	80.00	0.00	2177	22%
Unassigned			7823	78%
Total			10000	100%

Table 48: Cervus overview output for mother given known father simulation

Level	Confidence (%)	Critical Delta	Assignments	Assignments rate
Strict	95.00	1.68	1714	17%
Relaxed	80.00	0.00	2093	21%
Unassigned			7907	78%
Total			10000	100%

7.7.2 Simulation parameters

Table 49: Simulation parameters output from Cervus for simulation of mother

Input	
Number of offspring:	10000
Number of candidate mothers:	3
Proportion of candidate mothers sampled:	0.2000
Proportion of loci typed:	0.9
Proportion of loci mistyped:	0.05
Error rate in likelihood calculations:	0.05
Minimum number of typed loci:	7

Output	
Confidence determined using:	Delta
Relaxed confidence level:	80%
Strict confidence level:	95%

7.7.3 Delta distributions

Table 50: Delta distributions for mother alone

Identity of most likely candidates	N	Mean Delta	Standard deviation
True mother	1772	3.06	1.67
Non-mother (true mother sampled)	20	1.19	0.83
Non-mother (true mother unsampled)	385	1.19	0.83
None	7823		
Total	10000		

Table 51: Delta distributions for mother given known father

Identity of most likely candidates	N	Mean Delta	Standard deviation
True mother	1842	4.29	2.15
Non-mother (true mother sampled)	13	1.13	1.14
Non-mother (true mother unsampled)	238	1.43	1.22
None	7907		
Total	10000		

7.7.4 Breakdown of parentage assignments

Table 52: Mother alone breakdown of parentage assignments

Identity of most likely candidates	Confidence level		
	Strict	Relaxed	Most likely
True mother	1074 (95%)	1772 (81%)	1772 (81%)
Non-mother (true mother sampled)	2 (0%)	20 (1%)	20 (1)
Non-mother (true mother unsampled)	54 (5%)	385 (18%)	385 (5%)
Total assignments	1130	2177	2177
No assignments made	8870	7823	7283
Total tests	10000	10000	10000

Table 53: Mother given known father breakdown of parentage assignments

Identity of most likely candidates	Confidence level		
	Strict	Relaxed	Most likely
True mother	1629 (95%)	1842 (88%)	1842 (88%)
Non-mother (true mother sampled)	3 (0%)	13 (1%)	13 (1%)
Non-mother (true mother unsampled)	82 (%)	238 (11%)	238 (11%)
Total assignments	1714	2093	2093
No assignments made	8286	7097	7097
Total tests	10000	10000	10000

7.8 PARENTAL ANALYSIS OF MATERNITY

Table 54: Parental analysis of maternity (Spread over three pages), FPNP = first parent non-exclusion probability, SPNP = second parent non-exclusion probability

Offspring ID	Loci typed	FPNP	SPNP	Candidate mother ID	Loci typed	Pair loci compared	Pair loci mismatching	Pair LOD score	Pair Delta	Pair confidence
CL 238_024	12	9.14E-02	9.14E-02							
CL_238_370	12	2.12E-01	2.12E-01	CL_387	7	7	0	2.30E-02	2.30E-02	+
CL_238158	12	2.38E-02	2.38E-02							
CL_240279	12	2.25E-01	2.25E-01							
CL_279 No.1	12	1.37E-01	1.37E-01	CL_240279	12	12	0	1.12E+00	1.12E+00	+
CL_279_No.2	12	2.39E-01	2.39E-01	CL_922	10	10	0	1.71E+00	1.71E+00	+
CL_279_No.3	12	1.03E-01	1.03E-01	CL_240279	12	12	0	4.68E-01	4.68E-01	+
CL_240279_No.4	11	3.31E-01	3.31E-01	CL_922	10	10	0	2.12E+00	2.04E+00	+
CL_240279_No.4	11	3.31E-01	3.31E-01	CL_240279	12	11	0	8.36E-02	0.00E+00	
CL_279_No.5	12	1.04E-01	1.04E-01	CL_240279	12	12	0	1.21E+00	1.21E+00	+
CL_279_No.6	11	1.45E-01	1.45E-01	CL_240279	12	11	0	8.14E-01	8.14E-01	+
CL_279_No.7	12	1.40E-01	1.40E-01	CL_240279	12	12	0	2.37E-01	2.37E-01	+
CL 238_024	12	9.14E-02	9.14E-02							
CL_238_370	12	2.12E-01	2.12E-01	CL_387	7	7	0	2.30E-02	2.30E-02	+

Offspring ID	Loci typed	FPNP	SPNP	Candidate mother ID	Loci typed	Pair loci compared	Pair loci mismatching	Pair LOD score	Pair Delta	Pair confidence
CL_238158	12	2.38E-02	2.38E-02							
CL_240279	12	2.25E-01	2.25E-01							
CL_279 No.1	12	1.37E-01	1.37E-01	CL_240279	12	12	0	1.12E+00	1.12E+00	+
CL_279_No.2	12	2.39E-01	2.39E-01	CL_922	10	10	0	1.71E+00	1.71E+00	+
CL_279_No.3	12	1.03E-01	1.03E-01	CL_240279	12	12	0	4.68E-01	4.68E-01	+
CL_240279_No.4	11	3.31E-01	3.31E-01	CL_922	10	10	0	2.12E+00	2.04E+00	+
CL_240279_No.4	11	3.31E-01	3.31E-01	CL_240279	12	11	0	8.36E-02	0.00E+00	
CL_279_No.5	12	1.04E-01	1.04E-01	CL_240279	12	12	0	1.21E+00	1.21E+00	+
CL_279_No.6	11	1.45E-01	1.45E-01	CL_240279	12	11	0	8.14E-01	8.14E-01	+
CL_279_No.7	12	1.40E-01	1.40E-01	CL_240279	12	12	0	2.37E-01	2.37E-01	+
CL_337861	12	1.08E-01	1.08E-01	CL_387	7	7	0	2.47E+00	3.10E-01	+
CL_337861	12	1.08E-01	1.08E-01	CL_240279	12	12	0	2.16E+00	0.00E+00	
CL_387	7	5.24E-01	5.24E-01							
CL_387_No.1	11	1.13E-01	1.13E-01	CL_387	7	6	0	2.78E-01	2.78E-01	+
CL_387_No.2	10	3.24E-02	3.24E-02	CL_387	7	7	0	2.37E+00	2.37E+00	+
CL_387_No.3	8	1.03E-01	1.03E-01	CL_387	7	5	0	1.18E+00	1.18E+00	+
CL_387_No.4	11	3.72E-01	3.72E-01	CL_387	7	7	0	8.64E-01	8.64E-01	+

Offspring ID	Loci typed	FPNP	SPNP	Candidate mother ID	Loci typed	Pair loci compared	Pair loci mismatching	Pair LOD score	Pair Delta	Pair confidence
CL_387_No.5	12	7.57E-03	7.57E-03							
CL_387_No.6	11	4.61E-03	4.61E-03	CL_387	7	6	0	1.85E+00	1.85E+00	+
CL_387_No.7	12	2.56E-01	2.56E-01	CL_922	10	10	0	1.72E+00	7.65E-01	+
CL_387_No.7	12	2.56E-01	2.56E-01	CL_240279	12	12	0	9.58E-01	0.00E+00	
CL_387_No.8	11	1.10E-01	1.10E-01	CL_387	7	7	0	1.36E+00	1.36E+00	+
CL_387_No.9	11	2.53E-02	2.53E-02	CL_387	7	7	0	2.61E-01	2.61E-01	+
CL_4309	12	6.22E-02	6.22E-02							
CL_922	10	1.30E-01	1.30E-01							
CL_922 No.1	11	4.87E-02	4.87E-02	CL_922	10	10	0	3.41E+00	3.41E+00	*
CL_922 No.2	11	2.28E-03	2.28E-03	CL_922	10	10	1	2.13E+00	2.13E+00	+
CL_931248	9	1.54E-01	1.54E-01	CL_922	10	7	0	6.77E-01	6.77E-01	+
CL_931645	12	4.17E-03	4.17E-03							
CL_8022	12	2.79E-02	2.79E-02							

7.9 SIMULATION OF PATERNITY

7.9.1 Confidence level analysis for paternal parentage assignment

Simulation of paternity revealed very low assignment rates for a father alone, but an increase of 7% at the strict level and 6% can be seen when a known mother is given in the simulation (Table 10, Table 11). It important to note that at this stage, known mothers were not inputted in to Cervus as this was considered later.

Table 55: Cervus overview output for father alone simulation for simulation of paternity

Level	Confidence (%)	Critical Delta	Assignments	Assignments rate
Strict	95.00	3.83	583	6%
Relaxed	80.00	1.96	1547	15%
Unassigned			8453	85%
Total			10000	100%

Table 56: Cervus overview output for father given known mother simulation

Level	Confidence (%)	Critical Delta	Assignments	Assignments rate
Strict	95.00	3.12	1314	13%
Relaxed	80.00	1.06	2138	21%
Unassigned			7862	79%
Total			10000	100%

7.9.2 Simulation parameters

Table 57: Simulation parameters output from Cervus for simulation of paternity

Input	
Number of offspring:	10000
Number of candidate fathers	8
Proportion of candidate fathers sampled:	0.2000
Proportion of loci typed:	0.9
Proportion of loci mistyped:	0.05
Error rate in likelihood calculations:	0.05
Minimum number of typed loci:	7
Output	
Confidence determined using:	Delta
Relaxed confidence level:	80%
Strict confidence level:	95%

7.9.3 Delta distributions

Table 58: Delta distributions for father alone simulation

Identity of most likely candidates	N	Mean Delta	Standard deviation
True father	1763	3.03	1.77
Non-father (true father sampled)	83	0.83	0.87
Non-mother (true father unsampled)	1310	1.26	1.02
None	6844		
Total	10000		

Table 59: Delta distributions for father given known mother simulation

Identity of most likely candidates	N	Mean Delta	Standard deviation
True father	1859	4.27	2.26
Non-father (true father sampled)	29	1.11	0.76
Non-mother (true father unsampled)	809	1.36	1.08
None	7303		
Total	10000		

7.9.4 Breakdown of parentage assignments

Table 60: Breakdown of parentage assignments for father alone

Identity of most likely candidates	Confidence level		
	Strict	Relaxed	Most likely
True father	554 (95%)	1238 (80%)	1763 (56%)
Non- father (true father sampled)	2 (0%)	4 (0%)	83 (3%)
Non- father (true father unsampled)	27 (5%)	305 (20%)	1310 (42%)
Total assignments	583	1547	3156
No assignments made	9417	8453	6844
Total tests	10000	10000	10000

Table 61: Breakdown of parentage assignments for father given known mother

Identity of most likely candidates	Confidence level		
	Strict	Relaxed	Most likely
True father	1249 (95%)	1711 (80%)	1859 (69%)
Non- father (true father sampled)	0 (0%)	14 (1%)	29 (1%)
Non- father (true father unsampled)	65 (5%)	413 (19%)	809 (30%)
Total assignments	1314	2138	2697
No assignments made	8686	7862	7303
Total tests	10000	10000	10000

7.10 PARENTAL ANALYSIS OF PATERNITY

Table 62: Complete cervus output for parental analysis of paternity (no known mother provided)

Offspring ID	Loci typed	FPNP	SPNP	Candidate father ID	Loci typed	Pair loci compared	Pair loci mismatching	Pair LOD score	Pair Delta	Pair confidence
CL 238_024	12	9.14E-02	9.14E-02	CL_931248	9	9	0	9.43E-01	9.43E-01	-
CL_238_370	12	2.12E-01	2.12E-01							
CL_238158	12	2.38E-02	2.38E-02	CL_337861	12	12	0	3.35E+00	3.35E+00	+
CL_240279	12	2.25E-01	2.25E-01	CL_337861	12	12	0	2.16E+00	2.16E+00	+
CL_279 No.1	12	1.37E-01	1.37E-01	CL_8022	12	12	1	4.91E-01	4.91E-01	-
CL_279_No.2	12	2.39E-01	2.39E-01	CL_8022	12	12	1	1.02E-01	1.02E-01	-
CL_279_No.3	12	1.03E-01	1.03E-01	CL 238_024	12	12	1	3.26E-02	3.26E-02	-
CL_240279_No.4	11	3.31E-01	3.31E-01	CL_931248	9	8	0	1.61E-01	1.61E-01	-
CL_279_No.5	12	1.04E-01	1.04E-01							
CL_279_No.6	11	1.45E-01	1.45E-01	CL 238_024	12	11	0	1.64E+00	1.64E+00	-
CL_279_No.7	12	1.40E-01	1.40E-01	CL_8022	12	12	0	1.70E+00	1.70E+00	-
CL_337861	12	1.08E-01	1.08E-01	CL_238158	12	12	0	3.35E+00	3.35E+00	+
CL_387	7	5.24E-01	5.24E-01	CL_337861	12	7	0	2.47E+00	2.44E+00	+
CL_387	7	5.24E-01	5.24E-01	CL_238_370	12	7	0	2.30E-02	0.00E+00	
CL_387_No.1	11	1.13E-01	1.13E-01	CL_238_370	12	11	1	1.86E+00	1.86E+00	-
CL_387_No.2	10	3.24E-02	3.24E-02							
CL_387_No.3	8	1.03E-01	1.03E-01	CL_931645	12	8	1	3.98E-02	3.98E-02	-
CL_387_No.4	11	3.72E-01	3.72E-01	CL_238_370	12	11	0	3.02E+00	3.02E+00	+
CL_387_No.5	12	7.57E-03	7.57E-03							
CL_387_No.6	11	4.61E-03	4.61E-03							

Offspring ID	Loci typed	FPNP	SPNP	Candidate father ID	Loci typed	Pair loci compared	Pair loci mismatching	Pair LOD score	Pair Delta	Pair confidence
CL_387_No.7	12	2.56E-01	2.56E-01							
CL_387_No.8	11	1.10E-01	1.10E-01	CL_238_370	12	11	0	3.18E+00	3.18E+00	+
CL_387_No.9	11	2.53E-02	2.53E-02							
CL_4309	12	6.22E-02	6.22E-02							
CL_922	10	1.30E-01	1.30E-01	CL_931248	9	7	0	6.77E-01	6.77E-01	-
CL_922 No.1	11	4.87E-02	4.87E-02	CL_931248	9	8	0	4.78E-01	4.78E-01	-
CL_922 No.2	11	2.28E-03	2.28E-03							
CL_931248	9	1.54E-01	1.54E-01	CL 238_024	12	9	0	9.43E-01	9.43E-01	-
CL_931645	12	4.17E-03	4.17E-03							
CL_8022	12	2.79E-02	2.79E-02							

7.11 SIMULATION OF PAIRS

7.11.1 Confidence level analysis for paternity assignment

Table 63: Cervus overview output for mother alone simulation for simulation of pairs

Level	Confidence (%)	Critical Delta	Assignments	Assignments rate
Strict	95.00	2.31	1193	12%
Relaxed	80.00	0.00	2190	22%
Unassigned			7810	78%
Total			10000	100%

Table 64: Cervus overview output for mother alone simulation

Level	Confidence (%)	Critical Delta	Assignments	Assignments rate
Strict	95.00	3.41	753	8%
Relaxed	80.00	1.86	1617	16%
Unassigned			8383	84%
Total			10000	100%

7.11.2 Simulation parameters

Table 65: Simulation parameters output from Cervus for simulation of pairs (both mother and father)

Input	
Number of offspring:	10000
Number of candidate mothers:	3
Proportion of candidate mothers sampled:	0.2000
Number of candidate fathers:	8
Proportion of candidate fathers sampled	0.2000
Number of parent pairs:	24
Proportion of loci typed:	0.9
Proportion of loci mistyped:	0.05
Error rate in likelihood calculations:	0.05
Minimum number of typed loci:	7
Output	
Confidence determined using:	Delta
Relaxed confidence level:	80%
Strict confidence level:	95%

7.11.3 Delta distributions

Table 66: Delta distributions for mother alone simulation for simulation of pairs

Identity of most likely candidates	N	Mean Delta	Standard deviation
True mother	1783	3.07	1.73
Non-mother (true mother sampled)	36	1.17	1.08
Non-mother (true mother unsampled)	371	1.23	0.95
None	7810		
Total	10000		

Table 67: Delta distributions for father alone simulation for simulation of pairs

Identity of most likely candidates	N	Mean Delta	Standard deviation
True father	1788	3.08	1.78
Non-father (true father sampled)	102	0.95	0.85
Non-mother (true father unsampled)	1309	1.21	0.98
None	6801		
Total	10000		

Table 68: Delta distributions for Parent pair (sexes known) simulation for simulation of pairs

Identity of most likely candidates	N	Mean Delta	Standard deviation
True parent pair	364	6.10	2.93
Non-parent pair (True parent pair sampled)	20	1.42	1.42
Non-parent par (True mother sampled)	164	2.29	1.80
Non-parent pair (true father unsampled)	482	2.14	1.73
None parent pair (neither true parent sampled)	93	1.59	1.25
None	8877		
Total	10000		

7.11.4 Breakdown of parentage assignments

Table 69: Breakdown of parentage assignments for mother alone simulation

Identity of most likely candidates	Confidence level		
	Strict	Relaxed	Most likely
True mother	1134 (95%)	1783 (81%)	178. (81%)
Non-mother (true mother sampled)	5 (0%)	36 (2%)	36 (2%)
Non-mother (true mother unsampled)	54 (5%)	371 (17%)	371 (17%)
Total assignments	1193	2190	2190
No assignments made	8807	7810	7810
Total tests	10000	10000	10000

Table 70: Breakdown of parentage assignments for father alone simulation

Identity of most likely candidates	Confidence level		
	Strict	Relaxed	Most likely
True father	716 (95%)	1294 (80%)	1788 (56%)
Non- father (true father sampled)	2 (0%)	10 (1%)	102 (3%)
Non- father (true father unsampled)	35 (5%)	313 (19%)	1309 (41%)
Total assignments	753	1617	3199
No assignments made	9247	8383	6801
Total tests	10000	10000	10000

Table 71: Breakdown of parentage pair for parent pair (sexes known)

Identity of most likely candidates	Confidence level		
	Strict	Relaxed	Most likely
True parent pair	154 (95%)	224 (81%)	364 (32%)
Non-parent pair (true parent pair sampled)	0 (0%)	1 (0%)	20 (2%)
Non-parent pair (true mother unsampled)	2 (1%)	15 (5%)	164 (15%)
Non-parent pair (true father unsampled)	6 (4%)	37 (13%)	482 (43%)
Non-parent pair (neither true parent sampled)	0 (0%)	2 (1%)	938%)
Total assignments	162	279	1123
No assignments made	9838	9721	8877
Total tests	10000	10000	10000

7.12 PARENTAL ANALYSIS OF PAIRS

Table 72: Complete cervus output for parental analysis of pairs

Offspring ID	Loci typed	First parent non-exclusion probability	Parent pair non-exclusion probability	Candidate mother ID	Loci typed	Pair loci compared	Pair loci mismatching	Pair LOD score	Pair Delta	Pair confidence	Candidate father ID	Loci typed	Pair loci compared	Pair loci mismatching	Pair LOD score	Pair Delta	Pair confidence	Trio loci compared	Trio loci mismatching	Trio LOD score	Trio Delta	Trio confidence
CL 238_024	12	9.14 E-02	4.74 E-05																			
CL_238_370	12	2.12 E-01	2.60 E-05																			
CL_238158	12	2.38 E-02	5.55 E-07																			
CL_240279	12	2.25 E-01	1.33 E-03	CL_387	7	7	1	-8.41E-01	0.00E+00		CL_337861	12	12	0	2.16E+00	2.16E+00	+	12	1	2.07E-01	2.07E-01	-
CL_279 No.1	12	1.37 E-01	7.95 E-05	CL_240279	12	12	0	1.12E+00	1.12E+00	+	CL_8022	12	12	1	4.91E-01	4.91E-01	-	12	1	5.73E+00	5.73E+00	+
CL_279_No.2	12	2.39 E-01	4.21 E-04	CL_922	10	10	0	1.71E+00	1.71E+00	+	CL 238_024	12	12	1	-1.02E+00	0.00E+00		12	2	1.98E+00	9.75E-01	-
CL_279_No.2	12	2.39 E-01	4.21 E-04	CL_922	10	10	0	1.71E+00	1.71E+00	+	CL_4309	12	12	1	-3.84E-01	0.00E+00		12	2	1.00E+00	0.00E+00	
CL_279_No.2	12	2.39 E-01	4.21 E-04	CL_240279	12	12	0	-1.51E+00	0.00E+00		CL_8022	12	12	1	1.02E-01	1.02E-01	-	12	2	1.25E-01	0.00E+00	
CL_279_No.3	12	1.03 E-01	6.76 E-05	CL_240279	12	12	0	4.68E-01	4.68E-01	+	CL_8022	12	12	1	-1.56E-02	0.00E+00		12	2	2.42E+00	2.42E+00	-

CL_240279_No.4	11	3.31E-01	3.31E-03	CL_922	10	10	0	2.12E+00	2.04E+00	+	CL 238_024	12	11	0	-5.79E-01	0.00E+00		11	1	2.24E+00	6.66E-01	-
CL_240279_No.4	11	3.31E-01	3.31E-03	CL_922	10	10	0	2.12E+00	2.04E+00	+	CL_337861	12	11	0	-1.54E+00	0.00E+00		11	1	1.57E+00	0.00E+00	
CL_240279_No.4	11	3.31E-01	3.31E-03	CL_922	10	10	0	2.12E+00	2.04E+00	+	CL_931248	9	8	0	1.61E-01	1.61E-01	-	11	1	9.18E-01	0.00E+00	
CL_240279_No.4	11	3.31E-01	3.31E-03	CL_922	10	10	0	2.12E+00	2.04E+00	+	CL_238_370	12	11	0	-2.71E+00	0.00E+00		11	1	4.31E-01	0.00E+00	
CL_279_No.5	12	1.04E-01	6.52E-04	CL_240279	12	12	0	1.21E+00	1.21E+00	+	CL_8022	12	12	1	-1.13E+00	0.00E+00		12	1	3.07E+00	2.38E+00	-
CL_279_No.5	12	1.04E-01	6.52E-04	CL_240279	12	12	0	1.21E+00	1.21E+00	+	CL 238_024	12	12	1	-5.30E-01	0.00E+00		12	2	6.83E-01	0.00E+00	
CL_279_No.6	11	1.45E-01	3.58E-03	CL_922	10	10	1	-3.78E-01	0.00E+00		CL 238_024	12	11	0	1.64E+00	1.64E+00	-	11	1	1.97E+00	4.06E-01	-
CL_279_No.6	11	1.45E-01	3.58E-03	CL_240279	12	11	0	8.14E-01	8.14E-01	+	CL 238_024	12	11	0	1.64E+00	1.64E+00	-	11	1	1.57E+00	0.00E+00	
CL_279_No.6	11	1.45E-01	3.58E-03	CL_240279	12	11	0	8.14E-01	8.14E-01	+	CL_8022	12	11	1	-1.39E+00	0.00E+00		11	1	1.53E+00	0.00E+00	
CL_279_No.7	12	1.40E-01	6.03E-04	CL_240279	12	12	0	2.37E-01	2.37E-01	+	CL_8022	12	12	0	1.70E+00	1.70E+00	-	12	0	5.24E+00	4.00E+00	-
CL_279_No.7	12	1.40E-01	6.03E-04	CL_387	7	7	0	-9.64E-01	0.00E+00		CL_8022	12	12	0	1.70E+00	1.70E+00	-	12	1	1.24E+00	0.00E+00	
CL_337861	12	1.08E-01	1.15E-04	CL_387	7	7	0	2.47E+00	3.10E-01	+	CL_238158	12	12	0	3.35E+00	3.35E+00	+	12	0	5.76E+00	3.37E+00	-
CL_337861	12	1.08E-01	1.15E-04	CL_240279	12	12	0	2.16E+00	0.00E+00		CL_238158	12	12	0	3.35E+00	3.35E+00	+	12	2	2.38E+00	0.00E+00	

CL_387	7	5.24E-01	1.95E-02	CL_240279	12	7	1	-8.41E-01	0.00E+00		CL_4309	12	7	1	-1.93E+00	0.00E+00		7	1	8.85E-01	2.54E-01	-
CL_387	7	5.24E-01	1.95E-02	CL_240279	12	7	1	-8.41E-01	0.00E+00		CL_238_370	12	7	0	2.30E-02	0.00E+00		7	1	6.31E-01	0.00E+00	
CL_387	7	5.24E-01	1.95E-02	CL_922	10	6	1	-3.97E+00	0.00E+00		CL_337861	12	7	0	2.47E+00	2.44E+00	+	7	1	3.65E-01	0.00E+00	
CL_387_No.1	11	1.13E-01	8.40E-05																			
CL_387_No.2	10	3.24E-02	2.63E-05	CL_387	7	7	0	2.37E+00	2.37E+00	*	CL_931645	12	10	2	-2.15E-01	0.00E+00		10	2	3.01E+00	5.34E-01	-
CL_387_No.2	10	3.24E-02	2.63E-05	CL_387	7	7	0	2.37E+00	2.37E+00	*	CL_238_370	12	10	1	-5.03E-01	0.00E+00		10	1	2.48E+00	0.00E+00	
CL_387_No.3	8	1.03E-01	2.78E-04	CL_240279	12	8	1	-1.37E+00	0.00E+00		CL_931645	12	8	1	3.98E-02	3.98E-02	-	8	1	1.90E+00	1.76E+00	-
CL_387_No.3	8	1.03E-01	2.78E-04	CL_387	7	5	0	1.18E+00	1.18E+00	+	CL_931645	12	8	1	3.98E-02	3.98E-02	-	8	2	1.45E-01	0.00E+00	
CL_387_No.4	11	3.72E-01	8.48E-04	CL_387	7	7	0	8.64E-01	8.64E-01	+	CL_238_370	12	11	0	3.02E+00	3.02E+00	+	11	1	3.07E+00	3.07E+00	-
CL_387_No.5	12	7.57E-03	2.71E-07																			
CL_387_No.6	11	4.61E-03	7.59E-07																			
CL_387_No.7	12	2.56E-01	3.26E-03	CL_922	10	10	0	1.72E+00	7.65E-01	+	CL_931248	9	9	1	-5.61E-01	0.00E+00		12	1	1.58E+00	5.09E-01	-
CL_387_No.7	12	2.56E-01	3.26E-03	CL_922	10	10	0	1.72E+00	7.65E-01	+	CL_337861	12	12	0	-1.12E+00	0.00E+00		12	1	1.07E+00	0.00E+00	

CL_387_No.7	12	2.56 E-01	3.26 E-03	CL_240279	12	12	0	9.58E-01	0.00E+00		CL_8022	12	12	1	-2.40E+00	0.00E+00		12	1	9.49E-01	0.00E+00	
CL_387_No.8	11	1.10 E-01	1.50 E-04	CL_387	7	7	0	1.36E+00	1.36E+00	+	CL_238_370	12	11	0	3.18E+00	3.18E+00	+	11	0	5.10E+00	5.10E+00	-
CL_387_No.9	11	2.53 E-02	1.02 E-06																			
CL_4309	12	6.22 E-02	1.04 E-04	CL_922	10	10	1	-1.65E+00	0.00E+00		CL_8022	12	12	2	-1.37E+00	0.00E+00		12	2	7.26E-01	7.26E-01	-
CL_922	10	1.30 E-01	3.89 E-04																			
CL_922 No.1	11	4.87 E-02	5.64 E-04	CL_922	10	10	0	3.41E+00	3.41E+00	*	CL_931248	9	8	0	4.78E-01	4.78E-01	-	11	0	4.07E+00	3.79E+00	-
CL_922 No.1	11	4.87 E-02	5.64 E-04	CL_922	10	10	0	3.41E+00	3.41E+00	*	CL_238_370	12	11	1	-1.67E+00	0.00E+00		11	2	2.83E-01	0.00E+00	
CL_922 No.2	11	2.28 E-03	3.18 E-08																			
CL_931248	9	1.54 E-01	5.11 E-04	CL_922	10	7	0	6.77E-01	6.77E-01	+	CL 238_024	12	9	0	9.43E-01	9.43E-01	-	9	1	2.03E+00	2.03E+00	-

7.13 SIMULATION OF PATERNAL ASSIGNMENT WITH KNOWN MOTHERS

7.13.1 Confidence level analysis for paternal parentage assignment with known mothers

Table 73: Cervus overview output for father alone simulation (known mothers provided)

Level	Confidence (%)	Critical Delta	Assignments	Assignments rate
Strict	95.00	3.90	530	5%
Relaxed	80.00	1.84	1537	15%
Unassigned			8463	85%
Total			10000	100%

Table 74: Cervus overview output for father given known mother simulation

Level	Confidence (%)	Critical Delta	Assignments	Assignments rate
Strict	95.00	3.71	1088	11%
Relaxed	80.00	1.06	2135	21%
Unassigned			7865	79%
Total			10000	100%

7.13.2 Simulation parameters

Table 75: Simulation parameters output from Cervus for simulation of paternity (known mothers provided)

Input	
Number of offspring:	10000
Number of candidate fathers	8
Proportion of candidate fathers sampled:	0.2000
Proportion of loci typed:	0.9
Proportion of loci mistyped:	0.05
Error rate in likelihood calculations:	0.05
Minimum number of typed loci:	7
Output	
Confidence determined using:	Delta
Relaxed confidence level:	80%
Strict confidence level:	95%

7.13.3 Delta distributions

Table 76: Delta distributions for father alone simulation

Identity of most likely candidates	N	Mean Delta	Standard deviation
True father	1720	2.98	1.71
Non-father (true father sampled)	90	0.99	1.15
Non-mother (true father unsampled)	1283	124	0.97
None	6907		
Total	10000		

Table 77: Delta distributions for father given known mother simulation

Identity of most likely candidates	N	Mean Delta	Standard deviation
True father	1835	4.19	2.17
Non-father (true father sampled)	31	1.18	1.41
Non-mother (true father unsampled)	803	1.45	1.21
None	7331		
Total	10000		

7.13.4 Breakdown of parentage assignments

Table 78: Breakdown of parentage assignments for father alone

Identity of most likely candidates	Confidence level		
	Strict	Relaxed	Most likely
True father	504 (95%)	1230 (80%)	17603 (56%)
Non- father (true father sampled)	5 (1%)	15 (1%)	90 (3%)
Non- father (true father unsampled)	21 (4%)	292 (19%)	1283 (41%)
Total assignments	530	1537	3093
No assignments made	9470	8463	6907
Total tests	10000	10000	10000

Table 79: Breakdown of parentage assignments for father given known mother

Identity of most likely candidates	Confidence level		
	Strict	Relaxed	Most likely
True father	1034 (95%)	1708 (80%)	1835 (69%)
Non- father (true father sampled)	2 (0%)	11 (1%)	31 (1%)
Non- father (true father unsampled)	52 (5%)	416 (19%)	803 (30%)
Total assignments	1088	2135	2669
No assignments made	8912	7865	7331
Total tests	10000	10000	10000

7.14 PARENTAGE ANALYSIS OF PATERNITY GIVING KNOWN MOTHER

Table 80: Complete cervus output for parentage analysis of paternity giving known mother

Offspring ID	Loci typed	Mother ID	Loci typed	Pair loci compared	Pair loci mismatching	Pair LOD score	Candidate father ID	Loci typed	Pair loci compared	Pair loci mismatching	Pair LOD score	Pair Delta	Pair confidence	Trio loci compared	Trio loci mismatching	Trio LOD score	Trio Delta	Trio confidence
CL_279 No.1	12	CL_240279	12	12	0	1.12E+00	CL_8022	12	12	1	4.91E-01	4.91E-01	-	12	1	4.61E+00	4.61E+00	*
CL_279_No.2	12	CL_240279	12	12	0	-1.51E+00	CL_8022	12	12	1	1.02E-01	1.02E-01	-	12	2	1.63E+00	1.63E+00	+
CL_279_No.3	12	CL_240279	12	12	0	4.68E-01	CL_8022	12	12	1	-1.56E-02	0.00E+00		12	2	1.95E+00	1.95E+00	+
CL_240279_No.4	11	CL_240279	12	11	0	8.36E-02	CL 238_024	12	11	0	-5.79E-01	0.00E+00		11	2	-2.49E+00	0.00E+00	
CL_279_No.5	12	CL_240279	12	12	0	1.21E+00	CL_8022	12	12	1	-1.13E+00	0.00E+00		12	1	1.86E+00	1.86E+00	+
CL_279_No.6	11	CL_240279	12	11	0	8.14E-01	CL 238_024	12	11	0	1.64E+00	1.64E+00	-	11	1	7.51E-01	3.73E-02	-
CL_279_No.7	12	CL_240279	12	12	0	2.37E-01	CL_8022	12	12	0	1.70E+00	1.70E+00	-	12	0	5.00E+00	5.00E+00	*
CL_387_No.1	11	CL_387	7	6	0	2.78E-01	CL_238_370	12	11	1	1.86E+00	1.86E+00	-	11	3	-1.33E+00	0.00E+00	
CL_387_No.2	10	CL_387	7	7	0	2.37E+00	CL_931645	12	10	2	-2.15E-01	0.00E+00		10	2	6.42E-01	5.34E-01	-

CL_387_No.3	8	CL_387	7	5	0	1.18E+00	CL_931645	12	8	1	3.98E-02	3.98E-02	-	8	2	-1.03E+00	0.00E+00	
CL_387_No.4	11	CL_387	7	7	0	8.64E-01	CL_238_370	12	11	0	3.02E+00	3.02E+00	+	11	1	2.20E+00	2.20E+00	+
CL_387_No.5	12	CL_387	7	7	1	-1.74E+00	CL_337861	12	12	1	-1.87E+00	0.00E+00		12	2	-1.84E+00	0.00E+00	
CL_387_No.6	11	CL_387	7	6	0	1.85E+00	CL_238158	12	11	3	-3.62E+00	0.00E+00		11	4	-5.72E+00	0.00E+00	
CL_387_No.7	12	CL_387	7	7	0	-6.55E-01	CL_4309	12	12	1	-2.07E+00	0.00E+00		12	1	-4.86E-01	0.00E+00	
CL_387_No.8	11	CL_387	7	7	0	1.36E+00	CL_238_370	12	11	0	3.18E+00	3.18E+00	+	11	0	3.73E+00	3.73E+00	*
CL_387_No.9	11	CL_387	7	7	0	2.61E-01	CL_238_370	12	11	1	-1.06E+00	0.00E+00		11	3	-3.22E+00	0.00E+00	
CL_922 No.1	11	CL_922	10	10	0	3.41E+00	CL_931248	9	8	0	4.78E-01	4.78E-01	-	8	0	6.63E-01	6.63E-01	-
CL_922 No.2	11	CL_922	10	10	1	2.13E+00	CL_931248	9	8	1	-1.72E+00	0.00E+00		8	3	-3.77E+00	0.00E+00	

7.15 KINALZYER KINSHIP GROUPING OUTPUT

Table 81: Kinalyzer raw grouping output

Sibling set	Samples
0	CL_279_No.3 CL_387_No.6, CL_931645
1	CL_922 No.1 CL 238_024, CL_931248
2	CL_279_No.2 CL_387_No.5 CL_8022
3	CL_279_No.2,CL_279_No.3,CL_240279_No.4,CL_279_No.5,CL_279_No.6,CL_279_No.7 CL_387_No.7 CL_337861,
4	CL_279 No.1, CL_922 No.2
5	CL_387,CL_387_No.1,CL_387_No.2,CL_387_No.3,CL_387_No.4,CL_387_No.8,CL_387_No.9 CL_238_370, CL_240279
6	CL_238158, CL_4309, CL_922

7.16COMPLICATIONS WITH DNA EXTRACTIONS FROM HAIR

Hair samples were obtained from the population of free-ranging dogs in Italy to be used as part of the project with the intention to extract DNA and use it to establish the parentage for pups and kinship relationships in the population. There are a range of different methods available to use when extracting DNA from hair, however some of them are known for being laborious and prone to contamination, whilst some of the shorter protocols available

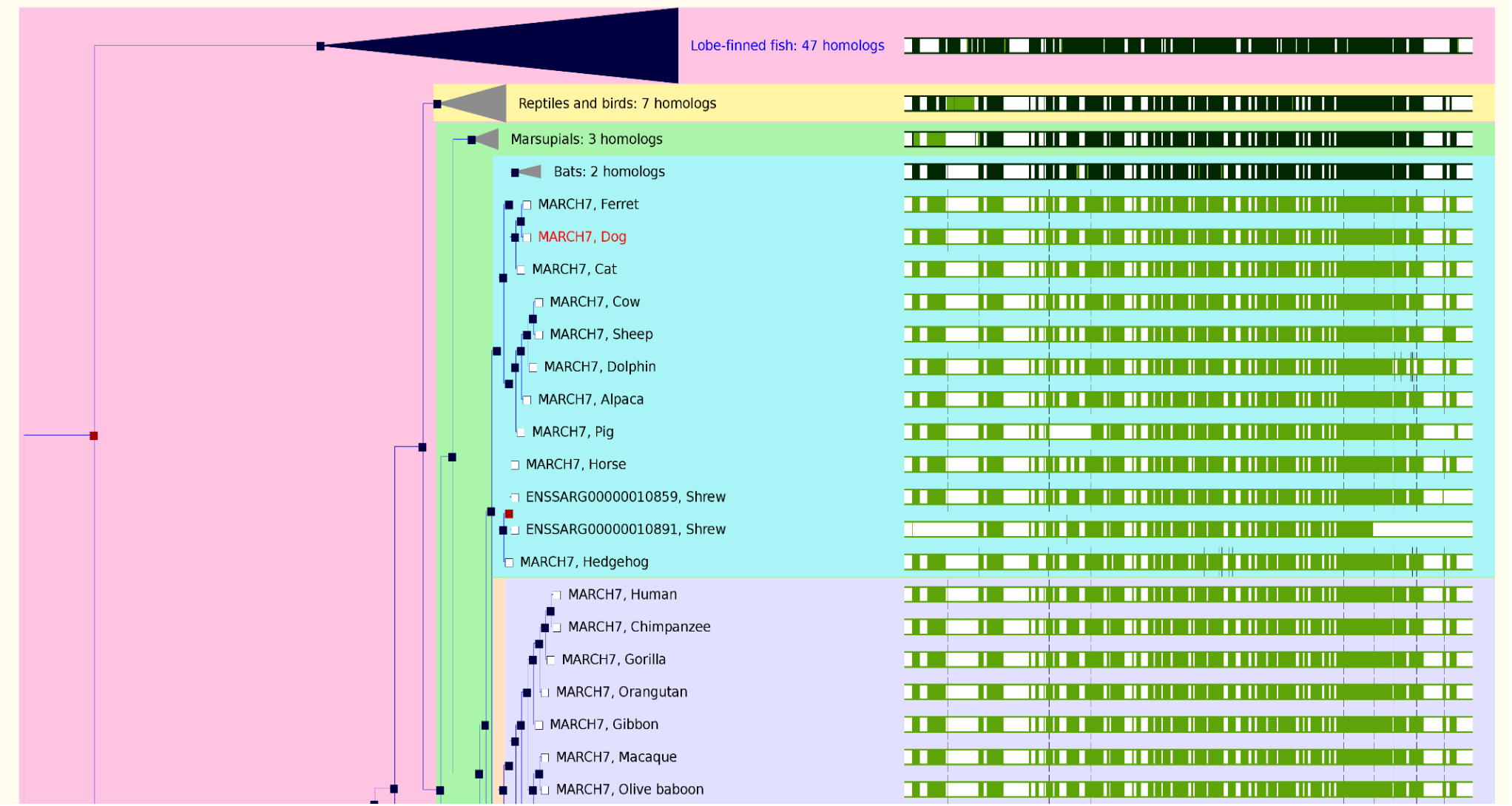
result in low DNA recovery yield (Almeida *et al.*, 2011). For this project, five different techniques were trailed, but DNA purity and quantity remained a consistent issue. Even when there was a sufficient level of DNA detected using the Nanodrop 1000 spectrophotometer, amplification by PCR failed to produce a visible band when ran on an agarose gel. From previous analysis of human hair samples it is known that hair shafts only contain minute amounts of genomic DNA and detectable mtDNA (Wetton *et al.*, 2003; Pfeiffer *et al.*, 2004). Research into human hair indicated the nuclear DNA from keratinised cells can be highly degraded and generally ~100bp in size (Takayanagi *et al.*, 2003). Hair samples by nature are protein-rich and the additional steps required to break down the shaft in order to release DNA result in exposing the samples to an increased risk of contamination (Graffy and Foran, 2005; Ghatak *et al.*, 2013). Failure or a relatively low success rate have been shown in existing animal hair DNA-extraction methods, if adherent root cells were absent (Pfeiffer et al, 2004).

Research has shown that successful amplification by PCR has not always been achieved, even when a sufficient quantity of DNA is present, suggesting PCR inhibitors may be present in the extracted hair samples (Suenaga and Nakamura, 2005). More specifically, previous work has identified that the hair pigmentation melanin is a strong inhibitor of the PCR process (Yoshii *et al.*, 1992; Yoshii *et al.*, 1993; Wilson *et al.*, 1995), and that hair-dying can have a strong effect on PCR (Yoshii *et al.*, 1992), although the latter should not cause complications in dog hair samples. Success rates for PCR is higher for mtDNA as there are between one hundred and one thousand mitochondria in every eukaryotic animal cell, but nuclear DNA was required for this project.

In multiple instances, repeats with alternations were ran in order to try and eliminate possible sources of errors. For instance, when considering the Chelex method the first set of set samples were incubated for 8 hours due to the methodology stating a "minimum of 6 hours". Upon failure, samples were reran and incubated for both 12 and 18 hours but both still resulted in failure. Any methodologies requiring buffers or solutions to be produced were repeated and recalculated each time they were attempt in order to reduce and avoid any continuation of error.

Extraction of tissue samples proved to be 100% effective, with minimal complications.

7.17 ENSEMBL PHYLOGENETIC TREE OF *MARCH7*



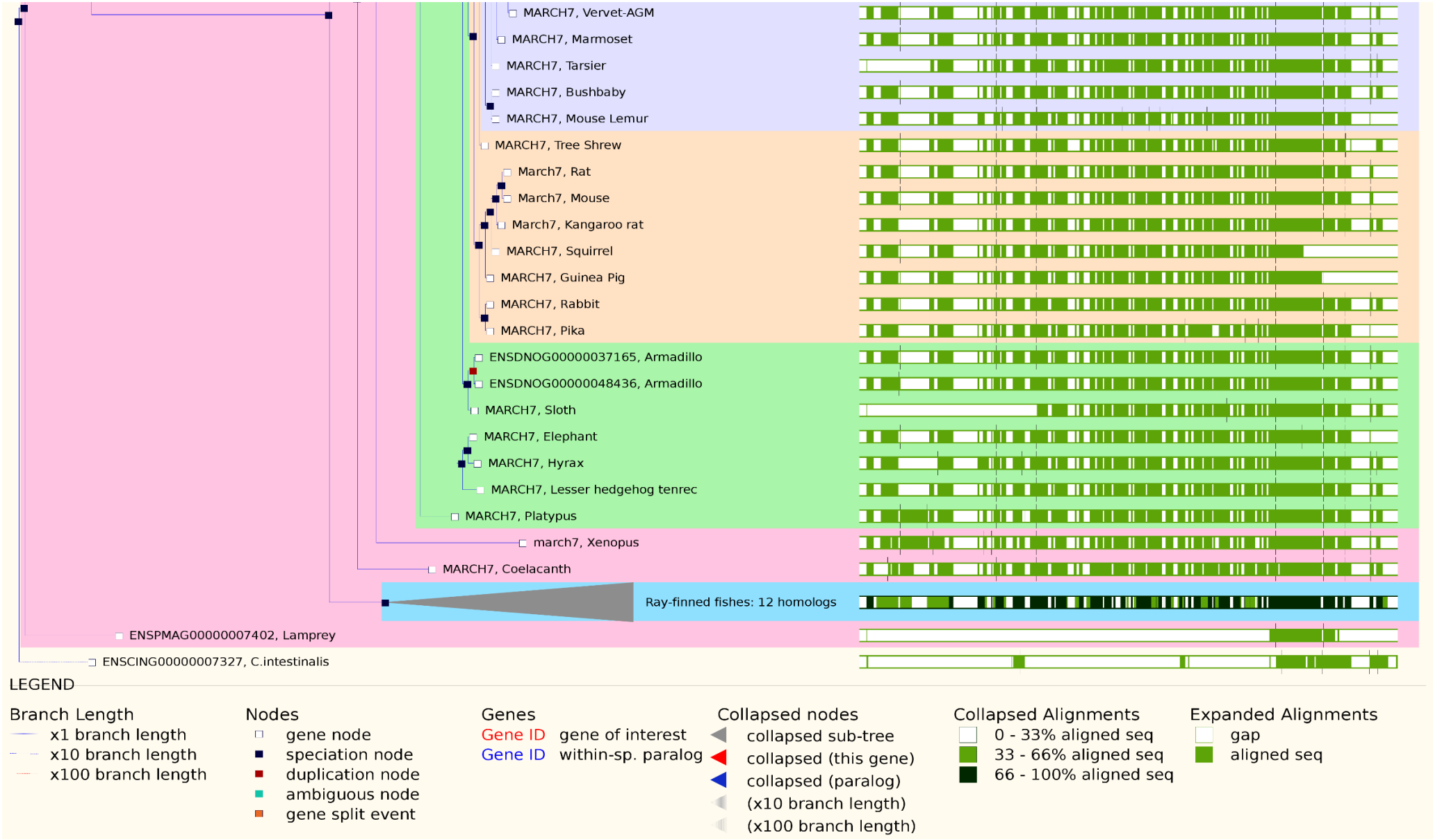


Figure 7-14: Ensembl Phylogenetic tree of MARCH7 including sequence similarity

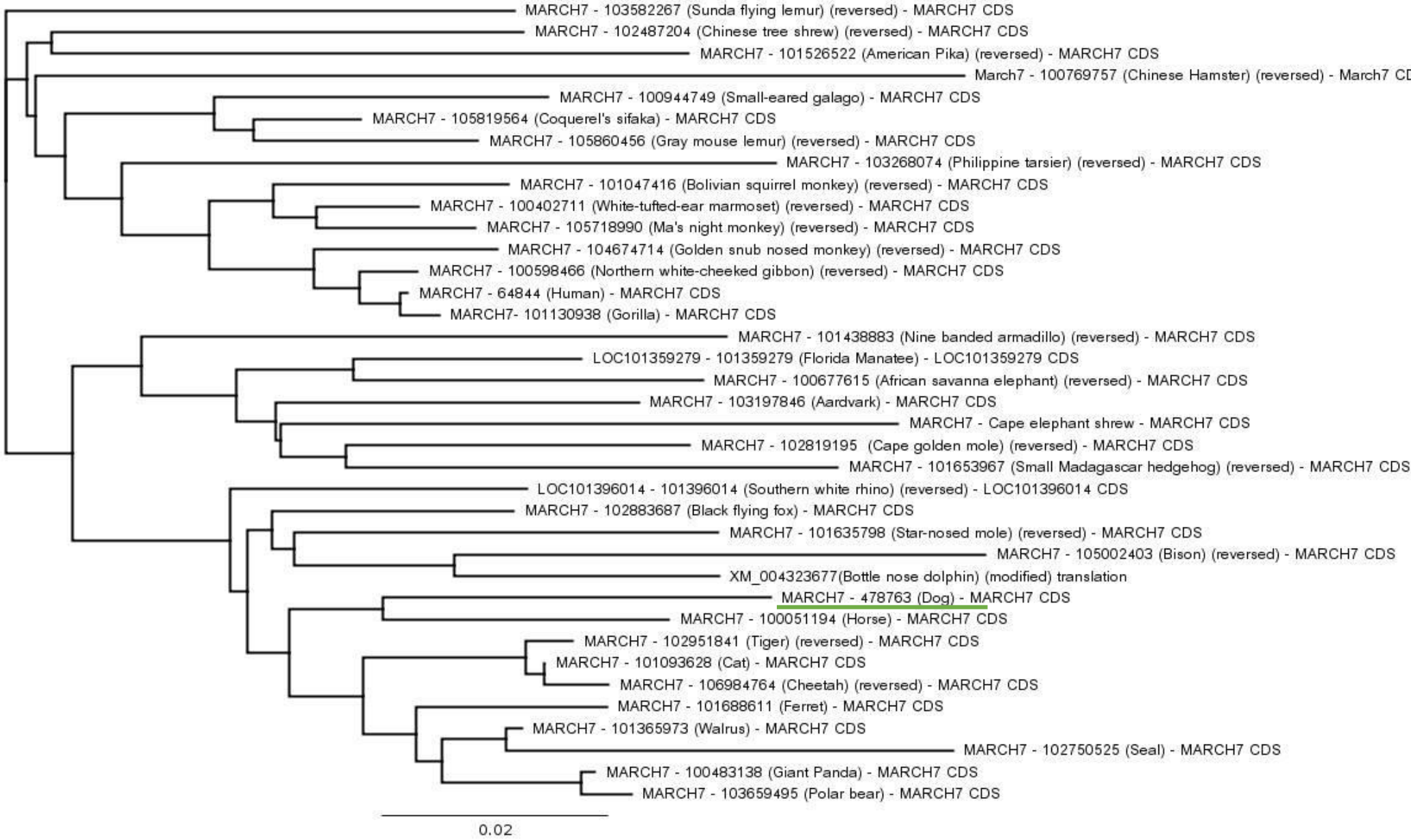


Figure 7-15: MrBayes Phylogenetic tree generated from Geneious. Dog is underlined in green.

7.18 DOG GENOME SNP DATABASE RESULTS FOR *MARCH7*

Table 82: List of single nucleotide polymorphisms present in Dog Genome SNP Database for MARCH7

Species	SNP ID	Chromosome	Position	Reference	Mutation	Region	Flank sequence
Gray Wolf					No result		
Chinese indigenous dog					No result		
Basenji	snp_cf0004002919958	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Dingo	snp_cf0007003138013	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG C/T GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Wolf					No result		
Wolf	snp_cf0006004028211	36	44,347	C	Y	Intron	GTATGATTTTGTAGTTTTTTCATTATTTTA **C/Y** CCCTAACTTTTTAAAGCAAAAGGCAAATAA
Wolf	snp_cf0006004028212	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog					No result		
Chinese indigenous dog	snp_cf0009003485508	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0010003513912	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0011003556583	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0012003496986	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0013003424899	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese	snp_cf0014003466774	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/**

indigenous dog							**T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0015003489817	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C**/**T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0016003502439	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C**/**T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0017003541432	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C**/**Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Belgian Malinois	snp_cf0018002903040	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C**/**T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
German Shepherd Dog	snp_cf0019002719110	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C**/**T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Tibetan Mastiff	snp_cf0020003466691	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C**/**T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0021003356206	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C**/**T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0023003341698	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C**/**Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
German Shepherd	snp_cf0024002755415	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C**/**T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
German Shepherd	snp_cf0025002722385	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C**/**T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
German Shepherd	snp_cf0026002732156	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C**/**T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
German Shepherd	snp_cf0027002735561	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C**/**T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
German Shepherd	snp_cf0028002642218	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C**/**T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
German Shepherd	snp_cf0029002583974	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C**/**T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG

German Shepherd	snp_cf0030002712319	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
German Shepherd	snp_cf0031002770680	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
German Shepherd	snp_cf0032002701285	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
German Shepherd	snp_cf0033002746191	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Wolf	snp_cf0034004036903	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0035003385324	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0036002767776	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0037002909701	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0038002803795	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0039002934216	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0040003034012	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0041003493373	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0042003642603	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0043002939157	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0044003079867	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG

Chinese indigenous dog	snp_cf0045003185860	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0046003255049	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0047003398963	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0048003333563	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0049003519025	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0050003252797	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog					No result		
Chinese indigenous dog	snp_cf0052003435336	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG C/Y GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0053003421833	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG C/Y GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0054003196256	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG C/Y GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog					No result		
Wolf	snp_cf0056004342143	36	44,347	C	Y	Intron	GTATGATTTTGTAGTTTTTTCATTATTTTA **C/Y** CCCTAACTTTTTAAAGCAAAAGGCAAATAA
Wolf	snp_cf0056004342144	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Wolf	snp_cf0057004133108	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Wolf	snp_cf0058004312318	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG

Wolf	snp_cf0058004312317	36	44,347	C	Y	Intron	GTATGATTTTGTAGTTTTTTCATTATTTTA **C/Y** CCCTAACTTTTTAAAGCAAAAGGCAAATAA
Tibetan Mastiff	snp_cf0059003495968	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Tibetan Mastiff	snp_cf0060003418123	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Tibetan Mastiff					no result		
Tibetan Mastiff	snp_cf0062003497794	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Tibetan Mastiff	snp_cf0063003456621	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Tibetan Mastiff					no result		
Tibetan Mastiff	snp_cf0065003484378	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Tibetan Mastiff	snp_cf0066003584807	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Tibetan Mastiff	snp_cf0067003452758	36	43,674	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Tibetan Mastiff	snp_cf0068003461207	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog					no result		
Chinese indigenous dog					no result		
Chinese indigenous dog	snp_cf0071003579565	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0072003589826	36	43,674	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese					no result		

indigenous dog							
Chinese indigenous dog					no result		
Chinese indigenous dog	snp_cf0075003583050	36	5525802	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0076003602498	36	5525802	C	T	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/T** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0077003565743	36	5525802	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG
Chinese indigenous dog	snp_cf0078003574304	36	5525802	C	Y	Intron	AAGGATGATTCTAGGGCCATACTATCCAAG **C/Y** GTGGTTAGCATATTTGTGGTAAAAAGAGAG

7.19 GENETIC VARIATION DATA FOR *MARCH7*; COMPARISON OF THE DOMESTIC DOG WITH CARNIVORES

Table 83: Table of SNPS present in MARCH7 for dog and carnivores

Positon	Consensus	Dog	Giant Panda	Cat	Walrus	Ferret	Seal	Tiger	Polar bear	Cheetah	Only in dogs
26,370	C	T	C	A	C	C	C	A	C	A	
26,371	C	T	C	T	C	C	C	T	C	T	
26,373	C	T	C	C	C	C	C	C	C	C	X
26,380	T	G	T	T	T	T	T	T	T	T	X
26,391	G	A	G	G	G	G	G	G	G	G	X
26,445	C	T	C	C	G	C	G	C	C	C	
56,692	T	C	T	T	T	T	T	T	T	T	X
52,725	G	A	A	G	G	G	G	G	A	G	
52,767	A	C	A	C	A	A	A	C	A	C	
52,785	G	C	G	G	G	G	G	G	G	G	X
52,824	C	T	C	T	C	C	C	T	C	T	
52,830	C	T	C	C	C	C	C	C	C	C	X
52,848	T	C	T	T	C	T	C	T	T	T	
52, 882	T	C	T	T	T	T	T	T	T	T	X
54,633	A	T	A	T	A	A	A	T	A	T	
54,639	A	G	A	G	A	A	A	G	A	G	

54,654	C	T	C	T	C	C	C	T	C	T	
58,212	T	G	T	T	T	T	T	T	T	T	X
58,303	C	T	C	C	C	C	C	C	C	C	X
58,309	G	A	G	G	A	A	A	G	G	G	
58,327	G	T	G	G	G	G	G	G	G	G	X
58,349	T	A	T	T	T	T	T	T	T	T	X
58,350	C	A	C	C	C	C	C	C	C	C	X
58,351	T	C	T	T	T	C	T	T	T	T	
58,363	T	G	T	T	T	T	T	T	T	T	X
58,397	A	G	A	A	A	A	A	A	A	A	X
58,402	C	A	C	C	C	C	C	C	C	C	X
58,417	A	G	A	G	A	A	A	G	A	G	
58,435	T	A	T	T	T	T	T	T	T	T	X
58,492	A	G	A	A	A	A	A	A	A	A	X
58,519	C	T	C	C	C	T	C	C	C	C	
58,523	G	A	G	G	G	G	G	G	G	G	X
58,534	A	G	T	A	A	A	G	A	T	A	
58,537	C	T	C	C	C	C	C	C	C	C	X

58,549	A	G	G	G	G	G	G	G	A	G	
58,588	M	T	G	C	A	A	A	C	G	C	
58,625	T	C	T	T	T	T	T	T	T	T	X
58,645	T	C	T	T	T	T	C	T	T	T	
58,666	A	C	A	A	A	A	A	A	A	A	X
58,702	T	C	C	T	T	T	T	T	T	T	X
58,711	C	T	T	C	C	T	C	C	T	C	
58,716	A	G	A	G	A	A	A	A	A	G	
58,720	G	A	G	G	G	G	G	G	G	G	X
58,774	C	A	C	C	C	C	C	A	C	A	
58,852	A	G	A	A	A	A	A	A	A	A	X
58,897	T	C	T	T	T	T	T	T	T	T	X
59,005	T	A	T	T	T	T	T	T	T	T	X
59,045	T	G	T	T	T	T	T	T	T	T	X
59,060	G	T	G	G	G	G	G	G	G	G	X
59,062	T	A	T	A	T	T	T	A	T	A	
59,065	G	A	G	G	A	G	G	G	G	G	
59,071	T	C	T	T	T	T	T	T	T	T	X

59,080	R	A	G	A	A	T	A	G	G	G	
59,090	A	C	A	A	A	A	A	A	A	A	X
59,134	G	A	G	G	G	G	G	G	G	G	X
59,189	T	C	T	T	T	T	T	T	T	T	X
59,209	C	T	T	C	T	C	T	C	C	C	
59,233	T	G	T	C	T	T	T	T	T	T	
59,287	C	T	C	C	C	C	C	C	C	C	X
62,138	C	T	C	C	C	T	N	C	C	C	
62,168	A	G	A	A	A	A	A	A	A	A	X
62,183	A	C	A	A	A	A	A	A	A	A	X
62,192	A	G	A	A	A	A	A	A	A	A	X
62,201	C	T	C	C	C	C	C	C	C	C	X
62,204	G	A	G	G	G	G	G	G	G	G	X
62,255	K	A	G	T	A	G	N	T	G	T	
62,271	T	C	T	T	T	T	T	T	T	T	X
62,291	T	C	T	T	C	C	N	T	T	T	
78,730	A	G	A	G	A	A	A	G	A	G	
79,177	T	C	T	T	T	T	T	T	T	T	X

7.20 GENETIC VARIATION DATA FOR MARCH7: TABLE OF SNPS FOR COMPARISON OF THE DOMESTIC DOG WITH PLACENTAL MAMMALS

Table 84: Table of SNPS present in MARCH7 for dog and placental mammals

Position	Consensus	SNP	Dog	Horse	Dolphin	Cat	Cheetah	Tiger	Walrus	Panda	Polar bear	Ferret	Black flying fox	Coquerels sifaka	Rhino	Bison	Gray mouse lemur	Small-eared galago	Chinese tree shrew	Sunda flying lemur	Human	Gorilla	Gibbon	Snug nosed monkey	White-tufted ear monkey	Ma's night monkey	Bolivian squirrel monkey	Amardillo	Manatee	Elephant	Aardvark	Star nosed mole	Mole	Tarsier	Pika	Shrew	Hedgehog	Seal	Chinese hamster
102	C	T/A/G	C	A	/	A	A	A	C	C	C	C	C	C	C	T	G	C	C	C	T	T	C	C	C	C	C	A	C	C	C	C	C	C	T	C	C	C
105	C	T/A/G	T	T	/	C	C	C	C	C	C	C	T	C	T	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	A	C	T	C	C	T
112	T	T/G	G	T	/	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
123	G	G/A	A	G	/	G	G	G	G	G	G	G	G	A	G	G	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	G
146	G	G/A/C	A	G	/	A	A	A	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	G	G	G	A	A	G	G	A	C	
177	C	C/T/A/G	T	C	/	C	C	C	G	C	C	C	C	C	C	C	C	A	T	C	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	G	C	
210	T	T/C/A	C	T	/	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	A	T	A	T	T	
243	G	G/A/T	A	G	/	G	G	G	G	A	A	G	G	G	G	A	A	G	A	G	A	A	A	A	G	A	A	G	T	T	G	G	A	A	G	G	G	G	G
303	G	G/A	A	G	/	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	A	G	G	G	
348	C	C/T	T	C	/	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	T	T	C	C	C	C	C	C	C	C	C	T	C	T
365	C	C/T	T	C	/	C	C	C	T	T	T	T	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	
366	T	T/C/G	C	G	/	T	T	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	T	
376	G	G/A	A	G	/	A	A	A	A	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	A	

382	A	C/A	C	C	/	C	C	C	C	C	C	C	C	C	A	C	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	C	A	A	A	A	A	C	A
400	C	C/T	T	C	/	T	T	T	T	T	T	T	T	T	T	T	T	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
422	G	G/A	A	G	/	A	A	A	A	A	A	A	A	G	G	G	G	A	A	G	G	G	G	G	A	A	A	G	G	G	G	A	G	G	G	G	G	A	G
432	C	C/T	T	C	/	T	T	T	C	C	C	C	C	C	C	C	C	C	C	C	C	T	T	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C	C
486	G	A/G	A	/	/	A	A	A	A	A	A	A	A	G	G	/	/	G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	A	G	G	A	A	A	G
495	G	A/G	A	A	/	A	A	A	A	A	A	A	A	G	/	/	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	A	G	G	A	A	A	G
513	C	T/C	T	C	/	T	T	T	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	T	C	T	C
575	T	G/T/C	C	T	/	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	T	T	T	T
624	G	T/G/A	T	G	/	T	T	T	T	T	T	T	G	T	A	G	T	T	G	A	A	A	A	A	A	A	A	G	G	G	G	G	G	G	G	G	G	T	G
666	C	C/T	T	C	/	C	C	C	C	C	C	C	T	C	C	C	C	T	C	C	C	C	/	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T
673	T	A/C/T	A	A	/	A	A	A	A	A	A	A	A	T	A	A	T	T	T	T	T	T	T	T	T	T	T	T	T	C	T	A	T	T	T	T	T	A	T
690	G	G/T/A	T	T	/	G	G	G	G	G	G	G	G	G	G	A	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	A	G	G	A
712	T	A/G/T	A	T	/	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		T	T	T	T	G	T	T	T	T	T	T	T
713	C	C/A	A	C	/	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	G	C	C	C	C	C	C	C	C	C
717	T	T/C	T	T	/	T	T	T	T	T	T	T	T	C	T	T	C	C	C	C	/	/	/	/	/	/	/	/	C	C	C	T	C	C	C	C	C	T	C
726	T	T/C	C	T	/	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	C	C	C	T	T	T	T	T	T	T	C	T	T	T	T
<u>750</u>	A	A/G	G	A	/	A	A	A	A	A	A	A	A	A	A	A	A	A	/	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
765	C	C/A/T	A	C	/	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	T	C	C	C	T	C	T	T	C	T
798	A	A/T/C	A	T	/	T	T	T	T	T	T	T	T	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	T	T	T	T
817	C	T/C/G	T	T	/	T	T	T	T	T	T	T	T	C	T	T	C	C	C	C	C	C	C	C	C	C	C	C	G	C	C	T	C	C	C	C	C	T	C
837	T	T/A	T	T	/	T	T	T	T	T	T	T	T	A	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	T	A
855	A	A/G/C	G	A	/	A	A	A	A	A	A	A	A	A	A	C	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G
<u>886</u>	G	A/G	A	G	/	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
891	G	G/A	A	A	/	A	A	A	A	A	A	A	A	G	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	G	A	G
895	A	A/T	T	A	/	T	T	T	T	T	T	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T

897	A	G/T/A	G	G	/	A	A	A	A	T	T	A	G	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	G	T
900	C	C/T/A/G	T	T	/	C	C	C	C	C	C	C	C	C	C	T	A	C	C	C	T	G	G	G	C	C	C	C	C	C	C	C	C	C	C	C	T	C	T
921	G	G/T/A	A	A	/	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	T	G	G	G	G	G	A	G	G	G
951	C	T/C	C	T	/	T	C	C	C	C	C	C	T	T	T	T	T	T	C	C	T	T	T	T	T	T	T	T	T	T	C	C	T	T	A	T	T	C	T
960	T	T/A/G/C	T	A	/	C	C	C	A	G	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
972	T	C/A	C	T	/	T	T	T	C	C	C	C	T	T	C	T	T	T	T	T	T	T	T	T	T	T	C	T	T	T	T	T	T	T	T	T	T	C	T
997	T	C/T	C	/	/	T	T	T	T	T	T	T	T	T	/	T	T	T	T	T	T	T	T	C	T	T	T	C	T	T	C	T	C	T	T	C	G	T	C
1017	T	C/T	C	C	/	T	T	T	T	T	T	T	T	C	C	T	C	C	T	C	C	C	C	C	T	T	T	T	C	T	C	T	C	C	C	T	T	C	T
1020	T	G/T/A	G	T	/	T	T	T	G	G	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	A	T	T	T	G	T
1038	A	A/C/G	C	A	/	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	/
1056	T	C/T	C	T	/	C	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	T
1074	T	C/T	C	T	/	T	T	T	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	T	T	T	T
1077	A	A/T/C	T	T	/	T	T	T	T	T	T	T	C	A	C	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	T	A
1083	C	T/C/A	T	T	/	C	C	C	C	T	T	T	C	C	T	T	C	T	T	T	C	T	C	C	C	C	C	A	A	A	A	T	A	T	C	A	A	C	T
1088	A	A/G/C	G	A	/	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G
1092	G	A/G	A	G	/	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	A	A	G	G
1224	A	A/G/T	G	A	/	A	A	A	A	A	A	A	G	A	A	A	A	A	T	G	A	A	A	A	A	G	A	A	G	A	A	A	A	A	A	G	G	A	G
1269	T	C/T	C	T	/	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	T	T	C	T	T	T	T
1278	G	A/G	A	G	/	A	A	A	A	A	A	A	A	G	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	A	G
1288	C	C/A	C	C	/	C	C	C	C	C	C	C	C	A	C	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	C	A	A	A	C	A	A	A
1335	/	TCA/GC	T	/	/	T	T	T	T	T	T	T	T	/	/	T	/	/	/	/	/	/	/	/	/	/	/	G	T	/	T	T	A	/	/	/	T	T	/
-	/	A	C	/	/	C	C	C	C	C	C	C	C	/	/	C	/	/	/	/	/	/	/	/	/	/	/	C	C	/	C	C	C	/	/	/	C	C	/
1338	/		A	/	/	A	A	A	A	A	A	A	A	/	/	A	/	/	/	/	/	/	/	/	/	/	/	A	A	/	A	A	A	/	/	/	A	A	/
1377	T	T/A/C/G	A	T	/	T	T	T	T	T	T	T	T	C	T	T	T	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	/ T

1388	R	G/A	G	G	/	G	G	G	G	G	G	G	G	A	G	G	A	G	G	A	A	A	A	A	A	G	G	A	A	A	A	A	A	A	A	A	G	G	G
1405	G	G/A	A	G	/	A	A	A	A	A	A	A	/	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	T	G	G	G	G	A	A	G
1417	C	C/G/T	G	C	/	T	T	T	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C
1432	G	T/G/C/A	T	G	/	G	G	G	G	G	G	G	G	G	G	G	G	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	C	G	G
1440	C	T/C	T	C	/	T	T	T	T	T	T	T	T	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	T	C
1442	G	G/A	A	G	/	G	G	G	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	G
1443	T	T/C	C	C	/	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	T	T
1462	A	C/A/G	C	A	/	A	A	A	A	A	A	A	A	A	A	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	G	A	G	G	A	A	A
1476	T	C/T	C	T	/	T	T	T	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	T
1506	G	G/A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	A	G	G	A	G	G
1509	T	T/C	C	T	/	C	C	C	C	C	C	C	T	T	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	C	T	T	T	T	C	T
1527	C	T/C	T	C	/	T	T	T	T	T	T	T	C	C	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	T	C
1556	G	G/A	A	A	/	A	A	A	A	A	A	A	A	G	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	G	A	G
1561	T	T/C/G	C	C	/	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	A	T	T	T	T	T	T	T	C
1581	C	T/C	T	C	/	C	C	C	T	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	C	T	C	C	T	T
1605	T	T/C/G	G	T	/	C	T	T	T	T	T	T	T	T	T	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C
1610	A	A/G	G	G	/	G	G	G	G	G	G	G	G	A	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	G	A
1704	C	T/C	T	C	/	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C
1734	A	G/A	G	A	/	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	/	A
1749	A	A/C/G	C	A	/	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
1758	A	G/A/C	G	A	/	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	C	A	G	C	T	A	T	A
1767	C	T/C	T	C	/	C	C	C	C	C	C	C	C	T	C	C	T	T	T	T	T	T	T	T	T	T	T	C	C	C	C	C	C	T	T	C	C	T	T
1770	G	A/G	A	G	/	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	A
1821	A	A/G/T	A	G	/	T	T	T	A	G	G	G	G	G	G	G	G	G	A	G	C	C	C	G	A	G	G	A	G	G	G	G	A	G	G	G	G	/	G

1837	T	C/T	C	T	/	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	/	T
1857	T	C/T/A	C	T	/	T	T	T	C	T	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	A	T	T	T	T	T	T	A	T
2133	T	GGT/TTC	G	G	T	T	T	T	T	T	T	T	T	T	T	T	G	T	T	T	T	T	T	T	T	T	G	T	T	T	T	T	T	T	T	T	T	G
-	T		G	G	T	T	T	T	T	T	T	T	T	T	T	T	G	T	T	T	T	T	T	T	T	T	G	T	T	T	T	T	T	T	T	T	T	G
2135	C		T	T	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	T
2137	G	A/C/G	A	A	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	A
2138	/	T	T	T	/	/	/	/	/	/	/	/	/	/	/	/	T	/	/	/	/	/	/	/	/	/	T	/	/	/	/	/	/	/	/	/	/	T
											/			/				/		/										/	/	/			
														/				/		/										/	/	/			
2140	G	G/A	A	A	G	G	G	G	G	G	G	G	G	G	G	A	A	G	G	G	G	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	A
2141	G	G/A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
2142	A	CCAGA	C	C	A	A	A	A	A	A	A	A	A	A	A	A	C	A	A	A	A	A	A	A	A	A	C	A	A	A	A	A	A	A	A	A	A	C
-	T		C	C	T	T	T	T	T	T	T	T	T	T	T	T	C	T	T	T	T	T	T	T	T	T	C	T	T	T	T	T	T	T	T	T	T	C
2146	G		A	A	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	A
	A		G	G	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	G
	T		A	A	T	T	T	T	T	T	T	T	T	T	T	T	A	T	T	A	T	T	T	T	T	T	A	T	T	T	T	T	T	G	T	T	T	A
2148	C	T/C	T	T	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	T
2150	G	T/G	T	T	G	G	G	G	G	G	G	G	G	G	G	G	T	G	G	G	G	G	G	G	G	G	T	G	G	G	G	G	G	G	G	G	G	T
2152	A	CTC/AG	C	C	A	A	A	A	A	A	A	A	A	A	A	A	C	A	A	A	A	A	A	A	A	A	C	A	A	A	A	A	A	A	A	A	A	C
-	G	A	T	T	G	G	G	G	G	G	G	G	G	G	G	G	T	G	G	G	G	G	G	G	G	G	T	G	G	G	G	G	G	T	G	G	G	T
2154	A		C	C	A	A	A	A	A	A	A	A	A	A	A	A	C	A	A	A	A	A	A	A	A	A	C	A	A	A	A	A	A	A	A	A	A	C
2157	A	A/G/C	G	C	A	A	A	A	A	A	A	A	A	A	A	A	C	A	A	A	A	A	A	A	A	A	C	A	A	A	A	A	A	/	A	A	A	C
2160	G	G/T/C	T	C	G	G	G	G	G	G	G	G	G	G	G	G	C	G	G	G	G	G	G	G	G	G	C	G	G	G	G	G	G	/	G	G	G	C
2161	A	AGA/GT	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	/	A	A	A	A
-	G	G	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G		G	G	G	G
2163	A		G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		A	A	A	A
2164	C	A/C/G	A	G	C	C	C	C	C	C	C	C	C	C	C	G	C	C	C	C	C	C	C	C	C	G	C	C	C	C	C	C	/	C	C	C	C	G
2165	C	G/C	G	G	C	C	C	C	C	C	C	C	C	C	C	G	C	C	C	C	C	C	C	C	C	G	C	C	G	C	C	C	/	C	C	C	C	G

2167	T	G/T	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	/	T	T	T	T	T
2168	A	A/G/C	C	G	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	/	A	A	A	A	G
2171	A	T/A/C	T	/	A	A	A	A	A	A	A	A	A	A	A	A	/	A	A	A	A	A	A	A	A	A	/	A	A	A	A	A	A	/	A	A	C	A	/
2172	G	T/G	T	/	A	A	A	A	A	A	A	A	A	A	A	A	/	A	A	A	A	A	A	A	A	A	/	A	A	A	A	A	A	/	A	A	A	A	/
2173	A	A/G	G	/	A	A	A	A	A	A	A	A	A	A	A	A	/	A	A	A	G	G	G	G	G	G	/	A	A	A	A	A	A	/	A	A	A	A	/
2174	A	A/C	C	/	A	A	A	A	A	A	A	A	A	A	A	A	/	A	A	A	A	A	A	A	A	A	/	A	A	A	A	A	G	/	A	A	G	A	/
2176	A	A/G/C/T	T	/	A	A	A	A	A	A	A	A	A	A	A	A	/	A	A	G	A	A	A	A	A	A	/	A	A	A	A	A	A	/	C	A	A	A	/
2178 - 2179	T T	TT/AA	A A	/	T T	T T	T T	T T	T T	T T	T T	T T	T T	T T	T T	T T	/	T T	T T	T T	T T	T T	T T	T T	T T	T T	/	T T	T T	T T	T T	T T	T T	/	T T	T T	T T	T T	/
2181	A	A/T	T	/	A	A	A	A	A	A	A	A	A	A	A	A	/	A	A	A	A	A	A	A	A	A	/	A	A	A	A	A	A	/	A	A	A	A	/
2183	A	C/A	C	/	A	A	A	A	A	A	A	A	A	A	A	A	/	A	A	A	A	A	A	A	A	A	/	A	A	A	A	A	A	/	A	A	A	A	/
2184	T	A/T	A	/	T	T	T	T	T	T	T	T	T	T	T	T	/	C	T	T	T	T	T	T	T	T	/	T	T	T	T	T	T	/	T	T	T	T	/
2186 - 2188	G C C	GCC/TTT	T T T	/	G C C	G C C	G C C	G C C	G C C	G C C	G C C	G C C	G C C	G C C	G C C	G C C	/	G C C	G C C	G C C	G C C	G C C	G C C	G C C	G C C	G C C	/	G C C	G C C	G C C	G C C	G C C	G C C	/	G C C	G C C	G C C	G C C	/
2190	A	T/A	T	/	A	A	A	A	A	A	A	A	A	A	A	A	/	A	A	A	A	A	A	A	A	A	/	A	A	A	A	A	A	/	A	A	A	A	/
2191	A	C/G/A	C	/	A	A	A	A	A	A	A	A	A	A	A	A	/	A	A	A	A	A	A	A	A	A	/	A	A	G	A	A	A	/	A	A	A	A	/

7.21 GENETIC VARIATION DATA FOR *MARCH7*; COMPARISON OF THE DOMESTIC DOG TO PLACENTAL MAMMALS - AMINO ACID SUBSTITUTIONS

Table 85: Amino acid substitutions comparing the domestic dog to the placental mammals

SNP Position	1st	2nd	3rd	Mammals	Dog	Only in dog	Gaps
102							
105							
112				L	V	X	
123							
146				N/S/G	N		
177				D/E	D		
210				S/C/V	S		
243				Q/P/R/H/E	Q		
303				Q	Q		
348				N	N		
365				A/T	V		
366				A/T	V		
376				I/V	I		
382				H/N	H		
400				S/A	P	X	
422				N/G/S	N		
432				P	P		X
486				R	R		
495				L/V	L		X
505				A/T/S	A		
513				S	S		
575				V/I	G	X	
624				D/E	D		
666				N/H	N		X
673				M/L	,		
690				H/Q	H		
712				N/A/S	N	X	
713				N/A/S	N	X	
717				S	S		X
726				S/P	S		
727				L/F	L		
750				R	R		X
765				P/S/L/F	P		
798				I/L/V/M/S/T	I		
817				S/P	S		X
837				D/E	D		
855				Q/H/R	Q		
886				S/G	S	X	

891				R	R			
895				S/T/I	S			
897				S/T/I	S			
900				T	T			
921				L	L			
951				T	T			
960				S/P	S			
972				S/N	S			
997				L/V	L			X
1017				Y/C	Y			
1018				V/I	V			
1020				V/I	V			
1038				T/A	T			X
1056				S/N/R	S			
1074				S/T	S			
1077				D/E/N	D			
1083				P/S/T/N	P			
1088				S/N/D/T/G	S			X
1092				R	R			
1224				P	P			
1269				S	S			
1278				R	R			
1288				R	R			
1335				S/A	S			X
1336-1337	X	X		H	H			X
1377				L	L			
1388				S/N	S			
1405				T/A	T			
1417				A/P/S	A	X		
1432				S/A/Q/E	S	X		
1440				S/G/T	S			
1442				N/S	N			
1443				N/S	N			
1462				P/T/I/A	P			
1476				A/V/T/P	A			
1506				G	G			
1509				I/V/F	I			
1527				F	F			
1556				N/S	N			
1561				L/M	L			
1581				I	I			
1605				G	G			
1610				S/N	S			
1704				D/G	D			
1734				R	R			X

1749				A	A		X
1758				S/F	S		
1767				N/A	A		
1770				L	L		
1821				E/D	E		X
1837				L	L		X
1857				S/P	S		
2133				T	T		
2134-2135	X	X		V/F/S	V		X
2137				I	I		X
2138				I	I		
2140				N/S/G/D	N	X	
2141				N/S/G/D	N	X	
2142				N/S/G/D	N	X	
2143-2145	X	X	X	Q/S/G/D/	Q		
2146				I/F/S	I		
2148				I/F/S	I		
2150				L/E	L		
2152-2154	X	X	X	L/R/E/D	L		
2157				R/S/D	R		
2160				C/F/W/G	C	X	
2161-2163	X	X	X	V/R/D	V	X	
2164				R/G/P/H/L	R	X	
2165				R/G/P/H/L	R	X	
2167				A/N	A	X	
2168				A/N	A	X	
2171				L/Q/R	L	X	
2172				L/Q/R	L	X	
2173				A/N/T	A		
2174				A/N/T	A		
2176				L/I/F	L	X	
2178				L/I/F	L	X	
2179				S/D	S	X	
2181				S/D	S	X	
2183				S/Y/I/T	S	X	
2184				S/Y/I/T	S	X	
2186-2187		X	X	F/C/A	F	X	
2188				F/L	F	X	
2190				F/L	F	X	
2191				*	L	X	

7.22 OUTPUT FROM DNASP

7.22.1Comparison of the domestic dog with carnivores

7.22.1.1Synonymous and NonSynonymous Substitutions

```
------------------------------------
Input Data File: C:\...\Translated-Carnivores.fas
Number of sequences: 9    Number of sequences used: 9
Selected region: 1-2124   Number of sites: 2124
Total number of sites (excluding sites with gaps / missing data): 1728

Number of codons analyzed: 576   ( 1728 sites )
  Total number of codons with alignment gaps or missing data: 132
Genetic Code: Nuclear Universal

Protein Coding, and Non-Coding Regions analyzed:
 Number of protein coding regions (exons): 1
 Number of noncoding regions (intronic and flanking regions): 0
  Protein coding region, from site: 1  to  2124

Nucleotide Diversity:
  Synonymous sites. Number of sites: 414.85
     Pi(s): 0.09703    Pi(s), Jukes & Cantor: 0.10521
     Theta(s) / Number of mutations: n.a.
  NonSynonymous sites. Number of sites: 1313.15
     Pi(a): 0.01696    Pi(a), Jukes & Cantor: 0.01718
     Theta(a) / Number of mutations: n.a.

Protein Coding Region.  Total Number of sites
SS, Synonymous sites.   NSS, NonSynonymous sites
  MARCH7_-_478763_(DogSS:  414.33       NSS: 1313.67
  MARCH7_-_100483138_(SS:  417.17       NSS: 1310.83
  MARCH7_-_103659495_(SS:  417.33       NSS: 1310.67
  XM_004394827       SS:  414.50       NSS: 1313.50
  MARCH7_-_101688611_(SS:  414.83       NSS: 1313.17
  MARCH7_-_101093628_(SS:  413.33       NSS: 1314.67
  MARCH7_-_106984764_(SS:  414.00       NSS: 1314.00
  MARCH7_-_102951841_(SS:  413.33       NSS: 1314.67
  MARCH7_-_102750525_(SS:  414.83       NSS: 1313.17
```

Table 86: Comparison of synonymous and nonsynonymous substitutions, output from DNAsp; dog and carnivores

Seq 1	Seq 2	SynDif	SynPos	Ks	N	SynDif	NSynPos	Ka	Ka/Ks
MARCH7_-_4	**MARCH7_-_1**	57.83	415.75	0.1539		24.17	1312.25	0.0186	0.120858
MARCH7_-_4	**MARCH7_-_1**	59	415.83	0.1573		26	1312.17	0.0201	0.127781
MARCH7_-_4	**XM_0043948**	44	414.42	0.1145		18	1313.58	0.0138	0.120524
MARCH7_-_4	**MARCH7_-_1**	64	414.58	0.1728		25	1313.42	0.0193	0.11169
MARCH7_-_4	**MARCH7_-_1**	60	413.83	0.1611		23	1314.17	0.0177	0.10987
MARCH7_-_4	**MARCH7_-_1**	60	414.17	0.161		27	1313.83	0.0208	0.129193
MARCH7_-_4	**MARCH7_-_1**	60	413.83	0.1611		27	1314.17	0.0208	0.129112
MARCH7_-_4	**MARCH7_-_1**	44	414.58	0.1144		23	1313.42	0.0177	0.15472
MARCH7_-_1	**MARCH7_-_1**	15	417.25	0.0368		4	1310.75	0.0031	0.084239
MARCH7_-_1	**XM_0043948**	26.83	415.83	0.0675		16.17	1312.17	0.0124	0.183704
MARCH7_-_1	**MARCH7_-_1**	44.83	416	0.1163		25.17	1312	0.0194	0.16681
MARCH7_-_1	**MARCH7_-_1**	48.83	415.25	0.1279		27.17	1312.75	0.021	0.164191
MARCH7_-_1	**MARCH7_-_1**	48	415.58	0.1254		31	1312.42	0.024	0.191388
MARCH7_-_1	**MARCH7_-_1**	48	415.25	0.1255		29	1312.75	0.0224	0.178486
MARCH7_-_1	**MARCH7_-_1**	26.83	416	0.0674		20.17	1312	0.0155	0.22997
MARCH7_-_1	**XM_0043948**	28	415.92	0.0705		18	1312.08	0.0138	0.195745
MARCH7_-_1	**MARCH7_-_1**	44	416.08	0.114		27	1311.92	0.0209	0.183333
MARCH7_-_1	**MARCH7_-_1**	50	415.33	0.1312		29	1312.67	0.0224	0.170732
MARCH7_-_1	**MARCH7_-_1**	49	415.67	0.1282		33	1312.33	0.0256	0.199688
MARCH7_-_1	**MARCH7_-_1**	49	415.33	0.1284		31	1312.67	0.024	0.186916
MARCH7_-_1	**MARCH7_-_1**	29	416.08	0.0732		22	1311.92	0.017	0.23224

XM_0043948 MARCH7_-_1	29	414.67	0.0734	17	1313.33	0.0131	0.178474
XM_0043948 MARCH7_-_1	38	413.92	0.0979	21	1314.08	0.0162	0.165475
XM_0043948 MARCH7_-_1	39	414.25	0.1006	25	1313.75	0.0193	0.191849
XM_0043948 MARCH7_-_1	39	413.92	0.1007	21	1314.08	0.0162	0.160874
XM_0043948 MARCH7_-_1	8	414.67	0.0195	9	1313.33	0.0069	0.353846
MARCH7_-_1 MARCH7_-_1	55	414.08	0.1462	27	1313.92	0.0208	0.142271
MARCH7_-_1 MARCH7_-_1	55	414.42	0.1461	29	1313.58	0.0224	0.15332
MARCH7_-_1 MARCH7_-_1	55	414.08	0.1462	27	1313.92	0.0208	0.142271
MARCH7_-_1 MARCH7_-_1	31	414.83	0.0787	22	1313.17	0.0169	0.21474
MARCH7_-_1 MARCH7_-_1	7	413.67	0.0171	4	1314.33	0.003	0.175439
MARCH7_-_1 MARCH7_-_1	10	413.33	0.0246	4	1314.67	0.003	0.121951
MARCH7_-_1 MARCH7_-_1	40	414.08	0.1034	26	1313.92	0.0201	0.194391
MARCH7_-_1 MARCH7_-_1	5	413.67	0.0122	8	1314.33	0.0061	0.5
MARCH7_-_1 MARCH7_-_1	41	414.42	0.1061	30	1313.58	0.0232	0.218662
MARCH7_-_1 MARCH7_-_1	41	414.08	0.1062	26	1313.92	0.0201	0.189266

7.22.2 Comparison of the domestic dog with placental mammals

7.22.2.1 Synonymous and NonSynonymous Substitutions

Input Data File: C:\...\FINSIHED-DNA-CODE.fasta
Number of sequences: 37 Number of sequences used: 37
Selected region: 1-2133 Number of sites: 2133
Total number of sites (excluding sites with gaps / missing data): 1599

Number of codons analyzed: 533 (1599 sites)
  Total number of codons with alignment gaps or missing data: 178
Genetic Code: Nuclear Universal

Protein Coding, and Non-Coding Regions analyzed:
 Number of protein coding regions (exons): 1
 Number of noncoding regions (intronic and flanking regions): 0
  Protein coding region, from site: 1 to 2133

Nucleotide Diversity:
  Synonymous sites. Number of sites: 384.09
    Pi(s): 0.19728 Pi(s), Jukes & Cantor: 0.23256
    Theta(s) / Number of mutations: n.a.
  NonSynonymous sites. Number of sites: 1214.91
    Pi(a): 0.04715 Pi(a), Jukes & Cantor: 0.04887
    Theta(a) / Number of mutations: n.a.

Protein Coding Region. Total Number of sites
SS, Synonymous sites. NSS, NonSynonymous sites
  MARCH7_-_478763_(DogSS: 384.50 NSS: 1214.50
  MARCH7_-_100051194_(SS: 385.50 NSS: 1213.50
  MARCH7_-_101093628_(SS: 383.67 NSS: 1215.33
  MARCH7_-_106984764_(SS: 384.33 NSS: 1214.67
  MARCH7_-_102951841_(SS: 383.67 NSS: 1215.33
  MARCH7_-_101365973_(SS: 384.83 NSS: 1214.17
  MARCH7_-_100483138_(SS: 387.17 NSS: 1211.83
  MARCH7_-_103659495_(SS: 387.67 NSS: 1211.33
  MARCH7_-_101688611_(SS: 384.50 NSS: 1214.50
  MARCH7_-_102883687_(SS: 382.67 NSS: 1216.33
  MARCH7_-_105819564_(SS: 382.83 NSS: 1216.17
  LOC101396014_-_10139SS: 381.17 NSS: 1217.83
  MARCH7_-_105002403_(SS: 387.67 NSS: 1211.33
  MARCH7_-_105860456_(SS: 385.17 NSS: 1213.83
  MARCH7_-_100944749_(SS: 385.33 NSS: 1213.67
  MARCH7_-_102487204_(SS: 385.50 NSS: 1213.50
  MARCH7_-_103582267_(SS: 383.67 NSS: 1215.33
  MARCH7_-_64844_(HumaSS: 382.00 NSS: 1217.00
  MARCH7-_101130938_(GSS: 382.17 NSS: 1216.83
  MARCH7_-_100598466_(SS: 383.00 NSS: 1216.00
  MARCH7_-_104674714_(SS: 384.67 NSS: 1214.33
  MARCH7_-_100402711_(SS: 384.83 NSS: 1214.17
  MARCH7_-_105718990_(SS: 384.67 NSS: 1214.33
  MARCH7_-_101047416_(SS: 385.00 NSS: 1214.00
  MARCH7_-_101438883_(SS: 381.17 NSS: 1217.83

LOC101359279_-_10135SS: 383.17 NSS: 1215.83
MARCH7_-_100677615_(SS: 382.50 NSS: 1216.50
MARCH7_-_103197846_(SS: 385.83 NSS: 1213.17
MARCH7_-_101635798_(SS: 383.50 NSS: 1215.50
MARCH7_-_102819195__SS: 380.00 NSS: 1219.00
MARCH7_-_103268074_(SS: 382.00 NSS: 1217.00

Table 87: Synonymous and nonsynonymous substitutions output from DNAsp; dog and placental mammals

Seq1	Seq2	SynDif	Syn Pos	Ks	NSynDif	NsynPos	Ka	Dn/Ds
MARCH7_-_4	MARCH7_-_1	62	385	0.1813	30	1214	0.0251	0.138445
MARCH7_-_4	MARCH7_-_1	60	384.08	0.1752	21	1214.92	0.0175	0.099886
MARCH7_-_4	MARCH7_-_1	59	384.42	0.1717	25	1214.58	0.0209	0.121724
MARCH7_-_4	MARCH7_-_1	59	384.08	0.1719	24	1214.92	0.02	0.116347
MARCH7_-_4	MARCH7_-_1	45	384.67	0.1272	15	1214.33	0.0125	0.09827
MARCH7_-_4	MARCH7_-_1	54	385.83	0.1549	18	1213.17	0.015	0.096837
MARCH7_-_4	MARCH7_-_1	56	386.08	0.1612	21	1212.92	0.0175	0.108561
MARCH7_-_4	MARCH7_-_1	60	384.5	0.175	20	1214.5	0.0167	0.095429
MARCH7_-_4	MARCH7_-_1	68.5	383.58	0.204	31.5	1215.42	0.0264	0.129412
MARCH7_-_4	MARCH7_-_1	80.5	383.67	0.2461	47.5	1215.33	0.0401	0.162942
MARCH7_-_4	LOC1013960	66.5	382.83	0.1976	34.5	1216.17	0.0289	0.146255
MARCH7_-_4	MARCH7_-_1	82	386.08	0.2497	59	1212.92	0.0503	0.201442
MARCH7_-_4	MARCH7_-_1	73.5	384.83	0.2204	51.5	1214.17	0.0437	0.198276
MARCH7_-_4	MARCH7_-_1	88.5	384.92	0.2746	56.5	1214.08	0.048	0.1748
MARCH7_-_4	MARCH7_-_1	76.5	385	0.2308	52.5	1214	0.0445	0.192808
MARCH7_-_4	MARCH7_-_1	65.5	384.08	0.1935	58.5	1214.92	0.0498	0.257364
MARCH7_-_4	MARCH7_-_6	76.5	383.25	0.2321	53.5	1215.75	0.0453	0.195174
MARCH7_-_4	MARCH7-_10	76.5	383.33	0.232	55.5	1215.67	0.0471	0.203017
MARCH7_-_4	MARCH7_-_1	83.5	383.75	0.257	54.5	1215.25	0.0462	0.179767
MARCH7_-_4	MARCH7_-_1	85	384.58	0.2618	62	1214.42	0.0529	0.202063
MARCH7_-_4	MARCH7_-_1	80	384.67	0.2436	50	1214.33	0.0423	0.173645
MARCH7_-_4	MARCH7_-_1	78	384.58	0.2365	53	1214.42	0.045	0.190275
MARCH7_-_4	MARCH7_-_1	85	384.75	0.2617	51	1214.25	0.0432	0.165075
MARCH7_-_4	MARCH7_-_1	95.5	382.83	0.3033	63.5	1216.17	0.0541	0.178371
MARCH7_-_4	LOC1013592	87.83	383.83	0.273	54.17	1215.17	0.046	0.168498
MARCH7_-_4	MARCH7_-_1	93	383.5	0.2929	56	1215.5	0.0475	0.162171

MARCH7_-_4	MARCH7_-_1	86	385.17	0.2651	54	1213.83	0.0459	0.173142
MARCH7_-_4	MARCH7_-_1	88.17	384	0.2741	37.83	1215	0.0318	0.116016
MARCH7_-_4	MARCH7_-_1	96.67	382.25	0.3084	63.33	1216.75	0.0539	0.174773
MARCH7_-_4	MARCH7_-_1	77.67	383.25	0.2362	71.33	1215.75	0.0611	0.258679
MARCH7_-_4	MARCH7_-_1	101.5	382.5	0.3275	70.5	1216.5	0.0603	0.184122
MARCH7_-_4	MARCH7_-_C	114	383.75	0.3782	77	1215.25	0.0662	0.17504
MARCH7_-_4	MARCH7_-_1	120.33	384.5	0.405	69.67	1214.5	0.0597	0.147407
MARCH7_-_4	MARCH7_-_1	58.42	386.33	0.1689	56.58	1212.67	0.0482	0.285376
MARCH7_-_4	March7_-_1	87.5	384.58	0.2711	71.5	1214.42	0.0613	0.226116
MARCH7_-_4	XM_0043236	68.5	385.25	0.2029	38.5	1213.75	0.0324	0.159685
MARCH7_-_1	MARCH7_-_1	60	384.58	0.1749	33	1214.42	0.0277	0.158376
MARCH7_-_1	MARCH7_-_1	60	384.92	0.1747	35	1214.08	0.0294	0.168288
MARCH7_-_1	MARCH7_-_1	60	384.58	0.1749	36	1214.42	0.0302	0.17267
MARCH7_-_1	MARCH7_-_1	52	385.17	0.1488	32	1213.83	0.0268	0.180108
MARCH7_-_1	MARCH7_-_1	65	386.33	0.1905	35	1212.67	0.0294	0.154331
MARCH7_-_1	MARCH7_-_1	63	386.58	0.1837	36	1212.42	0.0303	0.164943
MARCH7_-_1	MARCH7_-_1	63	385	0.1846	37	1214	0.0311	0.168472
MARCH7_-_1	MARCH7_-_1	60	384.08	0.1752	32	1214.92	0.0268	0.152968
MARCH7_-_1	MARCH7_-_1	61	384.17	0.1784	48	1214.83	0.0406	0.227578
MARCH7_-_1	LOC1013960	37.5	383.33	0.1048	31.5	1215.67	0.0264	0.251908
MARCH7_-_1	MARCH7_-_1	74	386.58	0.221	57	1212.42	0.0486	0.21991
MARCH7_-_1	MARCH7_-_1	59	385.33	0.1713	53	1213.67	0.045	0.262697
MARCH7_-_1	MARCH7_-_1	77	385.42	0.2323	59	1213.58	0.0503	0.21653
MARCH7_-_1	MARCH7_-_1	67	385.5	0.1977	55	1213.5	0.0468	0.236722
MARCH7_-_1	MARCH7_-_1	56	384.58	0.1619	56	1214.42	0.0476	0.294009
MARCH7_-_1	MARCH7_-_6	64	383.75	0.1886	49	1215.25	0.0414	0.219512
MARCH7_-_1	MARCH7-_10	64	383.83	0.1886	51	1215.17	0.0432	0.229056
MARCH7_-_1	MARCH7_-_1	69	384.25	0.2053	50	1214.75	0.0423	0.20604
MARCH7_-_1	MARCH7_-_1	68.5	385.08	0.203	58.5	1213.92	0.0498	0.24532
MARCH7_-_1	MARCH7_-_1	66	385.17	0.1945	49	1213.83	0.0415	0.213368

MARCH7_-_1	MARCH7_-_1	72	385.08	0.2151	54	1213.92	0.0459	0.213389
MARCH7_-_1	MARCH7_-_1	72	385.25	0.215	52	1213.75	0.0441	0.205116
MARCH7_-_1	MARCH7_-_1	75.5	383.33	0.2285	63.5	1215.67	0.0541	0.236761
MARCH7_-_1	LOC1013592	71.83	384.33	0.215	53.17	1214.67	0.0451	0.209767
MARCH7_-_1	MARCH7_-_1	74	384	0.2227	55	1215	0.0467	0.209699
MARCH7_-_1	MARCH7_-_1	78	385.67	0.2357	55	1213.33	0.0468	0.198557
MARCH7_-_1	MARCH7_-_1	76.67	384.5	0.2318	38.33	1214.5	0.0322	0.138913
MARCH7_-_1	MARCH7_-_1	79.67	382.75	0.2438	65.33	1216.25	0.0557	0.228466
MARCH7_-_1	MARCH7_-_1	65.5	383.75	0.1937	67.5	1215.25	0.0577	0.297883
MARCH7_-_1	MARCH7_-_1	96.5	383	0.307	68.5	1216	0.0586	0.190879
MARCH7_-_1	MARCH7_-_C	112	384.25	0.369	76	1214.75	0.0653	0.176965
MARCH7_-_1	MARCH7_-_1	108.33	385	0.3527	71.67	1214	0.0615	0.174369
MARCH7_-_1	MARCH7_-_1	71.25	386.83	0.2114	73.75	1212.17	0.0635	0.300378
MARCH7_-_1	March7_-_1	76	385.08	0.229	75	1213.92	0.0645	0.281659
MARCH7_-_1	XM_0043236	53	385.75	0.1518	36	1213.25	0.0303	0.199605
MARCH7_-_1	MARCH7_-_1	6	384	0.0158	4	1215	0.0033	0.208861
MARCH7_-_1	MARCH7_-_1	9	383.67	0.0238	3	1215.33	0.0025	0.105042
MARCH7_-_1	MARCH7_-_1	39	384.25	0.1091	20	1214.75	0.0166	0.152154
MARCH7_-_1	MARCH7_-_1	47	385.42	0.1331	23	1213.58	0.0192	0.144252
MARCH7_-_1	MARCH7_-_1	47	385.67	0.133	26	1213.33	0.0217	0.163158
MARCH7_-_1	MARCH7_-_1	53	384.08	0.1525	24	1214.92	0.02	0.131148
MARCH7_-_1	MARCH7_-_1	62	383.17	0.1823	33	1215.83	0.0276	0.151399
MARCH7_-_1	MARCH7_-_1	66	383.25	0.1956	47	1215.75	0.0397	0.202965
MARCH7_-_1	LOC1013960	55	382.42	0.1597	33	1216.58	0.0276	0.172824
MARCH7_-_1	MARCH7_-_1	74	385.67	0.2216	59	1213.33	0.0503	0.226986
MARCH7_-_1	MARCH7_-_1	60	384.42	0.175	48	1214.58	0.0406	0.232
MARCH7_-_1	MARCH7_-_1	82	384.5	0.2509	57	1214.5	0.0485	0.193304
MARCH7_-_1	MARCH7_-_1	68.5	384.58	0.2034	53.5	1214.42	0.0454	0.223206
MARCH7_-_1	MARCH7_-_1	59	383.67	0.1721	56	1215.33	0.0476	0.276583
MARCH7_-_1	MARCH7_-_6	68	382.83	0.2027	48	1216.17	0.0405	0.199803

MARCH7_-_1	MARCH7-_10	68	382.92	0.2027	50	1216.08	0.0423	0.208683
MARCH7_-_1	MARCH7_-_1	71	383.33	0.2127	51	1215.67	0.0432	0.203103
MARCH7_-_1	MARCH7_-_1	72.5	384.17	0.2174	59.5	1214.83	0.0507	0.233211
MARCH7_-_1	MARCH7_-_1	70	384.25	0.2087	49	1214.75	0.0415	0.19885
MARCH7_-_1	MARCH7_-_1	73	384.17	0.2191	50	1214.83	0.0423	0.193063
MARCH7_-_1	MARCH7_-_1	73	384.33	0.219	48	1214.67	0.0406	0.185388
MARCH7_-_1	MARCH7_-_1	80.5	382.42	0.2471	58.5	1216.58	0.0497	0.201133
MARCH7_-_1	LOC1013592	68.83	383.42	0.2052	53.17	1215.58	0.0451	0.219786
MARCH7_-_1	MARCH7_-_1	77	383.08	0.234	55	1215.92	0.0467	0.199573
MARCH7_-_1	MARCH7_-_1	76	384.75	0.2293	55	1214.25	0.0467	0.203663
MARCH7_-_1	MARCH7_-_1	74.17	383.58	0.2236	38.83	1215.42	0.0327	0.146243
MARCH7_-_1	MARCH7_-_1	77	381.83	0.2349	61	1217.17	0.0519	0.220945
MARCH7_-_1	MARCH7_-_1	72.5	382.83	0.2183	69.5	1216.17	0.0594	0.272103
MARCH7_-_1	MARCH7_-_1	94.5	382.08	0.3001	68.5	1216.92	0.0585	0.194935
MARCH7_-_1	MARCH7_-_C	108	383.33	0.3533	70	1215.67	0.0599	0.169544
MARCH7_-_1	MARCH7_-_1	111.83	384.08	0.3685	67.17	1214.92	0.0574	0.155767
MARCH7_-_1	MARCH7_-_1	57.58	385.92	0.1664	61.42	1213.08	0.0524	0.314904
MARCH7_-_1	March7_-_1	91	384.17	0.2847	68	1214.83	0.0582	0.204426
MARCH7_-_1	XM_0043236	61	384.83	0.1781	38	1214.17	0.032	0.179674
MARCH7_-_1	MARCH7_-_1	5	384	0.0131	7	1215	0.0058	0.442748
MARCH7_-_1	MARCH7_-_1	39	384.58	0.109	24	1214.42	0.02	0.183486
MARCH7_-_1	MARCH7_-_1	47	385.75	0.133	27	1213.25	0.0226	0.169925
MARCH7_-_1	MARCH7_-_1	47	386	0.1329	30	1213	0.0251	0.188864
MARCH7_-_1	MARCH7_-_1	53	384.42	0.1523	26	1214.58	0.0217	0.142482
MARCH7_-_1	MARCH7_-_1	63	383.5	0.1854	35	1215.5	0.0294	0.158576
MARCH7_-_1	MARCH7_-_1	66	383.58	0.1955	51	1215.42	0.0432	0.220972
MARCH7_-_1	LOC1013960	55	382.75	0.1595	37	1216.25	0.0311	0.194984
MARCH7_-_1	MARCH7_-_1	77	386	0.2319	63	1213	0.0538	0.231997
MARCH7_-_1	MARCH7_-_1	62	384.75	0.1814	52	1214.25	0.0441	0.243109
MARCH7_-_1	MARCH7_-_1	82	384.83	0.2507	61	1214.17	0.052	0.207419

MARCH7_-_1	MARCH7_-_1	68.5	384.92	0.2031	57.5	1214.08	0.0489	0.240768
MARCH7_-_1	MARCH7_-_1	59	384	0.1719	60	1215	0.0511	0.297266
MARCH7_-_1	MARCH7_-_6	68	383.17	0.2025	52	1215.83	0.044	0.217284
MARCH7_-_1	MARCH7-_10	68	383.25	0.2025	54	1215.75	0.0458	0.226173
MARCH7_-_1	MARCH7_-_1	71	383.67	0.2125	55	1215.33	0.0467	0.219765
MARCH7_-_1	MARCH7_-_1	72.5	384.5	0.2172	63.5	1214.5	0.0542	0.24954
MARCH7_-_1	MARCH7_-_1	70	384.58	0.2085	53	1214.42	0.045	0.215827
MARCH7_-_1	MARCH7_-_1	73	384.5	0.2189	54	1214.5	0.0458	0.209228
MARCH7_-_1	MARCH7_-_1	73	384.67	0.2188	52	1214.33	0.0441	0.201554
MARCH7_-_1	MARCH7_-_1	81.5	382.75	0.2505	62.5	1216.25	0.0532	0.212375
MARCH7_-_1	LOC1013592	70.83	383.75	0.2119	57.17	1215.25	0.0486	0.229353
MARCH7_-_1	MARCH7_-_1	79	383.42	0.2409	57	1215.58	0.0484	0.200913
MARCH7_-_1	MARCH7_-_1	76	385.08	0.229	59	1213.92	0.0502	0.219214
MARCH7_-_1	MARCH7_-_1	74.17	383.92	0.2234	42.83	1215.08	0.0361	0.161594
MARCH7_-_1	MARCH7_-_1	79	382.17	0.2418	65	1216.83	0.0554	0.229115
MARCH7_-_1	MARCH7_-_1	75.5	383.17	0.2286	73.5	1215.83	0.063	0.275591
MARCH7_-_1	MARCH7_-_1	94.5	382.42	0.2998	70.5	1216.58	0.0603	0.201134
MARCH7_-_1	MARCH7_-_C	110	383.67	0.3613	74	1215.33	0.0635	0.175754
MARCH7_-_1	MARCH7_-_1	113.83	384.42	0.3767	71.17	1214.58	0.061	0.161933
MARCH7_-_1	MARCH7_-_1	57.58	386.25	0.1662	65.42	1212.75	0.056	0.336943
MARCH7_-_1	March7_-_1	93	384.5	0.292	72	1214.5	0.0618	0.211644
MARCH7_-_1	XM_0043236	61	385.17	0.1779	40	1213.83	0.0337	0.189432
MARCH7_-_1	MARCH7_-_1	39	384.25	0.1091	21	1214.75	0.0175	0.160403
MARCH7_-_1	MARCH7_-_1	47	385.42	0.1331	26	1213.58	0.0217	0.163035
MARCH7_-_1	MARCH7_-_1	47	385.67	0.133	29	1213.33	0.0243	0.182707
MARCH7_-_1	MARCH7_-_1	53	384.08	0.1525	25	1214.92	0.0209	0.137049
MARCH7_-_1	MARCH7_-_1	62	383.17	0.1823	36	1215.83	0.0302	0.165661
MARCH7_-_1	MARCH7_-_1	66	383.25	0.1956	49	1215.75	0.0414	0.211656
MARCH7_-_1	LOC1013960	55	382.42	0.1597	36	1216.58	0.0302	0.189105
MARCH7_-_1	MARCH7_-_1	77	385.67	0.2321	62	1213.33	0.0529	0.227919

MARCH7_-_1	MARCH7_-_1	62	384.42	0.1816	50	1214.58	0.0423	0.23293
MARCH7_-_1	MARCH7_-_1	80	384.5	0.2437	59	1214.5	0.0502	0.205991
MARCH7_-_1	MARCH7_-_1	70.5	384.58	0.2102	56.5	1214.42	0.048	0.228354
MARCH7_-_1	MARCH7_-_1	59	383.67	0.1721	59	1215.33	0.0502	0.291691
MARCH7_-_1	MARCH7_-_6	66	382.83	0.1959	51	1216.17	0.0432	0.220521
MARCH7_-_1	MARCH7-_10	66	382.92	0.1958	53	1216.08	0.0449	0.229316
MARCH7_-_1	MARCH7_-_1	69	383.33	0.2058	54	1215.67	0.0458	0.222546
MARCH7_-_1	MARCH7_-_1	70.5	384.17	0.2105	62.5	1214.83	0.0533	0.253207
MARCH7_-_1	MARCH7_-_1	68	384.25	0.2018	52	1214.75	0.0441	0.218533
MARCH7_-_1	MARCH7_-_1	71	384.17	0.2122	53	1214.83	0.0449	0.211593
MARCH7_-_1	MARCH7_-_1	71	384.33	0.2121	51	1214.67	0.0432	0.203678
MARCH7_-_1	MARCH7_-_1	82.5	382.42	0.2544	61.5	1216.58	0.0523	0.205582
MARCH7_-_1	LOC1013592	70.83	383.42	0.2121	56.17	1215.58	0.0477	0.224894
MARCH7_-_1	MARCH7_-_1	79	383.08	0.2411	58	1215.92	0.0493	0.204479
MARCH7_-_1	MARCH7_-_1	76	384.75	0.2293	56	1214.25	0.0476	0.207588
MARCH7_-_1	MARCH7_-_1	74.17	383.58	0.2236	41.83	1215.42	0.0352	0.157424
MARCH7_-_1	MARCH7_-_1	79	381.83	0.2421	64	1217.17	0.0545	0.225114
MARCH7_-_1	MARCH7_-_1	75.5	382.83	0.2288	72.5	1216.17	0.0621	0.271416
MARCH7_-_1	MARCH7_-_1	95.5	382.08	0.304	71.5	1216.92	0.0612	0.201316
MARCH7_-_1	MARCH7_-_C	108	383.33	0.3533	73	1215.67	0.0626	0.177187
MARCH7_-_1	MARCH7_-_1	113.83	384.08	0.3771	69.17	1214.92	0.0592	0.156988
MARCH7_-_1	MARCH7_-_1	57.58	385.92	0.1664	62.42	1213.08	0.0533	0.320313
MARCH7_-_1	March7_-_1	93	384.17	0.2923	67	1214.83	0.0573	0.196031
MARCH7_-_1	XM_0043236	61	384.83	0.1781	41	1214.17	0.0346	0.194273
MARCH7_-_1	MARCH7_-_1	24	386	0.0649	13	1213	0.0108	0.16641
MARCH7_-_1	MARCH7_-_1	26	386.25	0.0705	16	1212.75	0.0133	0.188652
MARCH7_-_1	MARCH7_-_1	26	384.67	0.0708	15	1214.33	0.0125	0.176554
MARCH7_-_1	MARCH7_-_1	53.5	383.75	0.1542	30.5	1215.25	0.0255	0.16537
MARCH7_-_1	MARCH7_-_1	61.5	383.83	0.1803	46.5	1215.17	0.0393	0.21797
MARCH7_-_1	LOC1013960	46.5	383	0.1324	34.5	1216	0.0289	0.218278

MARCH7_-_1	MARCH7_-_1	73	386.25	0.2178	60	1212.75	0.0512	0.235078
MARCH7_-_1	MARCH7_-_1	56.5	385	0.1633	50.5	1214	0.0428	0.262094
MARCH7_-_1	MARCH7_-_1	72.5	385.08	0.2168	56.5	1213.92	0.0481	0.221863
MARCH7_-_1	MARCH7_-_1	60.5	385.17	0.1763	53.5	1213.83	0.0454	0.257516
MARCH7_-_1	MARCH7_-_1	51.5	384.25	0.1477	58.5	1214.75	0.0498	0.33717
MARCH7_-_1	MARCH7_-_6	62	383.42	0.1821	53	1215.58	0.0449	0.246568
MARCH7_-_1	MARCH7-_10	62	383.5	0.1821	55	1215.5	0.0467	0.256452
MARCH7_-_1	MARCH7_-_1	67	383.92	0.1986	54	1215.08	0.0458	0.230614
MARCH7_-_1	MARCH7_-_1	65.5	384.75	0.1931	61.5	1214.25	0.0524	0.271362
MARCH7_-_1	MARCH7_-_1	59	384.83	0.1715	50	1214.17	0.0424	0.24723
MARCH7_-_1	MARCH7_-_1	65	384.75	0.1914	53	1214.25	0.045	0.23511
MARCH7_-_1	MARCH7_-_1	68	384.92	0.2014	51	1214.08	0.0432	0.214499
MARCH7_-_1	MARCH7_-_1	70	383	0.2095	63	1216	0.0537	0.256325
MARCH7_-_1	LOC1013592	73.33	384	0.2204	53.67	1215	0.0455	0.206443
MARCH7_-_1	MARCH7_-_1	74.5	383.67	0.2247	53.5	1215.33	0.0454	0.202047
MARCH7_-_1	MARCH7_-_1	71.5	385.33	0.2132	53.5	1213.67	0.0454	0.212946
MARCH7_-_1	MARCH7_-_1	70.17	384.17	0.2093	38.83	1214.83	0.0327	0.156235
MARCH7_-_1	MARCH7_-_1	73.67	382.42	0.2226	63.33	1216.58	0.054	0.242588
MARCH7_-_1	MARCH7_-_1	64.5	383.42	0.1905	70.5	1215.58	0.0604	0.31706
MARCH7_-_1	MARCH7_-_1	86	382.67	0.2671	72	1216.33	0.0617	0.231
MARCH7_-_1	MARCH7_-_C	104.5	383.92	0.3382	77.5	1215.08	0.0667	0.197221
MARCH7_-_1	MARCH7_-_1	107.83	384.67	0.351	71.17	1214.33	0.061	0.173789
MARCH7_-_1	MARCH7_-_1	24.58	386.5	0.0665	45.42	1212.5	0.0384	0.577444
MARCH7_-_1	March7_-_1	80.5	384.75	0.2453	71.5	1214.25	0.0613	0.249898
MARCH7_-_1	XM_0043236	50.5	385.42	0.144	38.5	1213.58	0.0324	0.225
MARCH7_-_1	MARCH7_-_1	12	387.42	0.0316	3	1211.58	0.0025	0.079114
MARCH7_-_1	MARCH7_-_1	38	385.83	0.1056	20	1213.17	0.0167	0.158144
MARCH7_-_1	MARCH7_-_1	67.5	384.92	0.1998	34.5	1214.08	0.029	0.145145
MARCH7_-_1	MARCH7_-_1	70.5	385	0.2099	50.5	1214	0.0428	0.203907
MARCH7_-_1	LOC1013960	60.5	384.17	0.1768	35.5	1214.83	0.0298	0.168552

MARCH7_-_1	MARCH7_-_1	82	387.42	0.2487	61	1211.58	0.0521	0.209489
MARCH7_-_1	MARCH7_-_1	61.5	386.17	0.179	54.5	1212.83	0.0463	0.258659
MARCH7_-_1	MARCH7_-_1	82.5	386.25	0.2514	60.5	1212.75	0.0516	0.205251
MARCH7_-_1	MARCH7_-_1	68	386.33	0.2006	54	1212.67	0.0459	0.228814
MARCH7_-_1	MARCH7_-_1	62.5	385.42	0.1827	57.5	1213.58	0.0489	0.267652
MARCH7_-_1	MARCH7_-_6	72	384.58	0.2154	54	1214.42	0.0458	0.212628
MARCH7_-_1	MARCH7-_10	70	384.67	0.2084	56	1214.33	0.0476	0.228407
MARCH7_-_1	MARCH7_-_1	77	385.08	0.2326	54	1213.92	0.0459	0.197334
MARCH7_-_1	MARCH7_-_1	77.5	385.92	0.2337	60.5	1213.08	0.0516	0.220796
MARCH7_-_1	MARCH7_-_1	74	386	0.2214	51	1213	0.0433	0.195574
MARCH7_-_1	MARCH7_-_1	74	385.92	0.2215	53	1213.08	0.045	0.20316
MARCH7_-_1	MARCH7_-_1	77	386.08	0.2319	52	1212.92	0.0441	0.190168
MARCH7_-_1	MARCH7_-_1	85.5	384.17	0.264	63.5	1214.83	0.0542	0.205303
MARCH7_-_1	LOC1013592	78.83	385.17	0.239	52.17	1213.83	0.0443	0.185356
MARCH7_-_1	MARCH7_-_1	83	384.83	0.2543	58	1214.17	0.0494	0.194259
MARCH7_-_1	MARCH7_-_1	81.5	386.5	0.2476	58.5	1212.5	0.0499	0.201535
MARCH7_-_1	MARCH7_-_1	80.17	385.33	0.2437	40.83	1213.67	0.0344	0.141157
MARCH7_-_1	MARCH7_-_1	83.67	383.58	0.2577	66.33	1215.42	0.0567	0.220023
MARCH7_-_1	MARCH7_-_1	70.33	384.58	0.2096	71.67	1214.42	0.0615	0.293416
MARCH7_-_1	MARCH7_-_1	99.5	383.83	0.3181	74.5	1215.17	0.064	0.201195
MARCH7_-_1	MARCH7_-_C	110.5	385.08	0.3617	78.5	1213.92	0.0676	0.186895
MARCH7_-_1	MARCH7_-_1	118.83	385.83	0.3966	72.17	1213.17	0.062	0.156329
MARCH7_-_1	MARCH7_-_1	39.25	387.67	0.1088	53.75	1211.33	0.0457	0.420037
MARCH7_-_1	March7_-_1	93.5	385.92	0.2926	76.5	1213.08	0.0659	0.225222
MARCH7_-_1	XM_0043236	63.5	386.58	0.1854	39.5	1212.42	0.0333	0.179612
MARCH7_-_1	MARCH7_-_1	38	386.08	0.1055	23	1212.92	0.0192	0.181991
MARCH7_-_1	MARCH7_-_1	67.5	385.17	0.1996	37.5	1213.83	0.0315	0.157816
MARCH7_-_1	MARCH7_-_1	72.5	385.25	0.2167	53.5	1213.75	0.0454	0.209506
MARCH7_-_1	LOC1013960	61.5	384.42	0.1799	38.5	1214.58	0.0324	0.1801
MARCH7_-_1	MARCH7_-_1	83	387.67	0.2521	62	1211.33	0.053	0.210234

MARCH7_-_1	MARCH7_-_1	63.5	386.42	0.1855	57.5	1212.58	0.049	0.264151
MARCH7_-_1	MARCH7_-_1	82.5	386.5	0.2512	61.5	1212.5	0.0525	0.208997
MARCH7_-_1	MARCH7_-_1	68	386.58	0.2005	57	1212.42	0.0486	0.242394
MARCH7_-_1	MARCH7_-_1	62.5	385.67	0.1826	60.5	1213.33	0.0516	0.282585
MARCH7_-_1	MARCH7_-_6	72	384.83	0.2152	56	1214.17	0.0476	0.22119
MARCH7_-_1	MARCH7-_10	70	384.92	0.2083	58	1214.08	0.0494	0.237158
MARCH7_-_1	MARCH7_-_1	75	385.33	0.2253	56	1213.67	0.0476	0.211274
MARCH7_-_1	MARCH7_-_1	75.5	386.17	0.2265	62.5	1212.83	0.0534	0.235762
MARCH7_-_1	MARCH7_-_1	74	386.25	0.2212	54	1212.75	0.0459	0.207505
MARCH7_-_1	MARCH7_-_1	74	386.17	0.2213	56	1212.83	0.0477	0.215545
MARCH7_-_1	MARCH7_-_1	76	386.33	0.2282	55	1212.67	0.0468	0.205083
MARCH7_-_1	MARCH7_-_1	85.5	384.42	0.2638	66.5	1214.58	0.0569	0.215694
MARCH7_-_1	LOC1013592	82.83	385.42	0.2532	53.17	1213.58	0.0451	0.17812
MARCH7_-_1	MARCH7_-_1	87	385.08	0.2688	59	1213.92	0.0502	0.186756
MARCH7_-_1	MARCH7_-_1	83.5	386.75	0.2546	61.5	1212.25	0.0525	0.206206
MARCH7_-_1	MARCH7_-_1	79.17	385.58	0.2399	43.83	1213.42	0.037	0.154231
MARCH7_-_1	MARCH7_-_1	82.67	383.83	0.2539	69.33	1215.17	0.0593	0.233557
MARCH7_-_1	MARCH7_-_1	75.33	384.83	0.2269	72.67	1214.17	0.0624	0.275011
MARCH7_-_1	MARCH7_-_1	100.5	384.08	0.3218	74.5	1214.92	0.064	0.198881
MARCH7_-_1	MARCH7_-_C	112.5	385.33	0.3698	80.5	1213.67	0.0694	0.187669
MARCH7_-_1	MARCH7_-_1	117.83	386.08	0.3918	75.17	1212.92	0.0647	0.165135
MARCH7_-_1	MARCH7_-_1	43.58	387.92	0.1217	56.42	1211.08	0.0481	0.395234
MARCH7_-_1	March7_-_1	95.5	386.17	0.3001	76.5	1212.83	0.0659	0.219593
MARCH7_-_1	XM_0043236	63.5	386.83	0.1853	40.5	1212.17	0.0342	0.184566
MARCH7_-_1	MARCH7_-_1	68.5	383.58	0.204	36.5	1215.42	0.0306	0.15
MARCH7_-_1	MARCH7_-_1	72.33	383.67	0.2171	51.67	1215.33	0.0438	0.20175
MARCH7_-_1	LOC1013960	56.5	382.83	0.1643	39.5	1216.17	0.0332	0.202069
MARCH7_-_1	MARCH7_-_1	81.5	386.08	0.2479	61.5	1212.92	0.0525	0.211779
MARCH7_-_1	MARCH7_-_1	67.33	384.83	0.1992	55.67	1214.17	0.0473	0.23745
MARCH7_-_1	MARCH7_-_1	85.83	384.92	0.2646	63.17	1214.08	0.0539	0.203704

MARCH7_-_1	MARCH7_-_1	72.5	385	0.2168	58.5	1214	0.0498	0.229705
MARCH7_-_1	MARCH7_-_1	61.5	384.08	0.1801	60.5	1214.92	0.0515	0.285952
MARCH7_-_1	MARCH7_-_6	77.83	383.25	0.2368	57.17	1215.75	0.0486	0.205236
MARCH7_-_1	MARCH7-_10	77.83	383.33	0.2368	59.17	1215.67	0.0503	0.212416
MARCH7_-_1	MARCH7_-_1	80.83	383.75	0.2473	59.17	1215.25	0.0503	0.203397
MARCH7_-_1	MARCH7_-_1	81.33	384.58	0.2484	66.67	1214.42	0.057	0.229469
MARCH7_-_1	MARCH7_-_1	76.83	384.67	0.2323	55.17	1214.33	0.0469	0.201894
MARCH7_-_1	MARCH7_-_1	78.83	384.58	0.2394	58.17	1214.42	0.0495	0.206767
MARCH7_-_1	MARCH7_-_1	83	384.75	0.2544	56	1214.25	0.0476	0.187107
MARCH7_-_1	MARCH7_-_1	85	382.83	0.2633	66	1216.17	0.0563	0.213825
MARCH7_-_1	LOC1013592	81.33	383.83	0.249	55.67	1215.17	0.0473	0.18996
MARCH7_-_1	MARCH7_-_1	87.5	383.5	0.272	56.5	1215.5	0.048	0.176471
MARCH7_-_1	MARCH7_-_1	83.5	385.17	0.2559	56.5	1213.83	0.0481	0.187964
MARCH7_-_1	MARCH7_-_1	77.17	384	0.2339	42.83	1215	0.0361	0.154339
MARCH7_-_1	MARCH7_-_1	82.67	382.25	0.2551	62.33	1216.75	0.0531	0.208154
MARCH7_-_1	MARCH7_-_1	78.33	383.25	0.2386	75.67	1215.75	0.065	0.272422
MARCH7_-_1	MARCH7_-_1	104	382.5	0.3377	75	1216.5	0.0643	0.190406
MARCH7_-_1	MARCH7_-_C	118	383.75	0.3957	78	1215.25	0.0671	0.169573
MARCH7_-_1	MARCH7_-_1	119.33	384.5	0.4006	74.67	1214.5	0.0641	0.16001
MARCH7_-_1	MARCH7_-_1	40.58	386.33	0.1132	56.42	1212.67	0.048	0.424028
MARCH7_-_1	March7_-_1	90.5	384.58	0.2824	74.5	1214.42	0.064	0.226629
MARCH7_-_1	XM_0043236	61.5	385.25	0.1795	40.5	1213.75	0.0341	0.189972
MARCH7_-_1	MARCH7_-_1	69	382.75	0.2062	45	1216.25	0.0379	0.183802
MARCH7_-_1	LOC1013960	54.67	381.92	0.1588	34.33	1217.08	0.0288	0.18136
MARCH7_-_1	MARCH7_-_1	76.17	385.17	0.2296	59.83	1213.83	0.051	0.222125
MARCH7_-_1	MARCH7_-_1	62	383.92	0.1819	50	1215.08	0.0423	0.232545
MARCH7_-_1	MARCH7_-_1	79	384	0.2405	55	1215	0.0467	0.194179
MARCH7_-_1	MARCH7_-_1	73	384.08	0.2192	52	1214.92	0.0441	0.201186
MARCH7_-_1	MARCH7_-_1	55	383.17	0.1593	55	1215.83	0.0467	0.293158
MARCH7_-_1	MARCH7_-_6	68.5	382.33	0.2047	53.5	1216.67	0.0453	0.221299

MARCH7_-_1	MARCH7-_10	70.5	382.42	0.2116	55.5	1216.58	0.0471	0.22259
MARCH7_-_1	MARCH7_-_1	71.5	382.83	0.2148	54.5	1216.17	0.0462	0.215084
MARCH7_-_1	MARCH7_-_1	72	383.67	0.216	60	1215.33	0.0511	0.236574
MARCH7_-_1	MARCH7_-_1	67.5	383.75	0.2004	52.5	1215.25	0.0445	0.222056
MARCH7_-_1	MARCH7_-_1	72.5	383.67	0.2177	53.5	1215.33	0.0454	0.208544
MARCH7_-_1	MARCH7_-_1	76.5	383.83	0.2317	49.5	1215.17	0.0419	0.180837
MARCH7_-_1	MARCH7_-_1	76	381.92	0.2312	65	1217.08	0.0554	0.239619
MARCH7_-_1	LOC1013592	71.83	382.92	0.2159	50.17	1216.08	0.0424	0.196387
MARCH7_-_1	MARCH7_-_1	76	382.58	0.2308	55	1216.42	0.0466	0.201906
MARCH7_-_1	MARCH7_-_1	74	384.25	0.2226	56	1214.75	0.0476	0.213836
MARCH7_-_1	MARCH7_-_1	80.5	383.08	0.2466	35.5	1215.92	0.0298	0.120843
MARCH7_-_1	MARCH7_-_1	90.67	381.33	0.286	61.33	1217.67	0.0521	0.182168
MARCH7_-_1	MARCH7_-_1	73.5	382.33	0.2221	68.5	1216.67	0.0585	0.263395
MARCH7_-_1	MARCH7_-_1	93	381.58	0.2947	69	1217.42	0.0589	0.199864
MARCH7_-_1	MARCH7_-_C	106	382.83	0.3455	73	1216.17	0.0626	0.181187
MARCH7_-_1	MARCH7_-_1	111.33	383.58	0.367	71.67	1215.42	0.0614	0.167302
MARCH7_-_1	MARCH7_-_1	68.08	385.42	0.2014	70.92	1213.58	0.0608	0.301887
MARCH7_-_1	March7_-_1	85.5	383.67	0.2644	74.5	1215.33	0.064	0.242057
MARCH7_-_1	XM_0043236	54.5	384.33	0.1572	37.5	1214.67	0.0315	0.200382
MARCH7_-_1	LOC1013960	59.5	382	0.1746	48.5	1217	0.0409	0.23425
MARCH7_-_1	MARCH7_-_1	88.33	385.25	0.2737	73.67	1213.75	0.0633	0.231275
MARCH7_-_1	MARCH7_-_1	25	384	0.0681	12	1215	0.0099	0.145374
MARCH7_-_1	MARCH7_-_1	57	384.08	0.1654	30	1214.92	0.0251	0.151753
MARCH7_-_1	MARCH7_-_1	62	384.17	0.1817	41	1214.83	0.0345	0.189873
MARCH7_-_1	MARCH7_-_1	56	383.25	0.1625	42	1215.75	0.0354	0.217846
MARCH7_-_1	MARCH7_-_6	52.5	382.42	0.1516	32.5	1216.58	0.0272	0.17942
MARCH7_-_1	MARCH7-_10	54.5	382.5	0.158	33.5	1216.5	0.0281	0.177848
MARCH7_-_1	MARCH7_-_1	57.5	382.92	0.1676	32.5	1216.08	0.0272	0.162291
MARCH7_-_1	MARCH7_-_1	55.83	383.75	0.1617	38.17	1215.25	0.0321	0.198516
MARCH7_-_1	MARCH7_-_1	51.5	383.83	0.1478	30.5	1215.17	0.0255	0.17253

MARCH7_-_1	MARCH7_-_1	55.5	383.75	0.1607	33.5	1215.25	0.0281	0.17486
MARCH7_-_1	MARCH7_-_1	59.5	383.92	0.1736	33.5	1215.08	0.0281	0.161866
MARCH7_-_1	MARCH7_-_1	76.5	382	0.233	59.5	1217	0.0506	0.217167
MARCH7_-_1	LOC1013592	69.5	383	0.2078	47.5	1216	0.0401	0.192974
MARCH7_-_1	MARCH7_-_1	77.5	382.67	0.2361	52.5	1216.33	0.0445	0.188479
MARCH7_-_1	MARCH7_-_1	77	384.33	0.2331	52	1214.67	0.0441	0.189189
MARCH7_-_1	MARCH7_-_1	80.67	383.17	0.2471	55.33	1215.83	0.0469	0.189802
MARCH7_-_1	MARCH7_-_1	86	381.42	0.2682	52	1217.58	0.044	0.164057
MARCH7_-_1	MARCH7_-_1	62.5	382.42	0.1843	50.5	1216.58	0.0427	0.231687
MARCH7_-_1	MARCH7_-_1	85	381.67	0.2642	55	1217.33	0.0466	0.176382
MARCH7_-_1	MARCH7_-_C	109.5	382.92	0.3601	65.5	1216.08	0.0559	0.155235
MARCH7_-_1	MARCH7_-_1	101.33	383.67	0.3256	61.67	1215.33	0.0525	0.161241
MARCH7_-_1	MARCH7_-_1	78.08	385.5	0.2361	85.92	1213.5	0.0744	0.315121
MARCH7_-_1	March7_-_1	85.5	383.75	0.2644	63.5	1215.25	0.0542	0.204992
MARCH7_-_1	XM_0043236	64	384.42	0.1883	51	1214.58	0.0432	0.229421
LOC1013960	MARCH7_-_1	70.5	384.42	0.2103	55.5	1214.58	0.0471	0.223966
LOC1013960	MARCH7_-_1	57.5	383.17	0.1674	52.5	1215.83	0.0445	0.26583
LOC1013960	MARCH7_-_1	74.5	383.25	0.225	57.5	1215.75	0.0489	0.217333
LOC1013960	MARCH7_-_1	63	383.33	0.1855	56	1215.67	0.0475	0.256065
LOC1013960	MARCH7_-_1	53	382.42	0.1532	53	1216.58	0.0449	0.293081
LOC1013960	MARCH7_-_6	60	381.58	0.1765	54	1217.42	0.0457	0.258924
LOC1013960	MARCH7-_10	60	381.67	0.1764	56	1217.33	0.0475	0.269274
LOC1013960	MARCH7_-_1	66.5	382.08	0.198	53.5	1216.92	0.0453	0.228788
LOC1013960	MARCH7_-_1	66.67	382.92	0.1981	60.33	1216.08	0.0513	0.25896
LOC1013960	MARCH7_-_1	61.5	383	0.1807	53.5	1216	0.0453	0.250692
LOC1013960	MARCH7_-_1	67.5	382.92	0.2009	54.5	1216.08	0.0462	0.229965
LOC1013960	MARCH7_-_1	70	383.08	0.2094	52	1215.92	0.044	0.210124
LOC1013960	MARCH7_-_1	72	381.17	0.2176	60	1217.83	0.051	0.234375
LOC1013960	LOC1013592	67.33	382.17	0.2008	52.67	1216.83	0.0446	0.222112
LOC1013960	MARCH7_-_1	73	381.83	0.2207	56	1217.17	0.0475	0.215224

LOC1013960	MARCH7_-_1	73	383.5	0.2196	53	1215.5	0.0449	0.204463
LOC1013960	MARCH7_-_1	66.67	382.33	0.1985	37.33	1216.67	0.0313	0.157683
LOC1013960	MARCH7_-_1	77.67	380.58	0.2382	63.33	1218.42	0.0539	0.22628
LOC1013960	MARCH7_-_1	65.5	381.58	0.1949	68.5	1217.42	0.0585	0.300154
LOC1013960	MARCH7_-_1	92	380.83	0.2916	67	1218.17	0.0571	0.195816
LOC1013960	MARCH7_-_C	102.67	382.08	0.3327	74.33	1216.92	0.0637	0.191464
LOC1013960	MARCH7_-_1	107.25	382.83	0.3507	66.75	1216.17	0.057	0.162532
LOC1013960	MARCH7_-_1	66.08	384.67	0.1951	75.92	1214.33	0.0653	0.3347
LOC1013960	March7_-_1	79.67	382.92	0.2437	75.33	1216.08	0.0647	0.26549
LOC1013960	XM_0043236	47	383.58	0.1338	34	1215.42	0.0285	0.213004
MARCH7_-_1	MARCH7_-_1	76.33	386.42	0.2293	76.67	1212.58	0.0661	0.288269
MARCH7_-_1	MARCH7_-_1	92.33	386.5	0.2876	82.67	1212.5	0.0715	0.248609
MARCH7_-_1	MARCH7_-_1	87	386.58	0.2676	79	1212.42	0.0682	0.254858
MARCH7_-_1	MARCH7_-_1	79	385.67	0.2392	82	1213.33	0.0708	0.295987
MARCH7_-_1	MARCH7_-_6	80.83	384.83	0.2464	74.17	1214.17	0.0637	0.258523
MARCH7_-_1	MARCH7-_10	80.83	384.92	0.2464	76.17	1214.08	0.0655	0.265828
MARCH7_-_1	MARCH7_-_1	89.33	385.33	0.2773	75.67	1213.67	0.0651	0.234764
MARCH7_-_1	MARCH7_-_1	89.5	386.17	0.2772	77.5	1212.83	0.0668	0.240981
MARCH7_-_1	MARCH7_-_1	86.33	386.25	0.2654	74.67	1212.75	0.0642	0.241899
MARCH7_-_1	MARCH7_-_1	90.33	386.17	0.2804	75.67	1212.83	0.0651	0.232168
MARCH7_-_1	MARCH7_-_1	91.5	386.33	0.2846	70.5	1212.67	0.0605	0.212579
MARCH7_-_1	MARCH7_-_1	95	384.42	0.2998	88	1214.58	0.0762	0.254169
MARCH7_-_1	LOC1013592	92.33	385.42	0.2886	73.67	1213.58	0.0633	0.219335
MARCH7_-_1	MARCH7_-_1	97	385.08	0.3069	79	1213.92	0.0681	0.221896
MARCH7_-_1	MARCH7_-_1	87	386.75	0.2674	77	1212.25	0.0664	0.248317
MARCH7_-_1	MARCH7_-_1	85.5	385.58	0.2629	58.5	1213.42	0.0498	0.189426
MARCH7_-_1	MARCH7_-_1	96.17	383.83	0.3049	81.83	1215.17	0.0706	0.231551
MARCH7_-_1	MARCH7_-_1	83.33	384.83	0.2555	89.67	1214.17	0.0777	0.30411
MARCH7_-_1	MARCH7_-_1	108.17	384.08	0.3531	95.83	1214.92	0.0833	0.235911
MARCH7_-_1	MARCH7_-_C	118.67	385.33	0.3965	91.33	1213.67	0.0793	0.2

MARCH7_-_1	MARCH7_-_1	125.83	386.08	0.4276	96.17	1212.92	0.0838	0.195978
MARCH7_-_1	MARCH7_-_1	89.58	387.92	0.276	97.42	1211.08	0.0851	0.308333
MARCH7_-_1	March7_-_1	99.17	386.17	0.3144	96.83	1212.83	0.0844	0.268448
MARCH7_-_1	XM_0043236	45.5	386.83	0.1279	41.5	1212.17	0.035	0.273651
MARCH7_-_1	MARCH7_-_1	55	385.25	0.1584	27	1213.75	0.0226	0.142677
MARCH7_-_1	MARCH7_-_1	57.5	385.33	0.1664	45.5	1213.67	0.0385	0.23137
MARCH7_-_1	MARCH7_-_1	55	384.42	0.1587	46	1214.58	0.0389	0.245117
MARCH7_-_1	MARCH7_-_6	50	383.58	0.1432	35	1215.42	0.0294	0.205307
MARCH7_-_1	MARCH7-_10	52	383.67	0.1495	36	1215.33	0.0302	0.202007
MARCH7_-_1	MARCH7_-_1	56	384.08	0.1621	35	1214.92	0.0294	0.18137
MARCH7_-_1	MARCH7_-_1	54.5	384.92	0.1569	42.5	1214.08	0.0358	0.228171
MARCH7_-_1	MARCH7_-_1	52	385	0.1489	33	1214	0.0277	0.186031
MARCH7_-_1	MARCH7_-_1	52	384.92	0.149	36	1214.08	0.0303	0.203356
MARCH7_-_1	MARCH7_-_1	58	385.08	0.1681	36	1213.92	0.0303	0.18025
MARCH7_-_1	MARCH7_-_1	65.5	383.17	0.194	59.5	1215.83	0.0506	0.260825
MARCH7_-_1	LOC1013592	62	384.17	0.1817	49	1214.83	0.0415	0.228398
MARCH7_-_1	MARCH7_-_1	75	383.83	0.2264	54	1215.17	0.0458	0.202297
MARCH7_-_1	MARCH7_-_1	70	385.5	0.2079	54	1213.5	0.0459	0.220779
MARCH7_-_1	MARCH7_-_1	76.67	384.33	0.2319	58.33	1214.67	0.0496	0.213885
MARCH7_-_1	MARCH7_-_1	76.5	382.58	0.2326	56.5	1216.42	0.0479	0.205933
MARCH7_-_1	MARCH7_-_1	59.5	383.58	0.1738	52.5	1215.42	0.0445	0.256041
MARCH7_-_1	MARCH7_-_1	81	382.83	0.2486	59	1216.17	0.0502	0.201931
MARCH7_-_1	MARCH7_-_C	101.5	384.08	0.3258	63.5	1214.92	0.0542	0.16636
MARCH7_-_1	MARCH7_-_1	103.33	384.83	0.3324	63.67	1214.17	0.0544	0.163658
MARCH7_-_1	MARCH7_-_1	74.08	386.67	0.2212	91.92	1212.33	0.0799	0.361212
MARCH7_-_1	March7_-_1	85	384.92	0.2616	68	1214.08	0.0582	0.222477
MARCH7_-_1	XM_0043236	60	385.58	0.1744	55	1213.42	0.0468	0.268349
MARCH7_-_1	MARCH7_-_1	75.5	385.42	0.227	52.5	1213.58	0.0446	0.196476
MARCH7_-_1	MARCH7_-_1	69.5	384.5	0.2068	53.5	1214.5	0.0454	0.219536
MARCH7_-_1	MARCH7_-_6	67	383.67	0.1988	47	1215.33	0.0397	0.199698

MARCH7_-_1	MARCH7-_10	69	383.75	0.2056	49	1215.25	0.0414	0.201362
MARCH7_-_1	MARCH7_-_1	74	384.17	0.2226	44	1214.83	0.0371	0.166667
MARCH7_-_1	MARCH7_-_1	71.5	385	0.2134	55.5	1214	0.0472	0.221181
MARCH7_-_1	MARCH7_-_1	68	385.08	0.2013	45	1213.92	0.038	0.188773
MARCH7_-_1	MARCH7_-_1	72	385	0.2151	46	1214	0.0389	0.180846
MARCH7_-_1	MARCH7_-_1	74	385.17	0.222	46	1213.83	0.0389	0.175225
MARCH7_-_1	MARCH7_-_1	87	383.25	0.2704	66	1215.75	0.0564	0.20858
MARCH7_-_1	LOC1013592	83.5	384.25	0.2566	58.5	1214.75	0.0498	0.194076
MARCH7_-_1	MARCH7_-_1	91.5	383.92	0.2868	63.5	1215.08	0.0542	0.188982
MARCH7_-_1	MARCH7_-_1	87	385.58	0.2684	62	1213.42	0.0529	0.197094
MARCH7_-_1	MARCH7_-_1	96.67	384.42	0.3063	67.33	1214.58	0.0576	0.188051
MARCH7_-_1	MARCH7_-_1	96.67	382.67	0.308	66.33	1216.33	0.0566	0.183766
MARCH7_-_1	MARCH7_-_1	77	383.67	0.2336	63	1215.33	0.0537	0.22988
MARCH7_-_1	MARCH7_-_1	96.5	382.92	0.3071	65.5	1216.08	0.0559	0.182025
MARCH7_-_1	MARCH7_-_C	114.5	384.17	0.3799	73.5	1214.83	0.0631	0.166096
MARCH7_-_1	MARCH7_-_1	122.33	384.92	0.4134	75.67	1214.08	0.0651	0.157475
MARCH7_-_1	MARCH7_-_1	90.08	386.75	0.2789	97.92	1212.25	0.0855	0.306561
MARCH7_-_1	March7_-_1	95.25	385	0.3002	80.75	1214	0.0697	0.232179
MARCH7_-_1	XM_0043236	72	385.67	0.2147	60	1213.33	0.0512	0.238472
MARCH7_-_1	MARCH7_-_1	59	384.58	0.1716	53	1214.42	0.045	0.262238
MARCH7_-_1	MARCH7_-_6	59.5	383.75	0.1737	47.5	1215.25	0.0401	0.230858
MARCH7_-_1	MARCH7-_10	59.5	383.83	0.1737	49.5	1215.17	0.0419	0.24122
MARCH7_-_1	MARCH7_-_1	64.5	384.25	0.19	49.5	1214.75	0.0419	0.220526
MARCH7_-_1	MARCH7_-_1	65	385.08	0.1912	52	1213.92	0.0441	0.230649
MARCH7_-_1	MARCH7_-_1	60.5	385.17	0.1763	48.5	1213.83	0.0411	0.233125
MARCH7_-_1	MARCH7_-_1	64.5	385.08	0.1896	48.5	1213.92	0.0411	0.216772
MARCH7_-_1	MARCH7_-_1	69.5	385.25	0.2064	45.5	1213.75	0.0385	0.186531
MARCH7_-_1	MARCH7_-_1	78.5	383.33	0.2392	66.5	1215.67	0.0568	0.237458
MARCH7_-_1	LOC1013592	74.83	384.33	0.2254	60.17	1214.67	0.0512	0.227152
MARCH7_-_1	MARCH7_-_1	83	384	0.255	63	1215	0.0537	0.210588

MARCH7_-_1	MARCH7_-_1	78	385.67	0.2357	64	1213.33	0.0547	0.232075
MARCH7_-_1	MARCH7_-_1	84.17	384.5	0.2588	62.83	1214.5	0.0536	0.20711
MARCH7_-_1	MARCH7_-_1	84.5	382.75	0.2615	65.5	1216.25	0.0559	0.213767
MARCH7_-_1	MARCH7_-_1	60.5	383.75	0.177	66.5	1215.25	0.0568	0.320904
MARCH7_-_1	MARCH7_-_1	90	383	0.2819	61	1216	0.0519	0.184108
MARCH7_-_1	MARCH7_-_C	108.5	384.25	0.3543	81.5	1214.75	0.0703	0.198419
MARCH7_-_1	MARCH7_-_1	108.83	385	0.3548	74.17	1214	0.0637	0.179538
MARCH7_-_1	MARCH7_-_1	79.08	386.83	0.2387	93.92	1212.17	0.0818	0.34269
MARCH7_-_1	March7_-_1	85	385.08	0.2614	73	1213.92	0.0627	0.239862
MARCH7_-_1	XM_0043236	66	385.75	0.1942	59	1213.25	0.0503	0.259011
MARCH7_-_1	MARCH7_-_6	53	382.83	0.153	46	1216.17	0.0388	0.253595
MARCH7_-_1	MARCH7-_10	53	382.92	0.153	48	1216.08	0.0405	0.264706
MARCH7_-_1	MARCH7_-_1	60	383.33	0.1756	45	1215.67	0.038	0.216401
MARCH7_-_1	MARCH7_-_1	61.5	384.17	0.1801	51.5	1214.83	0.0436	0.242088
MARCH7_-_1	MARCH7_-_1	54	384.25	0.1556	46	1214.75	0.0389	0.25
MARCH7_-_1	MARCH7_-_1	56	384.17	0.1621	49	1214.83	0.0415	0.256015
MARCH7_-_1	MARCH7_-_1	65	384.33	0.1917	47	1214.67	0.0397	0.207094
MARCH7_-_1	MARCH7_-_1	74	382.42	0.2238	64	1216.58	0.0545	0.243521
MARCH7_-_1	LOC1013592	74.83	383.42	0.2261	54.17	1215.58	0.0459	0.203008
MARCH7_-_1	MARCH7_-_1	77	383.08	0.234	61	1215.92	0.0519	0.221795
MARCH7_-_1	MARCH7_-_1	79.5	384.75	0.2417	59.5	1214.25	0.0507	0.209764
MARCH7_-_1	MARCH7_-_1	77.67	383.58	0.236	62.33	1215.42	0.0531	0.225
MARCH7_-_1	MARCH7_-_1	89.67	381.83	0.2817	68.33	1217.17	0.0584	0.207313
MARCH7_-_1	MARCH7_-_1	67	382.83	0.1993	64	1216.17	0.0546	0.273959
MARCH7_-_1	MARCH7_-_1	82.5	382.08	0.2546	65.5	1216.92	0.0559	0.21956
MARCH7_-_1	MARCH7_-_C	107	383.33	0.3491	79	1215.67	0.068	0.194787
MARCH7_-_1	MARCH7_-_1	105.83	384.08	0.3434	70.17	1214.92	0.0601	0.175015
MARCH7_-_1	MARCH7_-_1	69.08	385.92	0.2045	99.92	1213.08	0.0872	0.426406
MARCH7_-_1	March7_-_1	77.5	384.17	0.235	73.5	1214.83	0.0631	0.268511
MARCH7_-_1	XM_0043236	60.5	384.83	0.1764	58.5	1214.17	0.0498	0.282313

MARCH7_-_6	MARCH7-_10	4	382.08	0.0105	3	1216.92	0.0025	0.238095
MARCH7_-_6	MARCH7_-_1	11	382.5	0.0293	7	1216.5	0.0058	0.197952
MARCH7_-_6	MARCH7_-_1	14.5	383.33	0.0388	15.5	1215.67	0.0129	0.332474
MARCH7_-_6	MARCH7_-_1	23	383.42	0.0625	20	1215.58	0.0166	0.2656
MARCH7_-_6	MARCH7_-_1	27	383.33	0.074	22	1215.67	0.0183	0.247297
MARCH7_-_6	MARCH7_-_1	32	383.5	0.0885	23	1215.5	0.0192	0.216949
MARCH7_-_6	MARCH7_-_1	81.5	381.58	0.2514	63.5	1217.42	0.0541	0.215195
MARCH7_-_6	LOC1013592	77.83	382.58	0.2373	52.17	1216.42	0.0442	0.186262
MARCH7_-_6	MARCH7_-_1	81	382.25	0.249	60	1216.75	0.051	0.204819
MARCH7_-_6	MARCH7_-_1	80	383.92	0.2441	60	1215.08	0.0511	0.20934
MARCH7_-_6	MARCH7_-_1	83	382.75	0.256	58	1216.25	0.0493	0.192578
MARCH7_-_6	MARCH7_-_1	91.67	381	0.2901	63.33	1218	0.0539	0.185798
MARCH7_-_6	MARCH7_-_1	54	382	0.1566	51	1217	0.0431	0.275223
MARCH7_-_6	MARCH7_-_1	84.5	381.25	0.2627	60.5	1217.75	0.0514	0.19566
MARCH7_-_6	MARCH7_-_C	109.5	382.5	0.3606	72.5	1216.5	0.0621	0.172213
MARCH7_-_6	MARCH7_-_1	111.33	383.25	0.3674	69.67	1215.75	0.0596	0.162221
MARCH7_-_6	MARCH7_-_1	81.08	385.08	0.2472	92.92	1213.92	0.0807	0.326456
MARCH7_-_6	March7_-_1	83.5	383.33	0.2573	65.5	1215.67	0.0559	0.217256
MARCH7_-_6	XM_0043236	61.5	384	0.1802	53.5	1215	0.0454	0.251942
MARCH7-_10	MARCH7_-_1	13	382.58	0.0348	8	1216.42	0.0066	0.189655
MARCH7-_10	MARCH7_-_1	16.5	383.42	0.0443	16.5	1215.58	0.0137	0.309255
MARCH7-_10	MARCH7_-_1	25	383.5	0.0682	21	1215.5	0.0175	0.256598
MARCH7-_10	MARCH7_-_1	29	383.42	0.0797	23	1215.58	0.0192	0.240903
MARCH7-_10	MARCH7_-_1	34	383.58	0.0943	24	1215.42	0.02	0.212089
MARCH7-_10	MARCH7_-_1	84.5	381.67	0.2624	65.5	1217.33	0.0558	0.212652
MARCH7-_10	LOC1013592	76.83	382.67	0.2337	54.17	1216.33	0.0459	0.196406
MARCH7-_10	MARCH7_-_1	80	382.33	0.2453	62	1216.67	0.0528	0.215247
MARCH7-_10	MARCH7_-_1	81	384	0.2477	62	1215	0.0528	0.213161
MARCH7-_10	MARCH7_-_1	83	382.83	0.2559	60	1216.17	0.051	0.199297
MARCH7-_10	MARCH7_-_1	92.67	381.08	0.2939	65.33	1217.92	0.0557	0.18952

MARCH7-_10	MARCH7_-_1	54	382.08	0.1566	52	1216.92	0.044	0.280971
MARCH7-_10	MARCH7_-_1	86.5	381.33	0.2701	62.5	1217.67	0.0532	0.196964
MARCH7-_10	MARCH7_-_C	110.5	382.58	0.3647	75.5	1216.42	0.0648	0.17768
MARCH7-_10	MARCH7_-_1	110.33	383.33	0.3631	70.67	1215.67	0.0605	0.166621
MARCH7-_10	MARCH7_-_1	81.08	385.17	0.2471	94.92	1213.83	0.0826	0.334278
MARCH7-_10	March7_-_1	84.5	383.42	0.2609	66.5	1215.58	0.0568	0.217708
MARCH7-_10	XM_0043236	61.5	384.08	0.1801	55.5	1214.92	0.0471	0.261521
MARCH7_-_1	MARCH7_-_1	15.5	383.83	0.0415	16.5	1215.17	0.0137	0.33012
MARCH7_-_1	MARCH7_-_1	28	383.92	0.0767	18	1215.08	0.015	0.195567
MARCH7_-_1	MARCH7_-_1	30	383.83	0.0825	20	1215.17	0.0166	0.201212
MARCH7_-_1	MARCH7_-_1	35	384	0.0972	21	1215	0.0175	0.180041
MARCH7_-_1	MARCH7_-_1	87	382.08	0.2714	62	1216.92	0.0528	0.194547
MARCH7_-_1	LOC1013592	80.83	383.08	0.2478	53.17	1215.92	0.0451	0.182002
MARCH7_-_1	MARCH7_-_1	85	382.75	0.2633	61	1216.25	0.0519	0.197114
MARCH7_-_1	MARCH7_-_1	85	384.42	0.262	61	1214.58	0.052	0.198473
MARCH7_-_1	MARCH7_-_1	86	383.25	0.2666	61	1215.75	0.0519	0.194674
MARCH7_-_1	MARCH7_-_1	93.67	381.5	0.2974	65.33	1217.5	0.0557	0.18729
MARCH7_-_1	MARCH7_-_1	57	382.5	0.1661	47	1216.5	0.0397	0.239013
MARCH7_-_1	MARCH7_-_1	88.5	381.75	0.2773	61.5	1217.25	0.0523	0.188604
MARCH7_-_1	MARCH7_-_C	116.5	383	0.3901	72.5	1216	0.0621	0.15919
MARCH7_-_1	MARCH7_-_1	113.33	383.75	0.3754	73.67	1215.25	0.0632	0.168354
MARCH7_-_1	MARCH7_-_1	86.08	385.58	0.265	92.92	1213.42	0.0808	0.304906
MARCH7_-_1	March7_-_1	89.25	383.83	0.2783	66.75	1215.17	0.057	0.204815
MARCH7_-_1	XM_0043236	67	384.5	0.1983	54	1214.5	0.0458	0.230963
MARCH7_-_1	MARCH7_-_1	29.5	384.75	0.0809	25.5	1214.25	0.0213	0.263288
MARCH7_-_1	MARCH7_-_1	31.5	384.67	0.0867	27.5	1214.33	0.023	0.265283
MARCH7_-_1	MARCH7_-_1	37.5	384.83	0.1044	27.5	1214.17	0.023	0.220307
MARCH7_-_1	MARCH7_-_1	85.5	382.92	0.2651	67.5	1216.08	0.0577	0.217654
MARCH7_-_1	LOC1013592	84.33	383.92	0.2599	58.67	1215.08	0.0499	0.191997
MARCH7_-_1	MARCH7_-_1	84.5	383.58	0.2608	65.5	1215.42	0.0559	0.21434

MARCH7_-_1	MARCH7_-_1	85.5	385.25	0.2631	65.5	1213.75	0.056	0.212847
MARCH7_-_1	MARCH7_-_1	89.5	384.08	0.2791	69.5	1214.92	0.0595	0.213185
MARCH7_-_1	MARCH7_-_1	92.67	382.33	0.2927	66.33	1216.67	0.0566	0.193372
MARCH7_-_1	MARCH7_-_1	59.5	383.33	0.1739	56.5	1215.67	0.048	0.276021
MARCH7_-_1	MARCH7_-_1	90.5	382.58	0.2842	66.5	1216.42	0.0568	0.199859
MARCH7_-_1	MARCH7_-_C	117	383.83	0.3912	76	1215.17	0.0653	0.166922
MARCH7_-_1	MARCH7_-_1	118.33	384.58	0.3961	73.67	1214.42	0.0633	0.159808
MARCH7_-_1	MARCH7_-_1	84.58	386.42	0.2588	99.42	1212.58	0.0868	0.335394
MARCH7_-_1	March7_-_1	90	384.67	0.2804	70	1214.33	0.06	0.21398
MARCH7_-_1	XM_0043236	68.5	385.33	0.2029	59.5	1213.67	0.0507	0.249877
MARCH7_-_1	MARCH7_-_1	18	384.75	0.0483	13	1214.25	0.0108	0.223602
MARCH7_-_1	MARCH7_-_1	21	384.92	0.0566	15	1214.08	0.0125	0.220848
MARCH7_-_1	MARCH7_-_1	82	383	0.2521	59	1216	0.0502	0.199127
MARCH7_-_1	LOC1013592	78.83	384	0.2399	54.17	1215	0.046	0.191747
MARCH7_-_1	MARCH7_-_1	83	383.67	0.2552	62	1215.33	0.0528	0.206897
MARCH7_-_1	MARCH7_-_1	75	385.33	0.2253	61	1213.67	0.052	0.230803
MARCH7_-_1	MARCH7_-_1	78.67	384.17	0.2392	59.33	1214.83	0.0505	0.21112
MARCH7_-_1	MARCH7_-_1	91.67	382.42	0.2888	65.33	1216.58	0.0557	0.192867
MARCH7_-_1	MARCH7_-_1	58	383.42	0.1689	44	1215.58	0.0371	0.219657
MARCH7_-_1	MARCH7_-_1	82.5	382.67	0.2542	62.5	1216.33	0.0532	0.209284
MARCH7_-_1	MARCH7_-_C	110	383.92	0.361	76	1215.08	0.0653	0.180886
MARCH7_-_1	MARCH7_-_1	108.33	384.67	0.3531	74.67	1214.33	0.0642	0.181818
MARCH7_-_1	MARCH7_-_1	80.08	386.5	0.2425	90.92	1212.5	0.079	0.325773
MARCH7_-_1	March7_-_1	78.5	384.75	0.2381	69.5	1214.25	0.0595	0.249895
MARCH7_-_1	XM_0043236	64	385.42	0.1877	55	1213.58	0.0467	0.248801
MARCH7_-_1	MARCH7_-_1	23	384.83	0.0623	14	1214.17	0.0116	0.186196
MARCH7_-_1	MARCH7_-_1	87	382.92	0.2707	59	1216.08	0.0502	0.185445
MARCH7_-_1	LOC1013592	80.83	383.92	0.2471	57.17	1215.08	0.0486	0.196682
MARCH7_-_1	MARCH7_-_1	86	383.58	0.2664	63	1215.42	0.0537	0.201577
MARCH7_-_1	MARCH7_-_1	77.5	385.25	0.2342	63.5	1213.75	0.0542	0.231426

MARCH7_-_1	MARCH7_-_1	81.17	384.08	0.2482	62.83	1214.92	0.0536	0.215955
MARCH7_-_1	MARCH7_-_1	90.67	382.33	0.2851	68.33	1216.67	0.0584	0.20484
MARCH7_-_1	MARCH7_-_1	64	383.33	0.1889	49	1215.67	0.0414	0.219164
MARCH7_-_1	MARCH7_-_1	90.5	382.58	0.2842	67.5	1216.42	0.0577	0.203026
MARCH7_-_1	MARCH7_-_C	114	383.83	0.3781	79	1215.17	0.068	0.179847
MARCH7_-_1	MARCH7_-_1	111.33	384.58	0.3658	76.67	1214.42	0.0659	0.180153
MARCH7_-_1	MARCH7_-_1	82.25	386.42	0.2504	92.75	1212.58	0.0807	0.322284
MARCH7_-_1	March7_-_1	82.5	384.67	0.2526	71.5	1214.33	0.0613	0.242676
MARCH7_-_1	XM_0043236	72	385.33	0.2149	56	1213.67	0.0476	0.221498
MARCH7_-_1	MARCH7_-_1	91	383.08	0.2856	57	1215.92	0.0484	0.169468
MARCH7_-_1	LOC1013592	80.83	384.08	0.247	55.17	1214.92	0.0468	0.189474
MARCH7_-_1	MARCH7_-_1	91	383.75	0.285	61	1215.25	0.052	0.182456
MARCH7_-_1	MARCH7_-_1	81.5	385.42	0.2484	60.5	1213.58	0.0516	0.207729
MARCH7_-_1	MARCH7_-_1	84.5	384.25	0.2603	58.5	1214.75	0.0498	0.191318
MARCH7_-_1	MARCH7_-_1	93	382.5	0.2939	66	1216.5	0.0563	0.191562
MARCH7_-_1	MARCH7_-_1	65	383.5	0.1921	51	1215.5	0.0432	0.224883
MARCH7_-_1	MARCH7_-_1	89	382.75	0.2783	65	1216.25	0.0554	0.199066
MARCH7_-_1	MARCH7_-_C	119	384	0.3998	72	1215	0.0617	0.154327
MARCH7_-_1	MARCH7_-_1	112.83	384.75	0.372	73.17	1214.25	0.0628	0.168817
MARCH7_-_1	MARCH7_-_1	87.08	386.58	0.2679	89.92	1212.42	0.0781	0.291527
MARCH7_-_1	March7_-_1	87.5	384.83	0.2709	67.5	1214.17	0.0578	0.213363
MARCH7_-_1	XM_0043236	71	385.5	0.2113	52	1213.5	0.0441	0.208708
MARCH7_-_1	LOC1013592	75	382.17	0.2275	56	1216.83	0.0475	0.208791
MARCH7_-_1	MARCH7_-_1	87.5	381.83	0.2735	58.5	1217.17	0.0497	0.181718
MARCH7_-_1	MARCH7_-_1	81.5	383.5	0.2499	59.5	1215.5	0.0506	0.202481
MARCH7_-_1	MARCH7_-_1	93.67	382.33	0.2966	68.33	1216.67	0.0584	0.196898
MARCH7_-_1	MARCH7_-_1	81.17	380.58	0.2509	66.83	1218.42	0.057	0.227182
MARCH7_-_1	MARCH7_-_1	83.67	381.58	0.2594	79.33	1217.42	0.0682	0.262914
MARCH7_-_1	MARCH7_-_1	103	380.83	0.3354	82	1218.17	0.0705	0.210197
MARCH7_-_1	MARCH7_-_C	116	382.08	0.3891	74	1216.92	0.0634	0.16294

MARCH7_-_1	MARCH7_-_1	112.33	382.83	0.3722	70.67	1216.17	0.0605	0.162547
MARCH7_-_1	MARCH7_-_1	91.58	384.67	0.2864	101.42	1214.33	0.0885	0.309008
MARCH7_-_1	March7_-_1	98	382.92	0.313	79	1216.08	0.0679	0.216933
MARCH7_-_1	XM_0043236	77	383.58	0.2336	62	1215.42	0.0528	0.226027
LOC1013592	MARCH7_-_1	56.5	382.83	0.1643	27.5	1216.17	0.023	0.139988
LOC1013592	MARCH7_-_1	65	384.5	0.1916	43	1214.5	0.0363	0.189457
LOC1013592	MARCH7_-_1	89	383.33	0.2778	55	1215.67	0.0467	0.168107
LOC1013592	MARCH7_-_1	70	381.58	0.2104	45	1217.42	0.0379	0.180133
LOC1013592	MARCH7_-_1	76.33	382.58	0.232	69.67	1216.42	0.0596	0.256897
LOC1013592	MARCH7_-_1	101.17	381.83	0.3269	69.83	1217.17	0.0597	0.182625
LOC1013592	MARCH7_-_C	88	383.08	0.2743	59	1215.92	0.0502	0.183011
LOC1013592	MARCH7_-_1	94.33	383.83	0.2978	56.67	1215.17	0.0481	0.161518
LOC1013592	MARCH7_-_1	90.92	385.67	0.283	95.08	1213.33	0.0828	0.29258
LOC1013592	March7_-_1	98.33	383.92	0.3134	77.67	1215.08	0.0668	0.213146
LOC1013592	XM_0043236	74.33	384.58	0.2235	47.67	1214.42	0.0403	0.180313
MARCH7_-_1	MARCH7_-_1	70	384.17	0.2087	42	1214.83	0.0354	0.169621
MARCH7_-_1	MARCH7_-_1	94.67	383	0.2999	61.33	1216	0.0522	0.174058
MARCH7_-_1	MARCH7_-_1	77.67	381.25	0.2377	51.33	1217.75	0.0434	0.182583
MARCH7_-_1	MARCH7_-_1	84.5	382.25	0.2619	76.5	1216.75	0.0657	0.250859
MARCH7_-_1	MARCH7_-_1	103.33	381.5	0.3361	74.67	1217.5	0.064	0.19042
MARCH7_-_1	MARCH7_-_C	96.5	382.75	0.3073	60.5	1216.25	0.0515	0.167589
MARCH7_-_1	MARCH7_-_1	105.83	383.5	0.3441	50.17	1215.5	0.0425	0.123511
MARCH7_-_1	MARCH7_-_1	89.08	385.33	0.2764	94.92	1213.67	0.0826	0.298842
MARCH7_-_1	March7_-_1	100	383.58	0.3203	82	1215.42	0.0707	0.220731
MARCH7_-_1	XM_0043236	80	384.25	0.2439	51	1214.75	0.0432	0.177122
MARCH7_-_1	MARCH7_-_1	88	384.67	0.2729	59	1214.33	0.0502	0.18395
MARCH7_-_1	MARCH7_-_1	62	382.92	0.1824	49	1216.08	0.0414	0.226974
MARCH7_-_1	MARCH7_-_1	84	383.92	0.2587	77	1215.08	0.0662	0.255895
MARCH7_-_1	MARCH7_-_1	107.33	383.17	0.3507	74.67	1215.83	0.0641	0.182777
MARCH7_-_1	MARCH7_-_C	85.5	384.42	0.2638	57.5	1214.58	0.0489	0.185368

MARCH7_-_1	MARCH7_-_1	94.83	385.17	0.2984	47.17	1213.83	0.0399	0.133713
MARCH7_-_1	MARCH7_-_1	86.08	387	0.2639	94.92	1212	0.0827	0.313376
MARCH7_-_1	March7_-_1	94.75	385.25	0.298	81.25	1213.75	0.0701	0.235235
MARCH7_-_1	XM_0043236	73	385.92	0.218	57	1213.08	0.0485	0.222477
MARCH7_-_1	MARCH7_-_1	94.33	381.75	0.2998	65.67	1217.25	0.056	0.186791
MARCH7_-_1	MARCH7_-_1	89.17	382.75	0.279	76.83	1216.25	0.066	0.236559
MARCH7_-_1	MARCH7_-_1	102.5	382	0.3321	78.5	1217	0.0674	0.202951
MARCH7_-_1	MARCH7_-_C	119.67	383.25	0.4038	74.33	1215.75	0.0638	0.157999
MARCH7_-_1	MARCH7_-_1	121	384	0.4087	73	1215	0.0626	0.153169
MARCH7_-_1	MARCH7_-_1	87.75	385.83	0.271	80.25	1213.17	0.0693	0.25572
MARCH7_-_1	March7_-_1	98.67	384.08	0.3145	78.33	1214.92	0.0674	0.214308
MARCH7_-_1	XM_0043236	72.5	384.75	0.217	37.5	1214.25	0.0315	0.145161
MARCH7_-_1	MARCH7_-_1	82.17	381	0.2543	81.83	1218	0.0704	0.276838
MARCH7_-_1	MARCH7_-_1	109.83	380.25	0.3648	78.17	1218.75	0.067	0.183662
MARCH7_-_1	MARCH7_-_C	101.17	381.5	0.3272	62.83	1217.5	0.0535	0.163509
MARCH7_-_1	MARCH7_-_1	96.5	382.25	0.3078	55.5	1216.75	0.0471	0.153021
MARCH7_-_1	MARCH7_-_1	90.25	384.08	0.2819	102.75	1214.92	0.0897	0.318198
MARCH7_-_1	March7_-_1	107.17	382.33	0.351	79.83	1216.67	0.0687	0.195726
MARCH7_-_1	XM_0043236	85.17	383	0.2638	63.83	1216	0.0544	0.206217
MARCH7_-_1	MARCH7_-_1	90.5	381.25	0.2854	77.5	1217.75	0.0665	0.233006

MARCH7_-_1	MARCH7_-_C	108.5	382.5	0.3564	88.5	1216.5	0.0765	0.214646
MARCH7_-_1	MARCH7_-_1	107.33	383.25	0.3506	88.67	1215.75	0.0767	0.218768
MARCH7_-_1	MARCH7_-_1	85.92	385.08	0.2648	112.08	1213.92	0.0985	0.371979
MARCH7_-_1	March7_-_1	91.58	383.33	0.2877	88.42	1215.67	0.0765	0.265902
MARCH7_-_1	XM_0043236	65	384	0.1918	72	1215	0.0617	0.321689
MARCH7_-_1	MARCH7_-_C	123.83	381.75	0.4249	84.17	1217.25	0.0725	0.170628
MARCH7_-_1	MARCH7_-_1	139.83	382.5	0.5012	81.17	1216.5	0.0699	0.139465
MARCH7_-_1	MARCH7_-_1	105.58	384.33	0.3421	111.42	1214.67	0.0978	0.285881
MARCH7_-_1	March7_-_1	109.5	382.58	0.3605	82.5	1216.42	0.0711	0.197226
MARCH7_-_1	XM_0043236	93.5	383.25	0.2951	69.5	1215.75	0.0595	0.201627
MARCH7_-_C	MARCH7_-_1	122.33	383.75	0.4151	69.67	1215.25	0.0596	0.14358
MARCH7_-_C	MARCH7_-_1	122.08	385.58	0.4113	116.92	1213.42	0.1031	0.250669
MARCH7_-_C	March7_-_1	126.08	383.83	0.4322	93.92	1215.17	0.0816	0.188801
MARCH7_-_C	XM_0043236	104.5	384.5	0.3375	76.5	1214.5	0.0658	0.194963
MARCH7_-_1	MARCH7_-_1	127.17	386.33	0.4334	111.83	1212.67	0.0984	0.227042
MARCH7_-_1	March7_-_1	120.58	384.58	0.406	82.42	1214.42	0.0711	0.175123
MARCH7_-_1	XM_0043236	106.33	385.25	0.3442	72.67	1213.75	0.0624	0.18129
MARCH7_-_1	March7_-_1	100.25	386.42	0.3184	111.75	1212.58	0.0983	0.308731
MARCH7_-_1	XM_0043236	71.08	387.08	0.2106	79.92	1211.92	0.069	0.327635
March7_-_1	XM_0043236	79	385.33	0.2395	74	1213.67	0.0636	0.265553

7.23 TOPALI OUTPUT

7.23.1Comparison of the domestic dog with carnivores

7.23.1.1Model selection 1 (CP1)

Table 88: Model selection 1 (CP1) Topali output for comparison of the domestic dog with carnivores

Model	K	-ℓ	AIC_1	AIC_2	BIC
JC	0	1340.74	2711.48 (0)	2711.71 (0)	2796.40 (0)
F81	3	1330.14	2696.29 (0)	2696.61 (0)	2798.19 (0)
JC+G	1	1339.86	2711.72 (0)	2711.98 (0)	2802.29 (0)
JC+I	1	1340.22	2712.45 (0)	2712.70 (0)	2803.02 (0)
F81+G	4	1329.24	2696.47 (0)	2696.83 (0)	2804.03 (0)
K80	1	1340.74	2713.48 (0)	2713.74 (0)	2804.06 (0)
F81+I	4	1329.61	2697.21 (0)	2697.58 (0)	2804.77 (0)
HKY	4	1330.14	2698.29 (0)	2698.65 (0)	2805.85 (0)
JC+I+G	2	1339.5	2713.00 (0)	2713.30 (0)	2809.24 (0)
K80+G	2	1339.86	2713.72 (0)	2714.01 (0)	2809.96 (0)
K80+I	2	1340.22	2714.45 (0)	2714.74 (0)	2810.68 (0)
F81+I+G	5	1328.86	2697.71 (0)	2698.11 (0)	2810.93 (0)
HKY+G	5	1329.24	2698.47 (0)	2698.87 (0)	2811.69 (0)
TrNef	2	1340.74	2715.48 (0)	2715.77 (0)	2811.72 (0)
K81	2	1340.74	2715.48 (0)	2715.77 (0)	2811.72 (0)
HKY+I	5	1329.61	2699.21 (0)	2699.61 (0)	2812.44 (0)
TrN	5	1330.14	2700.29 (0)	2700.69 (0)	2813.51 (0)
K81uf	5	1330.14	2700.29 (0)	2700.69 (0)	2813.51 (0)
K80+I+G	3	1339.5	2715.00 (0)	2715.33 (0)	2816.90 (0)
TrNef+G	3	1339.86	2715.72 (0)	2716.04 (0)	2817.62 (0)
K81+G	3	1339.86	2715.72 (0)	2716.04 (0)	2817.62 (0)
TrNef+I	3	1340.22	2716.45 (0)	2716.77 (0)	2818.34 (0)
K81+I	3	1340.22	2716.45 (0)	2716.77 (0)	2818.34 (0)
HKY+I+G	6	1328.86	2699.71 (0)	2700.15 (0)	2818.59 (0)
TrN+G	6	1329.24	2700.47 (0)	2700.91 (0)	2819.35 (0)

K81uf+G	6	1329.24	2700.47 (0)	2700.91 (0)	2819.35 (0)
TIMef	3	1340.74	2717.48 (0)	2717.81 (0)	2819.38 (0)
TrN+I	6	1329.61	2701.21 (0)	2701.65 (0)	2820.10 (0)
K81uf+I	6	1329.61	2701.21 (0)	2701.65 (0)	2820.10 (0)
TIM	6	1330.14	2702.29 (0)	2702.73 (0)	2821.17 (0)
TrNef+I+G	4	1339.5	2717.00 (0)	2717.37 (0)	2824.56 (0)
K81+I+G	4	1339.5	2717.00 (0)	2717.37 (0)	2824.56 (0)
TIMef+G	4	1339.86	2717.72 (0)	2718.08 (0)	2825.28 (0)
TIMef+I	4	1340.22	2718.45 (0)	2718.81 (0)	2826.01 (0)
TrN+I+G	7	1328.86	2701.71 (0)	2702.19 (0)	2826.25 (0)
K81uf+I+G	7	1328.86	2701.71 (0)	2702.19 (0)	2826.25 (0)
TIM+G	7	1329.24	2702.47 (0)	2702.95 (0)	2827.01 (0)
TVMef	4	1340.74	2719.48 (0)	2719.84 (0)	2827.04 (0)
TIM+I	7	1329.61	2703.21 (0)	2703.70 (0)	2827.76 (0)
TVM	7	1330.14	2704.29 (0)	2704.77 (0)	2828.83 (0)
TIMef+I+G	5	1339.5	2719.00 (0)	2719.40 (0)	2832.23 (0)
TVMef+G	5	1339.86	2719.72 (0)	2720.12 (0)	2832.94 (0)
TVMef+I	5	1340.22	2720.45 (0)	2720.84 (0)	2833.67 (0)
TIM+I+G	8	1328.86	2703.71 (0)	2704.24 (0)	2833.91 (0)
TVM+G	8	1329.24	2704.47 (0)	2705.00 (0)	2834.68 (0)
SYM	5	1340.74	2721.48 (0)	2721.88 (0)	2834.70 (0)
TVM+I	8	1329.61	2705.21 (0)	2705.74 (0)	2835.42 (0)
GTR	8	1330.14	2706.29 (0)	2706.81 (0)	2836.49 (0)
TVMef+I+G	6	1339.5	2721.00 (0)	2721.44 (0)	2839.89 (0)
SYM+G	6	1339.86	2721.72 (0)	2722.16 (0)	2840.60 (0)
SYM+I	6	1340.22	2722.45 (0)	2722.89 (0)	2841.33 (0)
TVM+I+G	9	1328.86	2705.71 (0)	2706.28 (0)	2841.58 (0)
GTR+G	9	1329.24	2706.47 (0)	2707.04 (0)	2842.34 (0)
GTR+I	9	1329.61	2707.21 (0)	2707.79 (0)	2843.08 (0)
SYM+I+G	7	1339.5	2723.00 (0)	2723.49 (0)	2847.55 (0)
GTR+I+G	10	1328.86	2707.71 (0)	2708.33 (0)	2849.24 (0)

7.23.1.2 Model selection 1 (CP2)

Table 89: Model selection 1 (CP2) Topali output for comparison of the domestic dog with carnivores

Model	K	-ℓ	AIC_1	AIC_2	BIC
HKY	4	1113.46	2264.91 (0)	2265.28 (0)	2372.47 (0)
HKY+G	5	1111.32	2262.64 (0)	2263.04 (0)	2375.86 (0)
F81	3	1119.37	2274.75 (0)	2275.07 (0)	2376.65 (0)
TrN	5	1112.29	2264.57 (0)	2264.97 (0)	2377.79 (0)
K80	1	1128.22	2288.45 (0)	2288.71 (0)	2379.03 (0)
HKY+I	5	1112.91	2265.82 (0)	2266.22 (0)	2379.04 (0)
F81+G	4	1117.25	2272.50 (0)	2272.86 (0)	2380.06 (0)
TrN+G	6	1110.12	2262.23 (0)	2262.67 (0)	2381.11 (0)
K81uf	5	1114.18	2268.36 (0)	2268.76 (0)	2381.58 (0)
K80+G	2	1126.08	2286.16 (0)	2286.45 (0)	2382.40 (0)
JC	0	1133.82	2297.65 (0)	2297.88 (0)	2382.57 (0)
HKY+I+G	6	1111.03	2264.06 (0)	2264.50 (0)	2382.95 (0)
F81+I	4	1118.83	2275.66 (0)	2276.02 (0)	2383.22 (0)
TrNef	2	1126.9	2287.80 (0)	2288.09 (0)	2384.04 (0)
TrN+I	6	1111.74	2265.47 (0)	2265.91 (0)	2384.36 (0)
TIM	6	1112.01	2266.02 (0)	2266.46 (0)	2384.90 (0)
K81uf+G	6	1112.02	2266.04 (0)	2266.48 (0)	2384.92 (0)
K80+I	2	1127.68	2289.36 (0)	2289.65 (0)	2385.59 (0)
JC+G	1	1131.69	2295.39 (0)	2295.65 (0)	2385.97 (0)
F81+I+G	5	1116.97	2273.93 (0)	2274.33 (0)	2387.15 (0)
TrNef+G	3	1124.74	2285.48 (0)	2285.80 (0)	2387.38 (0)
K81	2	1128.94	2291.88 (0)	2292.17 (0)	2388.12 (0)
K81uf+I	6	1113.63	2269.26 (0)	2269.69 (0)	2388.14 (0)
TrN+I+G	7	1109.82	2263.64 (0)	2264.12 (0)	2388.18 (0)
TIM+G	7	1109.84	2263.67 (0)	2264.16 (0)	2388.22 (0)
JC+I	1	1133.28	2298.56 (0)	2298.82 (0)	2389.13 (0)

K80+I+G	3	1125.79	2287.58 (0)	2287.90 (0)	2389.48 (0)
TrNef+I	3	1126.35	2288.70 (0)	2289.03 (0)	2390.60 (0)
TIMef	3	1126.56	2289.12 (0)	2289.44 (0)	2391.02 (0)
K81+G	3	1126.76	2289.53 (0)	2289.85 (0)	2391.42 (0)
TIM+I	7	1111.46	2266.92 (0)	2267.40 (0)	2391.46 (0)
K81uf+I+G	7	1111.72	2267.44 (0)	2267.92 (0)	2391.98 (0)
JC+I+G	2	1131.41	2296.81 (0)	2297.10 (0)	2393.05 (0)
TVM	7	1112.27	2268.54 (0)	2269.02 (0)	2393.09 (0)
TIMef+G	4	1124.39	2286.78 (0)	2287.14 (0)	2394.34 (0)
TrNef+I+G	4	1124.44	2286.89 (0)	2287.25 (0)	2394.45 (0)
K81+I	3	1128.38	2292.77 (0)	2293.09 (0)	2394.67 (0)
TIM+I+G	8	1109.54	2265.07 (0)	2265.60 (0)	2395.28 (0)
TVM+G	8	1110.13	2266.26 (0)	2266.78 (0)	2396.46 (0)
TIMef+I	4	1126.01	2290.02 (0)	2290.38 (0)	2397.58 (0)
GTR	8	1111.13	2268.25 (0)	2268.78 (0)	2398.45 (0)
K81+I+G	4	1126.45	2290.90 (0)	2291.26 (0)	2398.46 (0)
TVMef	4	1126.82	2291.64 (0)	2292.00 (0)	2399.20 (0)
TVM+I	8	1111.72	2269.44 (0)	2269.97 (0)	2399.65 (0)
TIMef+I+G	5	1124.09	2288.19 (0)	2288.59 (0)	2401.41 (0)
GTR+G	9	1108.97	2265.94 (0)	2266.51 (0)	2401.80 (0)
TVMef+G	5	1124.67	2289.34 (0)	2289.74 (0)	2402.56 (0)
TVM+I+G	9	1109.83	2267.66 (0)	2268.24 (0)	2403.53 (0)
SYM	5	1125.96	2291.91 (0)	2292.31 (0)	2405.13 (0)
GTR+I	9	1110.68	2269.35 (0)	2269.92 (0)	2405.22 (0)
TVMef+I	5	1126.26	2292.53 (0)	2292.93 (0)	2405.75 (0)
SYM+G	6	1123.45	2288.90 (0)	2289.34 (0)	2407.78 (0)
GTR+I+G	10	1108.69	2267.38 (0)	2268.00 (0)	2408.91 (0)
TVMef+I+G	6	1124.37	2290.73 (0)	2291.17 (0)	2409.62 (0)
SYM+I	6	1124.9	2291.79 (0)	2292.23 (0)	2410.67 (0)

SYM+I+G	7	1123.25	2290.50 (0)	2290.98 (0)	2415.04 (0)

7.23.1.3 Model selection 1 (CP3)

Table 90: Model selection 1 (CP3) Topali output for comparison of the domestic dog with carnivores

Model	K	-ℓ	AIC_1	AIC_2	BIC
HKY+G	5	1757.76	3555.51 (0)	3555.91 (0)	3668.73 (0)
K81uf+G	6	1756.92	3555.83 (0)	3556.27 (0)	3674.71 (0)
HKY+I+G	6	1757.76	3557.52 (0)	3557.96 (0)	3676.40 (0)
K81uf+I+G	7	1756.87	3557.74 (0)	3558.22 (0)	3682.28 (0)
TIM+G	7	1756.99	3557.98 (0)	3558.46 (0)	3682.52 (0)
K81uf+I	6	1761.83	3565.66 (0)	3566.10 (0)	3684.54 (0)
TVM+G	8	1756.43	3558.86 (0)	3559.39 (0)	3689.07 (0)
K81uf	5	1768.33	3576.66 (0)	3577.06 (0)	3689.88 (0)
TIM+I+G	8	1756.9	3559.81 (0)	3560.33 (0)	3690.01 (0)
TIM+I	7	1761.77	3567.53 (0)	3568.02 (0)	3692.08 (0)
TVM+I+G	9	1756.35	3560.70 (0)	3561.27 (0)	3696.56 (0)
GTR+G	9	1756.5	3561.00 (0)	3561.57 (0)	3696.86 (0)
TIM	6	1768.14	3578.28 (0)	3578.72 (0)	3697.16 (0)
TVM+I	8	1761.37	3568.74 (0)	3569.26 (0)	3698.94 (0)
TVM	7	1767.84	3579.68 (0)	3580.17 (0)	3704.23 (0)
GTR+I+G	10	1756.38	3562.77 (0)	3563.38 (0)	3704.29 (0)
GTR+I	9	1761.3	3570.60 (0)	3571.17 (0)	3706.47 (0)
GTR	8	1767.65	3581.29 (0)	3581.82 (0)	3711.50 (0)
TrN+G	6	1787.31	3616.62 (2)	3617.06 (2)	3735.50 (2)
TrN+I+G	7	1786.48	3616.95 (2)	3617.43 (2)	3741.49 (2)
TrN+I	6	1798.03	3638.05 (0)	3638.49 (0)	3756.93 (0)
HKY	4	1805.83	3649.67 (0)	3650.03 (0)	3757.23 (0)
TrN	5	1805.59	3651.18 (0)	3651.57 (0)	3764.40 (0)
F81+G	4	1822.98	3683.95 (0)	3684.31 (0)	3791.51 (0)
HKY+I	5	1820.98	3681.95 (2)	3682.35 (2)	3795.17 (2)

F81+I+G	5	1823.02	3686.04 (0)	3686.44 (0)	3799.26 (0)
F81+I	4	1827.2	3692.39 (0)	3692.75 (0)	3799.95 (0)
F81	3	1833.72	3703.44 (0)	3703.77 (0)	3805.34 (0)
K81+G	3	1845.63	3727.25 (0)	3727.58 (0)	3829.15 (0)
TIMef+G	4	1845.59	3729.17 (0)	3729.53 (0)	3836.73 (0)
K81+I+G	4	1845.63	3729.26 (0)	3729.63 (0)	3836.83 (0)
K81+I	3	1852.34	3740.67 (0)	3741.00 (0)	3842.57 (0)
TVMef+G	5	1844.74	3729.48 (0)	3729.88 (0)	3842.70 (0)
TIMef+I+G	5	1845.59	3731.18 (0)	3731.58 (0)	3844.41 (0)
K81	2	1859.82	3753.65 (0)	3753.94 (0)	3849.89 (0)
TIMef+I	4	1852.27	3742.53 (0)	3742.90 (0)	3850.09 (0)
SYM+G	6	1844.7	3731.41 (0)	3731.85 (0)	3850.29 (0)
TVMef+I+G	6	1844.73	3731.46 (0)	3731.90 (0)	3850.35 (0)
TVMef+I	5	1851.33	3742.65 (0)	3743.05 (0)	3855.87 (0)
TIMef	3	1859.74	3755.48 (0)	3755.81 (0)	3857.38 (0)
SYM+I+G	7	1844.69	3733.39 (0)	3733.87 (0)	3857.93 (0)
SYM+I	6	1851.26	3744.52 (0)	3744.95 (0)	3863.40 (0)
TVMef	4	1859.21	3756.41 (0)	3756.77 (0)	3863.97 (0)
SYM	5	1859.12	3758.25 (0)	3758.64 (0)	3871.47 (0)
K80+G	2	1878.42	3790.84 (2)	3791.13 (2)	3887.08 (2)
K80+I+G	3	1877.99	3791.99 (2)	3792.31 (2)	3893.89 (2)
TrNef+G	3	1879.28	3794.56 (2)	3794.89 (2)	3896.46 (2)
TrNef+I+G	4	1878.84	3795.68 (2)	3796.05 (2)	3903.24 (2)
K80+I	2	1886.79	3807.57 (2)	3807.86 (2)	3903.81 (2)
JC+G	1	1895.06	3822.12 (0)	3822.37 (0)	3912.69 (0)
TrNef+I	3	1887.71	3811.42 (2)	3811.74 (2)	3913.32 (2)
JC+I+G	2	1895.13	3824.26 (0)	3824.55 (0)	3920.49 (0)
K80	1	1900.31	3832.61 (2)	3832.87 (2)	3923.19 (2)
JC+I	1	1900.44	3832.88 (0)	3833.13 (0)	3923.45 (0)

JC	0	1907.62	3845.24 (0)	3845.47 (0)	3930.16 (0)
TrNef	2	1900.47	3834.94 (2)	3835.23 (2)	3931.18 (2)

7.23.2 Comparison of the domestic dog with placental mammals

7.23.2.1 Model Selection 1 (CP1)

Table 91: Model selection 1 (CP1) Topali output for comparison of the domestic dog with placental mammals

Model	K	-ℓ	AIC_1	AIC_2	BIC
TVMef+G	5	4010.58	8173.16 (2)	8178.85 (2)	8603.72 (0)
K81+G	3	4018.75	8185.49 (2)	8190.89 (2)	8604.73 (0)
SYM+G	6	4009.83	8173.65 (2)	8179.50 (2)	8609.88 (0)
K80+G	2	4025.18	8196.35 (0)	8201.60 (0)	8609.92 (2)
TVM+G	8	4002.24	8162.47 (0)	8168.63 (0)	8610.03 (2)
TIMef+G	4	4018.27	8186.54 (2)	8192.08 (2)	8611.43 (0)
TVMef+I+G	6	4012.76	8179.51 (2)	8185.36 (2)	8615.74 (0)
GTR+G	9	4001.36	8162.72 (0)	8169.03 (0)	8615.94 (2)
TrNef+G	3	4024.65	8197.31 (2)	8202.70 (2)	8616.54 (4)
K81+I+G	4	4020.86	8191.73 (2)	8197.27 (2)	8616.62 (0)
K80+I+G	3	4027.31	8202.63 (2)	8208.02 (2)	8621.86 (4)
SYM+I+G	7	4012.14	8180.28 (2)	8186.28 (2)	8622.17 (0)
TVM+I+G	9	4004.71	8169.41 (0)	8175.73 (0)	8622.64 (2)
TIMef+I+G	5	4020.51	8193.01 (2)	8198.70 (2)	8623.57 (0)
K81uf+G	6	4018.18	8190.36 (0)	8196.20 (0)	8626.59 (2)
GTR+I+G	10	4003.69	8169.38 (0)	8175.86 (0)	8628.27 (2)
TrNef+I+G	4	4026.9	8203.81 (2)	8209.35 (2)	8628.71 (4)
TIM+G	7	4017.26	8190.52 (0)	8196.52 (0)	8632.41 (2)
HKY+G	5	4026.7	8205.40 (2)	8211.09 (2)	8635.96 (4)
K81uf+I+G	7	4020.53	8197.07 (0)	8203.07 (0)	8638.96 (2)
TrN+G	6	4025.76	8205.52 (2)	8211.36 (2)	8641.75 (4)
TIM+I+G	8	4019.47	8196.94 (0)	8203.10 (0)	8644.50 (2)
HKY+I+G	6	4029.1	8212.21 (2)	8218.05 (2)	8648.44 (4)
TrN+I+G	7	4028.03	8212.06 (2)	8218.06 (2)	8653.95 (4)
TVMef+I	5	4053.45	8258.91 (4)	8264.60 (4)	8689.47 (2)

K81+I	3	4063.15	8274.30 (4)	8279.70 (4)	8693.54 (2)
TVM+I	8	4044.5	8246.99 (2)	8253.15 (2)	8694.55 (0)
SYM+I	6	4052.21	8258.43 (4)	8264.27 (4)	8694.66 (2)
TIMef+I	4	4061.96	8273.92 (4)	8279.46 (4)	8698.81 (2)
K80+I	2	4069.84	8285.68 (4)	8290.93 (4)	8699.25 (2)
GTR+I	9	4043.99	8247.97 (2)	8254.29 (2)	8701.19 (0)
TrNef+I	3	4068.65	8285.30 (4)	8290.70 (4)	8704.54 (2)
K81uf+I	6	4061.57	8277.13 (2)	8282.98 (2)	8713.36 (0)
TIM+I	7	4061.13	8278.25 (2)	8284.25 (2)	8720.14 (0)
HKY+I	5	4070.34	8292.67 (2)	8298.36 (2)	8723.23 (0)
TrN+I	6	4069.88	8293.77 (2)	8299.61 (2)	8729.99 (0)
JC+G	1	4110.39	8364.78 (2)	8369.88 (2)	8772.68 (0)
JC+I+G	2	4112.67	8371.35 (2)	8376.60 (2)	8784.91 (0)
F81+G	4	4112.39	8374.78 (2)	8380.32 (2)	8799.67 (0)
F81+I+G	5	4114.73	8381.46 (2)	8387.15 (2)	8812.02 (0)
JC+I	1	4154.32	8452.64 (4)	8457.75 (4)	8860.54 (2)
TVMef	4	4147.29	8444.58 (4)	8450.12 (4)	8869.47 (2)
K81	2	4156.58	8459.16 (4)	8464.40 (4)	8872.72 (2)
SYM	5	4146.41	8444.82 (4)	8450.51 (4)	8875.38 (2)
K80	1	4163.25	8470.50 (4)	8475.61 (4)	8878.40 (2)
TIMef	3	4155.68	8459.37 (4)	8464.76 (4)	8878.60 (2)
TrNef	2	4162.36	8470.72 (4)	8475.97 (4)	8884.29 (2)
F81+I	4	4155.42	8460.84 (6)	8466.38 (6)	8885.74 (4)
TVM	7	4145.89	8447.77 (2)	8453.77 (2)	8889.67 (0)
GTR	8	4145.35	8448.70 (2)	8454.86 (2)	8896.26 (0)
K81uf	5	4159.93	8471.86 (2)	8477.56 (2)	8902.42 (0)
TIM	6	4159.41	8472.82 (2)	8478.67 (2)	8909.05 (0)
HKY	4	4167.81	8485.62 (2)	8491.16 (2)	8910.52 (0)
TrN	5	4167.28	8486.57 (2)	8492.26 (2)	8917.13 (0)
JC	0	4244.9	8631.81 (10)	8636.77 (10)	9034.04 (8)
F81	3	4248.46	8644.92 (6)	8650.31 (6)	9064.15 (4)

7.23.2.2 Model selection 1 (CP2)

Table 92: Model selection 1 (CP2) Topali output for comparison of the domestic dog with placental mammals

Model	K	-ℓ	AIC_1	AIC_2	BIC
TrNef+G	3	2678.18	5504.35 (2)	5509.75 (2)	5923.59 (0)
TIMef+G	4	2676.27	5502.54 (2)	5508.09 (2)	5927.44 (0)
TrN+G	6	2668.93	5491.87 (0)	5497.71 (0)	5928.09 (2)
TIM+G	7	2666.18	5488.36 (2)	5494.36 (2)	5930.25 (0)
SYM+G	6	2672.28	5498.56 (2)	5504.41 (2)	5934.79 (0)
TrNef+I+G	4	2681.25	5512.51 (0)	5518.05 (0)	5937.41 (2)
GTR+G	9	2663.22	5486.44 (0)	5492.75 (0)	5939.66 (2)
TIMef+I+G	5	2679.35	5510.71 (2)	5516.40 (2)	5941.27 (0)
TrN+I+G	7	2672.05	5500.10 (0)	5506.10 (0)	5941.99 (2)
TIM+I+G	8	2669.3	5496.60 (0)	5502.76 (0)	5944.16 (2)
SYM+I+G	7	2675.12	5506.24 (2)	5512.24 (2)	5948.14 (0)
K81uf+G	6	2679.17	5512.35 (12)	5518.19 (12)	5948.58 (12)
GTR+I+G	10	2666.37	5494.73 (0)	5501.21 (0)	5953.62 (2)
TVMef+G	5	2687.52	5527.04 (6)	5532.73 (6)	5957.60 (6)
TVM+G	8	2676.72	5511.43 (6)	5517.59 (6)	5958.99 (6)
K81uf+I+G	7	2681.84	5519.68 (2)	5525.68 (2)	5961.57 (0)
K81+I+G	4	2694.9	5539.80 (6)	5545.35 (6)	5964.70 (6)
TVM+I+G	9	2679.33	5518.66 (2)	5524.98 (2)	5971.89 (0)
TVMef+I+G	6	2690.88	5535.76 (6)	5541.60 (6)	5971.99 (6)
TrNef+I	3	2703.76	5555.53 (6)	5560.92 (6)	5974.76 (6)
TIMef+I	4	2701.85	5553.69 (6)	5559.24 (6)	5978.59 (6)
K81+G	3	2705.92	5559.83 (14)	5565.22 (14)	5979.06 (14)
TrN+I	6	2696.25	5546.51 (6)	5552.35 (6)	5982.73 (6)
SYM+I	6	2697.06	5548.12 (6)	5553.96 (6)	5984.34 (6)
TIM+I	7	2693.47	5542.94 (6)	5548.94 (6)	5984.84 (6)
GTR+I	9	2688.73	5537.47 (6)	5543.78 (6)	5990.69 (6)
HKY+I	5	2706.96	5565.92 (12)	5571.62 (12)	5996.49 (12)

K81uf+I	6	2705.73	5565.46 (6)	5571.30 (6)	6001.68 (6)
K81+I	3	2718.22	5584.44 (6)	5589.83 (6)	6003.67 (6)
TVMef+I	5	2713.44	5578.87 (8)	5584.56 (8)	6009.43 (8)
TVM+I	8	2702.57	5563.15 (6)	5569.30 (6)	6010.70 (6)
HKY+G	5	2714.16	5580.32 (16)	5586.02 (16)	6010.88 (16)
K80+G	2	2726.95	5599.91 (12)	5605.16 (12)	6013.48 (12)
HKY+I+G	6	2728.87	5611.75 (8)	5617.59 (8)	6047.98 (6)
K80+I+G	3	2742.84	5633.68 (12)	5639.08 (12)	6052.91 (12)
TrNef	2	2762.3	5670.59 (6)	5675.84 (6)	6084.16 (6)
JC+G	1	2767.96	5679.91 (8)	5685.02 (8)	6087.81 (10)
TIMef	3	2760.34	5668.69 (6)	5674.08 (6)	6087.92 (6)
SYM	5	2754.57	5661.15 (12)	5666.84 (12)	6091.71 (12)
TrN	5	2755.25	5662.50 (6)	5668.20 (6)	6093.07 (6)
JC+I+G	2	2767.95	5681.90 (8)	5687.15 (8)	6095.47 (10)
TIM	6	2753.52	5661.05 (8)	5666.89 (8)	6097.27 (8)
F81+G	4	2761.5	5673.00 (8)	5678.54 (8)	6097.89 (10)
GTR	8	2749	5655.99 (12)	5662.15 (12)	6103.55 (12)
F81+I+G	5	2761.5	5675.00 (8)	5680.69 (8)	6105.56 (10)
K80+I	2	2773.38	5692.76 (6)	5698.01 (6)	6106.33 (6)
TVMef	4	2771.35	5692.70 (12)	5698.24 (12)	6117.60 (12)
K81uf	5	2769.23	5690.45 (12)	5696.14 (12)	6121.01 (12)
TVM	7	2764.58	5685.15 (12)	5691.15 (12)	6127.04 (12)
JC+I	1	2793.59	5731.18 (8)	5736.28 (8)	6139.08 (10)
K81	2	2791.41	5728.81 (14)	5734.06 (14)	6142.38 (14)
F81+I	4	2787.14	5724.29 (8)	5729.83 (8)	6149.18 (10)
K80	1	2826.93	5797.85 (6)	5802.95 (6)	6205.75 (6)
HKY	4	2818.66	5787.31 (6)	5792.86 (6)	6212.21 (6)
JC	0	2852.14	5846.27 (8)	5851.23 (8)	6248.51 (10)
F81	3	2847.26	5842.53 (8)	5847.92 (8)	6261.76 (10)

7.23.2.3 Model selection 1 (CP3)

Table 93: Model selection 1 (CP3) Topali output for comparison of the domestic dog with placental mammals

Model	K	-ℓ	AIC_1	AIC_2	BIC
K81uf+G	6	7186.17	14526.34 (0)	14532.18 (0)	14962.57 (0)
K81uf+I+G	7	7185.51	14527.02 (0)	14533.02 (0)	14968.91 (0)
TVM+G	8	7182.06	14522.12 (0)	14528.27 (0)	14969.68 (0)
TIM+G	7	7186.07	14528.15 (0)	14534.15 (0)	14970.04 (0)
TrN+G	6	7191.78	14537.56 (0)	14543.40 (0)	14973.78 (0)
TVM+I+G	9	7181.47	14522.95 (0)	14529.26 (0)	14976.17 (0)
TIM+I+G	8	7185.31	14528.61 (0)	14534.77 (0)	14976.17 (0)
GTR+G	9	7182.02	14524.04 (0)	14530.36 (0)	14977.27 (0)
GTR+I+G	10	7181.36	14524.73 (0)	14531.20 (0)	14983.61 (0)
K81uf+I	6	7226.43	14606.87 (10)	14612.71 (10)	15043.09 (10)
TIM+I	7	7225.7	14607.41 (10)	14613.41 (10)	15049.30 (10)
TVM+I	8	7222.24	14602.49 (10)	14608.64 (10)	15050.04 (10)
GTR+I	9	7221.49	14602.98 (10)	14609.30 (10)	15056.20 (10)
K81uf	5	7285.32	14722.63 (10)	14728.32 (10)	15153.19 (10)
TVM	7	7278.35	14712.69 (10)	14718.69 (10)	15154.59 (10)
GTR	8	7278.27	14714.54 (10)	14720.69 (10)	15162.09 (10)
TVMef+G	5	7316.57	14785.15 (10)	14790.84 (10)	15215.71 (10)
SYM+G	6	7316.5	14787.00 (10)	14792.85 (10)	15223.23 (10)
SYM+I+G	7	7317.59	14791.18 (10)	14797.18 (10)	15233.07 (10)
TIMef+G	4	7334.9	14819.81 (10)	14825.35 (10)	15244.71 (10)
TIMef+I+G	5	7335.61	14823.22 (10)	14828.91 (10)	15253.78 (10)
TVMef+I	5	7365	14882.00 (10)	14887.69 (10)	15312.56 (10)
SYM+I	6	7364.57	14883.13 (10)	14888.98 (10)	15319.36 (10)
TIMef+I	4	7377.42	14904.84 (10)	14910.38 (10)	15329.73 (10)
TVMef	4	7426.45	15002.90 (10)	15008.45 (10)	15427.80 (10)
SYM	5	7425.87	15003.75 (10)	15009.44 (10)	15434.31 (10)
TVMef+I+G	6	7425.21	15004.41 (10)	15010.26 (10)	15440.64 (10)
TIMef	3	7438.83	15025.67 (10)	15031.06 (10)	15444.90 (10)
TIM	6	7450.35	15054.70 (10)	15060.55 (10)	15490.93 (10)

K81+G	3	7504.66	15157.32 (10)	15162.71 (10)	15576.55 (10)
K81+I+G	4	7515.73	15181.46 (10)	15187.00 (10)	15606.35 (10)
K81+I	3	7552.09	15252.18 (10)	15257.57 (10)	15671.41 (10)
K81	2	7564.01	15274.03 (10)	15279.27 (10)	15687.59 (10)
K80+G	2	7635.53	15417.06 (4)	15422.31 (4)	15830.63 (4)
K80+I+G	3	7636.43	15420.85 (4)	15426.24 (4)	15840.08 (4)
TrNef+G	3	7681.94	15511.88 (4)	15517.28 (4)	15931.12 (4)
K80+I	2	7686.59	15519.18 (10)	15524.43 (10)	15932.75 (10)
TrNef+I+G	4	7682.79	15515.58 (4)	15521.12 (4)	15940.48 (4)
HKY+G	5	7682.83	15517.67 (2)	15523.36 (2)	15948.23 (2)
HKY+I+G	6	7682.84	15519.68 (2)	15525.53 (2)	15955.91 (2)
TrN+I+G	7	7712.99	15581.99 (2)	15587.99 (2)	16023.88 (2)
TrNef+I	3	7733.12	15614.24 (10)	15619.63 (10)	16033.47 (10)
K80	1	7754.52	15653.04 (10)	15658.14 (10)	16060.94 (10)
HKY+I	5	7745.02	15642.04 (14)	15647.73 (14)	16072.60 (14)
JC+G	1	7789.71	15723.43 (10)	15728.53 (10)	16131.33 (10)
TrN+I	6	7774.91	15703.82 (14)	15709.67 (14)	16140.05 (14)
JC+I+G	2	7790.49	15726.98 (10)	15732.23 (10)	16140.55 (10)
TrNef	2	7801.5	15748.99 (10)	15754.24 (10)	16162.56 (10)
F81+G	4	7795.87	15741.73 (4)	15747.28 (4)	16166.63 (4)
F81+I+G	5	7796.08	15744.16 (4)	15749.85 (4)	16174.72 (4)
JC+I	1	7830.37	15804.73 (12)	15809.83 (12)	16212.63 (12)
F81+I	4	7827	15804.01 (10)	15809.55 (10)	16228.90 (10)
HKY	4	7832.37	15814.74 (10)	15820.28 (10)	16239.64 (10)
TrN	5	7858.8	15869.60 (10)	15875.29 (10)	16300.16 (10)
JC	0	7888.87	15919.74 (12)	15924.70 (12)	16321.97 (12)
F81	3	7881.84	15911.67 (10)	15917.07 (10)	16330.91 (10)

7.24 EXON 1 SEQUENCES FOR COMPARISON OF DOG AND CAT

7.24.1 Dog (CanFam 3.1) Nucleotide Sequence

GAAATCTCAAGAATGGAGTCTAAACCTTCAAGGATTCCAAGAAGAATTTCTGTTCAACCTTCTAGC
TCTGTAAGTGCTAGAATGATGTCTGGAAGCAGAGGAAATAGTTTAAATGATACCTATCACTCAAGA
GATTCTTCATTTAGACTGGATTCTGAATATCAG

7.24.2 Cat (Felis catus 6.2) Nucleotide Sequence

GAAACTTTAAGAATGGAGTCTAAACCTTCAAGGATCCCAAGAAGAATTTCTGTTCAACCATCCAGC
TCTTTAAGTGCTAGGATGATGTCTGGAAGCAGAGGAAATAGTTTAAATGATACCTATCACTCAAGA
GACTCTTCATTTAGACTGGATTCTGAATATCAG

Figure 7-16: Comparison of nucleotide sequence of exon 1 for the domestic dog and domestic cat. Yellow highlighting represents differences in nucleotide between species, green highlighted text represents untranslated region

7.24.3 Dog (CanFam 3.1) Amino acid

MESKPSRIPRRISVQPSSSVSARMMSGSRGNSLNDTYHSRDSSFRLDSEYQ

7.24.4 Cat (Felis catus 6.2) Amino acid

MESKPSRIPRRISVQPSSSLSARMMSGSRGNSLNDTYHSRDSSFRLDSEYQ

Figure 7-17: Comparison of amino acid sequence of exon 1 for the domestic dog and domestic cat. Yellow highlighting represents differences in nucleotide between species, green highlighted text represents untranslated region

7.25 COMPARISON OF HUMAN AND GORILLA EXON 1

Comparison of the Human to the Gorilla was conducted when the reliability of QUARKS model prediction came in to question to test whether similarity scores between them were high when inputted in to PDBe Fold. This is further discussed in more detail later in chapter 4.7.6.

Submitted Primary Sequence

Gorilla:

ATGGAGTCTAAACCTTCAAGGATTCCAAGAAGAATTTCTGTTCAACCTTCCAGCTCCTTAAGTGCT
AGGATGATGTCTGGAAGCAGAGGAAGTAGTTTAAATGATACCTATCACTCAAGAGACTCTTCATGT
AGATTGGATTCTGAATATCAG

Human:

ATGGAGTCTAAACCTTCAAGGATTCCAAGAAGAATTTCTGTTCAACCTTCCAGCTCCTTAAGTGCT
AGGATGATGTCTGGAAGCAGAGGAAGTAGTTTAAATGATACCTATCACTCAAGAGACTCTTCATTT
AGATTGGATTCTGAATATCAG

Figure 7-18: Submitted primary sequence for the Gorilla and Human.

Predicted Secondary Structure:

Gorilla

MESKPSRIPRRISVQPSSSLSARMMSGSRGSSLNDTYHSRDSSCRLDSEYQ (Amino acid)
CCCCTTTTTTTEETTTCTTTTTTTTTTTTTTTTTTTTTTTCTTTC (Structure)

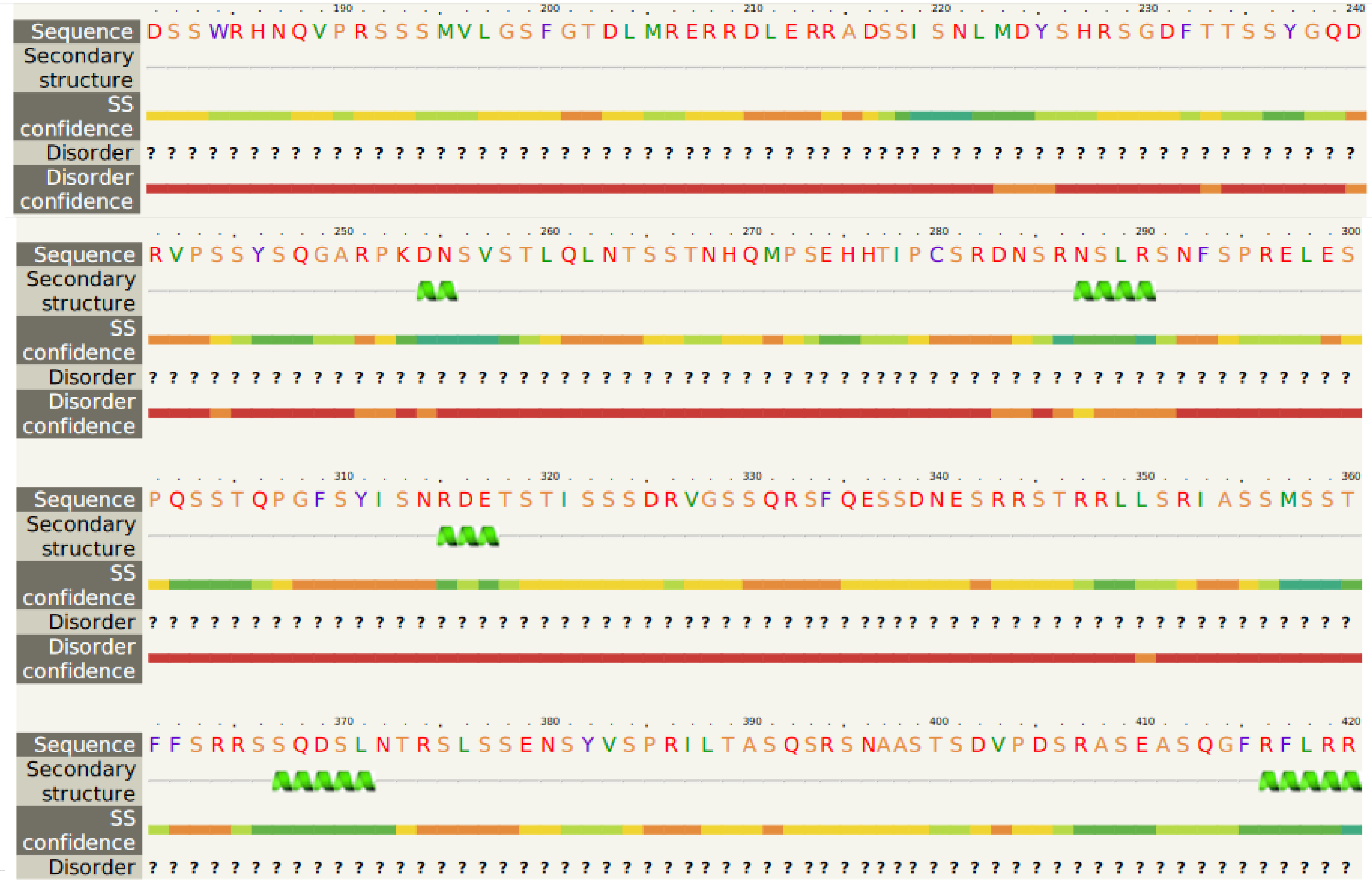
Human

MESKPSRIPRRISVQPSSSLSARMMSGSRGSSLNDTYHSRDSSFRLDSEYQ (Amino acid)
CCCCTTTTTTTEETTTCTTTTTTTTTTTTTTTTTTTTTTTCTTTC (Structure)

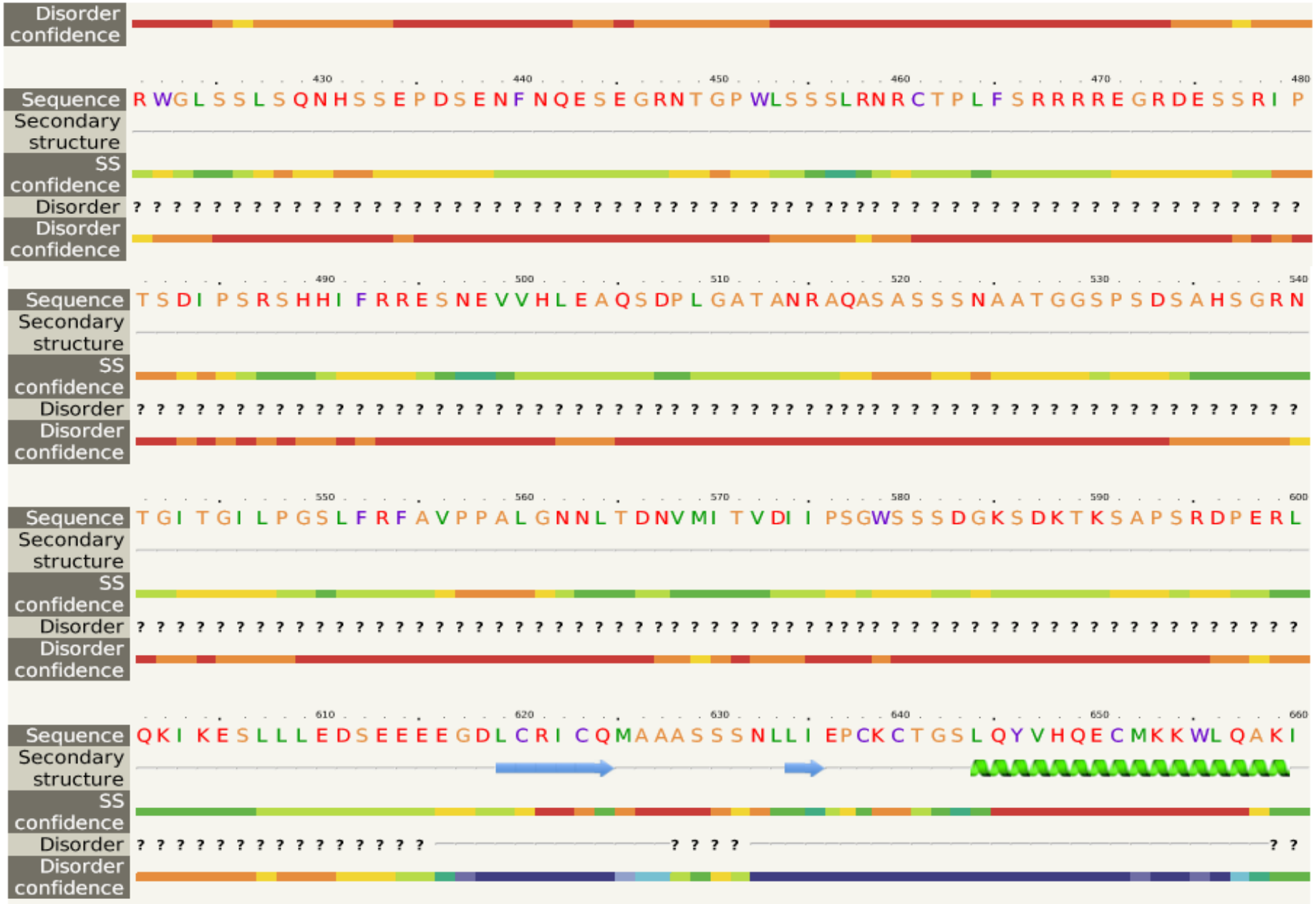
>C-coil;H-helix;E-sheet;T-beta turn

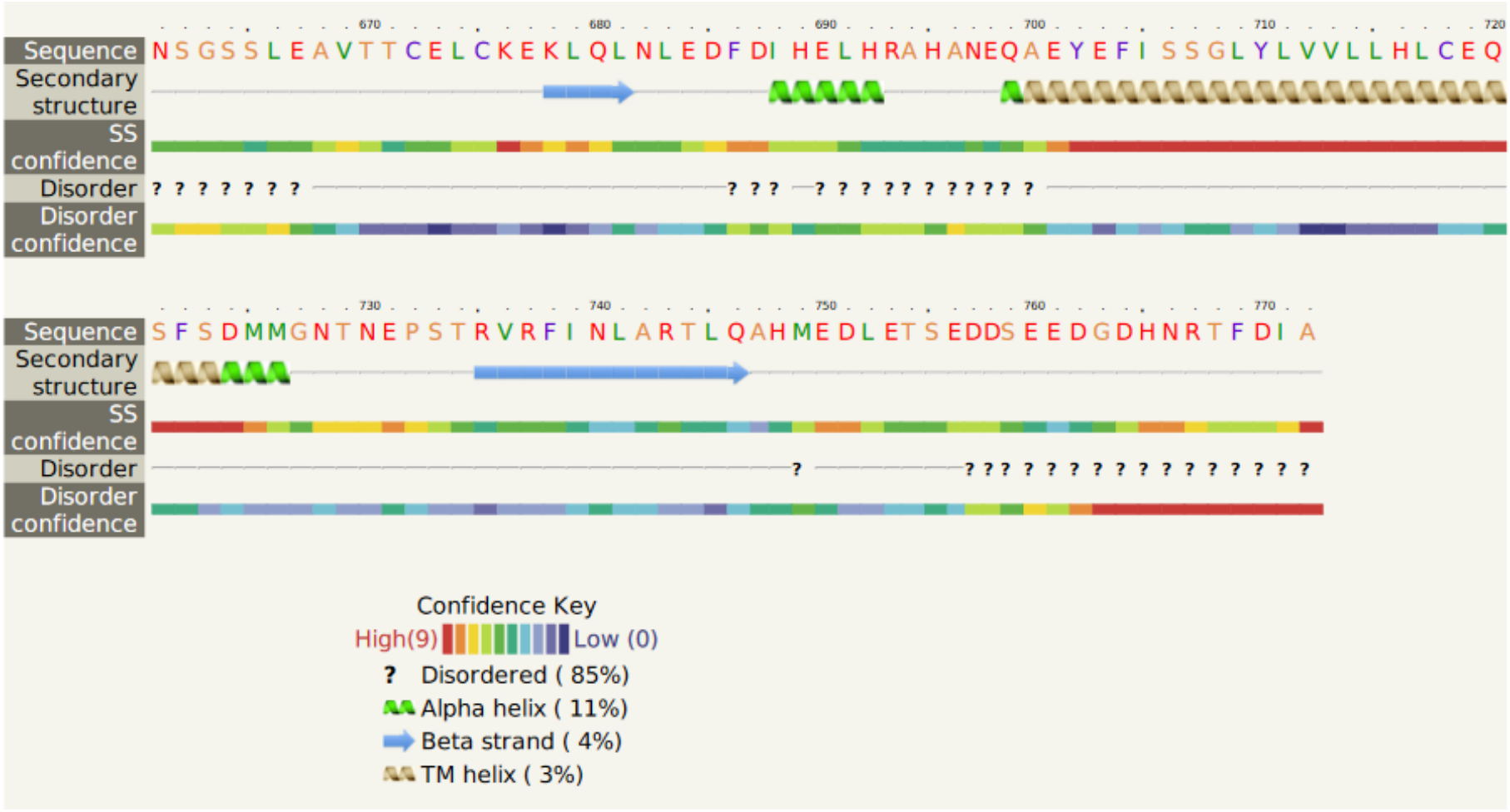
Figure 7-19: Predicted secondary structure for the Gorilla and Human.

7.26 PHYRE DISORDER PREDICTION OF MARCH7



190
200
210
220
230
240
Sequence
Secondary structure
SS confidence
Disorder
Disorder confidence
DSSWRHNQVPRSSSMVLGSFGTDLMRERRDLERRADSSISNLMDYSHRSGDFTTSSYGQD
250
260
270
280
290
300
RVPSSYSQGARPKDNSVSTLQLNTSSTNHQMPSEHHTIPCSRDNSRNSLRSNFSPRELES
310
320
330
340
350
360
PQSSTQPGFSYISNRDETSTISSSDRVGSSQRSFQESSDNESRRSTRRLLSRIASSMSST
370
380
390
400
410
420
FFSRRSSQDSLNTRSLSSENSYVSPRILTASQSRSNAASTSDVPDSRASEASQGFRFLRR


Disorder confidence
Sequence
Secondary structure
SS confidence
Disorder
Disorder confidence
R W G L S S L S Q N H S S E P D S E N F N Q E S E G R N T G P W L S S S L R N R C T P L F S R R R R E G R D E S S R I P
T S D I P S R S H H I F R R E S N E V V H L E A Q S D P L G A T A N R A Q A S A S S S N A A T G G S P S D S A H S G R N
T G I T G I L P G S L F R F A V P P A L G N N L T D N V M I T V D I I P S G W S S S D G K S D K T K S A P S R D P E R L
Q K I K E S L L L E D S E E E E G D L C R I C Q M A A A S S S N L L I E P C K C T G S L Q Y V H Q E C M K K W L Q A K I


Sequence
Secondary structure
SS confidence
Disorder
Disorder confidence
Confidence Key
High(9) Low (0)
? Disordered (85%)
Alpha helix (11%)
Beta strand (4%)
TM helix (3%)